Plant Physiology and Biochemistry 141 (2019) 380–387

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Research article

Investigation and development of maize fused network analysis with multi-omics

Jing Jiang[a], Fei Xing[a], Chunyu Wang[b], Xiangxiang Zeng[c,**], Quan Zou[d,*]

[a] School of Aerospace Engineering, Xiamen University, Xiamen, 361001, China
[b] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
[c] School of Information Science and Engineering, Hunan University, 410082, Changsha, China
[d] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, 610000, China

ARTICLE INFO

Keywords:
Maize
Omics
Yield
Fused network
Orphan

ABSTRACT

Maize is a critically important staple crop in the whole world, which has contributed to both economic security and food in planting areas. The main target for researchers and breeding is the improvement of maize quality and yield. The use of computational biology methods combined with multi-omics for selecting biomolecules of interest for maize breeding has been receiving more attention. Moreover, the rapid growth of high-throughput sequencing data provides the opportunity to explore biomolecules of interest at the molecular level in maize. Furthermore, we constructed weighted networks for each of the omics and then integrated them into a final fused weighted network based on a nonlinear combination method. We also analyzed the final fused network and mined the orphan nodes, some of which were shown to be transcription factors that played a key role in maize development. This study could help to improve maize production via insights at the multi-omics level and provide a new perspective for maize researchers. All related data have been released at http://lab.malab.cn/~jj/maize.htm.

1. Introduction

Maize (*Zea mays*) is an important staple food in many regions of the world with total production higher than other grain crops, such as wheat or rice. Despite its status as a food staple in many areas, most maize is used for animal feed and ethanol fuel. Maize is also an important model organism for plant development, physiology, and genetics studies in addition to its economic value. The genome of maize is approximately 2.4 gigabases, and the haploid chromosome number is 10 (Schnable et al., 2009; Zhang et al., 2009). Fig. 1 shows the 10 haploid chromosomes.

Maize is widely cultivated globally, thus it is necessary to explore the factors that influence its production, such as disease tolerance, drought, pests, and nutrition, as well as yield. Owing to the importance of exploring these factors to reveal how they react at the biomolecular level in maize through multi-omics, these issues have already investigated by some researchers. The DNA sequence data of the B73 maize genome is stored in GenBank, a consortium jointly established by the Department of Energy, US National Science Foundation, and Department of Agriculture in 2005. To study plant proteomics, protein extraction, involving 2DE-based gel maps, proteomic analysis and MS analysis, is required. Experimental data can reveal correlation between phenotype and omics, such as proteomics, to some extent, but this approach is expensive and often the role of biomolecules in the development of maize is difficult to discern. Thus, it is feasible to examine the biomolecule that influence maize production using one single omics.

In this context, combining multi-omics data to study the biomolecules of maize is a challenge. We showed how current methods could be successfully applied through omics technology to influence the yield of maize. In this paper, we use multi-omics data to construct weighted networks separately and integrated into one final fused network to provide insight into maize development, which made the results more reliable. We also analyzed the final fused network and mined the orphan nodes, some of which played a key role in maize development. These results could provide a new perspective for maize breeding and researchers.

* Corresponding author.
** Corresponding author.
E-mail addresses: xzeng@xmu.edu.cn (X. Zeng), zouquan@nclab.net (Q. Zou).

https://doi.org/10.1016/j.plaphy.2019.06.016
Received 20 February 2019; Received in revised form 12 June 2019; Accepted 12 June 2019
Available online 15 June 2019

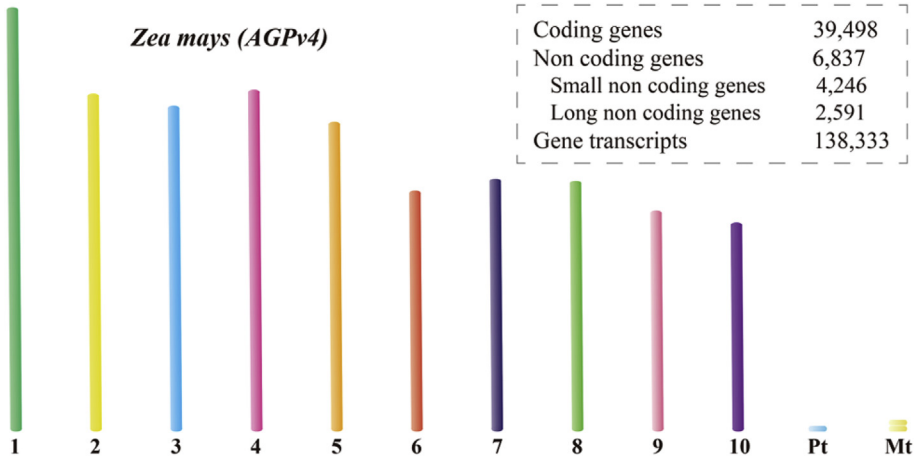

Fig. 1. The chromosomes of Zea mays (AGPv4). The lengths of the bars represent the chromosome lengths and the numbers are corresponding to the chromosome number.

