DR HUI LI (Orcid ID : 0000-0002-9021-8091)

DR JIANBING YAN (Orcid ID : 0000-0001-8650-7811)

Article type : Original Article

Genetic variants and underlying mechanisms influencing variance heterogeneity in maize

Hui Li[1*], Min Wang[2], Weijun Li[1], Linlin He[1], Yuanyuan Zhou[1], Jiantang Zhu[1], Ronghui Che[1], Marilyn L. Warburton[3], Xiaohong Yang[2] and Jianbing Yan[4]

[1]School of Biological Science and Technology, University of Jinan, Jinan 250022, China

[2]Key Laboratory of Crop Genomics and Genetic Improvement, National Maize Improvement Center of China, China Agricultural University, Beijing 100083, China

[3]USDA ARS Corn Host Plant Resistance Research Unit, Mississippi State, MS 39759, USA

[4]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

[*]For correspondence (e-mail bio_lih@ujn.edu.cn).

Hui Li, Min Wang and Weijun Li are the authors who contributed equally to this study.

Hui Li: bio_lih@ujn.edu.cn

Min Wang: B20173010039@cau.edu.cn

Weijun Li: lwj_bio@mail.ujn.edu.cn

Linlin He: helinlin2009@163.com

Yuanyuan Zhou: 15264106057@163.com

Jiantang Zhu: bio_zhujt@ujn.edu.cn

Ronghui Che: rhche39@126.com

Marilyn L. Warburton: mariwarby@hotmail.com

Xiaohong Yang: redyx2005@126.com

Jianbing Yan: yjianbing@mail.hzau.edu.cn

RUNNING TITLE: Variance heterogeneity GWAS in maize

KEYWORDS: maize, variance heterogeneity, vGWAS, mean-effect SNP, gene x gene interactions

SUMMARY

Traditional genetic studies focus on identifying genetic variants associated with the mean difference of a quantitative trait. Because genetic variants also influence phenotypic variation via heterogeneity, we conducted a variance-heterogeneity genome wide association study (vGWAS) to examine contribution of variance heterogeneity to oil-related quantitative traits. We identified 79 unique variance-controlling single nucleotide polymorphisms (vSNPs) from the sequences of 77 candidate variance heterogeneity genes for 21 oil-related traits using Levene' test ($P < 1.0 \times 10^{-5}$). About 30% of the candidate genes encode enzymes working in lipid metabolic pathways, and most of which define clear expression variance QTLs (evQTL). Of the vSNPs specifically associated with the genetic variance heterogeneity of oil concentration, 89% can be explained by additional linked mean-effects genetic variants. Furthermore, we demonstrated that gene x gene interactions play important roles in the formation of variance heterogeneity for fatty acid compositional traits. The interaction pattern was validated for one gene pair (*GRMZM2G035341* and *GRMZM2G152328*) using Yeast two-hybrid (Y2H) and Bimolecular fluorescent complimentary (BiFC) analyses. Our findings have implications for uncovering the genetic basis of hidden additive genetic effects, epistatic interaction effects, and we indicate opportunities to stabilize efficient high-oil maize (*Zea mays* L.) breeding and selection.

INTRODUCTION

In quantitative genetics, understanding the genetic architecture affecting a quantitative trait is key to unlocking future improvement strategies. Total phenotypic variance can be partitioned into additive (V_A), dominance (V_D), epistatic (V_E) and environmental (V_e) variance. The focus of genome-wide association studies (GWAS) has typically been to detect the additive effects of genetic variants and use them to explain the contribution of each candidate gene independently to

the narrow-sense heritability of a trait ($h^2=V_A/V_P$) (Leal, 1998). However, only a limited proportion of the genetic contribution to phenotypic variance can be detected in this additive approach. The remaining genetic contribution to phenotypic variation of quantitative traits is typically called "missing" heritability, which is caused by epistasis, phenotypic plasticity or rare genetic variants (Makowsky *et al.*, 2011, Shen *et al.*, 2012, Wood *et al.*, 2014, Ek *et al.*, 2018, Liu and Yan, 2019). Dissection the genetic variants associated with the variance heterogeneity contributing to the phenotypic variability is an alternative way to explore the "missing" heritability. Variance heterogeneity is a measure of how much the variance of trait differs between two genotypes at a locus in a population (Forsberg *et al.*, 2015).

In the 1980s, variance heterogeneity QTL (vQTL) had been observed with effects on the variance, not the mean, of a complex trait (Weller *et al.*, 1988). Recently, studies in human, plants and yeast confirmed that variance-heterogeneity GWAS (vGWAS) is an effective way to detect variance difference between genotype and finally identify unexplored genetic variations with non-additive effects contributing to broad-sense heritability (Struchalin *et al.*, 2010, Rönnegård and Valdar, 2011, Shen *et al.*, 2012, Shen *et al.*, 2014). What's even more exciting is that vGWAS is always performed to leverage existing GWAS datasets. It is now suggested that genetic variance heterogeneity due to extended linkage disequilibrium (LD) across a variance-controlling locus including multiple structural variants has been demonstrated at the *MOT1* locus (Forsberg *et al.*, 2015).

Phenotypic variance heterogeneity analysis is a useful method for detecting gene x gene or gene x environment interactions and has been used successfully in human studies (Struchalin *et al.*, 2010, Hothorn *et al.*, 2012, Wang *et al.*, 2014). A vSNP (rs7202116) for an *FTO* (fat mass and obesity) variant controlled variance heterogeneity of BMI (body mass index), which affected by the interaction between genetic factors and environments including physical activity, alcohol consumption and socioeconomic status (Yang *et al.*, 2012, Rask-Andersen *et al.*, 2017). Gene x gene interactions can explain the variance heterogeneity of RNA levels in humans (Wang *et al.*,

2014). However, the underlying mechanisms of variance heterogeneity are relatively unexplored in maize.

High-oil maize (with kernel oil concentration above 6%) is a popular resource for food, animal feed and bioenergy due to its high energy content and concentration of healthy polyunsaturated fatty acids. Detection of the genetic architecture of oil and fatty acid biosynthesis and accumulation will increase efficiency of selection gain for improvement of high oil levels and quality. In our previous study, 74 loci significantly ($P < 1.8 \times 10^{-6}$) associated with kernel oil-related traits were identified via GWAS in a maize association population including 500 inbred lines with 560,000 high quality SNPs (Li *et al.*, 2013). Whereas the broad-sense heritability for ten fatty acid traits was found to be > 90%, only approximately 7.3-83% of the phenotypic variance for each trait could be explained by the significantly associated loci using an additive model (Li *et al.*, 2013). Although using new statistical method, such as the A-D test which is particularly effective for abnormal phenotypes, or using polymorphic structural variants in GWAS, "missing" heritability of these oil-related traits still exist (Yang *et al.*, 2014, Yang *et al.*, 2019). A potential explanation for the "missing" heritability is that the traditional GWAS did not consider non-additive genetic contributions, such as genetic variance heterogeneity.

In this study, we performed a vGWAS of maize kernel oil concentration and 20 oil-related traits to identify loci associated with variance heterogeneity that contribute to phenotypic variability. Next, we evaluated whether the variance heterogeneity could be explained by linked mean-effect SNPs or gene x gene interactions. To demonstrate validation of potential gene interactions, we used Yeast two-hybrid (Y2H) and Bimolecular fluorescent complimentary (BiFC) analyses, incorporating maize protein-protein interaction information available online, to present a comprehensive description of the interaction pattern of a candidate gene pair.

RESULTS

Variance-heterogeneity loci associated with oil-related traits

To identify vSNPs in the maize genome, we used Levene's test to perform a vGWAS with ~560,000 polymorphisms for 21 oil-related traits, including oil concentration, ten fatty acid compositional traits and ten fatty acid ratio traits. This method provided satisfactory vGWAS results as determined by Q-Q plots of *p*-values for each trait (Supplementary Figure 1). We identified 188 vSNPs with a significant effect on the variance of oil concentration and /or at least one of the other twenty fatty acid compositional traits at $P < 1.0 \times 10^{-5}$ (Supplementary Figure 1; Supplementary Table 1). When we merged the significant vSNPs detected in all traits, 79 unique vSNP, located in 77 unique candidate genes remained (Figure 1; Table 1; genes within a 100-kb flanking region of the lead vSNPs are also listed in Supplementary Table 2). Among the 79 vSNP, two were associated with more than one oil-related trait at $P < 1.0 \times 10^{-5}$, reflecting the strong correlation between these traits (Table 1). We also re-analyzed the variance heterogeneity using the double generalized linear model (DGLM) for the 79 vSNPs, and 62 of the 79 SNPs (78%) still reached the same significance threshold (Supplementary Table 3). Of these, 42 variance heterogeneity loci each explained more than 5% of the phenotypic variance (Table 1).

To rule out false results due to population structure, we performed a vGWAS again using all 79 loci and the lead traits in the normal-oil lines only (those with oil content < 6%). Thirty vSNPs affecting variance heterogeneity of fatty acid compositional traits were still significantly associated at $P < 1.0 \times 10^{-5}$ in the normal-oil lines (Supplementary Table 3). However, none of the vSNPs associated with oil concentration variance heterogeneity were identified in the normal-oil lines. For these fatty acid compositional traits, the variance-heterogeneity association and MAF of variance-controlling SNPs did not change notably, indicating that compositional traits have not been the target of selective breeding (Supplementary Table 3).

Relationship between vSNPs and SNPs identified by vGWAS and GWAS

To further understand the relationship between vSNPs and other SNPs, we compared the genome-wide *p*-value distributions resulting from the vGWAS and the GWAS from Li *et al.*, (2013). There was little overlap; using the same genome wide significance threshold of $P < 1.0 \times$

10^{-5}, only seven SNPs were significantly associated in both analyses for all oil-related traits (Figure 2A). Only five of the 79 unique vSNPs had a significant effect on the mean of the traits at $P < 1.0 \times 10^{-5}$, so that only 6.2% of vSNPs are also GWAS SNPs for the same genes (Figure 2B). All five loci were positively correlated, as the variance of the oil-related trait increased with increasing mean (Supplementary Figure 2). These results indicate that vGWAS is an effective complement for GWAS analysis to identify a novel set of loci affecting the phenotypic variation by heterogeneity.

