



Original article

RGPDB: database of root-associated genes and promoters in maize, soybean, and sorghum

Gleb Moissejev^{5,†}, Kiyoul Park^{2,4,†}, Alix Cui^{5,†}, Daniel Freitas^{5,†},
Divith Rajagopal^{5,†}, Anji Reddy Konda^{2,4}, Madalayne Martin-Olenski⁵,
Mackenzie Mcham⁵, Kan Liu^{1,4}, Qian Du^{1,4}, James C. Schnable^{3,4},
Etsuko N. Moriyama^{1,4}, Edgar B. Cahoon^{2,4,*} and Chi Zhang^{1,4,*}

¹School of Biological Sciences, University of Nebraska, Lincoln, NE 68588 USA, ²Department of Biochemistry, University of Nebraska, Lincoln, NE 68588 USA, ³Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583 USA, ⁴Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588, USA and ⁵Young Nebraska Scientists Program, University of Nebraska (EPSCoR), Lincoln, NE 68588, USA

*Corresponding author: Tel: +1(402)472-1396 (CZ); Fax: +1(402)472-8722 (CZ); E-mail: czhang5@unl.edu
Correspondence may also be addressed to Edgar B. Cahoon. Email: ecahoon2@unl.edu

† Authors contributed equally to this work.

Citation details: Moissejev, G., Park, K., Cui, A. *et al.* RGPDB: database of root-associated genes and promoters in maize, soybean, and sorghum. *Database* (2020) Vol. 2020: article ID baaa038; doi:10.1093/database/baaa038

Received 22 August 2019; Revised 2 March 2020; Accepted 6 May 2020

Abstract

Root-associated genes play an important role in plants. Despite the fact that there have been studies on root biology, information on genes that are specifically expressed or upregulated in roots is poorly collected. There exist very few databases dedicated to genes and promoters associated with root biology, preventing effective root-related studies. Therefore, we analyzed multiple types of omics data to identify root-associated genes in maize, soybean, and sorghum and constructed a comprehensive online database of these genes and their promoter sequences. This database creates a pivotal platform capable of stimulating and facilitating further studies on manipulating root growth and development.

Database URL: <https://crrl.unl.edu/databases> and <http://sysbio.unl.edu/RGPDB/>.

Introduction

Roots are of critical importance for plant biology because they link below and above ground systems and extract water and nutrients from soil (1). The ability of roots to acquire minerals and water from the soil and to respond to the changing environments is dependent on the root system architecture, such as the root growth angle. It determines

the direction of root elongation in the soil affecting the area in which roots capture water and nutrients. The root system and the root system architecture are the results of continuous root growth and development (2, 3). The root growth regulation is a highly complicated process and is controlled by complex gene interaction networks in both time and space. Identification of root-associated genes, their

functions, and their interactions can reveal the physiological and molecular mechanisms that regulate the root growth and development and have the potential to improve crop production (4). Moreover, in plant biotechnology, synthetic promoters can provide precise spatial and temporal control of transgene expression to improve crop productivity (5, 6). The information of promoters in root-associated genes helps the synthetic biology tool development to generate plants with novel root traits to enhance plant performance (7, 8).

Due to the importance of roots, many root-associated genes have been studied. For example, some genes were discovered being involved in control of root-cell elongation in *Arabidopsis thaliana* mediated by the 1-aminocyclopropane-1-carboxylic acid (ACC) (9). Some others were found to have a function for regulation of root angle and gravitropism (10). Some root-associated genes were identified by genetic analysis of root response to drought stress and abscisic acid (11). Root-associated genes were also discovered and studied by gene expression profiling of the *Arabidopsis* root (12) and the *Arabidopsis* root transcriptome sequencing (13). These studies are, however, mainly for *Arabidopsis*. There are some but a limited number of studies on root-associated genes in maize, sorghum, and soybeans. For example, a few studies were conducted for genetic and genomic dissection of maize root development and architecture (14, 15). The expression of an expansin gene has been correlated with root elongation in soybean (16). For root-associated promoters, for example, *cis*-acting elements of the barley *IDS2* gene promoter were found to confer iron-deficiency-inducible, root-specific expression in heterogeneous tobacco plants (17). However, there is no comprehensive data collection for root-associated genes and their promoters in maize, sorghum, and soybeans. Analysis of existing gene expression data facilitates the discovery of root-associated genes. Combining with genome sequence information, one can predict the promoter sequences for these candidate root-associated genes. Such information is useful for understanding and manipulating root growth and development.

