BMC Bioinformatics

CrossMark

RiceMetaSys for salt and drought stress responsive genes in rice: a web interface for crop improvement

Maninder Sandhu[1,2†], V. Sureshkumar[1,3†], Chandra Prakash[1†], Rekha Dixit[2,4], Amolkumar U. Solanke[1], Tilak Raj Sharma[1], Trilochan Mohapatra[5] and Amitha Mithra S. V. [1*]

Abstract

Background: Genome-wide microarray has enabled development of robust databases for functional genomics studies in rice. However, such databases do not directly cater to the needs of breeders. Here, we have attempted to develop a web interface which combines the information from functional genomic studies across different genetic backgrounds with DNA markers so that they can be readily deployed in crop improvement. In the current version of the database, we have included drought and salinity stress studies since these two are the major abiotic stresses in rice.

Results: RiceMetaSys, a user-friendly and freely available web interface provides comprehensive information on salt responsive genes (SRGs) and drought responsive genes (DRGs) across genotypes, crop development stages and tissues, identified from multiple microarray datasets. 'Physical position search' is an attractive tool for those using QTL based approach for dissecting tolerance to salt and drought stress since it can provide the list of SRGs and DRGs in any physical interval. To identify robust candidate genes for use in crop improvement, the 'common genes across varieties' search tool is useful. Graphical visualization of expression profiles across genes and rice genotypes has been enabled to facilitate the user and to make the comparisons more impactful. Simple Sequence Repeat (SSR) search in the SRGs and DRGs is a valuable tool for fine mapping and marker assisted selection since it provides primers for survey of polymorphism. An external link to intron specific markers is also provided for this purpose. Bulk retrieval of data without any limit has been enabled in case of locus and SSR search.

Conclusions: The aim of this database is to facilitate users with a simple and straight-forward search options for identification of robust candidate genes from among thousands of SRGs and DRGs so as to facilitate linking variation in expression profiles to variation in phenotype.
Database URL: http://14.139.229.201

Keywords: Rice, Meta-analysis, Salinity, Drought, DNA markers

Background

Rice has the dual distinction of being a staple food crop for nearly 50% of world population and a genomic model crop for monocots which includes wheat and corn, the former being a staple cereal, and the latter a major source of animal nutrition [1]. In the last six decades, rice production has kept its growth in pace with the raising global food demand. However, rice production is supposed to further increase by 0.6 to 0.9% per year till 2050 to feed the additional 2 billion people expected to inhabit the earth by then [2–4]. Besides this major challenge of improving productivity, drought and salinity stress have emerged as the most important abiotic stresses that could endanger the sustainability of rice production. Since salinity and drought stress tolerance in rice are complex traits, in terms of their inheritance as well as molecular mechanism, researchers have

* Correspondence: amithamithra.nrcpb@gmail.com
†Equal contributors
[1]ICAR-National Research Centre on Plant Biotechnology, LBS Building, Pusa Campus, New Delhi 110012, India
Full list of author information is available at the end of the article

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 2 of 11

been trying to address this problem by using genetic and genomic approaches [5–8].

One of the major approaches followed for dissecting complex traits such as drought and salt tolerance is the identification of QTLs by preliminary genetic mapping followed by fine mapping and identification of the candidate gene(s). Though this is a robust approach, it is laborious and time-consuming. With the advances in genomics, the entire process can be accelerated, especially, the steps after coarse mapping, even in crops not traditionally amenable for map-based cloning such as oil palm [9, 10]. In species where high-quality genome sequence information is available such as human, rice and Arabidopsis, microarray hybridization based genome-wide expression analysis is a very popular and useful technique to understand functional genomics [1, 11]. Expression microarray studies have been effectively used to characterize mutants and transgenic plants by comparing them with wild type [12–15]. Microarray generally identifies a large number of differentially expressed genes (DEGs) even in closely related individuals such as isogenic lines contrasting for a single trait [12]. Hence, one of the proven and effective ways to dissect complex traits is to combine genetic mapping with genome-wide transcriptome profiling of the parental genotypes which can help to narrow down the candidate gene(s) underlying the functional polymorphism in the QTL [13]. When huge numbers of genes from different biological materials are implicated in expression of a trait, meta-analysis provides a cost effective way to identify robust candidate gene(s) for trait improvement through breeding.

Meta-analysis aims at identification of statistically robust candidate genes from the already existing information such as the expression microarray data available in the public domain. In rice, using the microarray data, several publically accessible databases like OryzaExpress ([16], http://plantomics.mind.meiji.ac.jp/OryzaExpress/), RicePLEX ([17], http://www.plexdb.org/plex.php?database=Rice), Rice Oligonucleotide Array database (ROAD) [18], RiceSRTFDB ([19], http://www.nipgr.res.in/RiceSRTFDB.html), Oryzabase ([20], http://shigen.nig.ac.jp/rice/oryzabase/), QlicRice ([21], http://nabg.iasri.res.in:8080/qlic-rice), OryGenesDB ([22], http://orygenesdb.cirad.fr/), RiceXPro ([23], http://ricexpro.dna.affrc.go.jp/) and qTeller ([24], http://qteller.com) and commercial platforms like Genevesigator and GeneMapper have been constructed. Of the freely available databases, ROAD is the most proficient and complete tool for meta-analysis of microarray data since it comprises of microarray data from multiple platforms, tissues, growth conditions and genotypes. Users can carry out gene expression analysis, co-expression and GO enrichment analysis and visualize the genes in a heat map. However, currently this database is not under maintenance and is not accessible. Orygene database is a functional genomic tool based on reverse genetics and hence offers flanking sequence tag

(FST) based search. Oryzabase is a genome browser which provides information about rice development and anatomy of rice varieties, especially, wild varieties of rice. The qTeller database gives the list of genes in a QTL or a particular genomic interval whereas QlicRice lists the QTLs for various abiotic stresses, and different QTLs intervals. Though ROAD is a very useful forward functional genomic tool for identifying candidate genes for the trait of interest, for a plant breeder, ROAD is either not directly useful or very complex to use. On the other hand, qTeller and QlicRice are user-friendly but have not integrated the microarray data with QTL intervals. The commercially available tools such as Genevestigator are though highly informative, again intensive like ROAD and expensive to use.

To fine map the large QTL regions, plant breeders primarily look for polymorphisms between the parents of the mapping population in that defined region, in addition to the search for candidate genes using expression and bioinformatics approaches. Though SNPs are the most abundant and routinely used markers in vogue with low cost per data point [6], for investigating a well-defined genomic region in a cost-effective manner in a mapping population, the co-dominant and PCR-based microsatellites markers (also known as simple sequence repeats; SSRs) and intron length polymorphisms (ILP) or intron spanning markers (ISM) are more suitable. A database is readily available for searching ILP and ISM polymorphisms in any given gene but not SSRs [25].