We then estimated the contribution of genetic variance heterogeneity for each vSNP using $V_p = V_M + V_V + V_R$ model (for more details see Methods). For 78% (62/79) of the loci, the genetic variance heterogeneity explained higher proportion than genetic mean effect of phenotypic variation, which means that variance heterogeneity is the primary factor affecting phenotypic variation for these loci (Figure 2C; Table 1; Supplementary Table 3). This result confirms that the vGWAS method is an effective way to detect novel loci involved in shaping the total phenotypic variance that contributes to the "missing" heritability and should be used as a compliment to GWAS.

Functions and genomic features of variance-heterogeneity controlling genes

To dissect the molecular mechanism of oil content, we annotated the 77 candidate genes containing the 79 vSNPs based on motif. There were 24 (31.17%) predicted genes involved in lipid metabolism in maize or *Arabidopsis* (Figure 3A; Table 1). The remaining 53 genes encode proteins including transcription factors, stress response factors and enzymes involved in non-oil biological pathways. Approximately a fifth of the identified genes don't have functional annotation information yet (Figure 3A; Table 1). Four of the candidate genes identified by vGWAS overlapped the previous GWAS results, including *DGAT1-2* (encoding diacylglycerol acyltransferase), *TAGL* (encoding triglyceride lipases), and PDPC (encoding pyruvate dehydrogenase phosphatase), all of which are involved in lipid metabolism (Table 1). The Gene Ontology (GO) analysis of these 77 candidate genes indicate that they are significantly enriched in regulation primarily via hormone levels and transport, seed germination and post-embryonic

development, and lipid and amino acid catabolic processes (Figure 3B). These results illustrate that vGWAS is an ideal supplement to GWAS to unravel the potential molecular basis of complex traits.

To identify genes that may regulate variance heterogeneity of oil-related traits at the expression level, we tested association between the polymorphisms at a genome-wide scale and the mRNA expression levels for the 77 candidate genes identified by vGWAS using Leven's test. At the $P < 1.0 \times 10^{-5}$ significance threshold, 73/77 genes defined clear expression variance QTLs (evQTL) (Table 1; Supplementary Table 4). This resulted in the identification of 221 associated gene-SNP pairs (Supplementary Table 4), as multiple evSNPs were identified for many genes. Among 73 candidate genes, 7 were associated with both *cis*- and *trans*-acting evQTLs. *Trans*-acting evQTL were more common (66/73 or 86%), indicating remote regulation mechanisms are the primary driver of expression variance heterogeneity (Table 1; Supplementary Table 3). At $P < 0.01$, expression levels of 40/73 genes were correlated with the phenotypic variation of the GWAS target traits or oil-related traits, suggesting that some of the genes affect phenotypic variance heterogeneity via transcriptional regulation (Supplementary Table 5).

Identification of linked mean-effect SNPs contributing to variance heterogeneity

Genetic variance heterogeneity can be explained by additional SNPs with mean effect linked to vSNPs in human (Forsberg and Carlborg, 2017, Ek *et al.*, 2018). To test the contribution of linked mean-effect SNPs to genetic variance heterogeneity, we performed association tests for SNPs located on the same chromosome as each vSNP with the target trait. For half of the vSNPs (41/79) we identified a linked mean-effect SNP at $P < 1 \times 10^{-5}$ (Table 2 & Supplementary Table 6). After adjusting for the significant mean-effect SNPs, we examined variance heterogeneity for these 41 vSNPs. This resulted in 9 vSNPs still significantly associated with variance heterogeneity. The variance heterogeneity of the remaining 32 vSNPs can now instead be explained by mean-effect SNPs (Table 2, Supplementary Table 6). The MAF of the 32 vSNPs tended to be higher than that of the mean-effect SNPs, with a slight positive correlation trend between the pairwise MAFs

(Supplementary Figure 3A). In addition, high LD between the vSNP and the corresponding mean-effect SNP as measured by squared correlation coefficient (Supplementary Figure 3B)

Most of the 32 vSNP alleles associated with larger variance always linked with minor allele of mean-effect SNP on target traits (Supplementary Figure 4). For example, vSNP chr7.S_10514965 on chromosome 7 in the vGWAS for oil concentration displayed a significant genome-wide genetic variance-heterogeneity ($-\log_{10}P$ =5.00, Table 2). This associated vSNPs is located in the exon region of the gene *GRMZM2G066290*, which encodes a pyruvate kinase and synthesizes pyruvate, the first step of lipid metabolism to generate Acetyl-CoA (Li-Beisson *et al.*, 2013). The variance heterogeneity for this vSNP is explained by one mean-effect SNP, chr7.S_9794647 with MAF = 0.05 (Table 2), which is associated with mean difference of oil concentration (Figure 4A). Four haplotypes can be constructed by combining alleles for the vSNP and linked mean-effect SNP (Figure 4B). The G-allele of the vSNP occurs most in the haplotype group carrying the major allele (G-allele) of the mean-effect SNP. The lines with the A-allele in the vSNP have two different haplotypes with similar numbers of genotypes (Figure 4B). Among the four haplotypes, the A-T showed the highest mean and moderate variance values for oil concentration, which can be used to conduct selection of favor allele combinations for oil concentration improvement and low phenotypic variability during high-oil maize selection.

SNPs interacting across the genome contribute to variable phenotype

Forty-seven vSNPs (9+38) were associated with variance heterogeneity that could not be explained by additional linked SNPs, and possibly represent biological interactions (Supplementary Table 7). To test such gene x gene interactions, we carried out a whole genome scanning to identify potential interaction SNPs (iSNPs) for each vSNP. A mixed linear model was used to identify the iSNPs, which are associated with oil phenotype among the maize inbred lines assigned to the large variance (L) group (see Materials and Methods). We observed highly significant statistical interactions for 36/47 vSNPs with at least one iSNP (Supplementary Table 7; Supplementary data).

The genes within a 100-kb flanking region of each iSNP are listed in the Supplementary data, and annotations provided potential gene functions (Supplementary data). We observed several biologically interesting gene pairs. For example, *GRMZM2G137961* encodes an acyl-CoA N-acyltransferase and *GRMZM2G096358* encodes a MYB domain protein (Supplementary data). Previous studies have validated that acyl-CoA N-acyltransferase plays a key role in fertility by regulating the lipid synthesis pathway in cotton (Fu *et al.*, 2015), and that the direct or indirectly target genes of MYB-type transcription factors participate in fatty acid elongation and cuticle wax biosynthesis (Raffaele *et al.*, 2008, Seo *et al.*, 2011). Thus, oil-related variability analysis can identify potential functional relationships between genes and shed light on molecular mechanisms of quantitative traits.

Interaction pattern validation of *ZmZF_RING_H2* and *ZmActin-1*

To validate the statistical interactions identified in the previous analysis, we compared our results with a protein-protein interaction database for maize (PPIM) and found two overlapping gene pairs (Supplementary Table 8). We chose one gene pair (*GRMZM2G035341* and *GRMZM2G152328*) that was confidently predicted to interact physically in the PPIM database, as a test case to validate in the laboratory. Figure 5A presents this high confidence interaction between vSNP chr8.S_8102492 from *GRMZM2G035341* and iSNP chr5.S_10231102 from *GRMZM2G152328*. Individuals with a T genotype for vSNP chr8.S_8102492 were further sorted by chr5.S_10231102 genotypes. The increased variability in C22:0/C24:0 for individuals with the chr8.S_8102492-T genotype was explained by the heterogeneity of chr5.S_10231102 genotypes. *GRMZM2G035341* has a plant-specific zinc finger motif and belonged to RING-H2 (C3H2C3-type) zinc finger protein subfamily, which is abbreviated to *ZmZF_RING_H2* in this study (Supplementary Figure 5). Phylogenetic analysis shown that it was clustered with Bradi2g00640 from *Brachypodium distachyon* (Supplementary Figure 5). Overexpression of a zinc finger encoding gene has been shown to activate lipid biosynthesis genes, thereby accelerating kernel oil accumulation (Li *et al.*, 2017). *GRMZM2G152328* encodes actin-1, here

called *ZmActin-1*. A homologous gene in cotton participates in cell expansion and cellulose synthesis during fiber elongation (Deng *et al.*, 2016).

To test whether the proteins (ZmZF_RING_H2 and ZmActin-1) interact, Y2H and BiFC analysis were performed. The interaction between ZmZF_RING_H2 and ZmActin-1 was observed in yeast two-hybrid assay (Figure 5B). We then used a BiFC assay to verified the interaction of ZmZF_RING_H2 and ZmActin-1. The BiFC assay demonstrated the presence of fluorescence in chloroplasts of tobacco cells co-transformed with ZmZF_RING_H2-YFP^N and ZmActin-1-YFP^C, while the chloroplasts of the control YFP^N/YFP^C transformants lacked fluorescence (Figure 5C). Together, these results indicate that ZmZF_RING_H2 interacts with ZmActin-1, and both are novel functional candidates involved in oil pathways via epistatic effect.

DISCUSSION

Phenotypic variation usually refers to the difference of phenotypic values among diverse genotypes. GWAS is a common method to elaborate the genetic mechanism of quantitative traits, and aims to identify loci, loci interactions, and locus x environmental interactions, that are associated with phenotype difference at a genome-wide significant level in a panel. Previously, we conducted GWAS and pathway analysis to explore the genetic architecture of oil-related traits in maize kernel (Li *et al.*, 2013, Li *et al.*, 2019). Although 26 genetic variants associated with oil concentration explained up to 83% of the phenotypic variation using an additive model, few loci associated with the fatty acid compositional traits were identified (Li *et al.*, 2013). Furthermore, epistatic interactions are notoriously difficult to detect via GWAS because of the huge number of genotypes needed for sufficient statistical power to reliably find these interactions. It was therefore not surprising that we detected no significant epistatic interactions among all the mean-effect SNPs identified in Li *et al.*, (2013). Thus, although GWAS is a powerful way to identify individual loci with the additive effects, it has limited power to detect non-additive genetic variance and variance heterogeneity which also contributing to phenotypic variation (Carlborg and S.Haley, 2004).

