

# PGSB PlantsDB: updates to the database framework for comparative plant genome research

Manuel Spannagl<sup>1,\*</sup>, Thomas Nussbaumer<sup>1,2,†</sup>, Kai C. Bader<sup>1,†</sup>, Mihaela M. Martis<sup>1,3</sup>, Michael Seidel<sup>1</sup>, Karl G. Kugler<sup>1</sup>, Heidrun Gundlach<sup>1</sup> and Klaus F.X. Mayer<sup>1</sup>

<sup>1</sup>Plant Genome and Systems Biology, Helmholtz Center Munich – German Research Center for Environmental Health, 85764 Neuherberg, Germany, <sup>2</sup>Division of Computational Systems Biology, Department of Microbiology and Ecosystem Science, University of Vienna, 1090 Vienna, Austria and <sup>3</sup>BILS (Bioinformatics Infrastructure for Life Sciences), Division of Cell Biology, Department of Clinical and Experimental Medicine, Linköping University, SE-558185 Linköping, Sweden

Received September 11, 2015; Revised October 14, 2015; Accepted October 15, 2015

## ABSTRACT

**PGSB (Plant Genome and Systems Biology; formerly MIPS) PlantsDB (<http://pgsb.helmholtz-muenchen.de/plant/index.jsp>) is a database framework for the comparative analysis and visualization of plant genome data. The resource has been updated with new data sets and types as well as specialized tools and interfaces to address user demands for intuitive access to complex plant genome data. In its latest incarnation, we have re-worked both the layout and navigation structure and implemented new keyword search options and a new BLAST sequence search functionality. Actively involved in corresponding sequencing consortia, PlantsDB has dedicated special efforts to the integration and visualization of complex triticeae genome data, especially for barley, wheat and rye. We enhanced CrowsNest, a tool to visualize syntenic relationships between genomes, with data from the wheat sub-genome progenitor *Aegilops tauschii* and added functionality to the PGSB RNASeqExpressionBrowser. GenomeZipper results were integrated for the genomes of barley, rye, wheat and perennial ryegrass and interactive access is granted through PlantsDB interfaces. Data exchange and cross-linking between PlantsDB and other plant genome databases is stimulated by the transPLANT project (<http://transplantdb.eu/>).**

## INTRODUCTION

With the availability of (cost-) efficient next-generation sequencing technologies as well as novel bioinformatic assembly and analysis strategies, plant genome sequencing has entered a new era. Recently, sequence drafts were gener-

ated for the large and complex genomes of some of the most economically important cereal crops including barley (1), wheat (2,3) and rye (4). This data significantly promotes our understanding of plant evolution and assists researchers in the much-needed improvement of crops (5). However, due to large genome sizes, high repeat content and polyploidy, genome sequence data generated within these initiatives typically remain unfinished and fragmented for the time being, with heterogeneous and high-volume data sets associated. These challenges strongly ask for dedicated and powerful data integration infrastructures which should also provide intuitive and comprehensive user access. Both data integration and easy access to the data together with the availability of genome sequence data from a variety of related species including models and industrial crops have shown their potential in addressing important biological questions (6). Unfortunately, database solutions and existing interfaces for (finished) plant model genome sequence data often cannot fully account for the storage and representation of these complex genome data, e.g. virtual gene orderings or assemblies of genes on representative model organisms.

Here, we describe the latest updates to the PGSB (Plant Genome and Systems Biology; formerly MIPS) PlantsDB database framework with a special focus on the integration and representation of complex cereal genome data as well as on tools for the (comparative) analysis of plant genomes. Further updates include simplified user navigation and search interfaces/options. PlantsDB is part of the EU-funded transPLANT network which facilitates the exchange, integration and virtual aggregation of plant genome data from distributed resources as well as the development of common standards and protocols.

PGSB PlantsDB can be accessed at <http://pgsb.helmholtz-muenchen.de/plant/index.jsp>.

\*To whom correspondence should be addressed. Email: manuel.spannagl@helmholtz-muenchen.de

†These authors contributed equally to the work as first authors.