Hence, we have constructed a database, named Rice-MetaSys, especially intended for breeders, which directly combines the rice microarray data for salt and drought tolerance from both stress tolerant and susceptible genotypes along with their physical location and marker data. Since crop improvement researchers mainly concentrate on one trait at a time, we made the database trait specific. Though the focus is on salt and drought tolerance in the current version of RiceMetaSys, we intend to add more such important traits namely tolerance to leaf and panicle blast and high temperature. The purpose of microarray technology which is to enable biologists to study expression variation at a whole-genome level and link it to phenotypic variation [26] can be assisted by such efforts.

Construction and content
Data source
Microarray meta-analysis involves combining multiple independent but related microarray datasets into a meaningful context based profiles. Two or more experiments run on the same crop and treatment is not a sufficient enough justification for combining such datasets. Reproducibility and homogeneity of results across laboratories and datasets is also necessary, and in this context, Affymetrix platforms are considered more robust than other platforms

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 3 of 11

[19]. Hence, the Affymetrix Microarray datasets comprising of 5 experiments (110 samples) pertaining to salinity and 6 experiments (131 samples) pertaining to drought treatment were retrieved from NCBI GEO database [27]. Expression data for salt stress was from nine varieties (Agami, M103, FL478, IR29, IR63731, Pokkali, CSR27, MI48 and IR64), representing vegetative and seedling growth stages and various sample tissues such as root, leaf and seedling (Additional file 1: Table S1). For drought stress, the expression datasets were from 10 different rice genotypes (Azucena, Bala, IRAT109, ZS97, IR64, Dhagaddeshi, IR20, Moroberekan, Nagina 22, and Nipponbare), representing eight different growth stages from seedling to panicle elongation, and seven different tissue samples covering vegetative to floral parts (Additional file 1: Table S1). The nature of the response of a genotype in terms of tolerance and sensitivity to a particular stress is also indicated in this table.

Data processing and gene expression analysis

Since the treatment across experiments is not uniform, pre-processing (background correction and removal of batch effects) was carried out prior to gene expression analysis. Pre-processing of the microarray raw data from drought datasets was done using RMA (Robust Multi-Array Average) method and salinity datasets were normalized by log 2 transformation using R script from GEO2R. Non-experimental variation (batch effects owing to inter-laboratory and inter-batch differences) was removed using ComBat [28] tool in R. We have divided our data into drought and salt groups and removed batch effects separately. For each dataset, gene expression analysis was done using limma package v.3.28.21 and the R script from GEO2R with some slight modifications [29]. We have kept adjusted *p*-value 0.01 (for drought) 0.05 (for salt), Log FC value <−1 to 1>, and Average Expression >8 for both drought and salt microarray data sets.