To identify variance-heterogeneity affecting loci using multiple statistical genetic models in GWAS and to further dissect underlying genetic architecture of mechanisms contributing to the broad-sense heritability that was missed in the traditional study, we re-analyzed GWAS dataset using Levene's test to detect variance heterogeneity. This test controls false-positive rates and is suitable for non-normalized data, such as our oil-related phenotypic data (Struchalin *et al.*, 2010, Li *et al.*, 2015). DGLM is also a valid statistical model to identify genetic-controlling loci associated with phenotypic variability in chicken F_2 crosses and in studies of human target gene expression levels (Rönnegård and Valdar, 2011, Hulse and Cai, 2013). We used DGLM to re-calculate significantly associated vSNPs as detected by Levene's test, and found that most of the vSNPs (78%) still reach the same significant threshold. There was little overlap in loci detected in the vGWAS and previous GWAS (Li *et al.*, 2013) for the same genome wide significant threshold, indicating that genetic variants influencing the mean and variance of oil-related trait are different. Thus, vGWAS is an effective tool to detect a novel set of genetic variants controlling the variance heterogeneity contributing to phenotype variation. The 79 unique vSNPs and corresponding 77 candidate genes identified by vGWAS had annotations of biological significance, uncovering the genetic architecture of oil biosynthesis. Further molecular biology experiments need to be carried out to verify the function mechanism of candidate genes affecting the phenotypic variance heterogeneity.

There are three different main genetic mechanisms causing stochasticity of a trait: Different alleles, structured LD between alleles at linked loci, and genetic interactions between loci. The last two could contribute to genetic variance heterogeneity as described in previous literatures (Shen *et al.*, 2012, Ayroles *et al.*, 2015, Ek *et al.*, 2018). When we tested the genetic mechanism of variance heterogeneity in maize, we found that ~40% (32/79) of variance heterogeneity loci are driven by additional mean-effect linked to the vSNPs (Table 2; Supplementary Figure 3). Thus, the LD between vSNP and linked mean-effect SNP is one of the causes of the phenotypic variance heterogeneity; these results are consistent with those found in human GWAS studies (Ek *et al.*,

2018).

To illustrate, we consider an association example of two bi-allelic SNPs, where the vSNP is fixed for allele A and the linked independent mean-effect SNP is fixed for allele B. When the major vSNP allele (A^+) is always combined with the major mean-effect SNP allele (B^+) in the haplotype group, lines with the A^+ allele have no haplotype variability at the vSNP. In contrast, lines carrying the A^- allele can display two different haplotypes (A^-B^- and A^-B^+), each with different mean phenotypic values. This is the case we found with vSNP chr7.S_10514965 in our study, which has one vSNP allele (G-allele, corresponding to small phenotypic variance) in the major haplotype group, and the other allele (A-allele, corresponding to large phenotypic variance) in the two less frequent haplotypes. So, some of the variance heterogeneity is caused by the LD between one allele in the vSNP and the two alleles in the mean-effect SNPs; identifying these cases in traditional GWAS would be difficult except with prohibitively large panel sizes.

It's worth noting that more than half of the vSNPs (17/32) associated with variance heterogeneity of oil concentration can be explained by more than one additional mean-effect SNP (Table 2). In such situations, the variance heterogeneity loci linked with additional mean-effect SNPs may contribute to narrow-sense heritability that is difficult to estimate using a traditional GWAS. Our results suggest that oil concentration, but not fatty acid compositional traits, is mainly inherited additively, which has also been observed in previous GWAS studies (Cook *et al.*, 2012, Li *et al.*, 2013).

The association 47 vSNP with the traits under study (including 45 associated with variance heterogeneity of fatty acid compositional traits) could not be explained by linked mean-effect SNPs, and were hypothesized to affect the oil related traits via gene x gene interactions. Under this scenario, 36 vSNPs were identified through the interacting effect of multiple SNPs to influence the final phenotypes (Supplementary Table 7), for a total of 132 vSNP-iSNP pairs (Supplementary Table 9). A previous review has neatly presented the connection between three different types

epistasis and the formation of genetic variance heterogeneity (Forsberg and Carlborg, 2017); two types of epistasis can be detected by vGWAS. Figure 6A illustrates one type of epistasis interaction between two loci, A and B, both of which have two alleles. In this theoretical example, only the A_2B_2 haplotype has a phenotypic effect (Forsberg and Carlborg, 2017). As a result, A and B loci affect both the mean and the variance of a quantitative trait and can be detected by conventional and vGWAS (Figure 6B). In our real data, SNP PZE-104040568 displayed C- and A- alleles and SNP chr4.S_142153507 displayed T- and A- alleles. Among the four haplotypes constructed by these two SNPs, only the A-A- haplotype (corresponding to A_2B_2 of the example) has an effect on the oil concentration phenotype (Figure 6C; Supplementary Table 9).

Another type of epistasis is shown in Figure 6D, where allele A_2 capacitates the effect of locus B, which means that alleles B_1 and B_2 display phenotypic effects only when combined with allele A_2 of locus A. In this case, locus A might only be identified in genetic variance heterogeneity analysis, but locus B can be identified in a traditional GWAS for additive effects (Figure 6E) (Forsberg and Carlborg, 2017). Figure 6F shows four haplotypes for two interacting loci from our study (chr1.S_287706446 with T- and C- alleles corresponding to A_1 and A_2, and chr8.S_75602135 with C- and T- alleles corresponding to B_1 and B_2). The C-allele of chr1.S_287706446 displayed phenotypic effects when combined with both alleles of chr8.S_75602135 (examples of A_2B_1 and A_2B_2; Figure 6F; Supplementary Table 9). When we classified the 132 vSNP-iSNP pairs by interaction categories as per Figure 6, we found that 124 fell into category I, and only 8 into category II (Supplementary Table 9). This clearly illustrates which kind of epistasis is important in fatty acid composition, and the value of vGWAS to interpret interactions between genetic variants.

In summary, this study has shown that variance heterogeneity can be attributed to vSNPs, and they are as common as mean-effect SNPs influencing oil-related traits identified by GWAS in maize. Most vSNP associated with variance heterogeneity of oil concentration were explained by additional mean-effect SNPs. This finding validated the additive manner in which oil

concentration is inherited, and can be used to guide the selection of haplotypes for oil concentration improvement even in populations with low phenotypic variability for high-oil maize selection. However, gene x gene interaction plays a leading role in the formation of variance heterogeneity of fatty acid compositional traits. This explains a significant portion of the missing heritability and will allow a modification of breeding selection plans to achieve the most efficient manner of creating maize lines with the desired fatty acids.

EXPERIMENTAL PROCEDURES

Genotype and phenotype data

The vGWAS was performed using an association mapping population including 368 inbred maize lines (Yang *et al.*, 2011) that had been genotyped with two platforms: an RNA-seq analysis resulted in gene expression data for 28,769 annotated genes and a SNP array provided 550,000 high quality SNPs with minor allele frequency (MAF) $\geq$ 0.05. These data were published previously (Fu *et al.*, 2013, Li *et al.*, 2013) and are available publicly (http://www.maizego.org/Resources.html). Oil concentration, the concentration levels of ten fatty acids, including palmitic (C16:0), palmitoleic (C16:1), stearic (C18:0), oleic (C18:1), linoleic (C18:2), linolenic (C18:3), arachidic (C20:0), gadoleic (C20:1), behenic (C22:0) and lignoceric (C24:0) acids, the ratios between several of these fatty acids, including C16:0/C16:1, C16:0/C18:0, C18:0/C18:1, C18:1/C18:2, C18:2/C18:3, C18:0/C20:0, C20:0/C20:1, C20:0/C22:0, C22:0:C24:0, and the ratio between saturated fatty acid and unsaturated fatty acid, referring to SFA/USFA=C16:0+C18:0+C20:0+C22:0+C24:0)/(C16:1+C18:1+C18:2+C18:3+C20:1), led to a total of 21 traits measured on all 368 inbred lines as per (Li *et al.*, 2013).

vGWAS analysis

The vGWAS was performed using a two-step approach. In the first step, to correct for population stratification, the trait was fit in a linear mixed model with kinship matrix, which calculated by the polygenic function in the R-package GenABEL (Aulchenko et al., 2007), to get Grammar + residuals. In the second step, the variance-heterogeneity between the Grammar + residuals and

SNPs were tested using Levene's test. And the Levene's test is based on an ANOVA of the absolute deviation from the median and detailed information is described in previous studies (Shen et al., 2012).

We also re-calculate all the vSNPs found to be significant using the double generalized linear model (DGLM) for homogeneity of variance in the DGLM R-package (Rönnegård and Valdar, 2011, Hulse and Cai, 2013) as follows:

$$y_i = \mu + x_i\beta + g_i\alpha + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2 \exp(g_i\theta))$$

where y_i indicates the phenotypic trait of the line i, x_i is the population structure, g_i is the SNP genotype, ε_i is the residual with variance σ^2, and θ is the corresponding vector of coefficients of genotype g_i on the residual variance.

Because of the strong LD among genome-wide SNPs, the number of independent SNPs were always used to assess the threshold for GWAS (Li et al., 2012; Yang et al., 2014). The parameters for independent maker determination have been reported in previous study using the same genotype data as us (Wang et al., 2016). Thus, the threshold was set to 1.0×10^{-5} (1/85,806) in present study.