2. Materials and methods

The accessibility of diverse omics data and the rapid improvement of analytical methods have enabled systems biology approaches to be extended directly to crop plant systems. While a single technique may be informative on its own, using multiple complementary approaches, such as a systems biology approach integrating multi-omics data (transcriptomic, metabolomic, genomic, and proteomic data) collected from the same plant material, will strengthen the overall analysis. Here, multi-omics technologies of each data type of maize were applied based on the hypothesis that the separate datasets may lack some information and their fusion may make the data more comprehensive (Fig. 2).

2.1. Genome

From a genetic standpoint, genome-wide association studies (GWAS), which study genetic variation genome-wide among disparate individuals, aim to elucidate genetic variants that are associated with a trait. The first successful GWAS was reported in 2005 (Klein et al., 2005; MacArthur et al., 2017; Visscher et al., 2012). Through the development of high-throughput sequencing technology, the number of publications on maize has increased over time. Along with the release of the B73 reference genome of maize and its wide use, GWAS has become a useful tool for effectively determining the relationships between genes and phenotypes (Schnable et al., 2009). The increase in the number of publications associated with the key words “maize” and “GWAS” is shown in Fig. 3.

Collections of gene expression profiles on the whole-genome scale are useful for gene discovery and functional characterization in metabolic pathways (Usadel et al., 2009; Marcotte et al., 1999), but these genes are always in uncorrelated expression profiles (Liu et al., 2009). Some studies have shown that the coregulatory pattern of two genes may be affected by genetic variation or expression levels of a third gene (Li et al., 2004, 2007; Li and Yuan, 2004; Li, 2002; Sun et al., 2008; Tai et al., 2010).

2.2. Transcriptome

The transcriptome is the overall set of transcripts in a particular biological context. To explore the development of maize by defining mRNAs and their abundance in tissues or organs is essential. A previous study revealed that combining the transcriptome and gene expression, 91 transcription factors that were important during seed development stages were identified (Chen et al., 2014). Moreover, mRNA clusters in

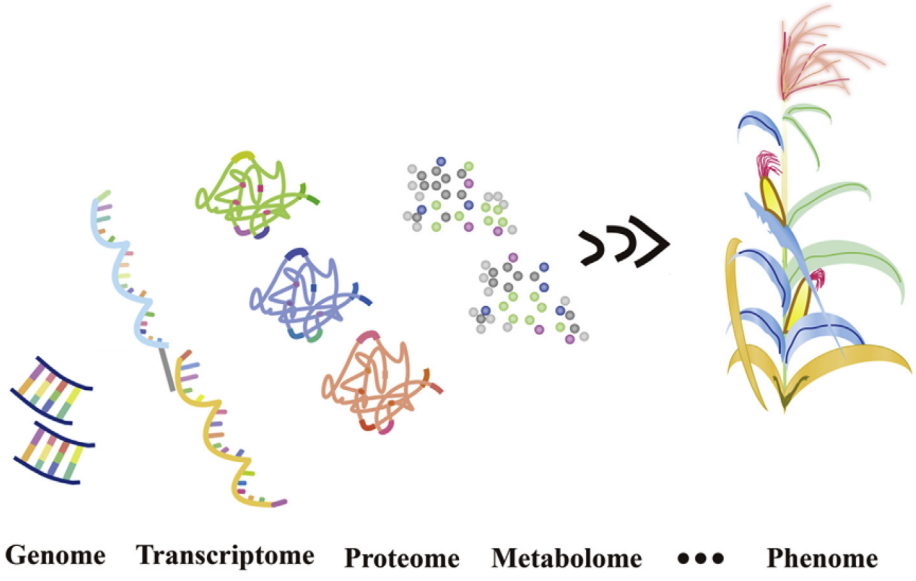

Fig. 2. Graphical representation of each omics in maize.

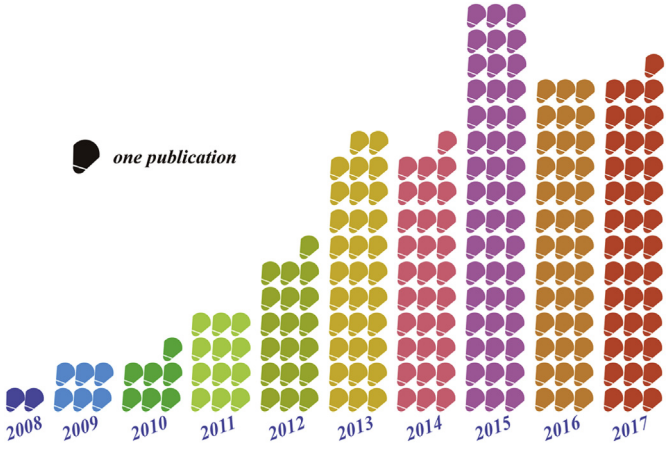

Fig. 3. Increasing number of publications on maize GWAS since the release of the B73 reference genome. The number of publications is represented by the number of maize kernels and the statistics are from the NCBI, obtained by searching PubMed using the terms "maize" and "genome-wide association study."

the embryo and main endosperm cell types in maize were discovered by applying RNA-Seq technology (Zhan et al., 2015). The above efforts revealed that certain mRNAs in maize had key roles during seed development stages.

Another critical factor regulating gene expression in plants is miRNA, which binds to target gene transcripts and inhibits the translation of transcripts or degrades the transcripts, and we can construct a network of ceRNAs, shared miRNA recognition elements, to identify mRNA-associated pathways for maize seed development.

