Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2021), pp. 1–3 doi: 10.1111/pbi.13634

Brief Communication

Mining novel kernel size-related genes by pQTL mapping and multi-omics integrative analysis in developing maize kernels

Qingqian Zhou[1,†], Zhiyuan Fu[1,*,†], Haijun Liu[2], Jingen Wang[1], Zhanyong Guo[1], Xuehai Zhang[1], Runmiao Tian[1], Yang Liu[1], Jianzhou Qu[3], Weihua Li[1], Jianbing Yan[2,*] and Jihua Tang[1,*]

[1]*Key Laboratory of Wheat and Maize Crops Science/Collaborative Innovation Center of Henan Grain Crops/College of Agronomy, Henan Agricultural University, Zhengzhou, China*

[2]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China*

[3]*The Key Laboratory of Biology and Genetic Improvement of Maize in Arid Area of Northwest Region, College of Agronomy, Northwest A&F University, Yangling, China*

Received 10 April 2021;
revised 8 May 2021;
accepted 10 May 2021.
**Correspondence* (Tel +86-037-56990336; fax +86 037 56990188; email fuzhiyuan2004@163.com (Z.F.); Tel +86-027-87280110; fax +86 27 87384670; email yjianbing@mail.hzau.edu.cn (J.Y.); Tel +86-037-56990336; fax +86 037 56990188; e-mail tangjihua1@163.com (J.T.))
[†]These authors contributed equally to this work.

Keywords: proteomics, protein quantitative trait locus, multi-omics integration, kernel-related traits, maize.

As a sink organ for starch, protein, oil and essential nutrients, the maize (*Zea mays*) kernel is not only the main target for yield and quality improvement but also a model system for genetic and molecular biology studies. We identified many candidate genes for maize kernel quality and size quantitative trait loci (QTLs) at the genomic, transcriptomic, metabolomic and phenomic levels by genome-wide association studies (GWAS) and joint-linkage mapping (Fu *et al*., 2013; Liu *et al*., 2017b; Liu *et al*., 2017a; Wen *et al*., 2014; Yang *et al*., 2014) using a widely adopted Chinese association panel (Yang *et al*. 2011) and five recombinant inbred line (RIL) populations (Liu *et al*., 2017b). However, maize kernel proteomics studies at the population scale have lagged behind.

Protein QTL (pQTL) analysis has proven to be useful in the diagnosis of various human diseases and has provided genome–proteome networks for clinical applications (Suhre *et al*., 2017). It is also necessary for elucidating the functional context of gene expression variation during modern maize breeding (Jiang *et al*., 2019). However, how pQTLs control maize kernel traits remain to be investigated. Here, we identified 468 clear and consistent protein spots in developing kernels of 210 inbred lines by combining 2-D gel electrophoresis with LC-MS/MS. These protein spots were translated from 283 unigenes, 84 of which encode proteins with post-transcriptional/translational modifications. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis indicated that 54% of the identified proteins were annotated and enriched in carbohydrate metabolism (13%), amino acid metabolism (16%) and genetic information processing (18%) (Figure 1a). Interestingly, 46% of the identified proteins could not be assigned to known KEGG pathways, suggesting that a considerable number of proteins in developing maize kernels are uncharacterized. A total of 297 protein spots were successfully retrieved and found to differentially accumulate among inbred lines. We generated regulatory networks based on the pairwise Pearson correlation coefficients ($r^2 \geq 0.25$, $P \leq 0.05$) of transcript and protein abundance using the modularity method implemented in Gephi0.9.2 (Figure 1b, c). Analysis with both networks revealed that genes involved in the same pathway do not always appear in the same module. Additionally, we noticed moderate or low correlations between the transcript level and the protein abundance of the same gene; examples included several previously reported genes for kernel development (Dai *et al*., 2021), GRMZM2G068506 (*Bt2*), GRMZM2G429899 (*Sh2*), GRMZM2G089713 (*Sh1*), GRMZM2G415359 (*Mdh4*), GRMZM2G306345 (*Pdk1*) and GRMZM2G097457 (*Pdk2*), among which only *Bt2* exhibited a strong correlation between the transcript and protein levels ($r = 0.78$, $P < 0.01$). These genes clustered into different subnetworks at the transcript level (Figure 1b) but into the same subnetwork at the protein level (Figure 1c). The results reveal that the transcript level alone does not always reliably predict protein abundance at the population scale, and protein abundance variation may play an important role in orchestrating the biological functions of genes involved in the same biological pathways. pQTL analysis is therefore necessary to fully elucidate the molecular basis of kernel-related phenotypes.