Phenotypic variation explained by multiple SNPs

We used the following model, $V_P = V_M + V_V + V_R$ to calculate the proportion of the variance heterogeneity-effect variance (V_V) to the phenotypic variance (V_P) (Hill and Mulder, 2010). In this model, V_M is the mean-effect variance, and V_R is the environmental variance. For two given homozygous genotypes in the population, $V_M = V_A = pq\alpha^2$, and the variance heterogeneity-effect variance is, $V_V = pq\phi^2$, where p and q are the frequencies for the two alleles and α and ϕ are the difference between the two homozygous genotypes in the mean and standard deviation, respectively.

Gene function annotation and GO enrichment analysis

We integrated two annotation resources, MaizeGDB (http://maizecyc.maizegdb.org) and InterPro

(http://www.ebi.ac.uk/interpro), to explore candidate gene functions (Zdobnov and Apweiler, 2001, Lawrence *et al.*, 2008). The agriGO v2 (http://bioinfo.cau.edu.cn/agriGO/) was performed to do GO enrichment analysis with SEA (Singular Enrichment Analysis) option (Blake, 2000, Du *et al.*, 2010). The *P* values were adjusted for multiple testing by controlling FDR.

Regional association analysis to identify mean-effect SNPs

For each significant vSNP, we performed association analysis to identify SNPs with mean effects on the target trait. First, only SNPs located on the same chromosome as the vSNP under study were considered. Then a mixed linear model controlling population structure and relative kinship was used to test for association between the target trait and the SNPs under consideration, in order to identify the mean-effect SNPs (Yu *et al.*, 2006). If there were no significantly associated SNPs at $P < 1.0 \times 10^{-5}$ on the same chromosome, we defined the vSNP as without any mean-effect SNP. If significant mean-effect SNPs were found, we ran the vGWAS analysis again for each significant vSNP conditioning on each significant mean-effect SNP, and selected the primary SNP as the one with the greatest impact on *P* value of the vSNP. Thus, if an associated SNP, when used as covariate in the vGWAS analysis, increased the *P* value of the vSNP to larger than 1×10^{-5}, it was defined as the primary mean-effect SNP. If the of *P* value of the vSNP after conditioning on the primary mean-effect SNP was still less than 1×10^{-5}, another mean-effect SNP with the second impact on *P* value of vSNP, like the primary SNP, is added to the vGWAS as a covariate. If the *P* value of the vSNP after conditioning on the primary mean-effect SNP and second mean-effect SNP still be less than 1×10^{-5}, it defined that this vSNP was still associated with variance heterogeneity of. If not, the second SNP also defined as mean-effect SNP.

Epistatic interactions

A two-stage method was used to identify interacting SNPs with each vSNP (Hulse and Cai, 2013). The inbred lines in association mapping panel were divided into two groups, L and S groups, based on genotypes of each vSNP, which were associated with large (L) and small (S) variances of target trait. Next, we performed a genome-wide scan to find SNPs via traditional GWAS with

mixed linear model using GWAS function in the rrBLUP R-library among lines of the L-group (Hulse and Cai, 2013).

Sequence feature and phylogenetic analysis

The sequences alignment was performed using ClustalX software (Thompson *et al.*,1997). Base on the conserved domain sequences, a phylogenetic tree was constructed using MEGA7.1 by the Neighbor Joining (NJ) method with 1000 bootstrap replicates in p-distance model.

Yeast two-hybrid assay

The ProQuest two-hybrid system (Invitrogen) was used in a yeast two-hybrid assay. The bait and prey plasmids were constructed by transferring the full length of *ZmZF_RING_H2* and *ZmActin-1*, respectively. Then, the construct pairs were co-transformed into yeast strain MaV203. Y2H screening was performed according to the protocol described by Lee *et al.*, 2014.

Bimolecular Fluorescence Complementation

BiFC assays were performed as previously described (Waadt *et al.*, 2010). The full-length cDNA of *ZmZF_RING_H2* was subcloned into the pSPYNE(R) vector (YFP^N), and the full-length cDNA of *ZmActin-1* was subcloned into the pSPYCE (R) vector (YFP^C). The plasmids were co-expressed in 5-week-old *Nicotiana benthamiana* leaf epidermis cells by *Agrobacterium*-mediated infiltration. YFP fluorescence was visualized with a confocal scanning microscope after infiltration for 72 h.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

ACKNOWLEGEMENTS

We thank Mr. Mauricio Carlos Kuki for reviewing the manuscript and providing excellent suggestions for improvement. We are grateful to the National Natural Science Foundation of China (31771797), the National Key Research and Development Program of China

(2016YFD0100503), and Joint Fund for Excellent Young Talents in Universities of Shandong Province (ZR201807061228).

AUTHOR CONTRIBUTIONS

LH, YJ, YX and WM conceived and designed the experiments. LH, WM and LW analyzed data. HL, ZY, ZJ and CR performed the experiments.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

Supplementary Figure 1. Manhattan (left) and quantile-quantile (right) plots resulting from vGWAS of 21 maize kernel oil-related traits. The dashed horizontal line represents the genome-wide significant threshold (1.0×10^{-5}). Oil, oil concentration; C16:0, palmitic acid; C16:1, palmitoleic acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, linolenic acid; C20:0, arachidic acid; C20:1, gadoleic acid; C22:0, behenic acid; C24:0, lignoceric acid; SFA, saturated fatty acid; USFA, unsaturated fatty acid.

Supplementary Figure 2. Violin figures of oil-related traits for the five loci having effects both on the mean and variance. Panels A-E indicate the five SNPs identified both by vGWAS and previous GWAS results (Li *et al.*, 2013).

Supplementary Figure 3. Minor allele frequency (MAF) and linkage disequilibrium (LD) for 32 vSNPs and additional linked mean-effect SNPs. (A) Pairwise MAF for vSNP and mean-effect SNP. (B) LD between each vSNP and corresponding mean-effect SNP.

Supplementary Figure 4. Violin figures of oil-related traits for the 31 vSNP with more than one mean-effect SNP.

Supplementary Figure 5. Sequence feature and phylogenetic analysis of GRMZM2G035341. (A) Amino acid sequence alignment of GRMZM2G035341, Bradi2g00640, AT5G20570, LOC_Os01g01700 and Glyma.12G100300. Characters highlighted with black indicate conserved

amino acids. (B) Phylogenetic analysis of GRMZM2G035341, Bradi2g00640, AT5G20570, LOC_Os01g01700 and Glyma.12G100300.

Supplementary Table 1. Summary of vSNPs significantly associated with variance heterogeneity of oil-related traits at $P < 1.0 \times 10^{-5}$ and $P < 1.8 \times 10^{-6}$.

Supplementary Table 2. List of possible additional candidate genes within a 100 kb flanking region of the 79 lead vSNPs identified as significantly ($P < 1.0 \times 10^{-5}$) associated with oil-related traits in this study.

Supplementary Table 3. Re-calculation the variance heterogeneity for 79 vSNP via double generalized linear model (DGLM).

Supplementary Table 4. Expression vQTL (evQTL) results for candidate genes identified by vGWAS in this study.

Supplementary Table 5. Correlation analysis between the trait phenotype and the expression of proposed candidate genes with evQTLs.

Supplementary Table 6. Nine vSNPs still associated with variance heterogeneity after adjusting for linked mean-effect SNPs.

Supplementary Table 7. Forty-seven vSNPs associated with variance heterogeneity which can't be explained by additional linked mean-effect SNPs and the number of interacting SNPs for each vSNP.

Supplementary Table 8. Interacting gene pairs which were validated by protein-protein interactions as identified in protein-protein interaction database for maize.

Supplementary Table 9. 145 vSNP-iSNP interaction pairs and epistasis categories.

REFERENCE

Aulchenko, Y.S., Stephan, R., Aaron, I., et al. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294-1296.

Ayroles, J.F., Buchanan, S.M., O'Leary, C., et al. (2015) Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc Natl Acad Sci U S A*, **112**, 6706-6711.

Blake, M.A.C.A.B.J.A. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**, 25-29.

Carlborg, Ö. and S.Haley, C. (2004) Epistasis:too often neglected in complex trait studies? *Nat Rev Genet* **5**, 618-625.

Cook, J.P., Mcmullen, M.D., Holland, J.B., et al. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol*, **158**, 824-834.

Deng, T., Yao, H., Wang, J., et al. (2016) GhLTPG1, a cotton GPI-anchored lipid transfer protein, regulates the transport of phosphatidylinositol monophosphates and cotton fiber elongation. *Sci Rep,* **6**, 26829.

Du, Z., Zhou, X., Ling, Y., et al. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res*, **38**, W64-70.

Ek, W.E., Rask-Andersen, M., Karlsson, T., et al. (2018) Genetic variants influencing phenotypic variance heterogeneity. *Hum Mol Genet*, **27**, 799-810.

Forsberg, S.K., Andreatta, M.E., Huang, X.Y., et al. (2015) The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. *PLoS Genet*, **11**, e1005648.

Forsberg, S.K.G. and Carlborg, O. (2017) On the relationship between epistasis and genetic variance heterogeneity. *J Exp Bot*, **68**, 5431-5438.

Fu, J., Cheng, Y., Linghu, J., et al. (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat Commun*, **4**, 2832.

Fu, W., Shen, Y., Hao, J., et al. (2015) Acyl-CoA N-acyltransferase influences fertility by regulating lipid metabolism and jasmonic acid biogenesis in cotton. *Sci Rep*, **5**, 11790.

Hill, W.G. and Mulder, H.A. (2010) Genetic analysis of environmental variation. *Genet Res (Camb)*, **92**, 381-395.

Hothorn, L.A., Libiger, O. and Gerhard, D.J.B.G. (2012) Model-specific tests on variance heterogeneity for detection of potentially interacting genetic loci. *BMC Genet*, **13**, 59-59.

Hulse, A.M. and Cai, J.J. (2013) Genetic variants contribute to gene expression variability in humans. *Genetics*, **193**, 95-108.