### PLANTSDB – new search options and tools

To facilitate intuitive and fast keyword searches on a set of plant gene (function) descriptions, a new search interface was created and linked from the PlantsDB front page ('PlantsDB Search & BLAST'; Figure 1). For a total of 13 different plant genomes and their corresponding gene predictions, human readable descriptions ('functional descriptions') were either derived from curated resources or computed using the AHRD ('Automatic assignment of human readable descriptions') tool (6). A filtering option enables searches for specific organisms, combinations of organisms or the entire set. From the results, links to the corresponding and more detailed gene reports, either within PlantsDB or to external reference databases, are provided.

We also implemented a completely new BLAST (7) interface to assist users with their search for homologous sequences across a broad range of plant species. Sequence similarity searches are possible against databases from 18 different species, including wheat (3), several wheat relatives/progenitor genomes (3,8,9), barley (1) and many more crop and model plants. Depending on the input sequence type and species, formatted BLAST databases were made available for coding sequences (CDS), protein sequences and genomic DNA sequence. After a search is completed, hits are visualized in a hierarchical view (Figure 2) and colored with respect to the query sequence identity. Furthermore, links to the corresponding databases with additional information about a specific gene are provided for all hits identified.

Available as a third search component, the transPLANT genome resources registry provides access to a comprehensive, curated and up-to-date collection of plant genome databases and resources. More details on this resource are given in the PlantsDB transPLANT section. Figure 1 shows a screenshot of the updated search functions and the BLAST interface available at PGSB PlantsDB.

In its latest version, both the layout and user navigation of PGSB PlantsDB have been completely re-worked to allow for easier, faster and more intuitive data access.

### PLANTSDB – *triticeae* instances

Many agronomically important crops such as barley, wheat and rye belong to the family of *triticeae* plants (10,11). The genomes of these crops are typically characterized by their large overall size, high repeat content and complex genetics. With ~17.1 GB in size the allo-hexaploid genome of bread wheat is more than five times larger than the human genome, exceeded also by the genome sizes of barley (~5.1 GB) and rye (~8 GB) (12,13). Assembly of these genomes is hampered by an exceptionally high repeat content (up to ~80% in wheat (3)) on the one hand and, for wheat, by the presence of three highly similar homologous sub-genomes on the other hand.

Nevertheless, draft genome sequences for barley (1), wheat (2,3) and rye (4) were released recently, each generated by dedicated sequencing efforts within international consortia. Novel bioinformatic strategies were used to analyze these genomes and derive the protein-coding gene complements (14). Within these projects, numerous heteroge-

neous data sets were generated and, in part, integrated with other data from the same or related organisms.

As a consequence, a number of specialized interfaces, viewers and tools were set up within PlantsDB in the past to accommodate novel data types, analysis results, raw data or integrated data sets for cereal genomes (15). This includes interfaces and tools to search, visualize and mine the 5x coverage 454 sequence of the bread wheat genome generated by an UK consortium in 2012 (2,15).

Recently, the International Wheat Genome Sequencing Consortium (IWGSC) released a chromosome-arm sorted whole-genome sequence draft for the bread wheat genome (cultivar Chinese Spring) together with the draft sequences of several wheat progenitors and wheat relatives (such as durum wheat) (3). Gene prediction and functional annotation for all these genomes were performed by PGSB and resulting data is stored within PlantsDB. Online access via PlantsDB standard interfaces and tools to all the IWGSC gene predictions including relatives/progenitor genomes is currently prepared and will be available soon. To assist the cereal research communities, we set up structured and comprehensive FTP download centers for both the barley and the wheat genome ([ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public\\_data/](ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/); <ftp://ftpmips.helmholtz-muenchen.de/plants/wheat/IWGSC/>). Within these, not only gene calls are available for download in several formats but also expression data, whole-genome sequence assemblies, physical and genetic maps plus their integration/anchoring, POPSEQ data (16), GenomeZipper data and repeat annotation.

