READS - Regulatory Element Analysis DatabaSe

A Resource For Plant Non-coding Regulatory Sequence Analysis

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

Abstract No. 39

6th International Plant Tissue Culture & Biotechnology Conference
Dec 3-5, 2010, Dhaka, Bangladesh

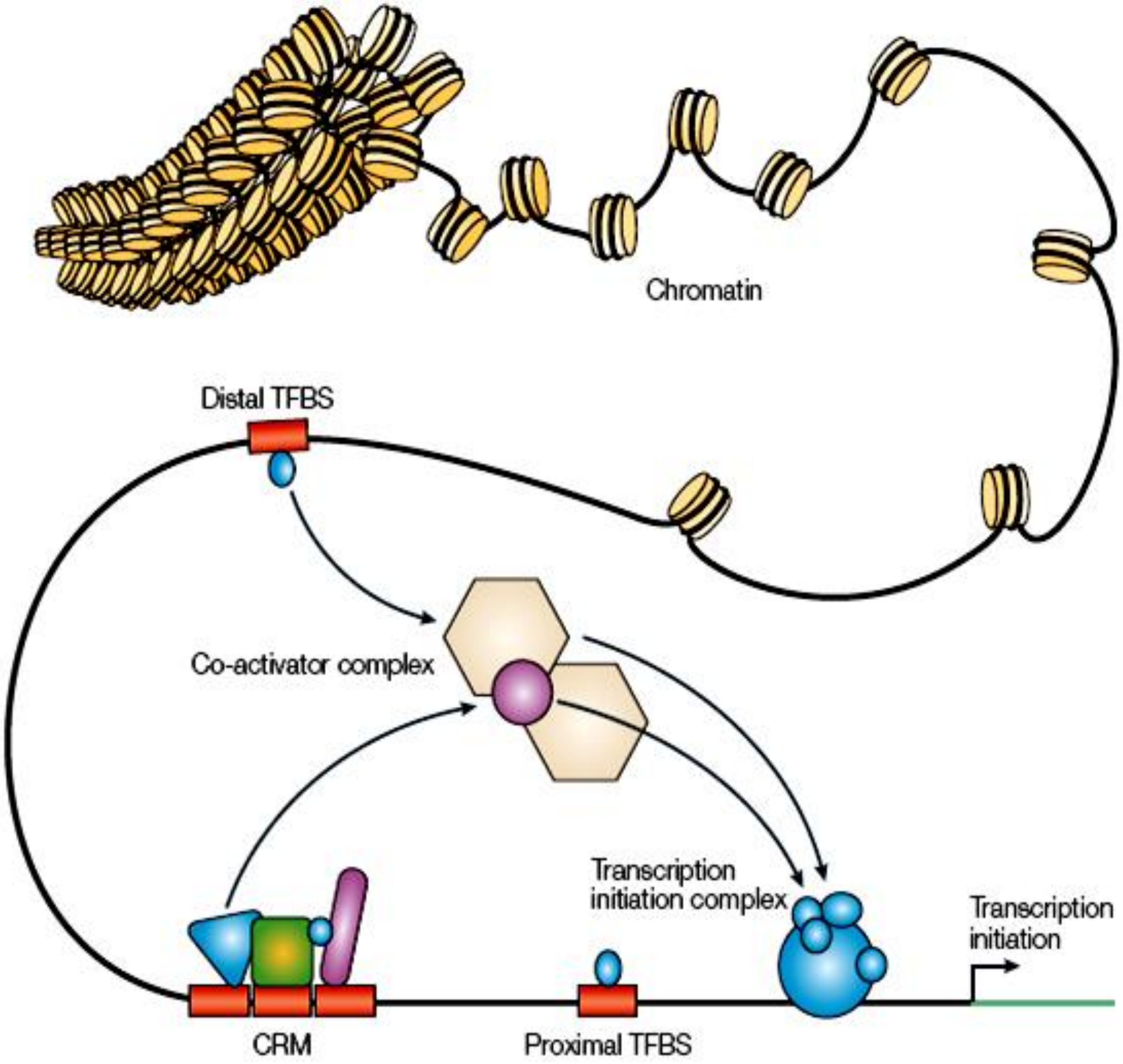
Plant Biotechnology Lab
Department of Biochemistry and Molecular Biology
University of Dhaka