Lawrence, C.J., Harper, L.C., Schaeffer, M.L., et al. (2008) MaizeGDB: The maize model organism database for basic, translational, and applied research. *Int J Genomics,* **2008**, 496957.

Leal, S.M. (1998) *Genetics and Analysis of Quantitative Traits*. *Am. J. Hum. Genet*., **68**,548-549.

Li-Beisson, Y., Shorrosh, B., Beisson, F., et al. (2013) Acyl-lipid metabolism. *The Arabidopsis Book.*

Li, H., Peng, Z., Yang, X., et al. (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*, **45**, 43-50.

Li, H., Thrash, A., Tang, J.D., et al. (2019) Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant J*, **98**, 853-863.

Li, Q.T., Lu, X., Song, Q.X., et al. (2017) Selection for a Zinc-Finger Protein Contributes to Seed Oil Increase during Soybean Domestication. *Plant Physiol*, **173**, pp.01610.02016.

Li, X., Qiu, W., Morrow, J., et al. (2015) A Comparative Study of Tests for Homogeneity of Variances with Application to DNA Methylation Data. *PLoS ONE,* **10**, e0145295.

Liu, H.J. and Yan, J. (2019) Crop genome-wide association study: a harvest of biological relevance. *Plant J*, **97**, 8-18.

Liu, X., Yue, Y., Li, B., et al. (2007) A G protein-coupled receptor is a plasma membrane receptor for the plant hormone abscisic acid. *Science*, **315**, 1712-1716.

Makowsky, R., Pajewski, N.M., Klimentidis, Y.C., Vazquez, A.I., Duarte, C.W., Allison, D.B. and de los Campos, G. (2011) Beyond missing heritability: prediction of complex traits. *PLoS Genet*, 7, e1002051.

Raffaele, S., Vailleau, F., Leger, A., et al. (2008) A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*. *Plant Cell*, **20**, 752-767.

Rask-Andersen, M., Karlsson, T., Ek, W.E., et al. (2017) Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet*, **13**, e1006977.

Rönnegård, L. and Valdar, W. (2011) Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics*, **188**, 435-447.

Seo, P.J., Lee, S.B., Suh, M.C., et al. (2011) The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in *Arabidopsis*. *Plant Cell*, **23**, 1138-1152.

Shen, X., De Jonge, J., Forsberg, S.K., et al. (2014) Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet*, **10**, e1004842.

Shen, X., Pettersson, M., Ronnegard, L., et al. (2012) Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsis thaliana*. *PLoS Genet*, **8**, e1002839.

Struchalin, M.V., Dehghan, A., Witteman, J.C., et al. (2010) Variance heterogeneity analysis for detection of

potentially interacting genetic loci: method and its limitations. *BMC Genet*, **11**, 92.

Thompson, J.D., Gibson, T.J. and Plewniak, F. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**.

Waadt, R., Schmidt, L.M., Hashimoto, K., et al. (2010) Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes in planta. *Plant J*, **56**, 505-516.

Wang, G., Yang, E., Brinkmeyer-Langford, C.L., et al. (2014) Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics*, **196**, 413-425.

Weller, J.I., Soller, M., and Brody, T. (1988) Linkage analysis of quantitative traits in an interspecific cross of tomato (lycopersicon esculentum x lycopersicon pimpinellifolium) by means of genetic markers. *Genetics,* **118**, 329-339.

Wood, A.R., Esko, T., Yang, J., et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*, **46**, 1173-1186.

Yang, J., Loos, R.J., Powell, J.E., et al. (2012) FTO genotype is associated with phenotypic variability of body mass index. *Nature*, **490**, 267-272.

Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., Huang, J., Deng, T., Luo, J., He, L., Wang, Y., Xu, P., Peng, Y., Shi, Z., Lan, L., Ma, Z., Yang, X., Zhang, Q., Bai, M., Li, S., Li, W., Liu, L., Jackson, D. and Yan, J. (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat Genet*, **51**, 1052-1059.

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W., Liu, J., Li, J. and Yan, J. (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet*, **10**, e1004573.

Yang, X., Gao, S., Xu, S., et al. (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breeding*, **28**, 511-526.

Yu, J., Pressoir, G., Briggs, W.H., et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, **38**, 203-208.

Zdobnov, E. and Apweiler, R.J.B. (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics,* **17**, 847-84

FIGURE LEGENDS

Figure 1. Chromosome distributions for unique significant vSNPs. The blue and red vertical lines represent unique significant vSNP at $P < 1.0 \times 10^{-5}$ and $P < 1.8 \times 10^{-6}$, respectively. Oil, oil concentration; C16:0, palmitic acid; C16:1, palmitoleic acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, linolenic acid; C20:0, arachidic acid; C20:1, gadoleic acid; C22:0, behenic acid; C24:0, lignoceric acid; SFA, saturated fatty acid; USFA, unsaturated fatty acid.

Figure 2. Correlation of *P* values and genetic contribution to phenotypic variance for loci identified by the vGWAS and GWAS. (A) Comparison of GC-corrected *P* values for genome-wide loci detected in previous GWAS results (Li *et al.*, 2013) and present in the current vGWAS for all oil-related traits. The black dashed lines indicate the genome-wide significant threshold of 10^{-5}. The red dots represent overlapping SNPs identified by both vGWAS and GWAS results. (B) Comparison of GC-corrected *P* values for 79 significant vSNP in vGWAS and GWAS results. The red dots represent vSNPs among 79 loci both having mean and variance effects. (C) Comparison of proportion of the phenotypic variance explained for 79 vSNP in vGWAS and GWAS.

Figure 3. Functional category annotations and their respective percentages (panel A) and GO analysis (panel B) for 77 unique candidate genes identified via vGWAS as significantly associated with the variance heterogeneity of oil related traits.

Figure 4. Associations and haplotype analysis of mean-effect SNPs and variance heterogeneity SNPs. (A) Top: The mean-effect SNP chr7.S_9794647 was associated with mean difference of oil concentration. Bottom, the blue triangle represents where the vSNP chr7.S_10514965 was associated with variance heterogeneity of oil concentration, and the red dot represents where the vSNP chr7.S_10514965 was no longer significantly associated with variance heterogeneity of oil concentration after adjusting for the mean-effect SNP chr7.S_9794647. (B) Top: Violin figure of oil concentration for vSNP chr7.S_10514965 with an effect on the variance. Bottom: Violin figure of oil concentration for the four different haplotypes.

Figure 5. Interactions between GRMZM2G035341 and GRMZM2G152328 contributing to variable phenotype. (A) Individuals with chr8.S_8102492-TT genotype are further sorted by

chr5.S_10231102 into two subgenotype groups, which are associated with different C22:0/C24:0 means. (B) Yeast two-hybrid assay. P22 or P32 plasmid containing GRMZM2G035341 and GRMZM2G152328 were transformed into yeast strain MAV203. P22, pEXP22; P32, pEXP32. L, Leucine; T, Threonine; U, Uracil; H, Histidine. WT (pEXP22/RalGDS-wt), m1 (pEXP22/RalGDS-m1), and m2 (pEXP22/RalGDS-m2) are control plasmids displaying a strong, weak or undetectable interaction with pEXP32-Krev1, respectively. (C) BiFC assay in tobacco leaves co-transformed with GRMZM2G152328-PXN and GRMZM2G035341-PXC. PXN: YFP N terminal region, PXC: YFP C terminal region. Scale bar: 50 μm.

Figure 6. Theoretical and real examples illustrate two types of epistasis. Panels A and D are two theoretical examples of epistasis interacting between two loci (A and D). Panels B and E are the theoretical phenotypic distributions corresponding to loci A and B. In B and E, the dark yellow fill represents individuals with one allele, the green fill represents individuals with the opposite allele. Panel C is the real example for a pairwise interaction between two SNPs (chr4.S_142153507 and SYN24171) affecting mean oil concentration. Panel F shows the real example for a pairwise interaction between two SNPs (chr8.S_75602135 and chr1.S_287706446) also affecting mean oil concentration.

TABLE LEGENDS

Table 1. vSNPs and linked candidate genes significantly ($P < 1.0 \times 10^{-5}$) associated with variance heterogeneity of 21 oil-related traits.

Table 2. *P* values for 32 vSNPs associated with variance heterogeneity after adjusting for the mean-effect SNPs.

Table 1. vSNPs and linked candidate genes significantly ($P < 1.0 \times 10^{-5}$) associated with variance heterogeneity of 21 oil-related traits.