While these data represent valuable resources for breeders and experimental plant researchers, data interpretation and analysis remains challenging in cereal genomes due to fragmented sequences and the absence of a physical position for a large number of genes. Here, the GenomeZipper concept (17) provides a powerful approach to create linearly ordered, information-rich scaffolds of cereal genomes. GenomeZippers are constructed by making use of gene orders found to be conserved over large portions ('synteny') between grass genomes (10,11) and incorporates chromosome sorting, next generation sequencing, array hybridization as well as fl-cDNAs, ESTs and genetic markers. Using this strategy, a total of 49 053 wheat genes could be assigned in the GenomeZipper on corresponding chromosomes recently. GenomeZippers have been constructed for barley (17), bread wheat (3), rye (4) and perennial ryegrass (18) so far and all data including raw sequences have been integrated into PlantsDB along with dedicated visualization interfaces and search options. GenomeZipper data can be queried online by any anchored gene model from one of the grass model organisms (Brachypodium, Sorghum and rice) as well as by any anchored genetic elements such as ESTs, reads or fl-cDNAs. All GenomeZipper results and data are also available for batch download in Excel and/or CSV format from PlantsDB to facilitate in-depth analyses (Barley: <http://pgsb.helmholtz-muenchen.de/plant/barley/gz/download/index.jsp>; Wheat: <ftp://ftpmips.helmholtz-muenchen.de/plants/wheat/IWGSC/genomeZipper/>).



## BLAST / Search

The PGSB plant genomics group focuses on the analysis of plant genomes, using bioinformatic techniques. To store and manage the data, we developed a database, **PlantsDB**, that aims to provide a data and information resource for individual plant species. In addition **PlantsDB** provides a platform for integrative and comparative plant genome research.

## Search in PlantsDB by keyword or gene identifier:

search string (e.g. "Leucine-rich", "AT3G25250.1")

Aegilops tauschii (BGI)  
Arabidopsis lyrata (PGSB V1.0)  
Arabidopsis thaliana (TAIR10)  
Brachypodium distachyon (PGSB V1.2)  
Hordeum vulgare (IBSC2012)  
Medicago truncatula (IMGAGv4.0)

(multiple selections possible)

Search Reset

## Plant genome databases BLAST search:

Please upload a file containing FASTA sequences (fasta):

Browse... No file selected.

insert <FASTA> sequences

Program: **blastn**

Data-sets: AegilopsSharonensis\_MIPS\_V2.0\_CDS.fa  
AegilopsSpeltoides\_MIPS\_V2.0\_CDS.fa  
AegilopsTauschii\_BGI\_CDS.fa  
AegilopsTauschii\_BGI\_GENOMICmasked.fa  
AegilopsTauschii\_MIPS\_V2.0\_CDS.fa

Max target sequences: 1

Word size: 5

E-value: 10e-5

Gap Costs: Linear

Reset Submit

## News &amp; Updates

**Science**

Major new findings on the structure and composition of the wheat genome, its evolution and the transcriptional interplay in a hexaploid cereal have been published in *Science* on July 18th, 2014. [\[Link\]](#)

An article about "Reticulate Evolution of the Rye Genome" has been published online in *The Plant Cell* in October 2013.

THE BARLEY GENOME GOES OKTOBERFEST! This [video](#) shows the cooperation of Gatersleben (IPK) and Munich (PGSB) in barley research to cope with climate change.

An article about "Flow sorting and sequencing meadow fescue chromosome 4F" has been published online in *Plant Physiology* in October 2013.

An article about "Analysing complex Triticeae genomes-concepts and strategies" has been published online in *Plant Methods* in September 2013.

**nature.com**

An article about "Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation" has been published online in *nature* on Apr. 04, 2013.

**nature.com**

An article about the repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres has been published online in *nature* on Dec. 20, 2012.

**Figure 1.** Re-worked PGSB (Plant Genome and Systems Biology; formerly MIPS) PlantsDB search interfaces, providing intuitive keyword searches for gene identifiers, functional gene descriptions and plant genome resources. A new BLAST interface enables sequence similarity searches against genome and CDS/protein sequences from 18 different plant species.