Promoters: Non-coding DNA Sequence

- Promoters are non-coding DNA sequences
- Located upstream of the transcription start site of a gene
- Contains short conserved DNA sequences known as cis-regulatory elements or motifs that control gene expression
- The cis-regulatory elements are recognized and bound by specific transcription factors

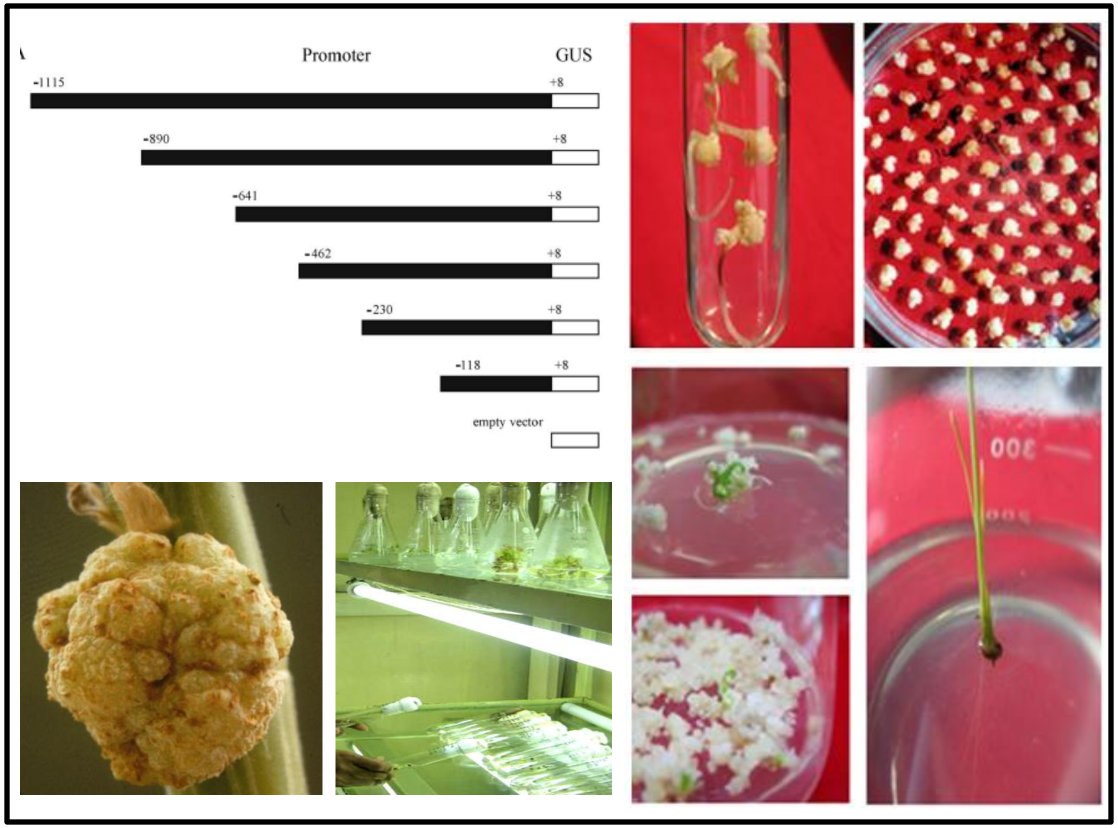
Cis-Regulatory Elements

- Cis-regulatory elements found around the Transcription Start Site (TSS) are known as core or proximal elements
- Cis-elements located further upstream are called distal elements or distal transcription factor binding sites



Identification of promoters and regulatory elements

- Traditional techniques such as promoter-reporter gene construct experiments are time consuming and labor-intensive
- The experiment results are not always reproducible