vSNP	Lead trait[a]	Other trait[b]	Chr.	Position[c]	Alleles[d]	MAF[e]	*P*[f]	*Vv*/*Vp*[g]	Gene_Id[h]	evQTL type[i]	Function description[j]
chr1.S_12352096	C22:0/C24:0		1	12352096	T/G	0.14	2.78E-07	10.44%	*GRMZM2G154211*	*Trans*	Sulfate transporter
chr1.S_23614308	SFA/USFA		1	23614308	T/G	0.48	2.66E-06	3.27%	*GRMZM2G153769*	*Trans*	Cop9 signalosome complex subunit 4
chr1.S_53423512	Oil		1	53423512	A/C	0.11	4.84E-06	5.83%	*GRMZM2G031001*	*Trans*	NAC-transcription factor 11
chr1.S_244088357	C22:0/C24:0		1	244088357	T/C	0.35	5.77E-06	2.25%	*GRMZM2G444692*	*Trans*	Regulator of MON1-CCZ1 complex
chr1.S_276308390	C20:0/C20:1		1	276308390	T/C	0.10	4.15E-06	6.11%	*GRMZM2G147687*	*Trans*	Glycoside hydrolase family 3 C-terminal domain
chr1.S_276361864	C20:0/C20:1		1	276361864	A/G	0.31	6.30E-06	4.57%	*GRMZM2G016546*	*Trans & Cis*	DNA methylase
chr1.S_287706446	C16:0/C16:1		1	287706446	T/C	0.19	4.95E-06	3.07%	*GRMZM2G448927*	*Trans*	Acyl-CoA N-acyltransferase
chr2.S_2110986	C22:0/C24:0		2	2110986	A/G	0.09	7.07E-06	7.39%	*GRMZM2G145825*	*Trans*	Unknown
chr2.S_3177741	SFA/USFA		2	3177741	A/C	0.07	9.87E-06	3.10%	*GRMZM2G119773*	*Trans & Cis*	Cell division protein FtsZ
chr2.S_144072332	C18:0/C18:1		2	144072332	T/G	0.11	5.67E-06	3.59%	*GRMZM2G055973*	*Trans*	Ring-H2 finger protein Atl7
chr2.S_197007535	C16:0/C18:0		2	197007535	T/C	0.07	6.63E-06	3.35%	*GRMZM2G042712*	*Trans*	Auxin-responsive protein saur61
chr2.S_204388944	Oil		2	204388944	A/C	0.15	8.53E-07	10.66%	*GRMZM2G144180*	*Trans*	Duf538 family protein
chr2.S_204470447	Oil		2	204470447	A/T	0.11	8.39E-07	7.47%	*GRMZM2G163233*	*Trans*	Male sterile32
chr2.S_224995610	C16:0/C18:0		2	224995610	C/G	0.06	9.09E-07	3.17%	*GRMZM2G082785*	*Trans*	Unknown
chr2.S_234229554	C18:0/C18:1		2	234229554	C/G	0.07	3.29E-06	3.01%	*GRMZM2G139467*	*Trans*	Cytochrome P450

chr3.S_1852151	C18:0/C18:1		3	1852151	C/G	0.07	4.04E-06	2.92%	*GRMZM2G093104*	*Trans & Cis*	Activating signal cointegrator 1 complex subunit 1
chr3.S_2876077	C18:0/C18:1		3	2876077	T/C	0.11	2.04E-09	6.22%	*GRMZM2G068217*	*Trans*	Ethylene insensitive 2
chr3.S_3887748	C22:0/C24:0		3	3887748	C/G	0.06	5.38E-06	5.28%	*GRMZM2G093278*	*Trans*	Phospho-N-acetylmuramoyl-pentapeptide-transferase homolog，catalyse
chr3.S_5568273	C18:1		3	5568273	C/G	0.30	8.59E-06	2.92%	*GRMZM2G143235*	*Trans*	Cytochrome P450
chr3.S_5580332	C20:0/C22:0		3	5580332	A/G	0.33	5.50E-06	4.80%	*GRMZM2G143235*	*Trans*	Cytochrome P450
chr3.S_8535639	C22:0/C24:0		3	8535639	A/G	0.05	1.83E-06	9.62%	*GRMZM2G353444*	*Trans*	Phospholipase A1
chr3.S_9950783	C18:2/C18:3		3	9950783	A/G	0.17	4.29E-06	5.74%	*GRMZM2G100260*	*Trans*	D-Tyr-Trunknown(Tyr) deacylase family protein
vSNP	Lead trait[a]	Other trait[b]	Chr.	Position[c]	Alleles[d]	MAF[e]	*P*[f]	Vv/Vp[g]	Gene_Id[h]	evQTL type[i]	Function description[j]
chr3.S_32410225	C18:0/C18:1		3	32410225	A/C	0.09	2.07E-07	3.69%	*GRMZM2G081719*	*NS*	Non-specific phospholipase C6
chr3.S_35663463	C16:0/C16:1		3	35663463	A/G	0.18	7.41E-06	2.89%	*GRMZM2G091119*	*Trans*	Importin subunit alpha
chr3.S_36226527	C18:0/C20:0		3	36226527	A/G	0.40	8.40E-06	3.73%	*GRMZM2G161658*	*Trans*	Epoxide hydrolase 2
chr3.S_130093718	C22:0/C24:0		3	130093718	T/C	0.07	9.73E-06	7.35%	*GRMZM2G022298*	*Trans*	F-Box hamily protein
SYN24171	Oil		3	135285424	A/G	0.15	5.82E-07	9.77%	*GRMZM2G004988*	*Trans*	Transcription coactivator activity
chr3.S_161573977	Oil	C18:2/C18:3	3	161573977	A/G	0.11	3.54E-06	7.37%	*GRMZM2G145346*	*Trans*	Unknown
chr3.S_166690078	Oil		3	166690078	A/T	0.09	4.86E-06	5.94%	*GRMZM2G176542*	*Trans*	Sn1-specific diacylglycerol lipase

chr3.S_169316286	C20:0/C20:1	3	169316286	T/C	0.14	5.71E-06	3.58%	*GRMZM2G346342*	*NS*	Mitogen-activated protein kinase 9
chr3.S_221918315	Oil	3	221918315	T/C	0.12	1.04E-06	7.60%	*GRMZM2G111123*	*NS*	B3 Domain-Containing Protein
chr3.S_224499613	C16:0/C16:1	3	224499613	A/C	0.08	2.03E-06	2.85%	*GRMZM2G137961*	*Trans*	Acyl-CoA N-acyltransferases superfamily protein
chr3.S_232019079	C16:0/C16:1	3	232019079	A/G	0.12	3.64E-06	3.27%	*GRMZM2G060811*	*Trans*	Unknown
chr4.S_2663528	C18:0/C18:1	4	2663528	A/C	0.12	7.36E-06	2.89%	*GRMZM2G106389*	*Trans*	Cysteine-rich receptor-like protein kinase 8
chr4.S_6601726	Oil	4	6601726	A/G	0.13	5.99E-06	7.04%	*GRMZM2G133675*	*Trans*	Transcription factor bHLH47
PZE-104040568	Oil	4	54552245	A/C	0.13	4.62E-06	7.87%	*GRMZM2G098496*	*Trans*	NSF attachment protein involved in vesicular transport
chr4.S_132404834	Oil	4	132404834	T/C	0.13	5.35E-06	8.21%	*GRMZM5G868917*	*Trans*	Unknown
chr4.S_178042468	C22:0/C24:0	4	178042468	T/G	0.09	5.06E-06	8.64%	*GRMZM2G158811*	*Trans & Cis*	Unknown
chr4.S_205809330	C16:1	4	205809330	A/G	0.06	7.25E-06	4.20%	*GRMZM2G103013*	*Trans*	Unknown
chr4.S_224911511	Oil	4	224911511	A/G	0.08	1.53E-06	7.04%	*GRMZM2G048733*	*Trans*	Abscisic acid receptor PYL9
chr4.S_228013669	Oil	4	228013669	T/C	0.12	2.95E-06	6.96%	*GRMZM2G092321*	*Trans*	Unknown
chr4.S_229539871	SFA/USFA	4	229539871	A/C	0.16	3.75E-06	2.35%	*GRMZM2G040452*	*Trans & Cis*	Pyruvate dehydrogenase phosphatase (PDPC)

chr5.S_1491470	C22:0/C24:0		5	1491470	A/T	0.07	7.82E-06	4.60%	*GRMZM2G125271*	*Trans*	Ribosomal Protein S4
vSNP	Lead trait[a]	Other trait[b]	Chr.	Position[c]	Alleles[d]	MAF[e]	*P*[f]	Vv/Vp[g]	Gene_Id[h]	evQTL type[i]	Function description[j]
chr5.S_17891972	Oil		5	17891972	C/G	0.16	4.90E-06	8.85%	*AC194158.3_FG005*	*Trans*	Fatty acid synthase
PZE-105079733	C18:0/C18:1		5	91265684	A/C	0.12	9.02E-06	2.37%	*GRMZM2G119571*	*Trans*	Autophagy-related protein 11
chr5.S_138158699	C16:1		5	138158699	A/G	0.09	1.88E-06	4.59%	*GRMZM2G109315*	*Trans*	Vacuolar protein sorting-associated protein 29
PZE-105128434	C20:0/C22:0		5	185645899	A/G	0.34	5.45E-06	4.94%	*GRMZM2G075255*	*Trans*	Fatty acid hydroxylase
chr6.S_73204004	C18:2/C18:3		6	73204004	A/G	0.09	3.80E-06	5.58%	*GRMZM2G062638*	*Trans*	ATP-dependent peptidases
chr6.S_104862142	Oil		6	104862142	A/G	0.13	4.00E-06	5.44%	*GRMZM2G169089*	*Trans*	DGAT1-2
chr6.S_104865747	C20:0/C22:0		6	104865747	A/C	0.24	9.64E-06	5.16%	*GRMZM2G169089*	*Trans*	DGAT1-2
chr6.S_138872466	C24:0		6	138872466	A/G	0.50	6.60E-07	3.25%	*GRMZM2G069713*	*Trans*	Probable protein phosphatase 2C 73
chr6.S_141864218	C22:0/C24:0		6	141864218	A/G	0.08	2.20E-06	9.09%	*GRMZM2G023105*	*Trans*	Putative VHS/GAT domain containing family protein
chr7.S_10514965	Oil		7	10514965	A/G	0.09	9.90E-06	6.32%	*GRMZM2G066290*	*Trans*	Pyruvate kinase
chr7.S_145764735	C20:0/C20:1		7	145764735	C/G	0.08	6.49E-06	5.42%	*GRMZM2G006416*	*Trans*	Probable Protein Phosphatase 2C 21
chr7.S_173072186	C22:0/C24:0		7	173072186	T/G	0.29	4.18E-06	8.35%	*GRMZM5G890815*	*Trans*	Unknown
chr8.S_1145487	C16:0		8	1145487	A/G	0.05	1.74E-06	3.03%	*GRMZM2G063244*	*Trans*	Peptidyl-prolyl cis-trans isomerase
chr8.S_8102492	C22:0/C24:0		8	8102492	A/T	0.11	4.92E-06	5.61%	*GRMZM2G035341*	*Trans*	Ring-box protein 1A