### PLANTSDB – comparative genomics and genome analysis tools

*CrowsNest synteny viewer updates.* CrowsNest is an interactive tool to visualize syntenic relationships at macro and micro levels in plant genomes, especially between the genomes of model species and the more complex crop grasses. CrowsNest features a comparative mapping and computation pipeline closely associated with PlantsDB which incorporates genetic, physical and hierarchical (fin-

gerprinted contigs) maps. The tool facilitates the identification and closer investigation of chromosomal rearrangements, inversions and deletions at different resolutions and between two or more chromosomes and/or genomes and therefore promotes the transfer of knowledge between several plant species. CrowsNest views have been computed between genes from the model grasses *Brachypodium*, *Sorghum* and rice as well as to the barley genome (15). Lately, we extended the browser with the draft genome from

### Plant genome databases BLAST search:



**Figure 2.** PGSB BLAST search results visualization. Homology matches and alignment visualization of one particular bread wheat gene (Traes\_7AL\_OF8FA764F.1) and homologous regions to selected progenitor genomes and relatives. The colors depict the sequence identity, whereas the numbers in brackets represent the nucleotide length of the transcripts.

*Aegilops tauschii* (9), which is the putative progenitor of the bread wheat D sub-genome. Because of its reduced genome size and complexity when compared to the allo-hexaploid bread wheat genome, the *Aegilops tauschii* genome serves as good model for its more complex relative. Furthermore, in *Aegilops tauschii* many genes could be anchored on a genetic map (9), providing a robust and comprehensive framework for synteny studies between triticeae genomes and grass model genomes. CrowsNest can be accessed from the ‘PlantsDB Tools’ section as well as from the gene reports of species included in CrowsNest.

*Accessing gene expression data with the RNASeqExpression-Browser.* The RNASeqExpressionBrowser, which is part of PlantsDB (access via ‘PlantsDB Tools’), is a web tool for browsing and visualizing gene expression information (19). It allows searching genes in several ways, such as by

gene identifier(s), functional annotations or sequence homology. Here, we present an updated version including a variety of novel features. After quantifying (genome-wide) expression, researchers often make use of clustering strategies for grouping interesting genes, e.g. by hierarchical clustering or supervised machine learning. One of these approaches is network biology, which aims at exploiting the interaction information between pairs of genes and is becoming more and more of a standard approach for analyzing huge data sets. Methods such as weighted correlation network analysis (WGCNA, (20)) allow establishing co-expression networks where groups of closely connected genes, referred to as modules, partition the data into smaller and easier to manageable portions. As the construction of gene-centric networks has become an important component in many studies, the RNASeqExpressionBrowser was extended to support re-

lated tasks. In addition to the existing search methods, now also network module membership can be used as a search criterion (Figure 3). Additionally, a filtering for highly connected ‘hub’ genes is possible, as the expression of these genes might be representative for the remaining module members. An alternative approach for narrowing down the list of gene candidates is to study patterns of differential expression. For RNA-seq several tools, e.g. Cufflinks (21) or DESeq (22), have become widely used standards for applying statistics to extract regulated genes. With the latest version of the RNASeqExpressionBrowser it is now also equipped with methods for including and searching information about differential expression.

During expression analysis, annotating or commenting on genes of interest is an important but burdensome task. Therefore, we now included an optional comment feature, which enables adding textual comments to genes. These text entries can then be searched later, e.g. for including references or flagging gene candidates. The entire list of stored comments can also be exported as a PDF.

RNASeqExpressionBrowser is available for download and standalone installation on the client side as well as for two PlantsDB projects using barley expression data (1,23).

#### MIPS repeat element database (mips-REdat) and catalog (mips-REcat)

In our previous description of PlantsDB (15) we provided access to the Repeat Element Database (pgsb-REdat) and Repeat Element Catalog (pgsb-REcat) and offered various methods for browsing and downloading repetitive elements. Whereas the pgsb-REdat provides sequences only, especially of importance during the pre-processing of sequences during gene annotation efforts where genomic sequences is normally masked by repetitive elements to prevent false gene calls, the repeat catalog offers a classification of repetitive elements into main groups and, if the order of domains is known, into more specific repeat families. Both resources enable researches to either use the existing classification and sequences for an *in silico*-based detection of repeats or for analyzing more complex and nested repeat insertions.

