Plant Tissue Cult. & Biotech. **20**(2): 211-223, 2010 (December)

PTC&B

READS - A Resource for Plant Non-coding Regulatory Sequence Analysis

Saima Shahid, Sabrina M. Elias, Sudip Biswas and Zeba I. Seraj*

Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh

Key words: Promoter database, Stress inducible, Constitutive, Gene expression, Cis-element

Abstract

Identification and analysis of regulatory sequences that control gene expression can be greatly facilitated by database-assisted bioinformatic approaches. READS (Regulatory Element Analysis DatabaSe) has been created as a web-accessible freely available database of plant non-coding regulatory sequences. It currently contains more than 300 known and putative promoters of constitutive as well as stress inducible genes belonging to diverse plants. The database has been manually curated with promoters collected mainly from scientific publications, thereafter cross-referenced with other resources (NCBI database, PubMed, PubMed Central). A user-friendly interface has been provided to allow easy access and analysis of data using different query options. A blast utility has also been provided, allowing users to search against all entries in the database. For each promoter, certain features such as expression data, GC content, core elements etc., were provided to assist in characterization of the regulatory sequences. To our knowledge, READS is the first plant promoter database that allows retrieval of sequences based on expression pattern. Thus the database can be utilized as a useful resource for identification of important putative regulatory cis-elements in promoters by analysis of upstream regions of hundreds of co-regulated or co-expressed genes. Such knowledge can also be of use for identifying minimal or stress inducible promoters for effective transgene expression. We aim to provide the most up-to-date collection of promoters of well-characterized stress inducible and constitutively expressed genes from many plant species. Hence, this resource will be updated regularly to incorporate new sequences. READs is available at http://www.pbtlabdu.net/READS/.

Introduction

Database assisted promoter analysis can greatly facilitate the study of gene regulatory networks. With the rapid advances in sequencing technology and the

*Author for correspondence. <zebai@univdhaka.edu>.

availability of complete genome sequences for several plant species, a number of promoter databases have been developed over the years to aid the study of transcriptional regulation. Most of the established plant promoter databases have been developed exclusively for a particular model species, such as *Arabidopsis* or rice. ATCISDB, AthaMap, Athena, Osiris, ppdb and Grassius are examples of this type of promoter database (Davuluri et al. 2003, O'Connor et al. 2005, Bulow et al. 2006, Morris et al. 2008, Yamamoto and Obokata 2008, Yilmaz et al. 2009). However, these databases hold general upstream region sequences and transcription factor binding site/cis-element information on a genome-scale basis, without focus on any specific expression pattern. There are some databases that are more specialized with respect to promoter data content, and have been developed to facilitate a specific purpose. For instance, PlantProm DB provides a collection of proximal promoter regions from 59 species with experimentally verified transcription start sites (Shahmuradov et al. 2003). This resource has been utilized by many researchers for developing core promoter prediction algorithms, TSSP is one among those (Shahmuradov et al. 2005). D$_o$OP is another database which provides clusters of orthologous promoters, mostly from *A. thaliana* and *Brassica oleracea* (Barta et al. 2005). Sprome and TGP are two less-established databases which report selective plant promoters that are biologically important for a certain process (Smirnova et al. 2006, Rajasekaran et al. 2008). Sprome houses around 50 putative promoters of salinity and drought stress inducible genes from rice, while TGP provides a small collection of reported transgene promoters with experimental observations.

However, none of the existing databases provides comprehensive resource of putative promoters from multiple organisms along with expression data, such as constitutivity or stress-inducibilty. Also most of the databases have not been updated as new genomic and expression data becomes available. It is therefore highly desirable to build a more comprehensive plant promoter database for *in silico* analysis of regulatory elements and functional genomics research. There is also a great need for resources that will allow promoter mining from halophytes, which can survive in extreme conditions. READS has been developed as an attempt to address such challenges. Currently, READS is the only plant promoter database which provides expression pattern information of more than 300 promoters from both glycophyte and halophyte species (Table 1).

Materials and Methods

The majority of the data provided by READS has been curated manually. Stress inducible and constitutive genes were identified primarily through literature mining and available microarray expression data analysis. Differentially expressed genes in cold, drought, salinity stresses and abscisic acid application have been identified previously in rice using cDNA microarray and RNA gel-blot analysis (Rabbani et al. 2003). In a more recent rice whole genome oligomer microarray experiment, drought and high salinity responsive genes in different organs were reported (Zhou et al. 2007). Data from these experiments were downloaded, and genes corresponding to stress responsive cDNA clones were identified by matching each cDNA sequence to a *O. sativa* cv. Japonica locus (MSU6 assembly), using the BLASTn search tool provided at the GRAMENE database (Liang et al. 2008).

Table 1. Comparison of READS with available plant promoter databases.