Database design

Affymetrix IDs of the salt and drought DEGs were converted to MSU7 IDs and RAP IDs by using OryzaExpress (http://bioinf.mind.meiji.ac.jp/OryzaExpress/ID_converter.php). A total of 1558 probe set IDs, either from salt responsive genes (SRGs) or drought responsive genes (DRGs) identified through analysis, did not have corresponding locus or gene IDs and hence were not considered for further processing. Physical positions and annotations were fetched from TIGR ([30], http://rice.-plantbiology.msu.edu/). Microsatellites present in DRGs and SRGs were identified using BatchPrimer3 tool ([31], http://batchprimer3.bioinformatics.ucdavis.edu/cgi-bin/batchprimer3/batchprimer3.cgi) which not only identifies the microsatellites but also designs primers for the amplifi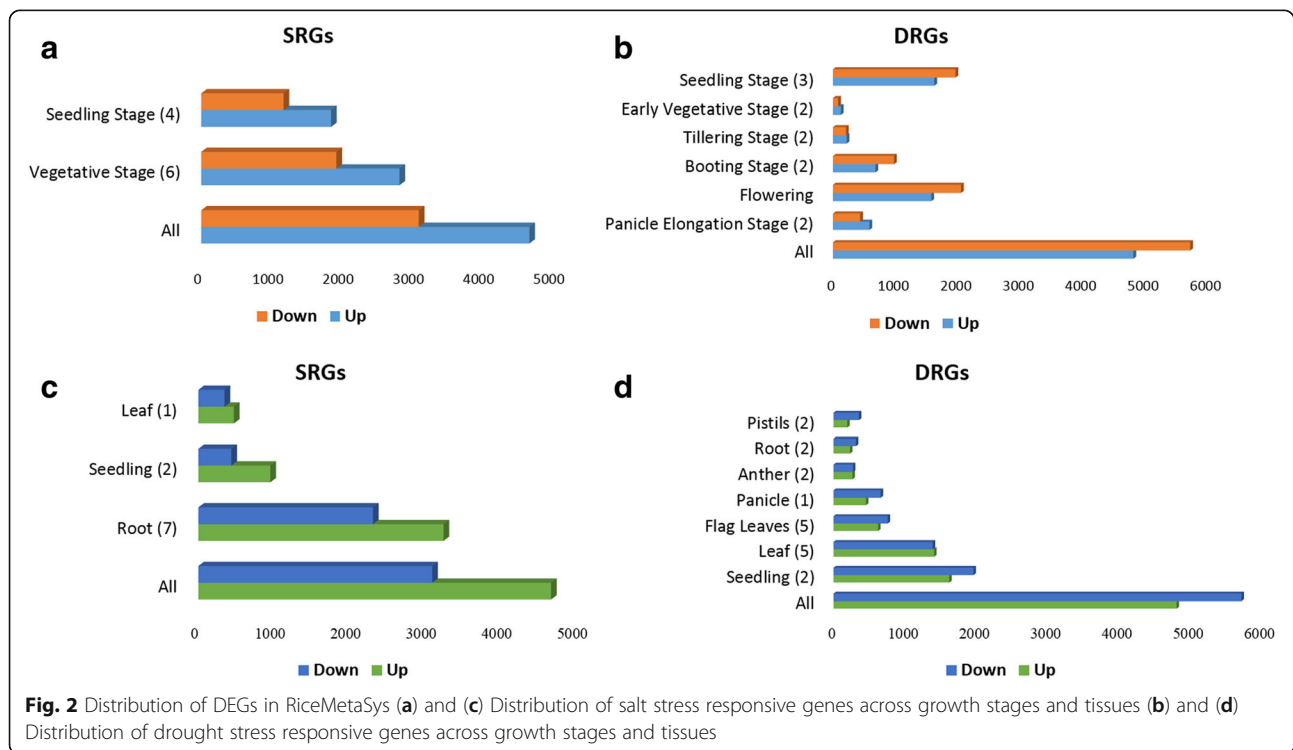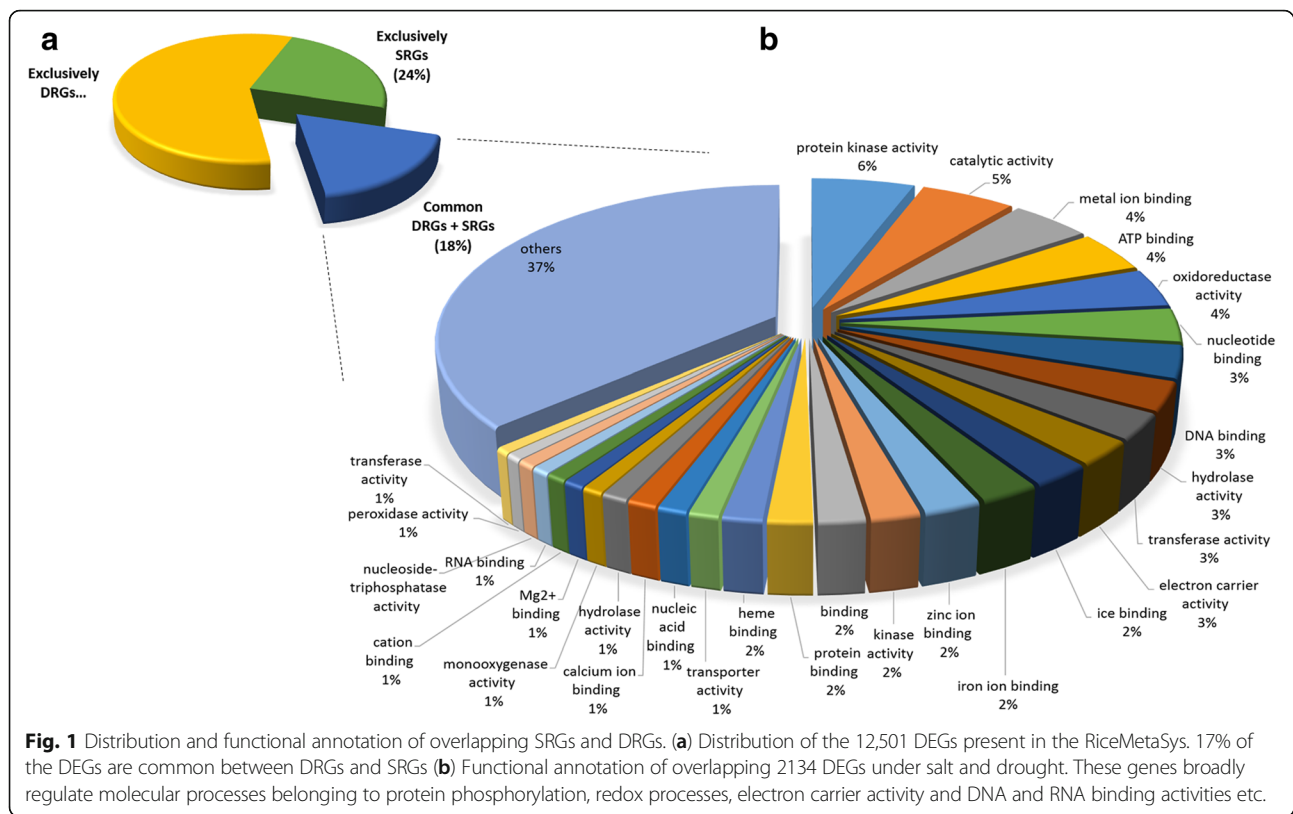cation of SSR fragments. The schematic representation of metadata analysis and RiceMetaSys design is given in Additional file 2: Figure S1. Server side scripting language used for RiceMetaSys was PHP with HTML5 in the front end and CSS with MySQL relational database at the backend. User interface framework employed was JQuery and JavaScript. Chart.js was used to generate graphs of expression profile of user-selected SRGs and DRGs in single or multiple rice genotypes. An external link option is provided in the SRG and DRG homepage to perform Gene Set Enrichment analysis (GSEA) and construct heat maps. Another external link enabled in the database is that of intron length based markers in rice. Database web server is XAMPP (Apache, MySQL, PHP, and Perl). The database is hosted in the server environment, FUJITSU PrimeRGY-Rx600S6 and Windows operating system. The database can be accessed at http://14.139.229.201.

Utility and discussion
Data statistics

RiceMetaSys contains a total of 3120 salt responsive genes (SRGs) identified from salt microarray datasets and 9381 drought responsive genes (DRGs) from drought microarray datasets, after removing the duplicate entries (genes) identified across different studies within an abiotic stress group. Since both drought and salinity stresses induce osmotic stress in plants [32, 33], we searched for the genes common to both SRG and DRG datasets and found 2134 such genes (Fig. 1a). Interestingly, SRG set had only 986 (31.6%) unique salt specific genes, suggesting that imparting drought tolerance to plants would more often than not enhance their salinity tolerance too. GO ontology functional annotation of the 2134 common genes revealed that the maximum number of genes encoded undefined expressed proteins followed by zinc finger domain containing proteins and cytochromes (Fig. 1b). Thus the undefined expressed proteins encoding genes are a major class of candidate genes to target in combatting abiotic stress tolerance. For all the three groups namely SRG, DRG and genes commonly regulated in both the stresses, separate links (tabs) have been provided in the homepage of RiceMetaSys.

The number of up and down regulated DEGs under salt and drought stress had a similar pattern i.e., the number of upregulated genes were more than downregulated genes (Fig. 2a). Based on the growth stage and tissue used in the experiments, the SRGs and DRGs were appropriately grouped. Comparison of DEGs among these groups revealed that this pattern was not true across the stages and tissues. The number of SRGs identified across tissues corresponded with the number of experiments conducted with a particular type of tissue (Fig. 2b). For instance, in salt microarray experiments, the root was the most often used tissue (7 times) and

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 4 of 11

Fig. 1 Distribution and functional annotation of overlapping SRGs and DRGs. (**a**) Distribution of the 12,501 DEGs present in the RiceMetaSys. 17% of the DEGs are common between DRGs and SRGs (**b**) Functional annotation of overlapping 2134 DEGs under salt and drought. These genes broadly regulate molecular processes belonging to protein phosphorylation, redox processes, electron carrier activity and DNA and RNA binding activities etc.