2.3. Proteome

Proteins have a key role in gene function and directly participate in cellular development and metabolism; thus, derived proteomics can be a useful omics measurement (Gong and Wang, 2013). The accuracy of proteomics used in describing the diversity of protein is increasing (Smith et al., 2013; Schluter et al., 2009). Conventionally, high-resolution two-dimensional polyacrylamide gel electrophoresis (2-DE) has been used to investigate the role of proteins with proteomics (Balsamo et al., 2011; Zolla et al., 2008; Vidal et al., 2015; Coll et al., 2011). In recent years, there is a new technology called isobaric tag for relative and absolute quantitation (iTRAQ), which has replaced 2-DE, because it identifies more proteins with more reliable quantitative information and is appropriate for multiple samples (Karp et al., 2010; Schulze and Usadel, 2010; Zi et al., 2013; Ma et al., 2014; Ross et al., 2004; Peng et al., 2003; Washburn et al., 2001).

Proteomic analyses is widely used in describing the response of plants to abiotic stresses (Wu and Wang, 2016); for example, the proteins of *Arabidopsis thaliana* rosette leaves were affected by heat stress treatments or short-term cold (Rocco et al., 2013). Maize is more sensitive to salt stress than other crops (Cui et al., 2015), so improving the salt tolerance of maize is important for breeding to increase its yield.

2.4. Phenomics

The phenotype of an individual is closely related to its genotype and can be affected by environment during its lifetime. Owing to their changing environment, plants have unique mechanisms to respond to abiotic and biotic stress (Bohnert et al., 1995). Artificial selection, which was based on the plant phenotype, keeping the seeds that had better adaptability to complex environments producing high-yield varieties for breeding (Ghalambor et al., 2015; Pigliucci, 2005).

In modern plant breeding, phenotyping is the one of the major bottlenecks, which can be solved in two ways. First, the plant breeders identify lines of interest having the greatest stress tolerance or highest yield in a given environment by analyzing the phenotypes of a large number of lines. Second, the breeders identify the genome region having deleterious or beneficial alleles based on detailed data from combining genotypes and phenotypes from different plants.

For decades, the breeders tried to identify genes related to climate adaptation or phenotypic variation by exploring genetic markers (Liu et al., 2015; Meyer and Purugganan, 2013; Hufford et al., 2012). As a result, the mutation or insertion of transposable elements may be associated with environmental adaptation of maize. Inserting a transposable element in the regulatory region of the tb1 gene affected branch growth and gene expression (Studer et al., 2011). The insertion of transposable element could inhibit the expression of ZmCCT, leading to reduce photoperiod sensitivity and allowing maize to adapt to long-day environments (Yang et al., 2013). The insertion of a miniature transposable element upstream of Vgt1, which is a known gene regulating the flowering time, strongly influenced the early flowering of maize (Castelletti et al., 2014). The above studies on insertion of transposable elements in maize revealed that insertion of transposable elements could influence the phenotype.

2.5. Data collection

The protein-protein interaction data for maize was downloaded in latest version of STRING. The gene interaction data, mRNA interaction data and the regulation association were obtained from Walleye et al. and can be downloaded from our website. For phenomics, profiles spanning the vegetative and reproductive stages of maize development as well as text mining in PubMed were used to validate genetic variation associated with phenotypes that influenced maize yield.

2.6. Construction of weighted network

One gene can be analyzed through its interactions with other genes, forming a co-expression pattern (for example, if they are involved in the same biological process), and this analysis can be extended to maize. The co-expression interactions can be visualized graphically as a weighted network, in which the nodes represent genes and link edges reflect the associations between genes. For each omics network, nodes stand for genes (mRNAs or proteins) and the weighted edges represent pairwise similarities, which can also be represented as a graph of G = (V, E). The vertices V stand for nodes and E corresponds to edges. The n × n adjacency matrix W of each network for W(i, j) indicates the weight between nodes.

To calculate the weighted matrix for each omics data type, it is essential to define a normalized similarity matrix to measure the similarities of nodes. The normalized similarity matrix is $P = D^{-1}W$, where D represents a diagonal matrix in which the element $D(i, i) = \sum_j W(i, j)$, therefore $\sum_j P(i, j) = 1$. Self-similarities existing in this computation may lead to a numerical instability, so the preferred normalization process is performed as follows:

$$P(i, j) = \begin{cases} W(i, j) \\ 2\sum_{k \neq i} W(i, k) \\ \frac{1}{2}, j = i \end{cases} \quad (1)$$

Given the graph G, let N_i stand for a set of v_i's directly connecting neighborhoods containing the v_i, the direct neighbors are used to compute intimate connections as follows:

$$S(i, j) = \begin{cases} W(i, j) \\ \sum_{k \in N_i} W(i, k), j \in N_i \\ 0, otherwise \end{cases} \quad (2)$$

2.7. Integration of weighted networks into one final fused network

With the development of sequencing technology, more data types can be used to effectively integrate the corresponding weighted networks into one fused network based on network iteration. The final fused network captures both shared and complementary information from multiple data sources, offering insight into how informative each data type is. Given four different data types, adjacency matrices $W^{(v)}$ are constructed for the v^{th} observation, v= 1,2,3,4. $P^{(v)}$ as well as $S^{(v)}$ are achieved from formulas (1) and (2), respectively.