Database	Represented species	Database characteristic	Website
ATCOCIS	*A. thaliana*	Genome-wide promoter database, provides motif and TSS information	http://bioinformatics.psb.ugent.be/ATCOECIS/
Athena	*A. thaliana*	Genome-wide promoter database, provides motif and TSS information	http://www.bioinformatics2.wsu.edu/Athena/
Osiris	*O. sativa*	Genome-wide promoter database, provides motif and TSS information	http://www.bioinformatics2.wsu.edu/Osiris/
Plant Prom	59 species	Proximal promoter region (251bp) with experimentally verified TSS	http://mendel.cs.rhul.ac.uk/mendel.php?topic=plantprom
ppdb	*A. thaliana*, *O. sativa* and *P. patens*	Genome-wide promoter database, provides core promoter element and TSSs	http://www.ppdb.gene.nagoya-u.ac.jp
D$_o$OP	269 species in total, but mainly *Arabidopsis* and *B. oleracea*	Clusters of orthologus promoters, with conserved motif predictions	http://doop.abc.hu/
Grassius	*O. sativa*	Genome-wide promoter database	http://grassius.org/grasspromdb.html
TGP	22 species	Published transgene promoters	http://wwwmgs.bionet.nsc.ru/mgs/dbases/tgp/
Sprome	*O. sativa*	Salinity and drought inducible promoters	http://www.btistnau.org/default1.aspx
READS	9 species	Constitutive and stress inducible promoters from glycophytes and halophytes, provides motif and TSS data	http://www.pbtlabdu.net/READS/

Information regarding the constitutive rice genes was obtained from a recently published report (Jiao et al. 2009), in which microarray experiments for 40 different cell types were performed. Briefly, putative constitutive genes were identified through statistical analysis (Lee et al. 2007) of microarray expression data based on the following criteria: (i) Ubiquitous expression and no more than two cell-type nulls, and (ii) uniform expression and a difference between the highest third and lowest third of log-transformed normalized value of < 2 (Jiao et al. 2009). Finally, promoter regions of all the stress-inducible/constitutive rice genes (1 kbp upstream of the translational start site) were extracted, using the bulk data retrieval tool available at OryGenesDB (Droc et al. 2006). Experimentally characterized individual promoter sequences from rice landraces, barley and halophyte species with stress specific expression pattern reported in scientific literatures were extracted by searching the NCBI nucleotide database with corresponding GenBank accessions. Several promoters from salt tolerant rice landraces that have been characterized in the Plant Biotechnology Lab., University of Dhaka and whose sequences have been deposited at GenBank also constituted a part of the final dataset.

To identify putative cis-elements/motifs in the collected promoters, the MotifScanner program provided by the TOUCAN 2 workbench (Aerts et al. 2003, Aerts et al. 2005) was used. MotifScanner searches for pre-defined motifs in DNA sequences based on a probabilistic sequence model (Thijs et al. 2001). The PlantCARE database (Rombauts et al. 1999) as the source of pre-defined plant specific motif matrices and 'plantprom (3rd order)' as the background model were selected for running MotifScanner. Putative transcription start sites were detected using the TSSP program available at softberry (Shahmuradov et al. 2005). A custom perl script was utilized for GC content calculation.

READS data are stored within a relational database management system, MySQL (http://www.mysql.com), using the MyISAM storage engine. The website was developed using the PHP language (http://www.php.net) and hosted at the Apache web server (http://www.apache.org). The 'IN BOOLEAN MODE' modifier of MySQL was used for Boolean full-text search in quick and advanced search options provided at the homepage, to facilitate data retrieval using specific key words. The www blast suite of NCBI (Altschul et al. 1997) was adapted and configured to allow the users blast against the READS promoter database. To enable fast updating of database, a manual curation utility has also been developed.

Results and Discussion

The main objective of the READS database is to provide a user-friendly way to retrieve plant promoter sequences according to their expression patterns and/or other functional groupings to facilitate further analysis. The current release of READS contains 218 constitutive and 98 stress-inducible promoters from nine different plant species (Tables 2 and 3). The database schema of READS is shown in Fig. 1.

Table 2. Statistics of READS data content (current release).

Category	Number
Total plant species	9
Monocots	6
Dicots	3
Halophytes	3
Total sequences	316
Constitutive promoters	218
Stress-inducible promoters	98

Table 3. Species information of READS (current release).

Plant species	Category
Oryza sativa japonica cv. Nipponbare	Monocot
O. sativa indica cv. Pokkali	Monocot
O. sativa indica cv. Horkuch	Monocot
O. sativa indica cv. Nagina 22	Monocot
O. sativa indica cv. Taichung Native 1	Monocot
Hordeum vulgare	Monocot
Atriplex centralasiatica	Dicot
Mesembryanthemum crystallinum	Dicot
Suaeda liaotungensis	Dicot

The READS data is stored in a MySQL database and freely accessible through a web interface at the following address: http://www.pbtlabdu.net/READS. Users can search the complete dataset of 316 plant promoters through three different entry points:

1. *Simple search form:* Searching by single or multiple keywords such as gene name, TIGR locus identifier, plant species, stress etc. using the 'quick search' option in the READS homepage (Fig. 2a).
2. *Advanced search:* Searching by single or combined query fields provided in the homepage (Fig. 2a), such as expression pattern, one or more stress types, plant species and/or Gene ID.