Using GWAS based on 1.25 M SNPs, we identified 421 independent significant SNPs for the abundance of 40 protein spots encoded by 38 unigenes using the recommended *P*-value (1/N, $P \leq 2.04 \times 10^{-6}$). Forty-six non-redundant pQTLs were defined within +/− 50 kb flanking regions of their lead SNPs based on the linkage disequilibrium of 50 kb ($r^2 \geq 0.1$) in genome-wide average of 210 inbred lines. These included 13 local pQTLs and 33 distant pQTLs that distributed unevenly across the ten maize chromosomes (Figure 1d). Chromosome 7 and 2 had the lowest and highest density of pQTLs, respectively

Please cite this article as: Zhou, Q., Fu, Z., Liu, H., Wang, J., Guo, Z., Zhang, X., Tian, R., Liu, Y., Qu, J., Li, W., Yan, J. and Tang, J. (2021) Mining novel kernel size-related genes by pQTL mapping and multi-omics integrative analysis in developing maize kernels. *Plant Biotechnol J*., https://doi.org/10.1111/pbi.13634

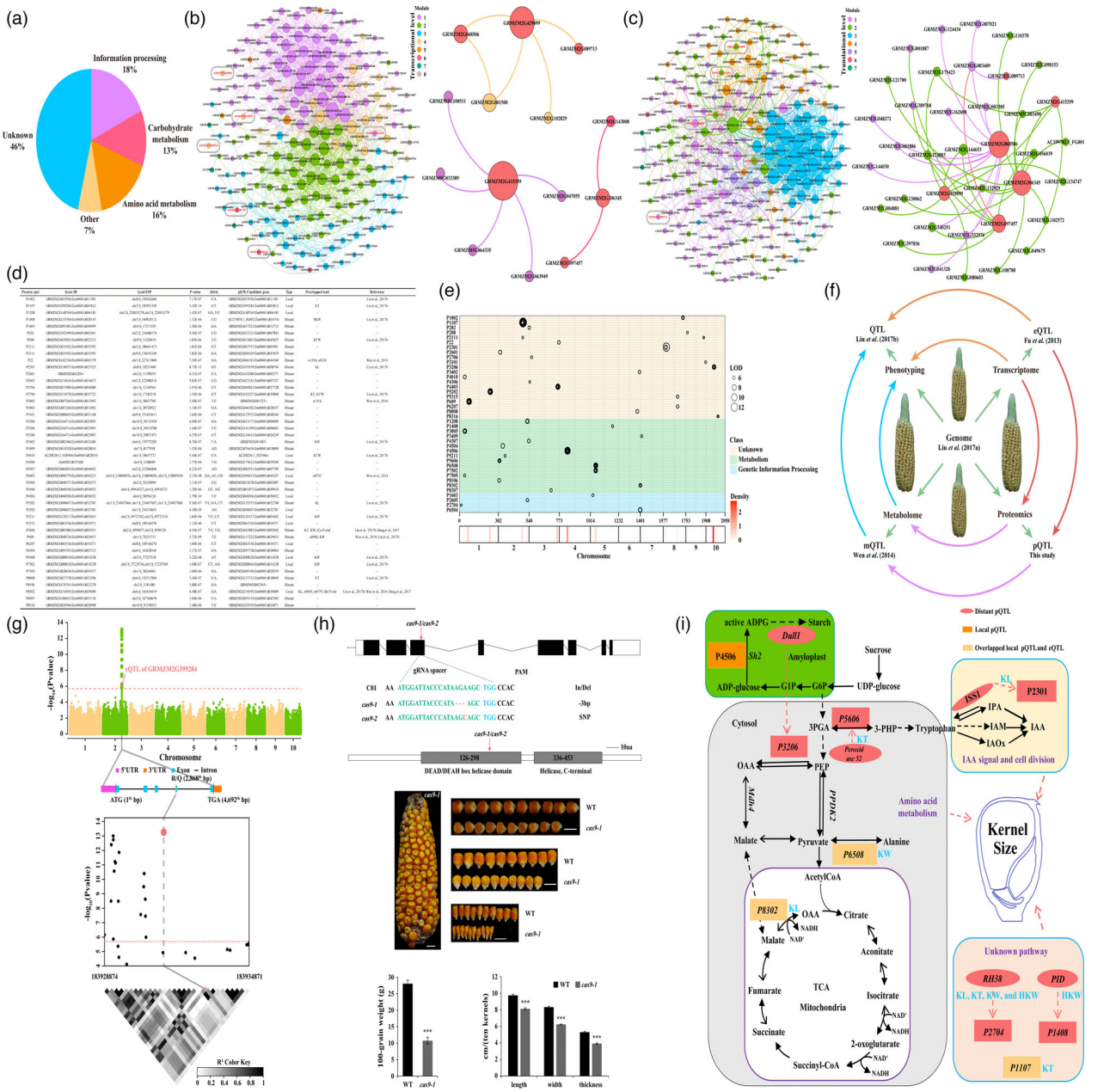

Figure 1 pQTL identification and multi-omics integration analysis in developing maize kernels. (a) Functional annotation of the identified protein spots. (b–c) Networks constructed based on the protein abundance of 297 protein spots and the expression levels of their corresponding transcripts. (d) The 46 pQTLs identified here and previously identified QTLs for kernel-related traits that coincided with these pQTLs. (e) Chromosomal distribution of the pQTLs. The *x*-axis indicates the physical positions (Mb) of the pQTLs across ten maize chromosomes, and