Bioinformatic approaches

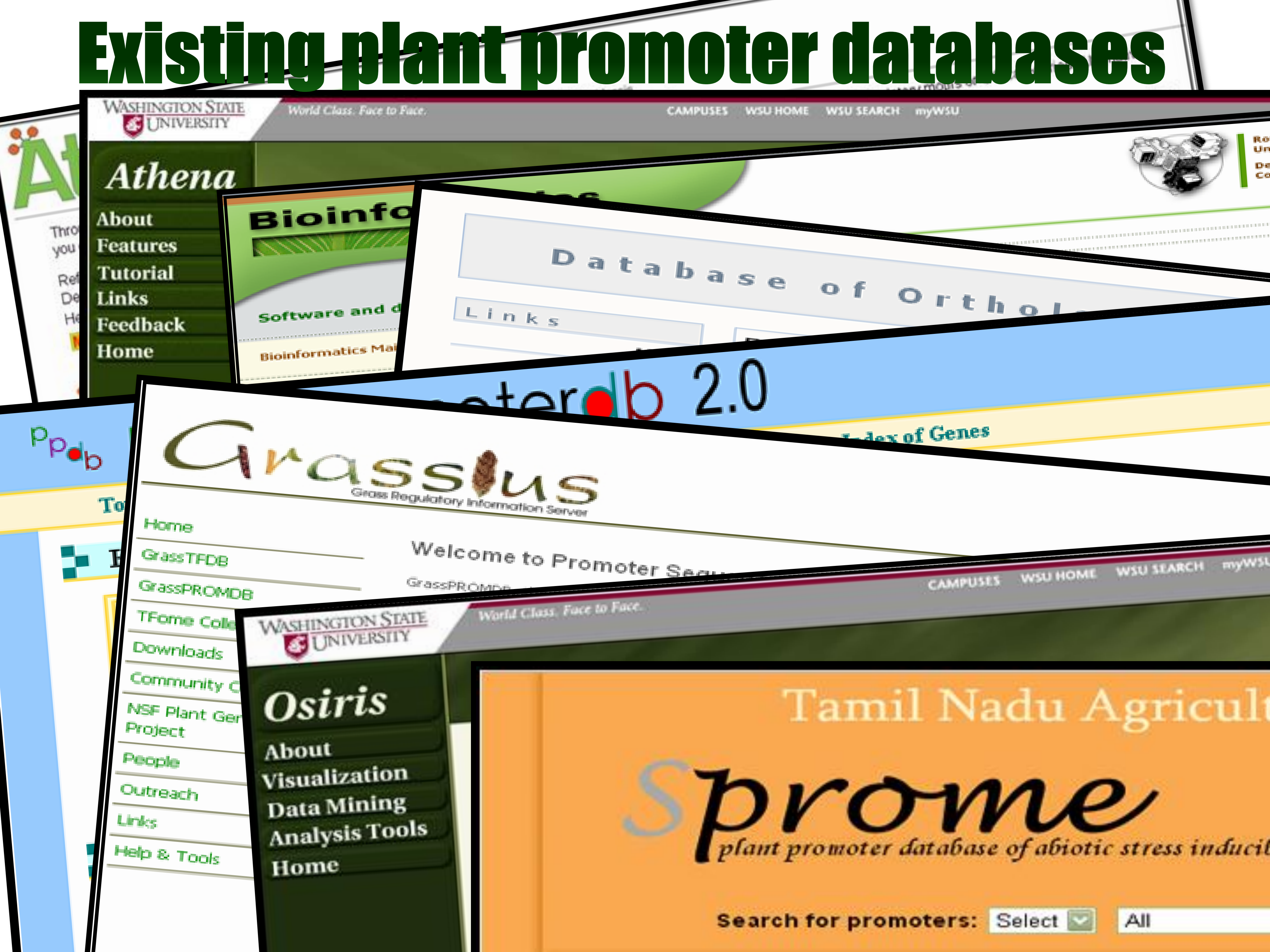
- Bioinformatic approaches can help to narrow down the possibilities and predict regulatory elements
- Still, identification of promoters and regulatory elements is a major challenge in bioinformatics because-
    - Regulatory elements are not well-conserved
    - Computational discovery of TSS in not straightforward
    - Promoter sequences do not have distinguishable statistical properties