In order to help researches in automatically detecting and quantifying repeats on their query sequences, we extended the BLAST server with a possibility to also compare query sequences against the 61 730 repeat sequences with known classification based on the latest database version 9.3. It represents an update to the previously reported version 9.0 and includes now additional 11 479 *Gossypium* repeats (24). As outcome of the BLAST search, homology-matching repeats are grouped into repeat families according to the repeat catalog. The alignments between query and repeats might then help to analyze the insertions of various repetitive elements. Access to pgsb-REdat and pgsb-REcat is granted through ‘PlantsDB tools’ or directly at <http://pgsb.helmholtz-muenchen.de/plant/recat/index.jsp>.

#### PLANTSDB – transPLANT

The EU-funded transPLANT (Trans-national Infrastructure for Plant Genomic Science) project is a collaborative effort to enhance data integration, management and analysis

of plant genome data and provides services to plant research communities by developing tailored visualization/analysis tools, data exchange protocols and technologies to interconnect major plant genome resources. A total of 11 different European institutions are organized within several transPLANT workpackages, with PlantsDB being part of the plant genome databases backbone. PGSB is also hosting a plant genome resources registry which is actively maintained and collects and annotates sequence-based databases and resources for both species of agricultural and economic importance and model plants. At date, a total of ~300 different plant genome resources are registered in the transPLANT registry giving access to associated information such as annotation version, database URL, tools available and data types stored.

Besides regular routine updates and ongoing manual collection and curation of publicly available plant genome database systems, users and database curators can now use an interactive submission and update system to either submit information on new resources or request updates and/or corrections on existing entries.

The registry can be queried both from the PGSB PlantsDB tools section (or directly at <http://pgsb.helmholtz-muenchen.de/plant/transplant/genomeResources.jsp>) and the official transPLANT web hub at EBI (<http://transplantdb.eu/>, synchronized with PlantsDB regularly). The transPLANT web hub at EBI also runs an integrated search engine which indexes a number of different PlantsDB services including direct links to CrowsNest views for a given gene identifier and gene reports.

Within transPLANT, a number of user training activities were conceived especially for the analysis and navigation through resources for complex cereal genomes. TransPLANT user training videos describing the use and analysis of distributed barley and wheat genome resources have been produced and are available from PlantsDB ([ftp://ftpmips.helmholtz-muenchen.de/plants/user\\_training/](ftp://ftpmips.helmholtz-muenchen.de/plants/user_training/)) and <http://www.transplantdb.eu/videos>.

#### CONCLUSION

The PGSB PlantsDB database framework has been significantly enhanced with new tools and interfaces over the last two years and plenty of new data have been integrated into the system, especially for the large and complex genomes of wheat, barley and rye. Close collaboration with user communities and ongoing commitments in several genome sequencing and analysis consortia have guided the development of both novel resources such as the GenomeZipper and facilities for the comparative analysis of data such as CrowsNest and RNASeqExpressionBrowser. We also completely re-worked the database design and the menu structure as well as numerous interfaces and tools such as our BLAST interface to provide a more intuitive and seamless user navigation.

The transPLANT project provides a framework for the integration of heterogeneous plant genome data as well as for data exchange and common services, e.g. integrated searches over multiple databases. We are planning to further utilize transPLANT services in PlantsDB, e.g. by aggregat-

**Figure 3.** Project entry page of the RNASeqExpressionBrowser with extended search features for gene-to-group mapping ('network modules') and for lists of differentially expressed genes.

ing external genome data from transPLANT partners via web services.

## ACKNOWLEDGEMENT

We want to thank all our collaboration partners and data contributors. For barley these include: all collaborators within the International Barley Genome Sequencing Project (IBSC), IPK (Nils Stein, Uwe Scholz), FLI (Matthias Platzer), JKI (Frank Ordon), James Hutton Institute (Robbie Waugh), IEB (Jaroslav Doležel), University of Udine (Michele Morgante), UCR (Tim Close), ACPFG (Peter Langridge) and many more. For wheat these include: all collaborators within the International Wheat Genome Sequencing Project (IWGSC), University of Liverpool (Neil Hall, Anthony Hall, Rachel Brenchley), University of Bristol (Keith J. Edwards, Gary L.A. Barker), John Innes Centre (Michael W. Bevan) and many more. We gratefully acknowledge Deutsche Forschungsgemeinschaft (DFG) funding to project SFB924 Molecular mechanisms regulating yield and yield stability in plants. We also want to thank all partners in the transPLANT project (esp. Paul Kersey&Dan Bolser, EBI, IPK and URGI INRA).