3. *BLAST utility:* Searching by sequence comparison against a local database of promoter sequences through the BLAST utility of NCBI.

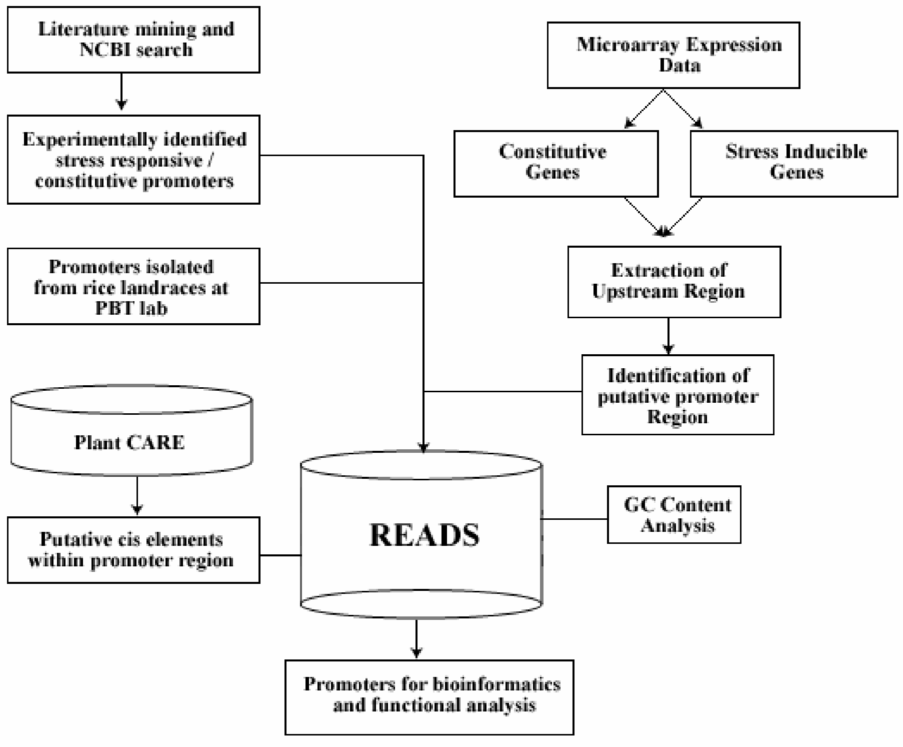
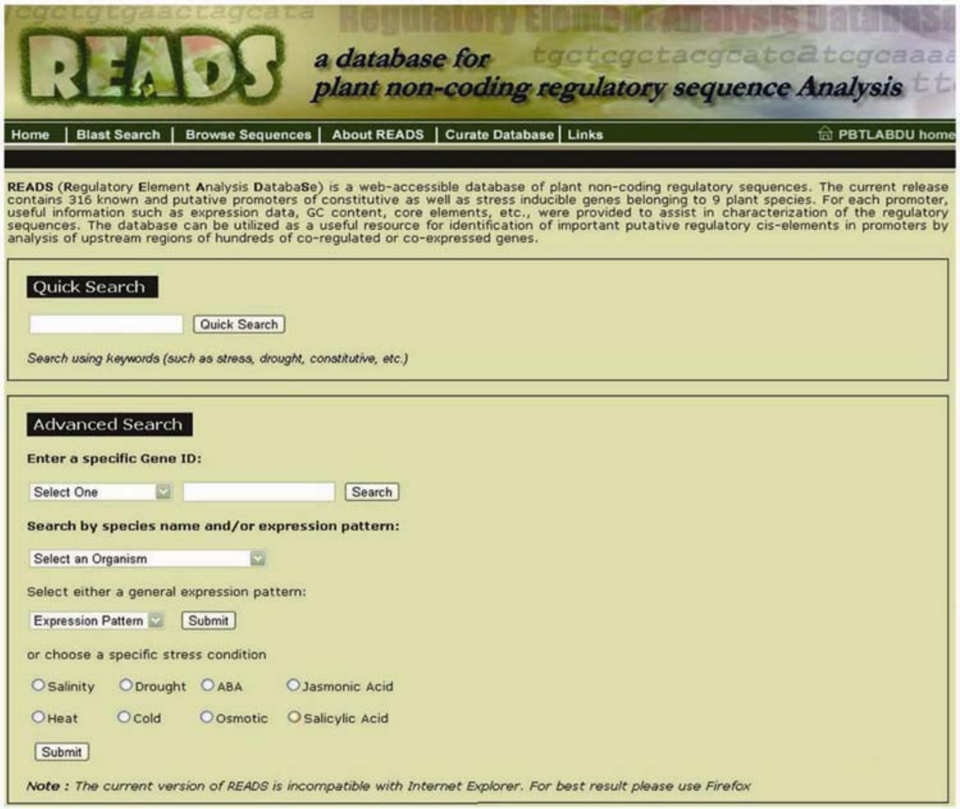