Fig. 2 Distribution of DEGs in RiceMetaSys (**a**) and (**c**) Distribution of salt stress responsive genes across growth stages and tissues (**b**) and (**d**) Distribution of drought stress responsive genes across growth stages and tissues

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 5 of 11

hence the number of SRGs from this tissue was more (Fig. 2c). Similarly, in DRGs, the DEGs were more in leaves collected at vegetative stage and entire seedling assays as the former was the most frequently sampled tissue (8 times) and the latter had the entire plant (Fig. 2d). Under drought, at flowering stage and in flag leaf and anther tissues, proportion of downregulated genes was slightly higher (53.65%, 52.75% and 63.15%; Fig. 2b and d). Similarly, under salinity, leaves had higher proportion of down regulated genes (65.5%; Fig. 2c). Comparison of DRGs in reproductive tissues revealed that the up and down regulated DEGs were nearly equal in pistils (51.3% and 48.7%) while in anthers, the number of down-regulated genes was nearly twice that of up regulated ones (63.15% and 36.85%).

GO annotation of the SRGs and DRGs revealed almost similar proportion of genes under cellular components and pathways. However, under molecular functions, and biological processes the abundance was more in the former than the latter for DRGs (46.22% and 29.42%) and vice-versa for SRGs (31.01% and 45.17%) (Fig. 3a and b). Comparison of known salt tolerant and susceptible genotypes (Additional file 1: Table S1, and Fig. 3a and b) revealed that more SRGs were from salt tolerant genotypes (143) than susceptible genotypes (116). In the case of DRGs, the trend was reverse with more number of DRGs found in drought sensitive genotypes (621 against 567). While under drought the number of up

and down regulated across tolerant and susceptible genotypes was comparable, in salinity the number of up-regulated genes were more in salt tolerant genotypes than all the other three classes. Under metabolic processes, the number of upregulated SRGs in tolerant genotypes was the highest (Fig. 3a). Under cellular processes, the number of downregulated DRGs in susceptible genotypes was the highest (Fig. 3b).

A total of 12,070 SSRs were found in DRGs (8451) and SRGs (3619) meeting the following parameters set for their mining: dinucleotide units repeated at least 6 times, trinucleotide motifs 5 times, tetranucleotide repeats 5 times, pentanucleotides repeats 3 times and hexanucleotide repeats 3 times. Trinucleotide motifs were the most abundant in both DRGs (51%) and SRGs (50%) as already reported in rice [34]. However, dinucleotide repeats in SRGs and DRGs were much lower (Fig. 4; 21.1% and 20.7%) as compared to previous reports [34, 35]. Tetranucleotides were the least abundant in both DRGs (2.5%) and SRGs (2.6%). Nearly, one-fifth (24.6%) of the repeats were class I microsatellites.