chr8.S_8615756	C22:0/C24:0		8	8615756	T/C	0.06	5.58E-06	9.82%	*GRMZM2G098015*	*Trans*	Unknown
chr8.S_9901457	C16:0/C16:1		8	9901457	A/G	0.07	1.10E-06	3.11%	*GRMZM2G095757*	*Trans*	CDP-diacylglycerol--serine O-phosphatidyltransferase 1
chr8.S_17422251	C18:1/C18:2		8	17422251	C/G	0.06	1.99E-06	5.02%	*GRMZM2G061969*	*Trans*	Phospholipase D2
chr8.S_66989244	Oil		8	66989244	A/G	0.11	1.50E-06	6.72%	*GRMZM2G156606*	*Trans*	Duf1639 Family Protein
chr8.S_111626169	C20:0/C22:0		8	111626169	T/C	0.10	7.36E-06	3.46%	*GRMZM2G416308*	*Trans*	Proline-rich receptor-like protein kinase PERK1
chr8.S_153259338	C18:2/C18:3		8	153259338	A/G	0.15	4.96E-06	6.51%	*GRMZM2G157043*	*Trans*	CRAL-TRIO lipid binding domain
chr8.S_153413361	C18:2/C18:3		8	153413361	T/G	0.14	9.06E-06	6.49%	*GRMZM2G469901*	*Trans & Cis*	Unknown
chr8.S_155978664	Oil		8	155978664	T/C	0.13	6.95E-06	6.67%	*GRMZM2G107570*	*Trans*	Oil Body-Associated Protein
chr8.S_161745997	C24:0		8	161745997	T/C	0.36	5.01E-06	3.69%	*GRMZM2G180335*	*Trans*	Dynamin-related protein 3A
vSNP	Lead trait[a]	Other trait[b]	Chr.	Position[c]	Alleles[d]	MAF[e]	*P*[f]	Vv/Vp[g]	Gene_Id[h]	evQTL type[i]	Function description[j]
chr8.S_166782652	C18:2		8	166782652	T/C	0.36	8.97E-06	2.40%	*GRMZM5G805609*	*Trans & Cis*	Glucan endo-13-beta-glucosidase 14
chr8.S_170708353	C20:0		8	170708353	A/G	0.09	3.04E-06	4.06%	*GRMZM5G805026*	*Trans*	Wuschel-related homeobox 13
chr9.S_17668908	C18:1/C18:2		9	17668908	C/G	0.14	5.15E-06	6.35%	*GRMZM2G071249*	*Trans*	Lipase
chr9.S_103956920	C16:0		9	103956920	T/C	0.16	1.94E-06	3.72%	*GRMZM2G054093*	*Trans*	Unknown
chr9.S_140343507	C20:0/C20:1		9	140343507	T/C	0.09	2.34E-07	5.61%	*GRMZM2G179336*	*Trans & Cis*	Duf3527 domain protein
chr9.S_141480266	C20:0/C20:1		9	141480266	T/G	0.06	3.02E-06	4.97%	*GRMZM2G146386*	*Trans & Cis*	ERAD-associated E3 ubiquitin-protein ligase

											component HRD3A
chr9.S_146459251	C18:1	C18:2, C18:1/C18:2	9	146459251	A/G	0.16	1.77E-06	3.01%	*GRMZM2G024718*	*Trans*	Heat shock 70 kDa protein 8
chr9.S_153916991	C22:0/C24:0		9	153916991	A/C	0.04	1.92E-06	11.03%	*GRMZM2G116681*	*Trans*	THO complex subunit 7B
chr10.S_16212078	Oil		10	16212078	A/G	0.09	1.65E-06	5.82%	*GRMZM2G181251*	*Trans*	Transcription factor binding
chr10.S_24577806	Oil		10	24577806	C/G	0.10	4.22E-06	5.90%	*GRMZM2G153206*	*Trans*	Rapid ALkalinization Factor
chr10.S_26138064	C18:0/C18:1		10	26138064	A/T	0.22	7.40E-06	4.18%	*GRMZM2G129457*	*Trans*	DNA-directed RNA polymerases
chr10.S_147448589	C22:0/C24:0		10	147448589	T/C	0.09	9.63E-07	8.82%	*GRMZM2G464157*	*Trans & Cis*	Phospholipase
chr10.S_148232465	C16:1		10	148232465	T/C	0.31	4.02E-06	3.66%	*GRMZM2G008714*	*NS*	Pyruvate kinase

[a]The oil trait associated with the highest *P* value among 21 oil-related traits. [b]Additional oil traits associated at $P < 1\times10^{-5}$. [c]Position in base pairs of the vSNP according to version 5b.60 of the maize reference sequence (http://ensembl.gramene.org/Zea_mays/Info/Index). [d] Minor allele/Major allele. [e]Minor allele frequency (MAF) for vSNP in this association mapping population. [f]*P* probability of vSNP associated with variance heterogeneity of the lead trait. [g]*Vv* is the variance due to heterogeneity between genotypes, and *Vp* is the phenotypic variance accounted for. [h]Gene identification of a plausible biological candidate gene in the locus or the nearest annotated gene to the lead vSNP according to version 5b.60 of the maize reference sequence (MaizeSequence, see URLs). [i]evQTL type *Cis* indicates that SNPs located within 100 kb region (50 kb upstream and downstream) of the candidate gene are significantly associated with the gene expression variance heterogeneity of this gene; *Trans* indicates that SNPs located outside the 100 kb region of the candidate gene are significantly associated with the gene expression variance heterogeneity of this gene; NS, not significant ($P > 1.0\times10^{-5}$). [j]Each candidate gene is annotated according to InterProScan

(http://www.ebi.ac.uk/interpro/).

Table 2. *P* values for 32 vSNPs associated with variance heterogeneity after adjusting for the mean-effect SNPs.

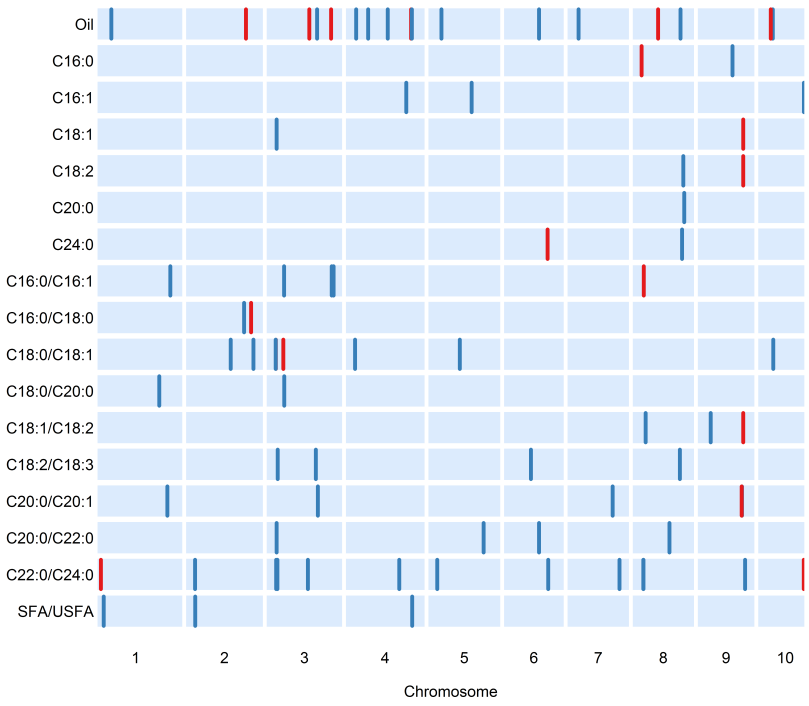
Trait	vSNP	Mean-effect SNP[a]	Mean-effect SNP_MAF[b]	-log$_{10}$[$P_{(dispersion)}$_before][c]	-log$_{10}$[$P_{(dispersion)}$_after][d]	-log$_{10}$[$P_{(dispersion)}$_difference][e]	-log$_{10}$*P*_mean[f]
Oil	chr1.S_53423512	chr1.S_55071145	0.09	5.31	2.01	3.30	5.88
C18:0/C18:1	chr2.S_144072332	chr2.S_43185558	0.13	5.25	3.96	1.28	5.03
Oil	chr2.S_204388944	chr2.S_149517635	0.06	6.07	2.43	3.64	6.33
Oil	chr2.S_204470447	chr2.S_149517635	0.06	6.08	2.36	3.72	5.23
Oil	chr3.S_161573977	chr3.S_178136002	0.07	5.46	1.60	3.86	6.07
Oil	chr3.S_166690078	chr3.S_167431166	0.08	5.31	1.58	3.73	5.95
C20:0/C20:1	chr3.S_169316286	chr3.S_9862488	0.46	5.24	3.40	1.84	5.10
C18:0/C18:1	chr3.S_1852151	chr3.S_158895417	0.08	5.39	3.79	1.61	6.45
Oil	chr3.S_221918315	chr3.S_166664152	0.07	5.98	1.73	4.25	6.34
C18:0/C18:1	chr3.S_32410225	chr3.S_156963535	0.06	6.68	5.24	1.45	6.26
C18:1	chr3.S_5568273	chr3.S_1552666	0.07	5.07	3.30	1.77	5.07
Oil	chr4.S_132404834	chr4.S_236185943	0.06	5.27	1.30	3.97	8.93