Literature mining and NCBI search
Experimentally identified stress responsive / constitutive promoters
Promoters isolated from rice landraces at PBT lab
Plant CARE
Putative cis elements within promoter region
Microarray Expression Data
Constitutive Genes
Stress Inducible Genes
Extraction of Upstream Region
Identification of putative promoter Region
READS
GC Content Analysis
Promoters for bioinformatics and functional analysis

Fig. 1. A schematic view of the READS database showing the source of dataset used and the sequence characterization.

Each of the search queries generates a list of entries that match the user specified criteria. By clicking on a single entry, information regarding that promoter is visualized in a specific page including gene name and description, species name, genomic location, promoter sequence, putative cis-elements, external links to other public resources and to Pubmed/Pubmed Central, GC content, putative or experimentally identified TSSs (Transcription Start Sites) and known expression pattern, for example constitutive expression, response to abiotic stresses, etc. A screen shot of a READS entry page is shown in Fig. 2b.

READS has been designed to allow easy retrieval of a group of sequences based on specific criteria like species name, expression pattern or other functional categories. For example, the 'Quick Search' form at the READS homepage was queried using the keyword 'salinity', which returned 79 promoters with salinity inducible expression pattern, as shown in Fig. 3a and b. To narrow down the results, the species '*Atriplex centralasiatica*' and expression pattern '*salinity*' were chosen from the dropdown menus in the 'Advanced Search' section at the homepage, which returned one promoter entry (Fig. 3c, d,e). For each promoter entry, features that are important in transcriptional regulations have been provided. Locations of putative elements have been presented graphically, and links for details of these elements have also been provided below the graphical representation (Fig. 2b). The sequence data of the READS database can be used for further analysis using existing tools, e.g. analysis of potential transcription factor binding sites or identifying conserved motifs in co-expressed genes using tools such as PLACE (Higo et al. 1999), TOUCAN(Aerts et al. 2005), MEME (Bailey et al. 2010) etc.



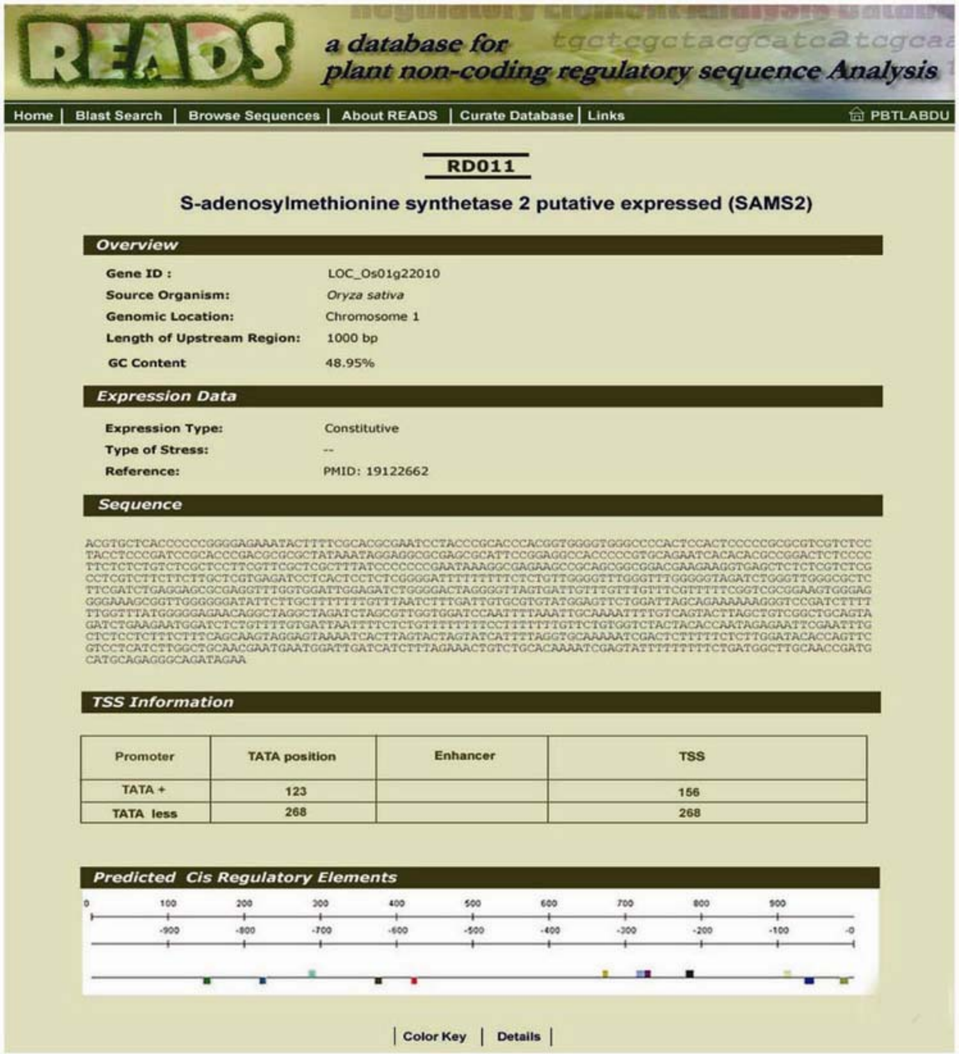
READS (**R**egulatory **E**lement **A**nalysis **D**ataba**S**e) is a web-accessible database of plant non-coding regulatory sequences. The current release contains 316 known and putative promoters of constitutive as well as stress inducible genes belonging to 9 plant species. For each promoter, useful information such as expression data, GC content, core elements, etc., were provided to assist in characterization of the regulatory sequences. The database can be utilized as a useful resource for identification of important putative regulatory cis-elements in promoters by analysis of upstream regions of hundreds of co-regulated or co-expressed genes.