## FUNDING

European Commission (framework 6 programme) within the Grain Legumes Integrated Project (GLIP) and EU-SOL project as well as within the framework 7 programme in the TriticeaeGenome project and transPLANT project (European Commission within its 7th Framework Programme under the thematic area 'Infrastructures', contract number 283496); German Ministry for Education and Research (BMBF) within the GABI and Plant for the future projects TRITEX and BARLEX. Funding for open access charge: Helmholtz Association Germany.

*Conflict of interest statement.* None declared.

## REFERENCES

- International Barley Genome Sequencing, C., Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K. *et al.* (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.
- Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D. *et al.* (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, **491**, 705–710.
- International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.
- Martis, M.M., Zhou, R., Haseneyer, G., Schmutzer, T., Vrana, J., Kubalakov, M., Konig, S., Kugler, K.G., Scholz, U., Hackauf, B. *et al.* (2013) Reticulate evolution of the rye genome. *Plant Cell*, **25**, 3685–3698.
- Mochida, K. and Shinozaki, K. (2013) Unlocking Triticeae genomics to sustainably feed the future. *Plant Cell Physiol.*, **54**, 1931–1950.
- Tomato Genome, C. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.
- Altschul, S.F., Gish, W., Miller, W., Myers, E. W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Ling, H.Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y. *et al.* (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*, **496**, 87–90.
- Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., Appels, R., Pfeifer, M., Tao, Y., Zhang, X. *et al.* (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, **496**, 91–95.
- Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Cereal genome evolution. Grasses, line up and form a circle. *Curr. Biol.*, **5**, 737–739.
- Devos, K.M. (2005) Updating the 'crop circle'. *Curr. Opin. Plant Biol.*, **8**, 155–162.
- Dolezel, J. and Bartos, J. (2005) Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot. (Lond)*, **95**, 99–110.
- Bartos, J., Paux, E., Kofler, R., Havrankova, M., Kopecky, D., Suchankova, P., Safar, J., Simkova, H., Town, C.D., Lelley, T. *et al.* (2008) A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. *BMC Plant Biol.*, **8**, 95.
- Spannagl, M., Martis, M.M., Pfeifer, M., Nussbaumer, T. and Mayer, K.F. (2013) Analysing complex Triticeae genomes - concepts and strategies. *Plant Methods*, **9**, 35.
- Nussbaumer, T., Martis, M.M., Roessner, S.K., Pfeifer, M., Bader, K.C., Sharma, S., Gundlach, H. and Spannagl, M. (2013) MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.*, **41**, D1144–D1151.
- Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., Munoz-Amatriain, M., Close, T.J., Wise, R.P., Schulman, A.H. *et al.* (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.*, **76**, 718–727.
- Mayer, K.F., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H. *et al.* (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, **23**, 1249–1263.

18. Pfeifer, M., Martis, M., Asp, T., Mayer, K.F., Lubberstedt, T., Byrne, S., Frei, U. and Studer, B. (2013) The perennial ryegrass GenomeZipper: targeted use of genome resources for comparative grass genomics. *Plant Physiol.*, **161**, 571–582.
19. Nussbaumer, T., Kugler, K.G., Bader, K.C., Sharma, S., Seidel, M. and Mayer, K.F. (2014) RNASeqExpressionBrowser—a web interface to browse and visualize high-throughput expression data. *Bioinformatics*, **30**, 2519–2520.
20. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
21. Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L. and Pachter, L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.*, **31**, 46–53.
22. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
23. Dey, S., Wenig, M., Langen, G., Sharma, S., Kugler, K.G., Knappe, C., Hause, B., Bichlmeier, M., Babaeizad, V., Imani, J. *et al.* (2014) Bacteria-triggered systemic immunity in barley is associated with WRKY and ETHYLENE RESPONSIVE FACTORS but not with salicylic acid. *Plant Physiol.*, **166**, 2133–2151.
24. Paterson, A.H., Wendel, J.F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K.C., Shu, S., Udall, J. *et al.* (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, **492**, 423–427.