For each omics weighted matrix, the final fused matrix and the iteratively updated process are determined as follows:

$$P_{t+1}^{(v)} = S^{(v)} \times \left(\frac{\sum_{k \neq v} P_t^{(k)}}{3} \right) \times (S^{(v)})^T, v = 1,2,3,4$$

where $P_{t+1}^{(v)}$ represents the status matrix of every omics after t iteration steps. The process updates the status matrices every step and then generates four matrices of the same type, interchanging diffusion at the same time. After t iteration steps, the final status matrix is calculated as follows:

$$P^{(c)} = \frac{\sum_v P_t^{(v)}}{4}, v = 1,2,3,4$$

3. Results

3.1. Multi-omics analysis and multi-omics information network

Using computational biology methods combined with multi-omics for selecting biomolecules of interest for maize breeding is receiving increased attention (Prioul et al., 2008). For example, in genomics and phenomics, variation analyses of one phenotype of maize is used to understand how genotype controls phenotype, thereby increasing or decreasing yields. Plant phenomics analysis is more challenging than genomics owing to the variety of phenotypes throughout the life of a plant.

Predicting the functional roles of individual genes using one omics approach is biased in biology. Thus, we collected the transcriptome, genome, and phenome interaction information of maize and then used these different omics interaction data to generate corresponding networks (Fig. 4). The gene regulatory network (GRN) was generated by using transcriptome interaction data (Krouk et al., 2013; Gardner and Faith, 2005; Bar-Joseph et al., 2003; De Smet and Marchal, 2010), and the co-expression network (van Noort et al., 2003; Stuart et al., 2003; Horvath and Dong, 2008) was generated by using genome interaction data. There was an assumption that mRNA measurements could be used as a proxy for measurements of protein abundance. However, there were only weak positive correlations between mRNAs and proteins (Schwanhausser et al., 2011; Vogel et al., 2010; Ghaemmaghami et al., 2003; Baerenfaller et al., 2008; Ghazalpour et al., 2011; Ponnala et al., 2014; Walley et al., 2013; Washburn et al., 2003), which showed that the interaction network built on genome data alone would be enhanced by adding transcriptome and phenome data.

For gene function, proteins were key players and were directly involved in cellular development and metabolism. The proteins acted not only as catalysts but also participated in intracellular regulatory processes, for instance, transcriptional regulation and signal transduction. Despite specific functions being assigned to proteins, they often remained inactive because they needed to undergo certain modification processes. Cooperation between proteins are called protein–protein interactions (PPI). Here, we combined the PPI network constructed from the information on maize with the above omics network for more comprehensive information on interactions.

3.2. Integration of maize data at four omics levels into a fused network

The four different omics data used here are from the work of Walleye et al. and the STRING database. In Walleye et al.'s study, 23 tissues were collected from the B73 inbred line for profiling spanning the vegetative and reproductive stages of maize development. The co-expression among genes and mRNAs was evaluated by computing Pearson correlations. Transcription factors (TFs) were defined using the GRASSIUS transcription factor list, and the corresponding targets were connected by their TFs in a ChIPseq assay. GRNs were constructed using the GENIE3 algorithm.

Identifying potential candidates is important with the availability of these rich datasets using integrative omics methods. Combining biological data through normalized methods is simple, but it will attenuate the low signal-to-noise ratio already present in each data type. The common solution is to analyze each dataset independently and then combine the data, but often the results from integration are inconsistent making conclusions difficult. Prompted by the approach proposed by Wang et al., we integrated different networks derived from various data sources into a fused network. The main step involved is constructing a weighted network for every omics dataset and then integrating these weighted networks into one final fused network through nonlinear combination approach. The final fused network captures multi-omics information enhancing the ability to obtain insight into maize development and selection of biomolecules of interest.

First, a network is created for each of the four types of data. The normalization process and the computation of intimate connections were based on the reasonable assumption that the nearest cluster similarities are more credible than distant ones; thus, the weight of the undirected neighbors were assigned by graph diffusion in the normalization-weighted network. Hence, P represents each node having similarities to all other nodes, while S represents each node only having similarities to its directly connected neighbors in the weighted network. The calculations started from the initial state matrix P while using the directed neighbors matrix S in the fusion process, which facilitated both the ability to capture the computational efficiency and the local structure of networks.

Next, four weighted networks are fused into one final weighted network. In each network, we used a message passing theory as the basis of the nonlinear method at every iteration step to make them similar to each other. Then, they were fused into one weighted network

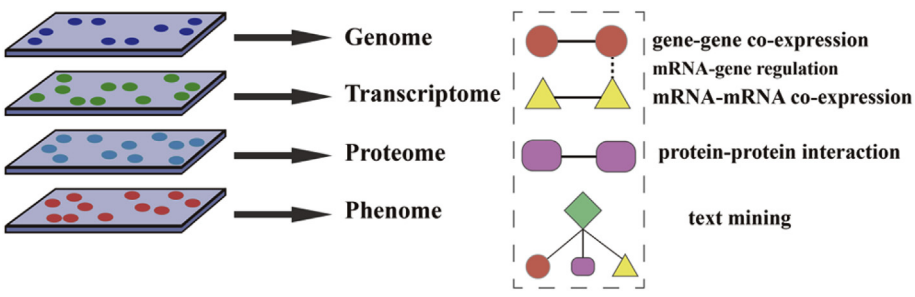

Fig. 4. The omics levels and the relationship among them in maize.