Database features vis-à-vis available datasets

More often than not, researchers focus on a specific trait and aim to understand the molecular mechanisms governing that trait. Further, crosstalk at the molecular level is extremely well known across stress responses [36, 37]. Hence, besides separate links for SRGs and DRGs,
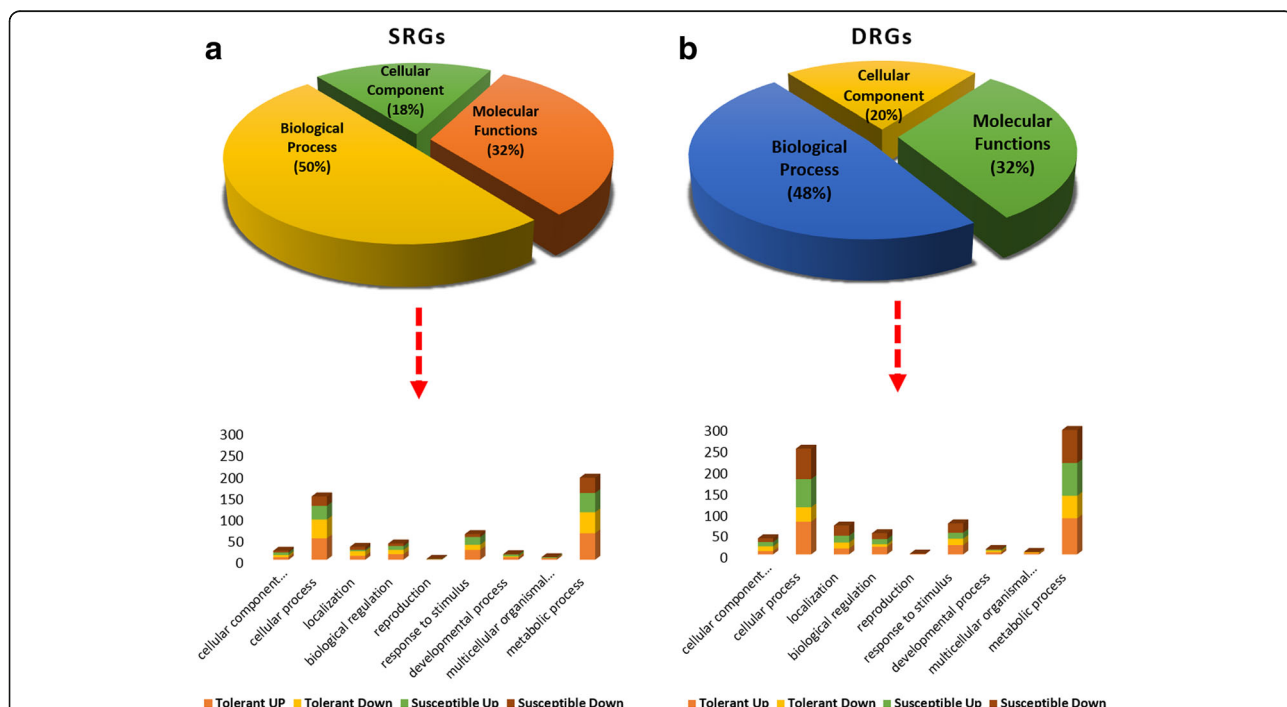
Fig. 3 Gene Ontology of the identified stress responsive genes (**a**) Majority of the identified SRGs corresponds to biological process (45.17%) followed by molecular function (31%) (**b**) The distribution pattern was vice-versa for DRGs with major proportion of the identified genes in the category molecular function (46.2%) followed by biological process (29.4%)
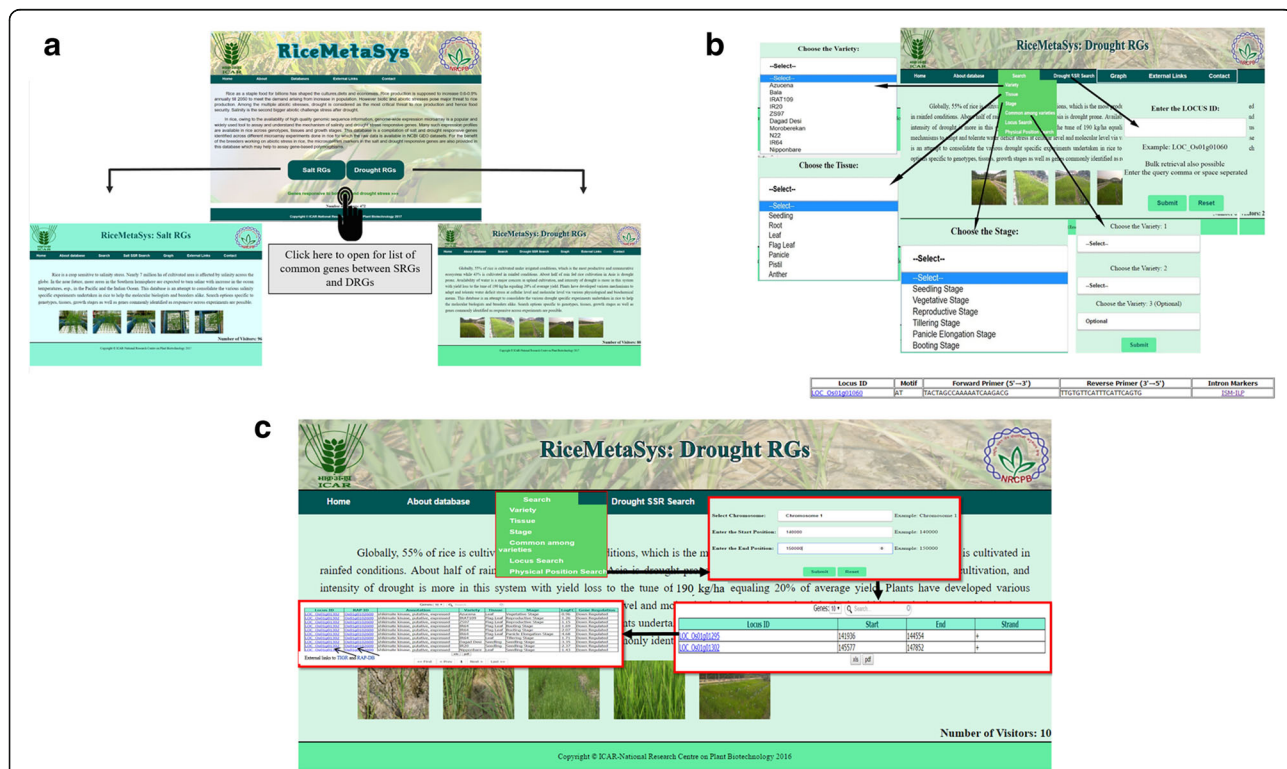
Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 6 of 11
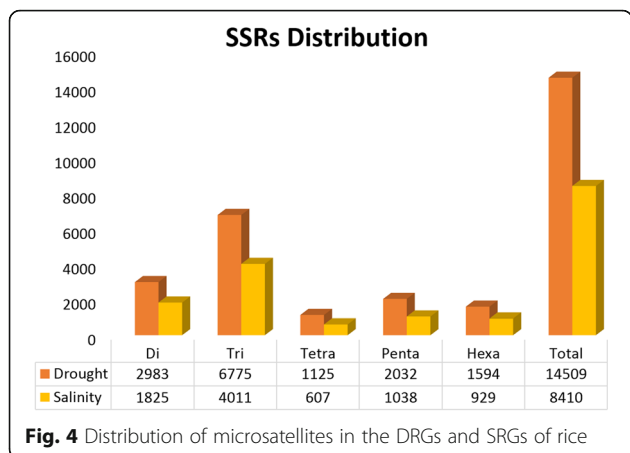
Fig. 4 Distribution of microsatellites in the DRGs and SRGs of rice

with genotype specific search, both growth stage and tissues specific searches were enabled in our RiceMetaSys database in all the three links (Fig. 5b). Any desired stages/tissue/variety can easily be selected from the drop down menu by the user under appropriate search option. The output gives a list of stress responsive genes with their gene IDs (LOC_ID), annotation, log fold change (FC) and the direction of regulation (up or down) specific to the search option (Fig. 5b). Data can be sorted according to FC values or direction of regulation of DEGs by clicking on each heading as per user's requirement. To enable this, the output format has been kept simple and in text format with limited graphics. Another important feature enabled in the RiceMetaSys web interface is the nature of output from stage and tissue specific search: rather than just a list of DEGs, complete information on the gene across genotypes is given with other details so that the importance of the gene can be easily deciphered (Fig. 5b). Visualization of output in multiples of 10 genes from 10 to 50 of genes (SRG/DRG) has also been enabled. In addition, the user has the choice of downloading the results in MS-Excel and PDF format.

another link for genes common to SRG and DRG has been provided in the home page of RiceMetaSys (Fig. 5a). Biologically, it is well known that the response to any stress is genotype, stage and tissue specific. For instance, the well-known salt tolerant QTL in chromosome 1 (Saltol) of rice confers tolerance only at the vegetative stage but not at reproductive stage [7]. Hence, along

Fig. 5 An overview of RiceMetaSys (**a**) Snapshot of the RiceMetaSys database showing the homepage with links to SRGs, DRGs and common genes between SRGs and DRGs. (**b**) Search options such as variety, tissue, stage, commonly expressed genes among varieties and SSRs. (**c**) Physical position search option and its output. Selecting the 'Physical position" search opens a window in which chromosome number and the genomic interval (start and end point) are to be provided as input by the user. This lists the stress responsive genes in the interval in another window. Selecting individual genes from this list provides detailed information on its stress responsiveness

From the breeders' and farmers' perspective, the stress incidence at the reproductive stage is more important than that at vegetative stage since the former affects both economic yield and quality of the produce more severely. Interestingly, from the available data, it was apparent that there were no microarray datasets available from reproductive stage or tissues for salt stress whereas in drought four of the six experiments analyzed had data from reproductive tissues or stage. Of late, QTLs for reproductive stage salinity stress tolerance have been mapped in rice [6, 7]. Thus, generating genome-wide expression data at reproductive stage would be very useful for fine mapping of the QTLs identified in those studies.