Search using keywords (such as stress, drought, constitutive, etc.)

***Note :** The current version of READS is incompatible with Internet Explorer. For best result please use Firefox*

Fig. 2a. Web interface of READS.

READS offers several critical advantages over the other available plant promoter databases. Firstly, all the promoter sequences in READS are annotated with specific expression data as well as putative motifs. Hence READS can be greatly useful to researchers for identifying unknown functional cis-elements/motifs in non-coding regions, designing minimal promoters for synthetic constructs and also in engineering transgenic crop plants, where the gene of interest is only expressed at the desired stage of plant life cycle, or in the desired tissue and/or in response to specific environmental cues. Secondly, READS is the only database to-date that allows easy retrieval of a group of sequences with similar expression pattern and thus facilitate comparative analysis among sets of promoters with differential expression pattern. To our knowledge, no available plant promoter database except READS provides easy retrieval of comprehensive promoter dataset for constitutive genes, even in a single species. Besides this, READS also allows sequence retrieval based on other expression patterns, such as drought, salinity, abscisic acid, cold inducibility etc. Comparative cis-element analyses among differentially expressed promoters have greatly benefitted identification of discriminative cis-elements involved in


READS
a database for
plant non-coding regulatory sequence Analysis
tgctcgctacgcatcatcgca
Home | Blast Search | Browse Sequences | About READS | Curate Database | Links
PBTLABDU

RD011

S-adenosylmethionine synthetase 2 putative expressed (SAMS2)

Overview

Gene ID :	LOC_Os01g22010
Source Organism:	*Oryza sativa*
Genomic Location:	Chromosome 1
Length of Upstream Region:	1000 bp
GC Content	48.95%

Expression Data

Expression Type:	Constitutive
Type of Stress:	--
Reference:	PMID: 19122662

Sequence

```
ACGTGCTCACCCCCCGGGGAGAAATACTTTTCGCACGCGAATCCTACCCGCACCCACGGTGGGGTGGGCCCCACTCCACTCCCCCGCGCGTCGTCTCC
TACCTCCCGATCCGCACCCGACGCGCGCTATAAATAGGAGGCGCGAGCGCATTCCGGAGGCCACCCCCGTGCAGAATCACACACGCCGGACTCTCCCC
TTCTCTCTGTCTCGCTCCTTCGTTCGCTCGCTTTATCCCCCCCGAATAAAGGCGAGAAGCCGCAGCGGCGGACGAAGAAGGTGAGCTCTCTCGTCTCG
CCTCGTCTTCTTCTTGCTCGTGAGATCCTCACTCCTCTCGGGGATTTTTTTTTCTCTGTTGGGGTTTGGGTTTGGGGGTAGATCTGGGTTGGGCGCTC
TTCGATCTGAGGAGCGCGAGGTTTGGTGGATTGGAGATCTGGGGACTAGGGGTTAGTGATTGTTTGTTTGTTTCGTTTTTCGGTCGCGGAAGTGGGAG
GGGAAAGCGGTTGGGGGGATATTCTTGCTTTTTTTGTTTAATCTTTGATTGTGCGTGTATGGAGTTCTGGATTAGCAGAAAAAAAGGGTCCGATCTTTT
TTGGTTTATGGGGGAGAACAGGCTAGGCTAGATCTAGCGTTGGTGGATCCAATTTTAAATTGCAAAATTTTGTCAGTACTTAGCTGTCGGCTGCAGTA
GATCTGAAGAATGGATCTCTGTTTTGTGATTAATTTTCTCTGTTTTTTTTCCTTTTTTTTGTTCTGTGGTCTACTACACCAATAGAGAATTCGAATTTG
CTCTCCTCTTTCTTTCAGCAAGTAGGAGTAAAATCACTTAGTACTAGTATCATTTTAGGTGCAAAAATCGACTCTTTTTCTCTTGGATACACCAGTTC
GTCCTCATCTTGGCTGCAACGAATGAATGGATTGATCATCTTTAGAAACTGTCTGCACAAAATCGAGTATTTTTTTTTCTGATGGCTTGCAACCGATG
CATGCAGAGGGCAGATAGAA
```