the heatmap shows the density of these pQTLs across the maize genome (window size = 10 Mb). (f) Flowchart of data analysis. (g) Candidate gene-based association mapping and pairwise linkage disequilibrium analysis of the local pQTL for P1107. GRMZM2G399284 is on the reverse strand, and the bar chart shows differences in protein abundance associated with the lead SNP. (h) Validation of GRMZM2G000823 by genetic transformation in KN5585. The top panel shows two T_1 independent transgenic lines. The middle panel shows the segregation of the 3-bp deletion on T_2 ear. Scale bar, 1 cm. The lower panel is the statistic analysis of kernel size and hundred-kernel weight between WT and mutant kernels. ***$P < 0.001$, Student's *t*-test, three biological replicates. (i) A network for maize kernel formation constructed based on multi-omics integration.

(Figure 1e). Five protein spots, P3206, P4506, P3005, P5202 and P2111, were underlain by two or more pQTLs, and the remaining protein spots by only one pQTL. Two pQTLs were found to regulate proteins that function in post-transcriptional/translational modifications (P5315 and P6207, P6508 and P7502).

We performed population-scale multi-omics integration to achieve a comprehensive understanding of the mechanistic basis of maize kernel development (Figure 1f). Eighty-seven previously reported eQTLs (Liu *et al.*, 2017a) were found to regulate the transcript levels of genes that encode 30 protein spots, whereas co-localization for eQTL and pQTL was only observed for protein