A comparative analysis of the available databases along with RiceMetaSys has been carried out based on multiple parameters such as general features, expression type, co-expression analysis, trait specificity, and marker type and output format (Table 1). ROAD database is the best tool available for expression analysis and covers most of the microarray experiments for salinity and drought. However, RiceMetaSys has more microarray experiment datasets (Affymetrix) for salinity and drought as ROAD database has not been updated since 2012 and is currently unavailable. Although ROAD database includes all biotic and abiotic traits for rice, expression analysis can be done with only one experiment at a time. Consequently, the meta-analysis in ROAD is not trait specific. The same issue exists with RicePLEX database as well. We have not enabled co-expression, pathway analysis and protein-protein interactions in our database because we wanted to keep it simple and user-friendly for the breeders. Still, an external link has been provided for Gene Set Enrichment analysis (GSEA) and construction of heat maps. Results (output of gene IDs) obtained from search performed with our database can be directly given as input to GSEA.

Common genes, locus and physical position search
Molecular mechanisms that impart tolerance to any abiotic stress can be either universal or genotype specific. The possibility of allelic diversity, epistasis and GXE interactions complicate the expression profile further. Thus, the robust candidate genes for tolerance could be the ones that have a similar pattern of expression in tolerant genotypes as against sensitive genotypes. Hence, comparison of SRGs and DRGs, up to three genotypes, has been enabled in RiceMetaSys which gives the list of commonly regulated genes across the genotypes selected (Fig. 5b). This search provision is also useful for short-listing of genes for their functional characterization. The 'common genes search across varieties' is a unique feature of RiceMetaSys.

For a researcher interested in a specific gene, for its plausible role in imparting salinity or drought stress tolerance, the 'Locus search' option is a convenient tool

(Fig. 5b). The LOC IDs have been hyperlinked with the genome browser for access to more information. Bulk retrieval of data is also possible in 'Locus search' without any limit on number of genes but per page view is restricted to a maximum of 50 genes for the sake of clarity.

For the analysis of genes present in the known and novel QTLs, it would be very useful if the stress responsive genes present in a given genomic interval are known. This would help in both fine mapping and gene validation (to pick the right candidate). RiceMetaSys makes this possible with the 'physical position search' tool (Fig. 5c). The workflow for using this option is explained in Additional file 3: Figure S2. Graphical representation of expression profiles of selected candidate genes, up to 10, in a single or multiple genotypes is also available in the database. The input required for this option is a list of locus IDs. This is a very useful tool to check whether a given candidate gene is functioning in a universal or variety specific manner (Fig. 6).

Once the list of stress responsive genes is available, the next logical and immediate step is to look for locus specific DNA markers in that interval so as to test for polymorphisms in the parents of the QTL mapping population for those markers. Though SNPs are the makers of choice [38, 39], fine mapping programs prefer simple-to-genotype markers that are also amenable to large scale genotyping. Both SSRs and Intron length spanning markers or intron length polymorphisms (ISM-ILP) fit this description perfectly [25, 40]. Hence, a separate tab for SSR search has been provided in the database. By submitting the list of LOC IDs found in a given physical interval, SSRs present, if any, in the genes would be displayed along with the SSR motif and primer information so that the polymorphisms can be surveyed by the researcher (Fig. 5c). If the researcher wants to look for ISM-ILP polymorphism in the SRGs or DRGs, an external link to ISM-ILP database (http://webapp.cab-grid.res.in/ismdb/database.html) has been provided with each LOC ID, under the SSR search tab. The marker polymorphisms identified can also be directly used for marker assisted selection in both back cross and recombinant breeding programs.

RiceMetaSys: Utility for rice breeders
Universal and robust candidate genes are preferred by breeders for exploitation in crop improvement. Using 'common variety search' tool and graphics tab for visualization of expression profile across varieties, it is possible for breeders to select the robust candidates (Fig. 5b). Further, they can select the DEGs in the known major QTL intervals by using the 'physical position search' option (Fig. 5C and Additional file 3:

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 8 of 11

Table 1 Comparison of main features of different rice expression databases

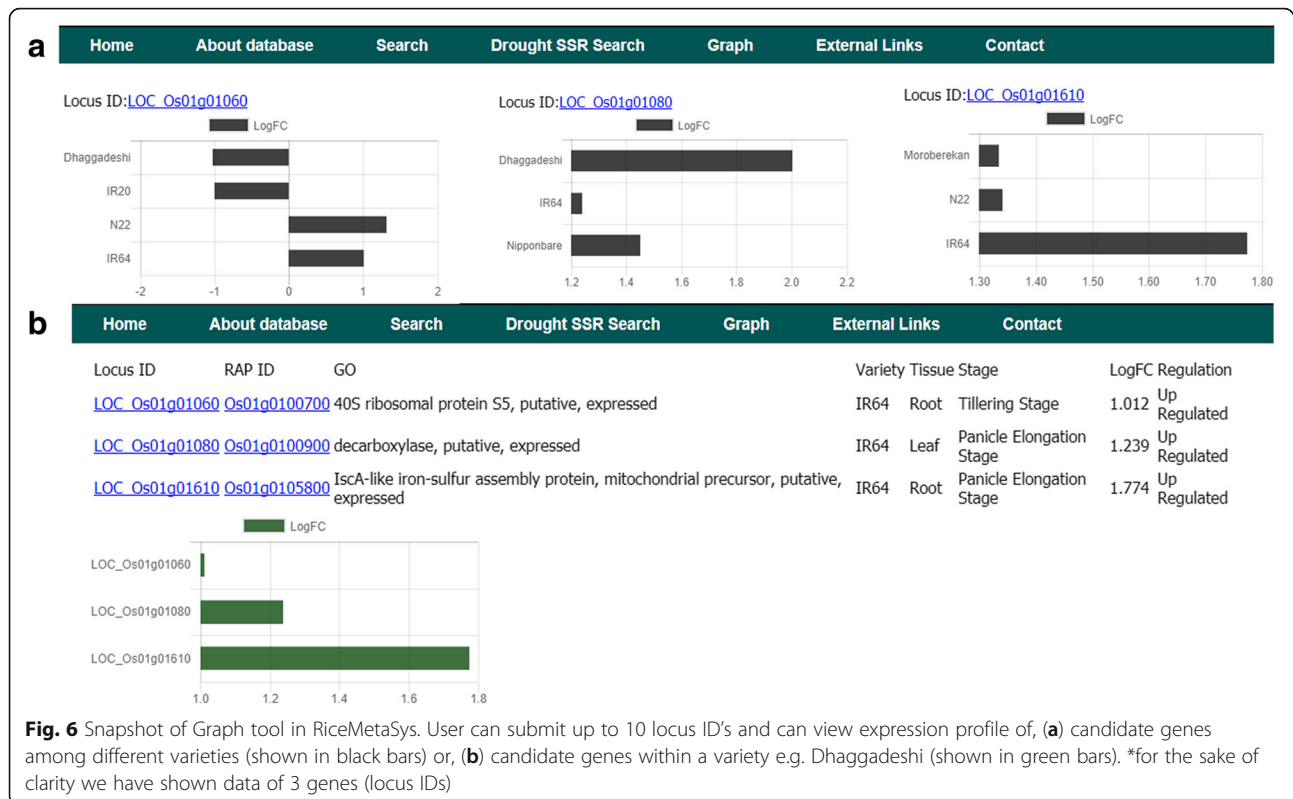| Parameter	RiceMetaSys	ROAD*	Rice SRTFdb	Rice-Plex	RiceXpro	Qteller	QlicRice
Tissue/stage/genotype specific expression	Yes	Yes	No	No (individual datasets)	Yes	Yes (need to select experiment)	No
Co-expression analysis	No(external link provided)	Yes	No	No	No	No	No
Trait specific search	Yes	Yes (but meta-analysis is not trait specific)	Yes	Yes	No	No	Yes (For QTLs)
Output format	Table and graphs	Heatmap, table and graphs	Table	Heatmap, table and graphs	Map chart and table	Table	Table
Marker information	SSRs, ISM-ILP	No	No	No	No	No	No
Bulk Acceptance/Retrieval	Yes	Yes	Yes	Yes	Yes	Retrieval possible but not acceptance	Yes
Other Details	Various search options for better comparison; Genes common between traits as well as among varieties can be retrieved; DEGs between two markers can be retrieved	Single and multiple platform probe search; Meta profiling possible	Focus on TFs; Common genes between traits can be retrieved	Based on rice and 15 other plant species; Homology among various species	Genes can be viewed from field/development and plant hormone microarray datasets	Based on expression studies in major crop species; Genes between two physical coordinates can be retrieved	QTL specific database; Genes in the QTL interval can be retrieved