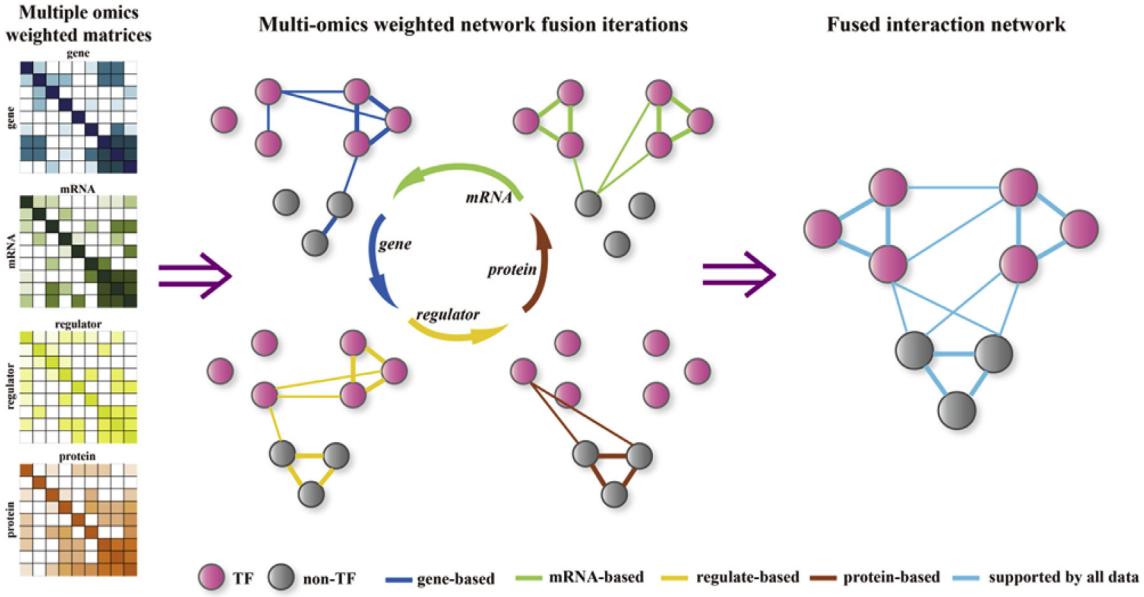

Fig. 5. Iterative process of integrating four omics weighted networks into a single network. (a) The weighted adjacency matrices for each omics. (b) Network fusion by iteratively updating each of the networks with information from the other networks, making them more similar with each step. The colors of the nodes represent whether they are TFs or not. The color of edges indicates which data type has contributed to the given similarity. (c) The iterative network fusion results in convergence to the final fused network. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

after several iteration steps. The advantage of the integrative procedure is that lower weighted edges in some networks are kept according to how closely they connected to their neighbors in all weighted networks. This nonlinearity fully considered the local structure of networks, while merging the complementary and common information across networks.

After t = 150 iterations, we obtained a fused interaction network from the weighted network for the four omics levels, as shown in Fig. 5. In each network, the nodes and edges in the network are shown in different colors corresponding to the different definitions. The pink colored nodes are defined as TFs, while the others (i.e., gray) are not TFs. In the weighted networks for the four omics levels, the color of weighted edges of every network stands for the corresponding data type. After the iteration process, the information of every edge in the final weighted interaction network contained the four data types.

3.3. Network analysis

Owing to the fusion process, there are nodes that do not connect to any other nodes in one or more data type network. These are called orphan nodes, which are included in the final fused network because they have relationships with their neighbors in the other data type networks. Importantly, some of the orphan nodes were shown to be TFs and play a key role in maize development. The orphan nodes of each weighted network in the different data types are shown in Table 1.

A Gene Ontology (GO) annotation is a statement about the function of a particular gene. The Gene Ontology defines all concepts associated with gene functions ("GO terms"), as well as how these functions are relevant to each other ("relations"). We obtained functional annotation information of maize from the PPIM database. The number of GOs associated with the 17 TFs, which are also orphan nodes, is shown in Fig. 6. GRMZM2G385622 is an orphan node in the gene and regulatory network that is associated with the most GO terms. GRMZM2G149040, GRMZM2G019446, GRMZM2G161009, and GRMZM2G180847 are four TFs enriched for identical GO terms (GO: 0043565 ~ sequence-specific DNA binding, GO: 0003700 ~ DNA-binding transcription factor activity, and GO: 0006355 ~ regulation of transcription, DNA-templated), while GRMZM2G153594 and GRMZM2G341747 are enriched in terms completely different from those above (GO: 0005515 ~ protein binding, GO: 0005634 ~ nucleus, and GO: 0003677 ~ DNA binding). Obviously, the TF genes are related to a class of GO terms that are associated with DNA binding. The binding DNAs are short DNA sequences, 4 to 30 base pairs long, which are specifically bound by one or more DNA-binding proteins or protein complexes. The proteins that bind to the DNA typically modify or pack the DNA or regulate gene expression.