spots P1107, P1208, P6508, P8302 and P5211. We then compared the identified pQTLs with previously reported quality QTLs and normal QTLs for kernel (Liu *et al*., 2017b; Wen *et al*., 2014; Yang *et al*., 2014). Seventeen pQTLs coincided with ten QTLs from the association panel and ten QTLs from the five RIL populations (Figure 1d). In only four cases did a phenotypic QTL co-segregate with its eQTL and pQTL. For example, the abundance of P6508 and P7502 proteins significantly correlated with their transcript levels, and their pQTL co-localized with a kernel width (KW) QTL (LOD = 7.03, R^2 = 9.37%) identified in the ZHENG58 × SK RIL population. The pQTL for P8302 coincided with its local eQTL, a QTL for (Glutamine and Glutamic acid)/total amino acid levels, and with an mQTL for metabolites n0565 and n0579 from the association panel, as well as with a kernel length (KL) QTL (LOD = 3.19, R^2 = 4.81%) from the ZHENG58 × SK RIL population. The P1107's pQTL coincided with a local eQTL and a kernel thickness (KT) QTL (LOD = 3.16, R^2 = 5.11%) from the BY815 × KUI3 RIL population (Figure 1g). The abundance of P5211 significantly correlated with its local eQTL, which co-localized with a previously reported KW QTL (LOD = 4.28, R^2 = 7.22%) from the K22 × CI7 RIL population. Taken together, these results demonstrate the power of multi-omics integration to uncover novel relationships between genetic variations and maize kernel traits.

However, the correlation among different levels is low because the flow of information from DNA to phenotype is a signal propagation process. Therefore, this integrative strategy does not apply to functionally unrelated QTLs that co-segregate, such as distant pQTLs. One such example is the candidate gene for P2704's distant QTL (GRMZM2G000823), which encodes DEAD-box ATP-dependent RNA helicase 38; it did not coincide with any of the previously reported kernel size QTLs. We validated this gene by CRISPR–Cas9 genetic transformation (Figure 1h), and the T_2 ears derived from a positive heterozygous T_1 line that carried a 3-bp deletion causing a lysine (K) deletion displayed a segregation ratio of roughly 3:1 for normal to small kernels ($\chi^2 = 0.06 < \chi^2_{0.05} = 3.84$). The 3-bp deletion was confirmed to be causal because all T_2 kernels with a mutant phenotype were found to carry the 3-bp deletion (–/–), whereas normal T_2 kernels exhibited either heterozygous (+/–) or homozygous WT (+/+) genotypes. In addition, kernel size components, including KL, KW and KT, all significantly reduced in mutant kernels compared with WT kernels harvested from the same ear, with HKW showing a nearly 50% reduction. These results substantiate the link between GRMZM2G000823 and kernel size variation and demonstrate the usefulness of pQTL analysis for the identification of novel candidate genes for kernel-related traits.

Altogether, our integrated multi-omics approach involving pQTL analysis adds a new layer of functional context to protein abundance variation in maize kernel formation. Our results suggest that genes involved in amino acid metabolism, signal transduction and a yet unknown mechanism may regulate maize kernel size (Figure 1i).

Acknowledgements

This research was supported by the National Natural Science Foundation of China (U2004144 and 31971893), the Key Technologies R & D Program of Henan Province (212102110061 and 202102110012), and the China Scholarship Council (201808410096).

Conflicts of interest

This study did not involve human participants and/or animals. The authors have no conflict of interest to declare.

Author contributions

Z.F., J.T. and J.Y. designed and supervised this study. Z.F., Q.Z., J.W., R.T. and Y.L. performed the experiments. Z.F., H.L., Z.G., X.Z., J.Q. and W.L. analysed the data. Z.F. wrote the manuscript with input from all authors.

REFERENCES

Dai, D., Ma, Z. and Song, R. (2021) Maize kernel development. *Mol. Breed.* **41**, 2.

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

Jiang, L.G., Li, B., Liu, S.X., Wang, H.W., Li, C.P., Song, S.H., Beatty, M. *et al*. (2019) Characterization of proteome variation during modern maize breeding. *Mol. Cell Proteomics*, **18**, 263–276.

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

Liu, J., Huang, J., Guo, H., Lan, L., Wang, H., Xu, Y., Yang, X. *et al*. (2017b) The conserved and unique genetic architecture of kernel size and weight in maize and rice. *Plant Physiol.* **175**, 774–785.

Suhre, K., Arnold, M., Bhagwat, A.M., Cotton, R.J., Engelke, R., Raffler, J., Sarwath, H. *et al*. (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J. *et al*. (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**, 3438.

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W. *et al*. (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., Zhang, Z., Prasanna, B.M., Li, L., Li, J. *et al*. (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* **28**, 511–526.