Database assisted promoter analysis

- New Biotechnologies (automatic sequencing, microarray, DNA chips protein identification, mass specs, etc) produce large quantities of biological data
- It is impossible to analyze large datasets by manual inspection
- Database assisted promoter analysis can facilitate identification of important cis-elements and regulatory networks

READS

Existing plant promoter databases



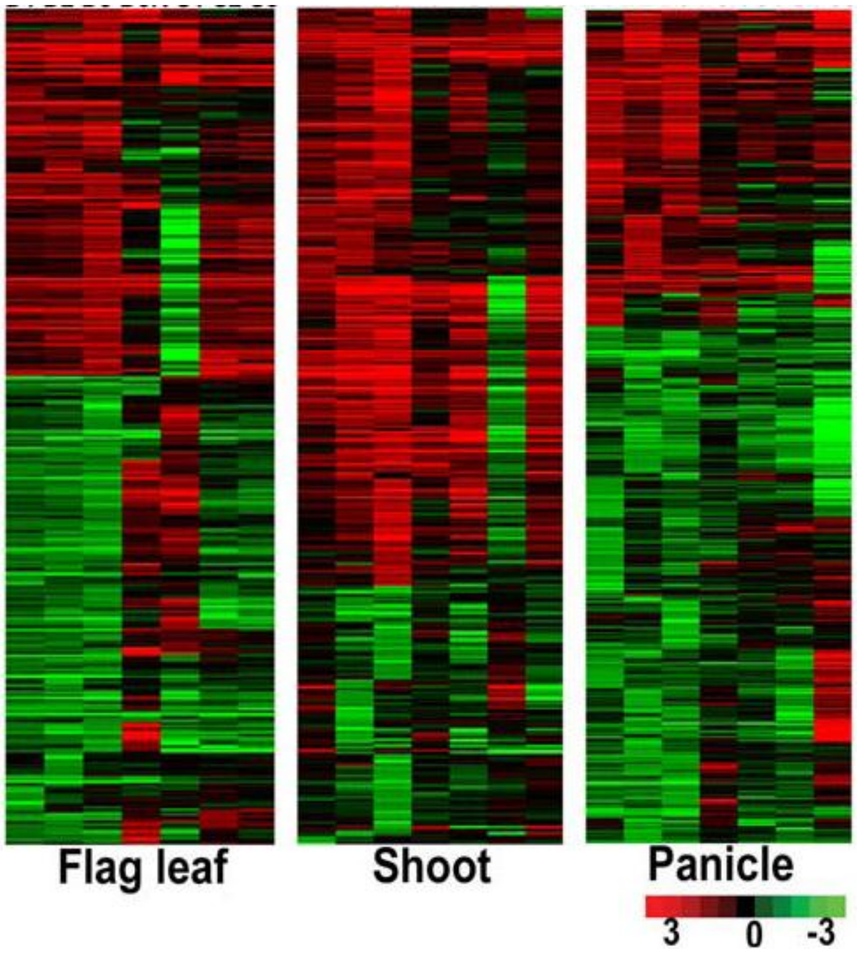
Existing plant promoter databases

- Genome specific promoter databases
  - Athena (Arabidopsis)
  - Osiris (Rice)
- Specialized promoter databases
  - DoOP (Database of Orthologous Promoters)
  - PlantProm (251bp region surroundingTSS)
- Most of the available databases do not provide expression pattern
- Challenges in mining promoter sequences expressed under specific condition