Currently, there is only one root-associated gene related database—iRootHair (18). It includes information about 153 root hair-related genes that have been identified in dicots and monocots along with their putative orthologs in higher plants with sequenced genomes. There are some databases with information on *cis*-acting elements that control the transcription initiation by binding corresponding nuclear factors. They include TRANSFAC (19), JASPAR (20), TRANSCompel (21), PlantCARE (22) and PLACE (23). A more complete plant gene promoter database is PlantProm (24). However, the gene promoter data included in PlantProm are mainly for *Arabidopsis*. No databases

have been developed specifically for root-associated genes and promoters, and especially for non-*Arabidopsis* plants.

To fulfill the important needs, we collected omics data and using these data, identified root-associated genes and their promoter sequences in maize, soybean, and sorghum. To maximize the value and usability of these types of data for efficient and effective data mining, we developed a web-based comprehensive database of root-associated genes and promoters. Our database, RGPDB, provides detailed information of both root-associated genes and the sequences of their promoter regions in maize, soybean, and sorghum.

Data collection, root-associated gene prediction and validation

We collected multiple types of omics datasets for maize, soybean, and sorghum, including tissue transcriptomic and proteomic data. For transcriptomic data, data measured under different stresses were collected as well. More than 200 datasets were collected and analyzed. For maize genes, the following datasets were used: maize gene expression atlas by RNA-seq (25), a developmental atlas of maize (26), tissue-specific proteomics data (27), and gene expression in root with low nitrogen (28) and drought stress (29). For sorghum, RNA-seq data of Sorghum 9d seedlings in response to osmotic stress and abscisic acid (30), transcriptomics of sorghum tissues (31), gene expression profiles in root with nitrogen stress (32), and sorghum transcriptome database (33) were included. For soybean, transcriptomic data were mainly obtained from the collection in SoyBase (34), as well as from the RNA-Seq atlas of *Glycine max*, which has data for the soybean transcriptome in different tissues (35).

Root-associated genes were predicted as follows. An in-house tool, written in Perl, was used to integrate all gene expression data and identify root-associated gene candidates. To find a root-associated gene, we looked for a gene whose root expression level is at least 10 times larger than its maximal expression levels in all other tissues. If there are multiple studies for the root tissue, the maximal expression level of genes in all studies was used. When it was possible, protein products of the candidate genes and/or orthologs of these genes in other plants were examined. For example, maize proteomes data (27) were employed to validate candidate genes in maize to see if their protein products are also high in root. If the gene product is root-specific and its orthologs have high gene expression levels in roots as well, the corresponding gene is considered a high-confidence candidate. Orthologs of maize genes were obtained from MaizeGDB (36), Rice Genome Annotation Project (37), and References (38, 39). For soybean genes, their orthologs



Figure 1. The searching page of the database.

were obtained from OrthoDB (40) and the reference (41). Sorghum genes' orthologs in Arabidopsis were obtained from Sorghum Functional Genomics Database (42), and orthologs in maize and rice are same for the maize data set. The gene expression atlas for other plants was obtained from various datasets, such as Schmid *et al.*'s work for Arabidopsis (43) and Rice Expression Database (44) for rice. For candidate genes, we extracted genomic sequences of the 2 kb regions upstream of a chosen TSS as promoter sequences because most transcriptional activity sets are within this region (45, 46). The information of each gene's TSS was obtained from the gene annotation files. If there are multiple TSSs for a given gene, the position for the longest transcript was taken. If a gene has multiple alleles, only a single entry is included in the database.

To validate root-specific promoters that we identified in plant systems, four candidate maize genes were selected (*GRMZM2G027098*, *GRMZM2G477685*, *GRMZM2G125023* and *GRMZM2G133475*). Their upstream 2 kb regions were cloned and fused with β -glucuronidase (*GUS*) reporter gene and these *promoter:GUS* constructs were introduced into the rice. As shown in Figure 3, histochemical *GUS* data is in agreement with our prediction, suggesting that predicted maize gene promoters

are subject to control by the root developmental process in monocots.