4. Discussion

Using computational biology methods combined with multi-omics data for selecting biomolecules of interest for maize breeding is getting more attention. The rapid development of high-throughput sequence data provides the opportunity to explore the biomolecules that influence maize yield by collecting multitudinous types of diverse omics data. In research, we typically investigate the merits of fused diverse data types, such as one data type may supplement the disadvantages of the others, and thereby improve the prediction accuracy. In this context, the combination of four omics technologies, namely, genomics, proteomics, transcriptomics, and phenomics, may offer a valuable reference for the development of maize. Additionally, we introduced current methods for integrating these diverse data types with better performance and successfully applied some of them to predict the yield-related biomolecules of maize.

Inspired by Wang et al.'s method and thanks to the large amounts of data available, integrative methods have become more important for identifying genes that are associated with maize development. Here, we integrated four different weighted networks into a fused network using the nodes that appeared in at least one network. A nonlinear combination method was used to construct networks for each omics data and then integrated these networks into one final fused weighted network. The final fused weighted network included the complementary information of the different data types as well as those in common, providing the contribution of each data resource to the final weight. After the fusion process, the orphan nodes were included in the final fused network, some of which were shown to be TFs that play key roles in maize.

The orphan nodes listed in Table 1 had strong associations with

Table 1
Orphan nodes in each data type network.

Type	The names of genes								
gene	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM5	GRMZM2	GRMZM2G
	G104504	G060611	G146292	G385622	G346639	G038953	G815165	G130085	043453
	GRMZM2	GRMZM2	GRMZM2	GRMZM5	GRMZM2	GRMZM2	GRMZM2	GRMZM2	
	G358618	G104847	G095219	G897376	G010801	G107089	G096037	G040511	
mRNA	GRMZM2	GRMZM2	GRMZM2						
	G016250	G040477	G326472						
regulate	GRMZM2	AF546188	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	
	G385622	.1_FG005	G429899	G101271	G038338	G038536	G092497	G136680	
	GRMZM2	GRMZM2	GRMZM2	GRMZM5	GRMZM2	GRMZM2	AC209877.	GRMZM2	GRMZM2G
	G063163	G017254	G043493	G887286	G138976	G150212	3_FG002	G038338	098594
	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2G
	G033130	G048276	G114930	G002320	G108865	G040467	G134329	G048194	149040
	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	AC196465.3
	G019446	G149543	G390221	G070038	G019738	G026918	G135322	G360688	_FG006
	GRMZM5	GRMZM2	GRMZM5	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2G
	G807019	G101390	G823750	G074377	G161009	G088083	G302913	G308083	032423
	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2G
protein	G341747	G136700	G078178	G143714	G104047	G154301	G148518	G057973	126812
	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2G
	G092018	G152526	G139691	G016546	G104377	G463726	G354053	G077124	004140
	GRMZM5	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	EF517601.	GRMZM2	GRMZM2G
	G819523	G440785	G096352	G049687	G021885	G324705	1_FG015	G030744	145905
	GRMZM2	GRMZM2	GRMZM2	GRMZM5	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2G
	G386643	G126170	G148555	G881803	G152768	G180847	G109480	G116592	358238
	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2G
	G439950	G001777	G153594	G033566	G003424	G370155	G024641	G057237	050166
	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2	GRMZM2		
	G152688	G045392	G125342	G026558	G092120	G114461	G092648		

*The red font represents TFs and background shading indicates that the TF occurs in two networks. The first column is for the weighted network for the four data types.

maize developmental processes according to their functional annotations, making them highly promising candidates. In the future, we hope that the orphan nodes will save time and reduce costs for plant breeding.

5. Conclusions

In summary, this study analyzed the current state of development and research at each omics level. Given the importance of integrating diverse omics levels, we fused four multi-omics similarity networks and analyzed the results of the fused network. In the future, we will add other omics of maize, such as metabolomics; our work will contribute to improving maize production and provide new perspectives for maize researchers.

Author contribution

QZ designed the research; XZ performed the research; FX and CW analyzed the data; and JJ wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The work was financially supported by the Natural Science Foundation of China (No.91735306, No.61771331).

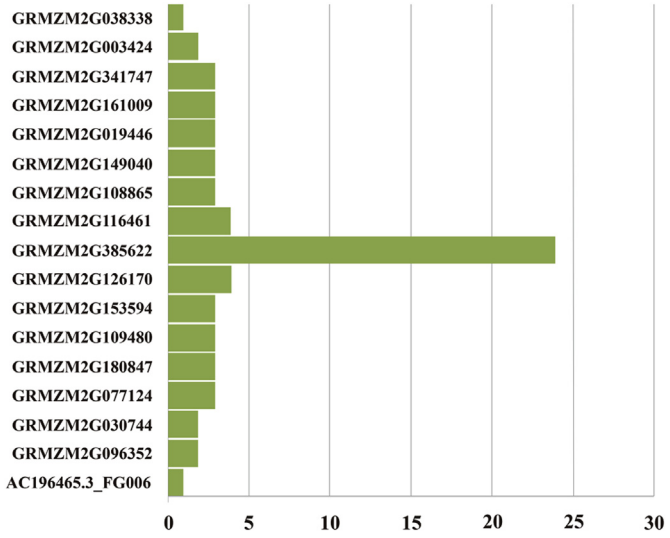

Fig. 6. Barplot of the numbers of GO terms enriched for 17 TFs. The abscissa represents the number of GO terms.