*Currently not available

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 9 of 11

Fig. 6 Snapshot of Graph tool in RiceMetaSys. User can submit up to 10 locus ID's and can view expression profile of, (**a**) candidate genes among different varieties (shown in black bars) or, (**b**) candidate genes within a variety e.g. Dhaggadeshi (shown in green bars). *for the sake of clarity we have shown data of 3 genes (locus IDs)

Figure S2). If desired, visualization of expression profile of DEGs in QTL intervals can also be done. Since growth stage specific tolerance is established in rice for both drought and rice, breeders might be interested in stage specific DEG option enabled in the database. For precise breeding applications, breeders can use the SSR and ISM-ILP polymorphism links and straightaway use the primers as PCR based markers (Fig. 5c). Since the database is simple in construction, breeders can use it intuitively without any guidance.

Conclusions

Meta-analysis of multiple microarray datasets provides a means for identification of robust candidate genes for the trait of interest. RiceMetaSys is a user-friendly web interface mainly intended for rice breeders for identification of salt and drought responsive genes in QTL intervals and those common to multiple stages, tissues and genetic backgrounds in rice. The SSR and ISM-ILP marker information provided is expected to help the molecular geneticists and breeders alike in their breeding and fine mapping efforts. Our purpose of developing RiceMetaSys is to provide a separate link for each and every economically important biotic and abiotic stress in rice. In the current version, we have accomplished it for salt and drought tolerance. In the next, we would be adding more important

traits like extreme temperature tolerance and leaf and panicle blast resistance. We will be integrating the RNA-seq data for these traits as well in the future.

Additional files

Additional file 1: Table S1. Detailed information about the microarray datasets retrieved from NCBI GEO database. (DOCX 17 kb)

Additional file 2: Figure S1. Schematic diagram of the RiceMetaSys database. Datasets were downloaded from the NCBI GEO and then were analyzed using GEO2R based script for the identification of DEGs. A comprehensive web based interface was developed to provide useful search information related to DEGs like commonly expressed genes, common genes across genotypes and DEGs in given physical intervals and genic microsatellites. (PPTX 610 kb)

Additional file 3: Figure S2. Detailed workflow for Physical position search (DOCX 36 kb)

database design, data analysis, decision to publish, or preparation of the manuscript.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]ICAR-National Research Centre on Plant Biotechnology, LBS Building, Pusa Campus, New Delhi 110012, India. [2]Shobhit University, Modipuram, Meerut 250110, Uttar Pradesh, India. [3]Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore 641003, India. [4]Current address: Department of biotechnology, Keralverma faculty of science, Swami Vivekanand Subharti University, Meerut 250005, Uttar Pradesh, India. [5]Indian Council of Agricultural Research, Krishi Bhawan, New Delhi 110001, India.