Database content

The current version of the database contains more than 1200 candidates of root-associated genes and their corresponding promoter sequences for maize (592), sorghum (363) and soybean (400). To store and query these data, we constructed an online database, RGPDB. In the online database, RGPDB, gene promoter sequences and other relevant resources are provided. For each gene, its normalized gene expression levels in different tissues are displayed, which were normalized by using DESeq package (47). These tissue-specific expression data can help users identify the most significant root-associated genes in the database for their interests. Other related information for each gene includes gene ID, description of the function, gene ontology (GO) annotations, Pfam IDs, orthologous genes in other plants and links to other databases, such as eFP browser (48), MaizeGDB (36) and SoyBase (34). For many genes, reference information (PubMed IDs) are also recorded. All information of a given gene can be downloaded as an XML or pdf file.

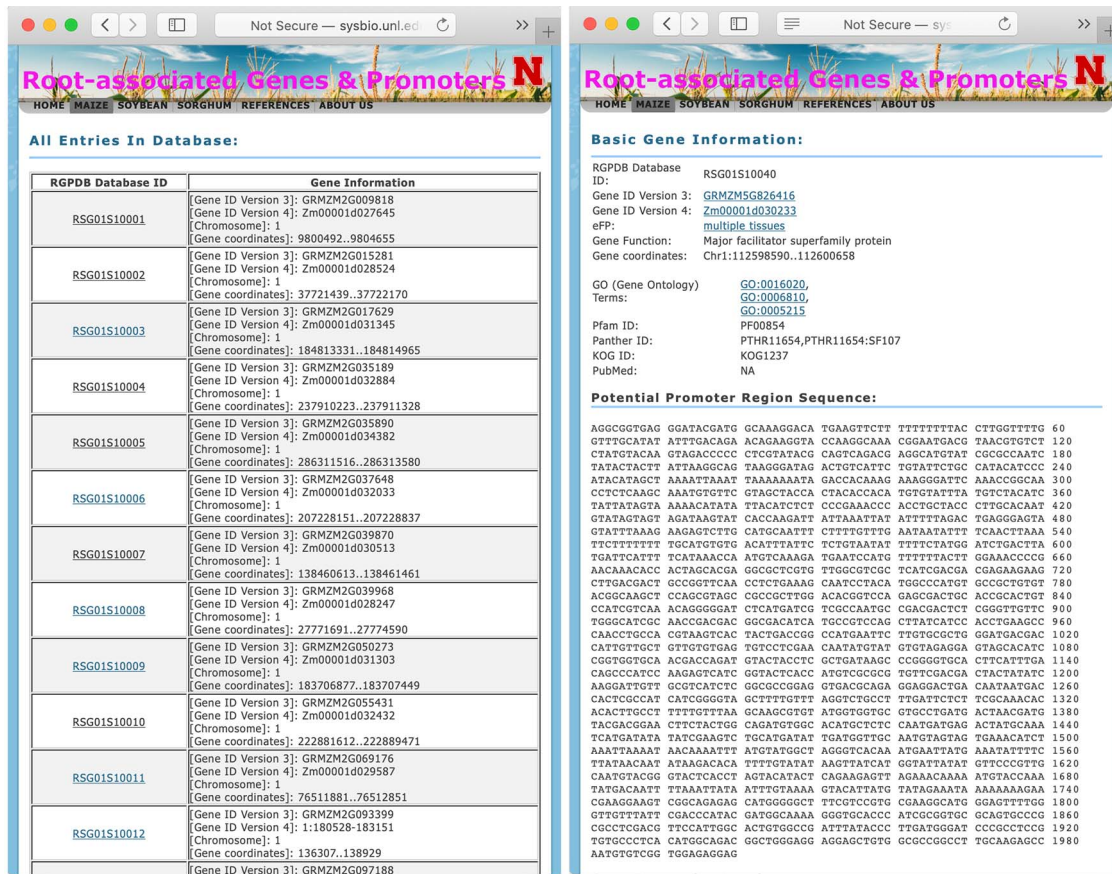


Figure 2. The display pages of searching result (left panel) and information of each gene (right panel).