Oil	chr4.S_224911511	chr4.S_6601732	0.07	5.82	1.19	4.63	6.74
Oil	chr4.S_228013669	chr4.S_236185943	0.06	5.53	1.51	4.02	8.93
SFA/USFA	chr4.S_229539871	chr4.S_191765041	0.49	5.43	3.90	1.52	6.18
C18:0/C18:1	chr4.S_2663528	chr4.S_162256670	0.46	5.13	3.31	1.82	5.05
Oil	chr4.S_6601726	chr4.S_6601732	0.06	5.22	2.01	3.21	8.93
Oil	chr5.S_17891972	chr5.S_15800012	0.06	5.31	1.72	3.59	12.42
Oil	chr6.S_104862142	chr6.S_104848924	0.10	5.40	3.15	2.25	11.75
C20:0/C22:0	chr6.S_104865747	chr6.S_104858442	0.15	5.02	2.86	2.16	19.67
C24:0	chr6.S_138872466	chr6.S_104865691	0.17	6.18	4.80	1.38	9.39
Oil	chr7.S_10514965	chr7.S_9794647	0.05	5.00	1.20	3.81	5.16
C20:0/C22:0	chr8.S_111626169	chr8.S_38520871	0.06	5.13	1.76	3.38	5.04
C18:2/C18:3	chr8.S_153259338	chr8.S_34664222	0.16	5.30	4.27	1.04	5.03
C18:2/C18:3	chr8.S_153413361	chr8.S_34664222	0.16	5.04	4.05	1.00	5.03
Oil	chr8.S_155978664	chr8.S_21615641	0.07	5.16	0.23	4.93	11.57
C18:2	chr8.S_166782652	chr8.S_113302105	0.09	5.04	3.85	1.19	7.01

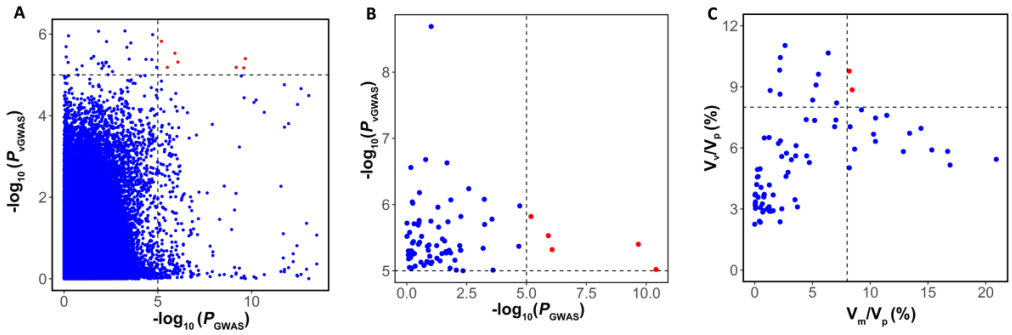
Trait	vSNP	Mean-effect SNP[a]	Mean-effect SNP_MAF[b]	$-\log_{10}[P_{(dispersion)}\_before]$[c]	$-\log_{10}[P_{(dispersion)}\_after]$[d]	$-\log_{10}[P_{(dispersion)}\_difference]$[e]	$-\log_{10}P\_mean$[f]

C20:0	chr8.S_170708353	chr8.S_118559286	0.18	5.52	3.97	1.55	5.30
Oil	chr8.S_66989244	chr8.S_38489921	0.06	5.82	0.48	5.34	13.47
C16:0	chr9.S_103956920	chr9.S_20476304	0.14	5.71	2.00	3.71	10.97
Oil	chr10.S_16212078	chr10.S_16487751	0.06	5.78	0.64	5.14	6.81
Oil	chr10.S_24577806	chr10.S_19069742	0.06	5.37	1.11	4.27	6.35

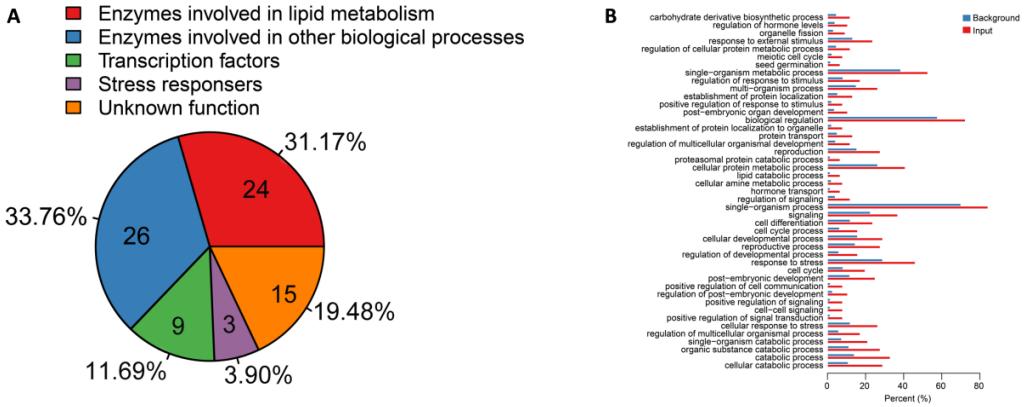
[a]Mean-effect SNP which can explain the variance heterogeneity for the vSNP. [b]Minor allele frequency (MAF) for mean-effect SNP in our maize association population; MAF for the primary mean-effect SNP/MAF for the second mean-effect SNP. [c]Negative $\log_{10}$ form of *P* value for vSNP associated with variance heterogeneity of the target trait before adjusting for the mean-effect SNPs. [d]Negative $\log_{10}$ form of *P* value for vSNP associated with variance heterogeneity of the target trait after adjusting for the primary mean-effect SNP or negative $\log_{10}$ form of *P* value for vSNP associated with variance heterogeneity of the target trait after adjusting for the primary and second mean-effect SNPs. [e]Negative $\log_{10}$ form of *P* value difference for vSNP of the target trait before and after adjusting for the significant mean-effect SNPs. [f]Negative $\log_{10}$ form of *P* value for mean-effect SNP associated with mean value difference of the target trait.



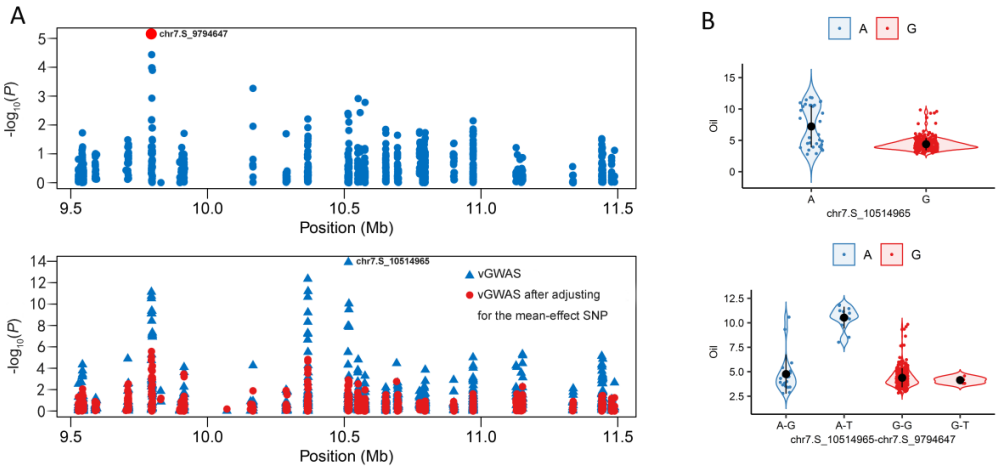
tpj_14786_f1.tif



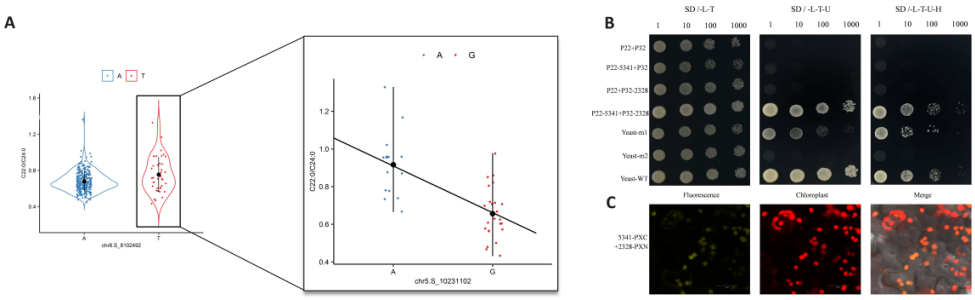
tpj_14786_f2.tif



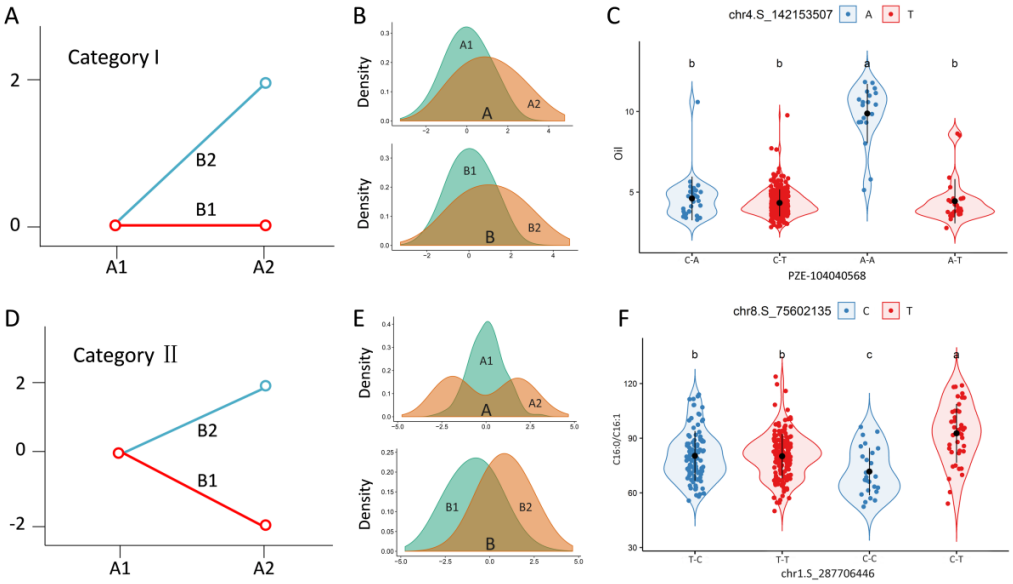
tpj_14786_f3.tif



tpj_14786_f4.tif



tpj_14786_f5.tif



tpj_14786_f6.tif