References

Baerenfaller, K., Grossmann, J., Grobei, M.A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., Baginsky, S., 2008. Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science 320, 938–941.

Balsamo, G.M., Cangahuala-Inocente, G.C., Bertoldo, J.B., Terenzi, H., Arisi, A.C., 2011. Proteomic analysis of four Brazilian Mon810 maize varieties and their four non-genetically-modified isogenic varieties. J. Agric. Food Chem. 59, 11553–11559.

Bar-Joseph, Z., Gerber, G.K., Lee, T.I., Rinaldi, N.J., Yoo, J.Y., Robert, F., Gordon, D.B., Fraenkel, E., Jaakkola, T.S., Young, R.A., Gifford, D.K., 2003. Computational discovery of gene modules and regulatory networks. Nat. Biotechnol. 21, 1337–1342.

Bohnert, H.J., Nelson, D.E., Jensen, R.G., 1995. Adaptations to environmental stresses. Plant Cell 7, 1099–1111.

Castelletti, S., Tuberosa, R., Pindo, M., Salvi, S., 2014. A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL Vgt1. G3 4, 805–812.

Chen, J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A., Lai, J., 2014. Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiol. 166, 252–264.

Coll, A., Nadal, A., Rossignol, M., Puigdomenech, P., Pla, M., 2011. Proteomic analysis of Mon810 and comparable non-GM maize varieties grown in agricultural fields. Transgenic Res. 20, 939–949.

Cui, D., Wu, D., Liu, J., Li, D., Xu, C., Li, S., Li, P., Zhang, H., Liu, X., Jiang, C., Wang, L., Chen, T., Chen, H., Zhao, L., 2015. Proteomic analysis of seedling roots of two maize inbred lines that differ significantly in the salt stress response. PLoS One 10, e0116697.

De Smet, R., Marchal, K., 2010. Advantages and limitations of current network inference methods, Nature reviews. Microbiology 8, 717–729.

Gardner, T.S., Faith, J.J., 2005. Reverse-engineering transcription control networks. Phys. Life Rev. 2, 65–88.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741.

Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N., Hughes, K.A., 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature 525, 372–375.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P.Z., Brewer, H., Weitz, K., Camp 2nd, D.G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D.J., Smith, R.D., Lusis, A.J., 2011. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet. 7, e1001393.

Gong, C.Y., Wang, T., 2013. Proteomic evaluation of genetically modified crops: current status and challenges. Front. Plant Sci. 4, 41.

Horvath, S., Dong, J., 2008. Geometric interpretation of gene coexpression network analysis. PLoS Comput. Biol. 4, e1000117.

Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhajarvi, T., Chia, J.M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., Lai, J., Morrell, P.L., Shannon, L.M., Song, C., Springer, N.M., Swanson-Wagner, R.A., Tiffin, P., Wang, J., Zhang, G., Doebley, J., McMullen, M.D., Ware, D., Buckler, E.S., Yang, S., Ross-Ibarra, J., 2012. Comparative population genomics of maize domestication and improvement. Nat. Genet. 44, 808–811.

Karp, N.A., Huber, W., Sadowski, P.G., Charles, P.D., Hester, S.V., Lilley, K.S., 2010. Addressing accuracy and precision issues in iTRAQ quantitation. Mol. Cell. Proteom. : MCP 9, 1885–1897.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., Hoh, J., 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308, 385–389.

Krouk, G., Lingeman, J., Colon, A.M., Coruzzi, G., Shasha, D., 2013. Gene regulatory networks in plants: learning causality from time and perturbation. Genome Biol. 14, 123.

Li, K.C., 2002. Genome-wide coexpression dynamics: theory and application. Proc. Natl. Acad. Sci. U. S. A 99, 16875–16880.

Li, K.C., Yuan, S., 2004. A functional genomic study on NCI's anticancer drug screen. Pharmacogenomics J. 4, 127–135.

Li, K.C., Liu, C.T., Sun, W., Yuan, S., Yu, T., 2004. A system for enhancing genome-wide coexpression dynamics study. Proc. Natl. Acad. Sci. U. S. A 101, 15561–15566.

Li, K.C., Palotie, A., Yuan, S., Bronnikov, D., Chen, D., Wei, X., Choi, O.W., Saarela, J., Peltonen, L., 2007. Finding disease candidate genes by liquid association. Genome Biol. 8, R205.

Liu, C.T., Yuan, S., Li, K.C., 2009. Patterns of co-expression for protein complexes by size in Saccharomyces cerevisiae. Nucleic Acids Res. 37, 526–532.

Liu, H., Wang, X., Warburton, M.L., Wen, W., Jin, M., Deng, M., Liu, J., Tong, H., Pan, Q., Yang, X., Yan, J., 2015. Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. Mol. Plant 8, 871–884.

Ma, C., Zhou, J., Chen, G., Bian, Y., Lv, D., Li, X., Wang, Z., Yan, Y., 2014. iTRAQ-based quantitative proteome and phosphoprotein characterization reveals the central metabolism changes involved in wheat grain development. BMC Genomics 15, 1029.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z.M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., Parkinson, H., 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901.

Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O., Eisenberg, D., 1999. A combined algorithm for genome-wide prediction of protein function. Nature 402, 83–86.

Meyer, R.S., Purugganan, M.D., 2013. Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. 14, 840–852.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P., 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43–50.

Pigliucci, M., 2005. Evolution of phenotypic plasticity: where are we going now? Trends Ecol. Evol. 20, 481–486.

Ponnala, L., Wang, Y., Sun, Q., van Wijk, K.J., 2014. Correlation of mRNA and protein abundance in the developing maize leaf. Plant J. : Cell Mol. Biol. 78, 424–440.

Prioul, J.L., Mechin, V., Lessard, P., Thevenot, C., Grimmer, M., Chateau-Joubert, S., Coates, S., Hartings, H., Kloiber-Maitz, M., Murigneux, A., Sarda, X., Damerval, C., Edwards, K.J., 2008. A joint transcriptomic, proteomic and metabolic analysis of maize endosperm development and starch filling. Plant Biotechnol. J. 6, 855–869.

Rocco, M., Arena, S., Renzone, G., Scippa, G.S., Lomaglio, T., Verrillo, F., Scaloni, A., Marra, M., 2013. Proteomic analysis of temperature stress-responsive proteins in Arabidopsis thaliana rosette leaves. Mol. Biosyst. 9, 1257–1267.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J., 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteom. : MCP 3, 1154–1169.

Schluter, H., Apweiler, R., Holzhutter, H.G., Jungblut, P.R., 2009. Finding one's way in proteomics: a protein species nomenclature. Chem. Cent. J. 3, 11.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., Wilson, R.K., 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326, 1112–1115.

Schulze, W.X., Usadel, B., 2010. Quantitation in mass-spectrometry-based proteomics. Annu. Rev. Plant Biol. 61, 491–516.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M., 2011. Global quantification of mammalian gene expression control. Nature 473, 337–342.

Smith, L.M., Kelleher, N.L., Consortium for Top Down, P., 2013. Proteoform: a single term describing protein complexity. Nat. Methods 10, 186–187.
Stuart, J.M., Segal, E., Koller, D., Kim, S.K., 2003. A gene-coexpression network for global discovery of conserved genetic modules. Science 302, 249–255.
Studer, A., Zhao, Q., Ross-Ibarra, J., Doebley, J., 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43, 1160–1163.
Sun, W., Yuan, S., Li, K.C., 2008. Trait-trait dynamic interaction: 2D-trait eQTL mapping for genetic variation study. BMC Genomics 9, 242.
Tai, S.K., Wu, G., Yuan, S., Li, K.C., 2010. Genome-wide expression links the electron transfer pathway of Shewanella oneidensis to chemotaxis. BMC Genomics 11, 319.
Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., Provart, N.J., 2009. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ. 32, 1633–1651.
van Noort, V., Snel, B., Huynen, M.A., 2003. Predicting gene function by conserved co-expression. Trends Genet. : TIG (Trends Genet.) 19, 238–242.
Vidal, N., Barbosa, H., Jacob, S., Arruda, M., 2015. Comparative study of transgenic and non-transgenic maize (Zea mays) flours commercialized in Brazil, focussing on proteomic analyses. Food Chem. 180, 288–294.
Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90, 7–24.
Vogel, C., Abreu Rde, S., Ko, D., Le, S.Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., Penalva, L.O., 2010. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. 6, 400.
Walley, J.W., Shen, Z., Sartor, R., Wu, K.J., Osborn, J., Smith, L.G., Briggs, S.P., 2013. Reconstruction of protein networks from an atlas of maize seed proteotypes. Proc. Natl. Acad. Sci. U. S. A 110, E4808–E4817.
Washburn, M.P., Wolters, D., Yates 3rd, J.R., 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247.
Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., Plouffe, D., Deciu, C., Winzeler, E., Yates 3rd, J.R., 2003. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A 100, 3107–3112.
Wu, X., Wang, W., 2016. Increasing confidence of proteomics data regarding the identification of stress-responsive proteins in crop plants. Front. Plant Sci. 7, 702.
Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., Li, K., Yang, N., Li, Y., Zhong, T., Li, J., Chen, Y., Yan, J., Yang, X., Xu, M., 2013. CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. U. S. A 110, 16969–16974.
Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M., Burnett, W.J., Logan, K.O., Wang, D., Wang, X., Drews, G.N., Yadegari, R., 2015. RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. Plant Cell 27, 513–531.
Zhang, L., Chia, J.M., Kumari, S., Stein, J.C., Liu, Z., Narechania, A., Maher, C.A., Guill, K., McMullen, M.D., Ware, D., 2009. A genome-wide characterization of microRNA genes in maize. PLoS Genet. 5, e1000716.
Zi, J., Zhang, J., Wang, Q., Zhou, B., Zhong, J., Zhang, C., Qiu, X., Wen, B., Zhang, S., Fu, X., Lin, L., Liu, S., 2013. Stress responsive proteins are actively regulated during rice (Oryza sativa) embryogenesis as indicated by quantitative proteomics analysis. PLoS One 8, e74229.
Zolla, L., Rinalducci, S., Antonioli, P., Righetti, P.G., 2008. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. J. Proteome Res. 7, 1850–1861.