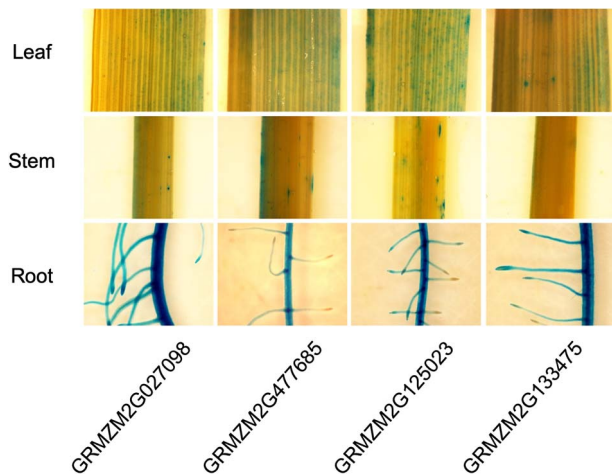


Figure 3. The histochemical GUS assay in a promoter:GUS transgenic rice.

User interface

Browsing and searching

The database system provides interactive access to all of the collected data, and users may connect to the database using a web browser. Figure 1 shows a screen snapshot of the user

interface for users to browse or search the database. The ‘browse’ button allows users to get a list of all records in one table. Search options are provided to conveniently locate genes of interest by using, for example, gene IDs, RGPDB Database IDs, chromosome numbers, keywords of gene annotations or DNA sequences of gene promoter regions.

Displaying the database content

When a user browses the whole database or searches with a specific option, the database first returns a table of related records with RSG Database IDs and their annotations as shown in Figure 2. For maize, soybean, and sorghum genes, their RSG Database IDs start with RSG01S, RSG07S, and RSG05S, respectively. The detailed information of each gene can be displayed by clicking the link of their RGPDB Database IDs. The information for each gene includes the basic information, promoter DNA sequences, GO annotation, Pfam ID, Panther ID, and their orthologs in various plant organisms. The basic information consists of gene IDs in different versions of the original database and function description. For maize, soybean, and sorghum genes, their gene IDs are linked to MaizeGDB (36), SoyBase (34) or

Ensembl database (49), respectively. The PubMed information links to all related publications. If the protein coded by a gene has GO annotation, Pfam domain information, Panther domain information and/or EuKaryotic Orthologous Groups IDs, links to those domain IDs are provided. If a gene has orthologs from other plants, such as Arabidopsis or rice, the link to these orthologs' databases and gene expression atlas databases are also provided, such as link to Rice Expression Database (44) for rice genes.

Implementation

We adopted the LAMP (Linux, Apache, MySQL, PHP) platform to construct the online database system. The user interface additionally accepts parameters via a URL for direct searching. This feature facilitates a link to the database from external sites allowing users to bookmark and to cite directly specific results.

Accessibility

The database is freely available to all users without restriction at <https://crri.unl.edu/databases> and <http://sysbio.unl.edu/RGPDB>. The source codes and other detailed information are available upon request.

Materials and Methods

Rice transformation

Rice (*Oryza sativa* L.) japonica variety 'Kitaake' was used in this study. Transgenic rice plants were generated by Agrobacterium-mediated transformation method. Dry rice seeds were soaked in 70% ethanol for 3 min, treated with 0.3% NaClO solution for 20 min and rinsed with sterile water. Seeds were plated on MSD media (1x Murashige and Skoog (MS) medium including vitamins, 3% sucrose, 2 mg/ml 2, 4-D and 0.2% gelite, pH 5.8) to induce callus formation. Promoter-GUS vector-containing Agrobacterium was co-incubated with 7-day-old calli for 2 days on MSD media supplemented with 5% sorbitol and 200 M acetosyringone. After co-incubation, calli were rinsed extensively with distilled water containing 400 mg/l carbenicillin and plated on MSD media containing 30 mg/l hygromycin for selection. Selected calli were transferred to regeneration media (1x MS medium, 3% sucrose, 2% sorbitol, 2.5 mg/l kinetin, 0.1 mg/l NAA, 30 mg/ml hygromycin and 0.2% gelite, pH 5.8) for shooting. Plantlets that shoot were 3 ~ 5 cm long were transferred to rooting media (1x MS medium, 3% sucrose, 30 mg/ml hygromycin and 0.2% gelite, pH 5.8) and fully regenerated rice plants were transferred to the soil for further growth.