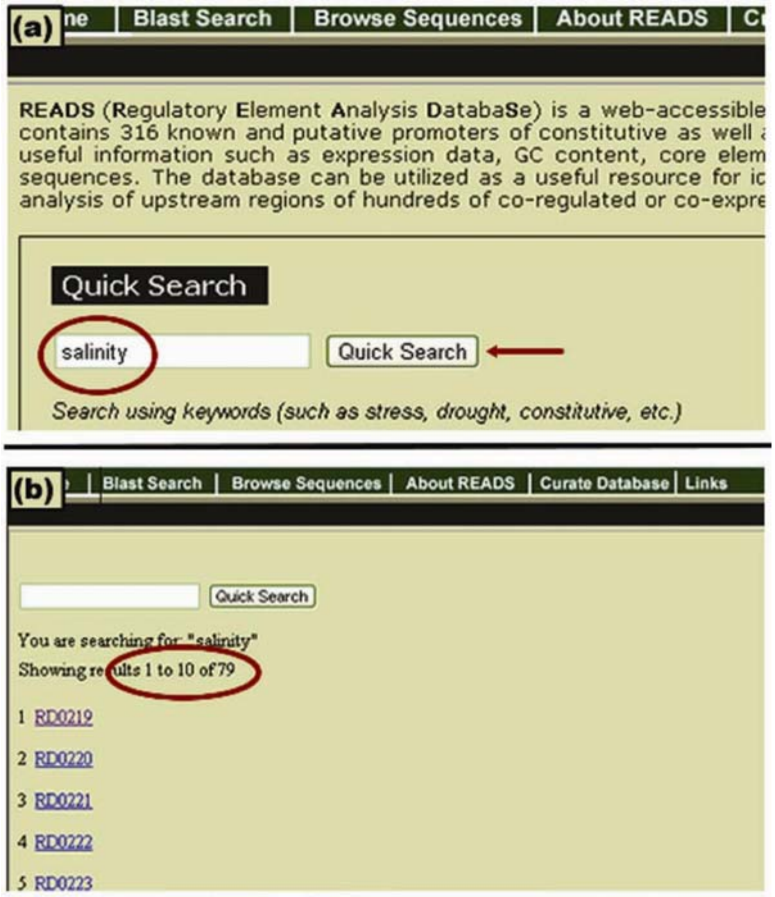
TSS Information

Promoter	TATA position	Enhancer	TSS
TATA +	123		156
TATA less	268		268

Predicted Cis Regulatory Elements

Fig. 2b. Screenshot of an entry page in READS.

stress/tissue specific gene regulation in many previous studies (Lindlof et al. 2009, Shi et al. 2010). Such information can also be utilized to detect common modules of cis-elements involved in regulation of a group of genes. Hence promoter databases that allow sequence retrieval based on expression pattern can be very useful for discriminative motif identification purposes, and READS has been developed as an attempt to address such necessities. For instance, the



READS (Regulatory Element Analysis DatabaSe) is a web-accessible contains 316 known and putative promoters of constitutive as well useful information such as expression data, GC content, core elem sequences. The database can be utilized as a useful resource for ic analysis of upstream regions of hundreds of co-regulated or co-expre

Search using keywords (such as stress, drought, constitutive, etc.)