References
1. Ma L, Chen C, Liu X, Jiao Y, Su N, Li L, Wang X, Cao M, Sun N, Zhang X, Bao J, Li J, Pedersen S, Bolun L, Zhao H, Yuan L, Wong GS, Wang J, Deng XW, Wang J. A microarray analysis of the rice transcriptome and its comparison to Arabidopsis. Genome Res. 2005;15(9):1274–83.
2. Carriger S, Vallee D. More crop per drop. Rice Today. 2007;6:10–3.
3. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2010 Revision, 2011; Volume I: Comprehensive Tables. ST/ESA/SER.A/313.
4. Shanmugavadivel PS, Amitha Mithra SV, Prakash C, Ramkumar MK, Tiwari RK, Mohapatra T, Singh NK. High resolution mapping of QTLs for heat tolerance in rice using a 5K SNP array. Rice. 2017;10:28.
5. Dixit S, Singh A, Sta Cruz MT, Maturan PT, Amante M, Kumar A. Multiple major QTL lead to stable yield performance of rice cultivars across varying drought intensities. BMC Genet. 2014;15:16.
6. Kumar V, Singh A, Mithra SVA, Krishnamurthy SL, Parida SK, Jain S, et al. Genome-wide association mapping of salinity tolerance in rice (Oryza sativa). DNA Res. 2015;22:133–45.
7. Tiwari S, SL K, Kumar V, Singh B, Rao A, Mithra SVA, et al. Mapping QTLs for Salt Tolerance in Rice (Oryza sativa L.) by Bulked segregant analysis of recombinant inbred lines using 50K SNP Chip. PLoS One. 2016;11:e0153610.
8. Chandra P, Amitha Mithra SV, Singh PK, Mohapatra T, Singh NK. Unraveling the molecular basis of oxidative stress management in a drought tolerant rice genotype Nagina 22. BMC Genomics. 2016;17:774.
9. Salvi S, Tuberosa R. Genomics-based approaches to improve drought tolerance of crops. Trend Plant Sci. 2006;11(8):405–12.
10. Singh R, Ong-Abdullah M, Low E-TL, Abdul MA, et al. Oil palm genome sequence reveals divergence of inter-fertile species in Old and New worlds. Nature. 2013;500:335–9.
11. Clarke JD, Zhu T. Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives. Plant J. 2006;45:630–50.
12. Lima JM, Nath M, Dokku P, Raman KV, Kulkarni KP, Vishwakarma C, Sahoo SP, Mohapatra UB, AmithaMithra SV, Chinnusamy V, Robin S, Sarla N, Seshashayee M, Singh K, Singh AK, Singh NK, Sharma RP, Mohapatra T. Physiological, anatomical and transcriptional alterations in a rice mutant leading to enhanced water stress tolerance. AoB Plants. 2015;7:1–19.
13. Pandit A, Rai V, Bal S, Sinha S, Kumar V, Chauhan M, Gautam RK, Singh R, Sharma PC, Singh AK, Gaikwad K, Sharma TR, Mohapatra T, Singh NK. Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (Oryza sativa L.). Mol Gen Genomics. 2013;284:121–36.
14. Byeon Y, Park S, Kim YS, Back K. Microarray analysis of genes differentially expressed in melatonin-rich transgenic rice expressing a sheep serotonin N-acetyltransferase. J Pineal Res. 2013;55:357–63.
15. Jangam AP, Pathak RR, Raghuram N. Microarray Analysis of Rice d1 (RGA1) Mutant Reveals the Potential Role of G-Protein Alpha Subunit in Regulating Multiple Abiotic Stresses Such as Drought, Salinity, Heat, and Cold. Front Plant Sci. 2016;7:11.
16. Kazuki H, Hongo K, Suwabe K, Shimizu A, Nagayama T, Abe R, Kikuchi S, Yamamoto N, Fujii T, Yokoyama K, Tsuchida H, Sano K, Mochizuki K, Oki N, Horiuchi Y, Fujita M, Watanabe M, Matsuoka M, Kurata N, Yano K. OryzaExpress: An integrated database of gene expression networks and omics annotations in rice. Plant Cell Physiol. 2011;52:220–9.
17. Dash S, Van HJ, Hong L, Wise RP, Dickerson JA. PLEXdb: gene expression resources for plants and plant pathogens. Nucl. Acids Res. 2012;40(D1):D1194–201.
18. Cao P, Jung KH, Choi D, Hwang D, Zhu J, Ronald PC. The Rice Oligonucleotide Array Database: an atlas of rice gene expression. Rice. 2012;5:171.
19. Priya P, Jain M. RiceSRTFDB: A database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. Database (Oxford). 2013;2013:bat027.
20. Kurata N, Yamazaki Y. Oryzabase. An Integrated Biological and Genome Information Database for Rice. Plant Physiol. 2006;140:12–7.
21. Smita S, Lenka SK, Katiyar A, Jaiswal P, Preece J, Bansal KC. QlicRice: A web interface for abiotic stress responsive QTL and loci interaction channels in rice. Database (Oxford). 2011;2011:bar037.
22. Droc G, Ruiz M, Larmande P, Pereira A, Piffanelli P, Morel JB, Dievart A, Courtois B, Guiderdoni E, Perin C. OryGenesDB.a database for rice reverse genetics. Nucl. Acids Res. 2008;34:D736–40.
23. Sato Y, Antonio BA, Namiki N, Takehisa H, Minami H, Kamatsuki K, et al. RiceXPro: a platform for monitoring gene expression in japonica rice grown under natural field conditions. Nucleic Acids Res. 2011;39(Database issue):D1141–8.
24. qTeller: Simple Tool to identify genes under your QTL. http://www.qteller.com/. Accessed 27 Dec 2016.
25. Badoni S, Sayal DS, Gopalakrishnan S, Singh AK, Rao AR, Tyagi AK. Genomewide generation and use of informative intron-spanning and intron-length polymorphism markers for high-throughput genetic analysis in rice. Sci Rep. 2016;6:23765.
26. Walia H, Wilson C, Ismail AM, Close TJ, Cui X. Comparing genomic expression patterns across plant species reveals highly diverged transcriptional dynamics in response to salt stress. BMC Genomics. 2009;10:398.
27. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. Nucl Acids Res. 2009;37:D885–90.
28. Johnson WE, Rabinovic A, Li C. Adjusting batch effects in microarray expression data using Empirical Bayes methods. Biostatistics. 2007;8(1):118–27.
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucl. Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007.
30. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, WR MC, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Ocampo LQ, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.
31. Frank MY, Huo N. GuYQ. Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J and Anderson OD. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics. 2008;9:253.

Sandhu *et al. BMC Bioinformatics* (2017) 18:432

Page 11 of 11

32. Zhu JK. Salt and drought stress signal transduction in plants. Annu Rev Plant Biol. 2002;53:247–73.
33. Golldack D, Li C, Mohan H, Probst N. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. Front Plant Sci. 2014;5:151.
34. Parida SK, Dalal V, Singh AK, Singh NK, Mohapatra T. Genic non-coding microsatellites in the rice genome: characterization, marker design and use in assessing genetic and evolutionary relationships among domesticated groups. BMC Genomics. 2009;10:140.
35. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch SR. Computational and Experimental Analysis of Microsatellites in Rice (*Oryza sativa* L.). Frequency, Length Variation, Transposon Associations, and Genetic Marker Potential. Genome Res. 2013;11:1441–52.
36. Chinnusamy V, Schumaker K, Zhu JK. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. J Exp Bot. 2004;55:225–36.
37. Sharma R, Vleesschauwer DD, Sharma MK, Ronald PC. Recent Advances in Dissecting Stress-Regulatory Crosstalk in Rice. Mol Plant. 2013;6(2):250–60.
38. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 2002; 5: 94-100.
39. Dixit N, Dokku P, Amitha Mithra SV, Parida SK, Singh NK, Mohapatra T. Haplotype structure in grain weight gene *GW2* and its association with grain characteristics in rice. Euphytica. 2013;192(1):55–61.
40. Tiwari KK. Singh. A., Pattnaik S. et al. Identification of a diverse mini-core panel of Indian rice germplasm based on genotyping using microsatellite markers. Plant Breed. 2015;134(2):164–71.