Histochemical GUS assay

Transgenic rice plants were immersed in GUS staining buffer (50 mM sodium phosphate, 2 mM cyclohexyl ammonium salt, 0.5 mM $K_3Fe(CN)_6$ and 0.5 mM $K_4Fe(CN)_6$, pH 7.2) and incubated 8 h at 37°C with the dark condition. To destain, 70% ethanol were treated for 6 h.

Acknowledgements

C.Z. and E.C. initialized this project and curated all data. G.M., A.C., D.F., D.R., M.M. and M.M. constructed the database and web pages. K.P. and A.K. conducted experiments for validation. K.L. and Q.D. collected omics datasets. C.Z., E.C., E.M. and J.S. supervised this project and drafted the manuscript.

Funding

The National Science Foundation (award # OIA-1557417 to C.Z., E.C., J.S. and E.M.) and the Nebraska Soybean Board (to C.Z. and E.C.).

Conflict of interest. None declared.

References

- Bardgett,R.D. and Wardle,D.A. (2010) *Aboveground–Belowground Linkages : Biotic Interactions, Ecosystem Processes, and Global Change*. Oxford University Press, Oxford.
- Slovak,R., Ogura,T., Satbhai,S.B. *et al.* (2016) Genetic control of root growth: from genes to networks. *Ann. Bot.*, **117**, 9–24.
- Wachsmann,G., Sparks,E.E. and Benfey,P.N. (2015) Genes and networks regulating root anatomy and architecture. *New Phytol.*, **208**, 26–38.
- Uga,Y., Kitomi,Y., Ishikawa,S. *et al.* (2015) Genetic improvement for root growth angle to enhance crop production. *Breed. Sci.*, **65**, 111–119.
- Ali,S. and Kim,W.C. (2019) A fruitful decade using synthetic promoters in the improvement of transgenic plants. *Front. Plant Sci.*, **10**, 1433.
- Bashor,C.J. and Collins,J.J. (2018) Understanding biological regulation through synthetic biology. *Annu. Rev. Biophys.*, **47**, 399–423.
- Liu,W. and Stewart,C.N. Jr. (2015) Plant synthetic biology. *Trends Plant Sci.*, **20**, 309–317.
- Liu,W., Yuan,J.S. and Stewart,C.N. Jr. (2013) Advanced genetic tools for plant biotechnology. *Nat. Rev. Genet.*, **14**, 781–793.
- Markakis,M.N., De Cnodder,T., Lewandowski,M. *et al.* (2012) Identification of genes involved in the ACC-mediated control of root cell elongation in Arabidopsis thaliana. *BMC Plant Biol.*, **12**, 208.
- Toal,T.W., Ron,M., Gibson,D. *et al.* (2018) Regulation of root angle and gravitropism. *G3*, **8**, 3841–3855.
- Xiong,L., Wang,R.G., Mao,G. *et al.* (2006) Identification of drought tolerance determinants by genetic analysis of root response to drought stress and abscisic acid. *Plant Physiol.*, **142**, 1065–1074.