You are searching for: "salinity"
Showing results 1 to 10 of 79

1 RD0219

2 RD0220

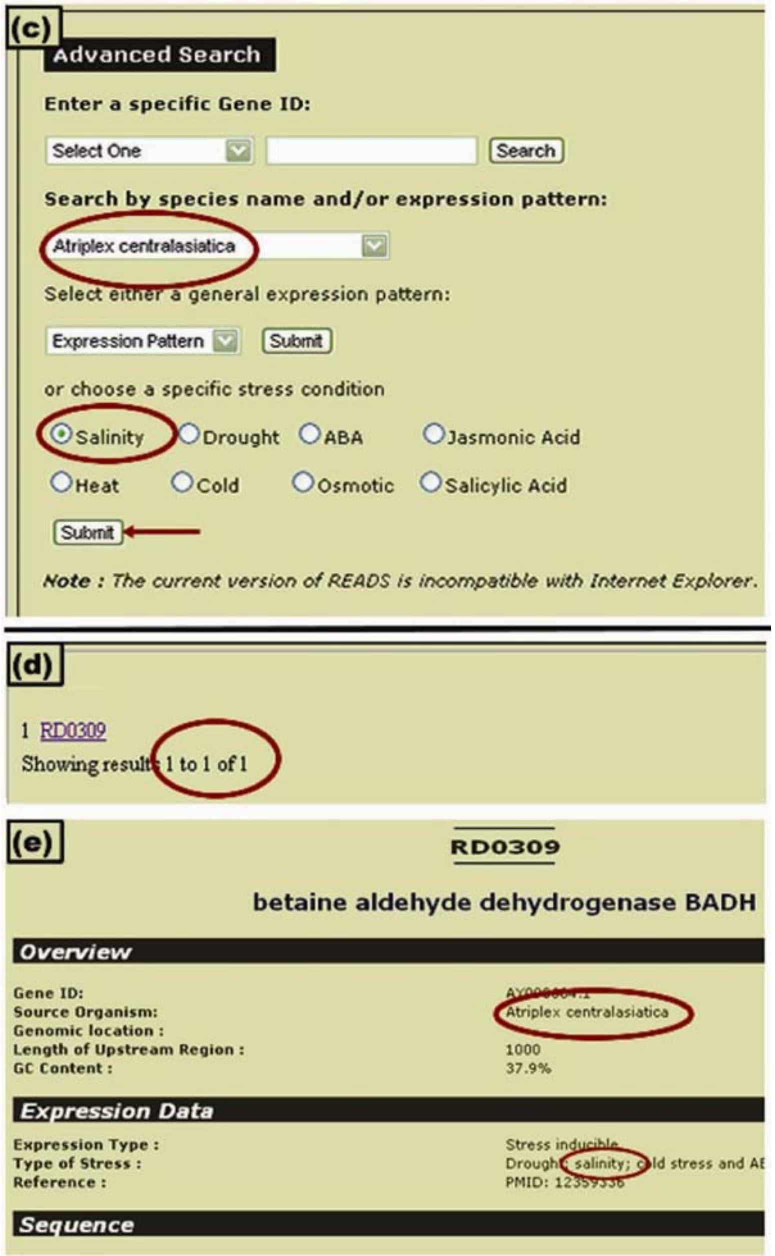
3 RD0221

4 RD0222

5 RD0223

Fig. 3a-b. Data Mining: A working example with the screenshots from READS search pages. (a) The Quick search option allows inputting any keyword to query the READS Database. (b) Search result for the keyword 'salinity' showing 79 results in the database.

information resource provided by READS can be utilized for *in silico* comparative analysis between constitutive and stress inducible promoters, and thus help researchers better understand the regulatory networks involved in plant stress adaptation. However, resources provided by READS will not necessarily facilitate motif analysis in orthologous promoters through phylogenetic footprinting, a well-known approach for conserved motif identification. For such analysis, the existing D$_o$OP (Database of Orthologous Promoters) may be more suitable.



Enter a specific Gene ID:

Search by species name and/or expression pattern:

Atriplex centralasiatica

Select either a general expression pattern:

or choose a specific stress condition

● Salinity ○ Drought ○ ABA ○ Jasmonic Acid

○ Heat ○ Cold ○ Osmotic ○ Salicylic Acid

Note : The current version of READS is incompatible with Internet Explorer.

1 RD0309

Showing result 1 to 1 of 1

RD0309

betaine aldehyde dehydrogenase BADH

Gene ID:	AY00[illegible]
Source Organism:	Atriplex centralasiatica
Genomic location :	
Length of Upstream Region :	1000
GC Content :	37.9%

Expression Type :	Stress inducible
Type of Stress :	Drought; salinity; cold stress and AE
Reference :	PMID: 12359336

Fig. 3c-e. (c) The Advanced search option showing the utility to choose a specific organism along with a specific stress condition, e.g 'salinity'. (d) The utility 'Advanced search' allowed narrowing down the previous results to only salinity inducible promoter from *Atriplex centralasiatica*. (e) The entry page showing details of the advanced search result, with circles pointing the specific stress condition and organism.

READS has also been developed to mine non-coding regulatory sequences from halophytes, which can survive and grow in extreme conditions. To-date, there is no publicly available promoter database for such purpose. Analysis of the regulatory regions of halophytes can be particularly significant to understand the dynamics of gene regulation in such organism and engineer crop plants tolerant to harsh environments. As genome sequences for halophytes are not available yet, the promoter data from such species in the current release of READS is limited. Nevertheless, continued efforts will be made to update the promoter sequences and expression data and other useful information. In future, promoters from additional plant species as well as tools for analyzing the data will be provided to enhance utility of the resource at READS. The manual curation section of the READS will also be made accessible to the community to enable improved annotation of non-regulatory sequences.

Acknowledgements

The authors would like to acknowledge Sharmin Jahan, Department of Biochemistry and Molecular Biology, University of Dhaka for her help with the promoter data extraction.