12. Birnbaum, K., Shasha, D.E., Wang, J.Y. *et al.* (2003) A gene expression map of the Arabidopsis root. *Science*, **302**, 1956–1960.
13. Fizames, C., Munos, S., Cazettes, C. *et al.* (2004) The Arabidopsis root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol.*, **134**, 67–80.
14. Hochholdinger, F. and Tuberosa, R. (2009) Genetic and genomic dissection of maize root development and architecture. *Curr. Opin. Plant Biol.*, **12**, 172–177.
15. Hochholdinger, F., Woll, K., Sauer, M. and Dembinsky, D. (2004) Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. *Ann. Bot.*, **93**, 359–368.
16. Lee, D.K., Ahn, J.H., Song, S.K. *et al.* (2003) Expression of an expansin gene is correlated with root elongation in soybean. *Plant Physiol.*, **131**, 985–997.
17. Kobayashi, T., Nakayama, Y., Itai, R.N. *et al.* (2003) Identification of novel *cis*-acting elements, IDE1 and IDE2, of the barley IDS2 gene promoter conferring iron-deficiency-inducible, root-specific expression in heterogeneous tobacco plants. *Plant J.*, **36**, 780–793.
18. Kwasniewski, M., Nowakowska, U., Szumera, J. *et al.* (2013) iRootHair: a comprehensive root hair genomics database. *Plant Physiol.*, **161**, 28–35.
19. Wingender, E., Chen, X., Fricke, E. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, **29**, 281–283.
20. Khan, A., Fornes, O., Stigliani, A. *et al.* (2018) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, **46**, D260–D266.
21. Kel-Margoulis, O.V., Kel, A.E., Reuter, I. *et al.* (2002) TRANSCOMP: a database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Res.*, **30**, 332–334.
22. Lescot, M., Dehais, P., Thijs, G. *et al.* (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, **30**, 325–327.
23. Higo, K., Ugawa, Y., Iwamoto, M. *et al.* (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.*, **27**, 297–300.
24. Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M. *et al.* (2003) PlantProm: a database of plant promoter sequences. *Nucleic Acids Res.*, **31**, 114–117.
25. Sekhon, R.S., Briskine, R., Hirsch, C.N. *et al.* (2013) Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. *PLoS One*, **8**, e61005.
26. Walley, J.W., Sartor, R.C., Shen, Z. *et al.* (2016) Integration of omic networks in a developmental atlas of maize. *Science*, **353**, 814–818.
27. Marcon, C., Malik, W.A., Walley, J.W. *et al.* (2015) A high-resolution tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axes. *Plant Physiol.*, **168**, 233–246.
28. Yu, P., Baldauf, J.A., Lithio, A. *et al.* (2016) Root type-specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. *Plant Physiol.*, **170**, 1783–1798.
29. Opitz, N., Marcon, C., Paschold, A. *et al.* (2016) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. *J. Exp. Bot.*, **67**, 1095–1107.
30. Dugas, D.V., Monaco, M.K., Olsen, A. *et al.* (2011) Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics*, **12**, 514.
31. Davidson, R.M., Gowda, M., Moghe, G. *et al.* (2012) Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J.*, **71**, 492–502.
32. Gelli, M., Duo, Y., Konda, A.R. *et al.* (2014) Identification of differentially expressed genes between sorghum genotypes with contrasting nitrogen stress tolerance by genome-wide transcriptional profiling. *BMC Genomics*, **15**, 179.
33. Makita, Y., Shimada, S., Kawashima, M. *et al.* (2015) MOROKOSHI: transcriptome database in *Sorghum bicolor*. *Plant Cell Physiol.*, **56**, e6.
34. Grant, D., Nelson, R.T., Cannon, S.B. *et al.* (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.*, **38**, D843–D846.
35. Severin, A.J., Woody, J.L., Bolon, Y.T. *et al.* (2010) RNA-Seq atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.*, **10**, 160.
36. Portwood, J.L., 2nd, Woodhouse, M.R., Cannon, E.K. *et al.* (2019) MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.*, **47**, D1146–D1154.
37. Kawahara, Y., de la Bastide, M., Hamilton, J.P. *et al.* (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)*, **6**, 4.
38. Schnable, J.C., Springer, N.M. and Freeling, M. (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 4069–4074.
39. Schnable, J.C., Freeling, M. and Lyons, E. (2012) Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol. Evol.*, **4**, 265–277.
40. Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F. *et al.* (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.*, **47**, D807–D811.
41. Jung, C.H., Wong, C.E., Singh, M.B. *et al.* (2012) Comparative genomic analysis of soybean flowering genes. *PLoS One*, **7**, e38250.
42. Tian, T., You, Q., Zhang, L. *et al.* (2016) SorghumFDB: sorghum functional genomics database with multidimensional network analysis. *Database (Oxford)*, **2016**: baw099.
43. Schmid, M., Davison, T.S., Henz, S.R. *et al.* (2005) A gene expression map of Arabidopsis thaliana development. *Nat. Genet.*, **37**, 501–506.
44. Xia, L., Zou, D., Sang, J. *et al.* (2017) Rice expression database (RED): an integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genomics*, **44**, 235–241.
45. Majewski, J. and Ott, J. (2002) Distribution and characterization of regulatory elements in the human genome. *Genome Res.*, **12**, 1827–1836.

46. Abeel,T, Saeys,Y, Bonnet,E. *et al.* (2008) Generic eukaryotic core promoter prediction using structural features of DNA. *Genome Res.*, **18**, 310–323.
47. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
48. Winter,D., Vinegar,B., Nahal,H. *et al.* (2007) An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One*, **2**, e718.
49. Zerbino,D.R., Achuthan,P., Akanni,W. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.