References

Aerts S, Thijs G, Coessens B, Staes M, Moreau Y and **De Moor B** (2003) Toucan: deciphering the cis-regulatory logic of coregulated genes. Nucleic Acids Res. **31:** 1753-1764.

Aerts S, Van Loo P, Thijs G, Mayer H, de Martin R, Moreau Y and **De Moor B** (2005) TOUCAN 2: the all-inclusive open source workbench for regulatory sequence analysis. Nucleic Acids Res. **33:** W393-396.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and **Lipman DJ** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389-3402.

Bailey TL, Boden M, Whitington T and **Machanick P** (2010) The value of position-specific priors in motif discovery using MEME. BMC Bioinformatics **11:** 179.

Barta E, Sebestyen E, Palfy TB, Toth G, Ortutay CP and **Patthy L** (2005) DoOP: Databases of Orthologous Promoters, collections of clusters of orthologous upstream sequences from chordates and plants. Nucleic Acids Res. **33:** D86-90.

Bulow L, Steffens NO, Galuschka C, Schindler M and **Hehl R** (2006) AthaMap: from in silico data to real transcription factor binding sites. In Silico Biol. **6:** 243-252.

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M and **Grotewold E** (2003) AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics **4:** 25.

Droc G, Ruiz M, Larmande P, Pereira A, Piffanelli P, Morel JB, Dievart A, Courtois B, Guiderdoni E and **Perin C** (2006) OryGenesDB: a database for rice reverse genetics. Nucleic Acids Res. **34:** D736-740.

Higo K, Ugawa Y, Iwamoto M and **Korenaga T** (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. **27:** 297-300.

Jiao Y, Tausta SL, Gandotra N, Sun N, Liu T, Clay NK, Ceserani T, Chen M, Ma L, Holford M, Zhang HY, Zhao H, Deng XW and **Nelson T** (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat. Genet. **41:** 258-263.

Lee S, Jo M, Lee J, Koh SS and **Kim S** (2007) Identification of novel universal housekeeping genes by statistical analysis of microarray data. J. Biochem. Mol. Biol. **40:** 226-231.

Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler ES, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A, Ravenscroft D, Ren L, Spooner W, Tecle I, Thomason J, Tung CW, Wei X, Yap I, Youens-Clark K, Ware D and **Stein L** (2008) Gramene: a growing plant comparative genomics resource. Nucleic Acids Res. **36:** D947-953.

Lindlof A, Brautigam M, Chawade A, Olsson O and **Olsson B** (2009) In silico analysis of promoter regions from cold-induced genes in rice (*Oryza sativa* L.) and Arabidopsis thaliana reveals the importance of combinatorial control. Bioinformatics **25:** 1345-1348.

Morris RT, O'Connor TR and **Wyrick JJ** (2008) Osiris: an integrated promoter database for *Oryza sativa* L. Bioinformatics **24:** 2915-2917.

O'Connor TR, Dyreson C and **Wyrick JJ** (2005) Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. Bioinformatics **21:** 4411-4413.

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K and **Yamaguchi-Shinozaki K** (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol. **133:** 1755-1767.

Rajasekaran S, Loganathan A, Pothiraj N and **Ponnusamy B** (2008) Sprome: A database on promoters of abiotic stress inducible genes in rice. International J. Integrative Biol. **2:** 153-156.

Rombauts S, Dehais P, Van Montagu M and **Rouze P** (1999) PlantCARE, a plant cis-acting regulatory element database. Nucleic Acids Res. **27:** 295-296.

Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM and **Solovyev VV** (2003) PlantProm: a database of plant promoter sequences. Nucleic Acids Res **31:** 114-117.

Shahmuradov IA, Solovyev VV and **Gammerman AJ** (2005) Plant promoter prediction with confidence estimation. Nucleic Acids Res. **33:** 1069-1076.

Shi R, Sun YH, Li Q, Heber S, Sederoff R and **Chiang VL** (2010) Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes. Plant Cell Physiol **51:** 144-163.

Smirnova OG, Ibragimova SS, Grigorovich DA and **Kochetov AV** (2006) TGP (Transgene Promoters): A database of biotechnologically important plant gene promoters. *In:* Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure, Vol 1, Novosibirsk, Russia.

Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P and **Moreau Y** (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics **17:** 1113-1122.

Yamamoto YY and **Obokata J** (2008) ppdb: a plant promoter database. Nucleic Acids Res. **36:** D977-981.

Yilmaz A, Nishiyama MY, Jr., Fuentes BG, Souza GM, Janies D, Gray J and **Grotewold E** (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiol. **149:** 171-180.

Zhou J, Wang X, Jiao Y, Qin Y, Liu X, He K, Chen C, Ma L, Wang J, Xiong L, Zhang Q, Fan L and **Deng XW** (2007) Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. Plant Mol. Biol. **63:** 591-608.