READS

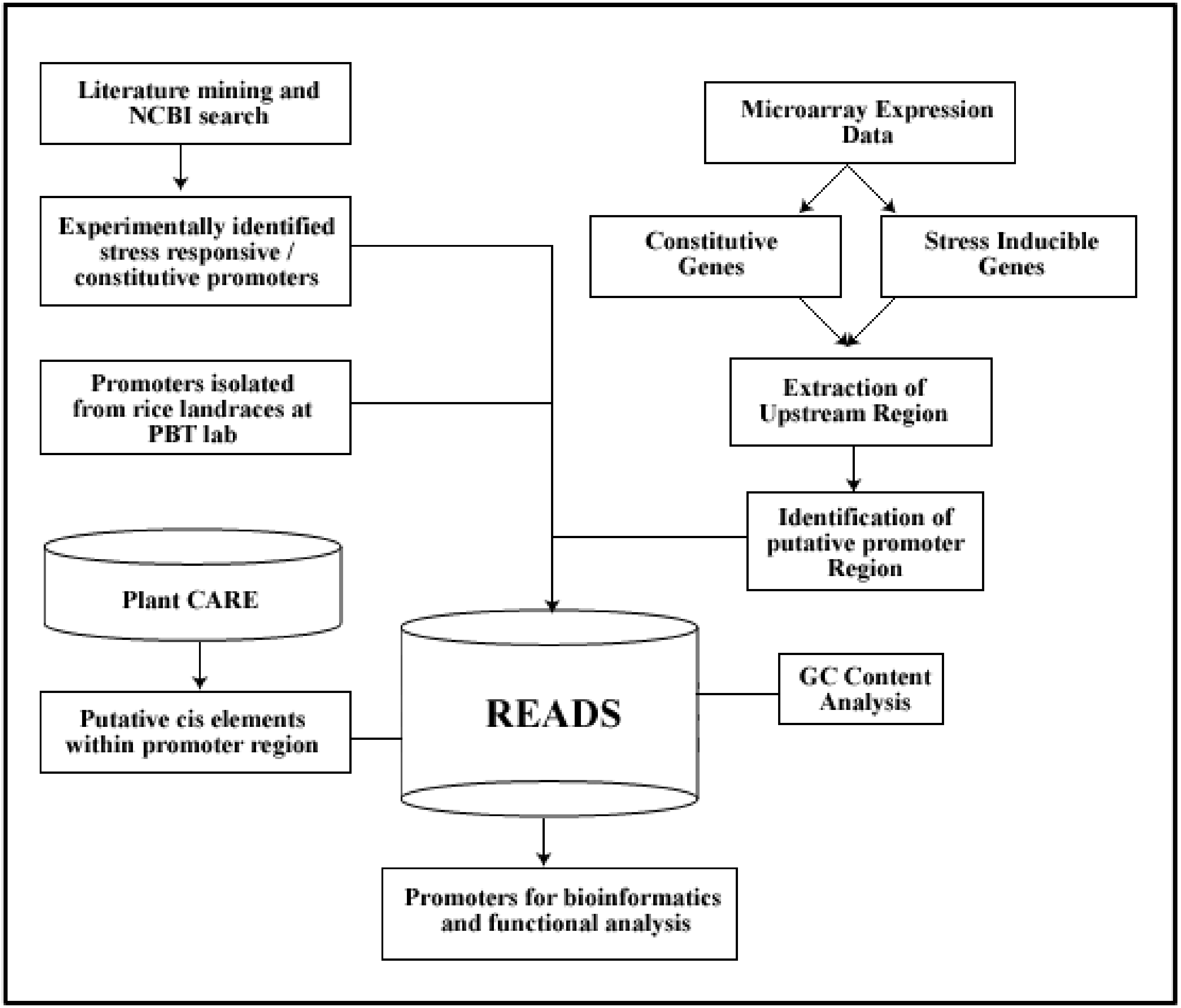
- READS is a specialized plant promoter database that contains more than 300 known and putative promoter sequences manually curated with expression data and other important biological features
- It allows easy retrieval of promoters based on expression pattern
- Prior knowledge about the expression pattern of a specific gene is unnecessary for data mining using READS

Construction of READS Dataset



- Literature mining
- Microarray data analysis
- Sequencing of upstream regions from rice landraces at PBT lab
- Promoter annotation with expression pattern, GC content, putative motifs and linked with references

READS Database Scheme



Statistics of READS Data Content (Current Release)

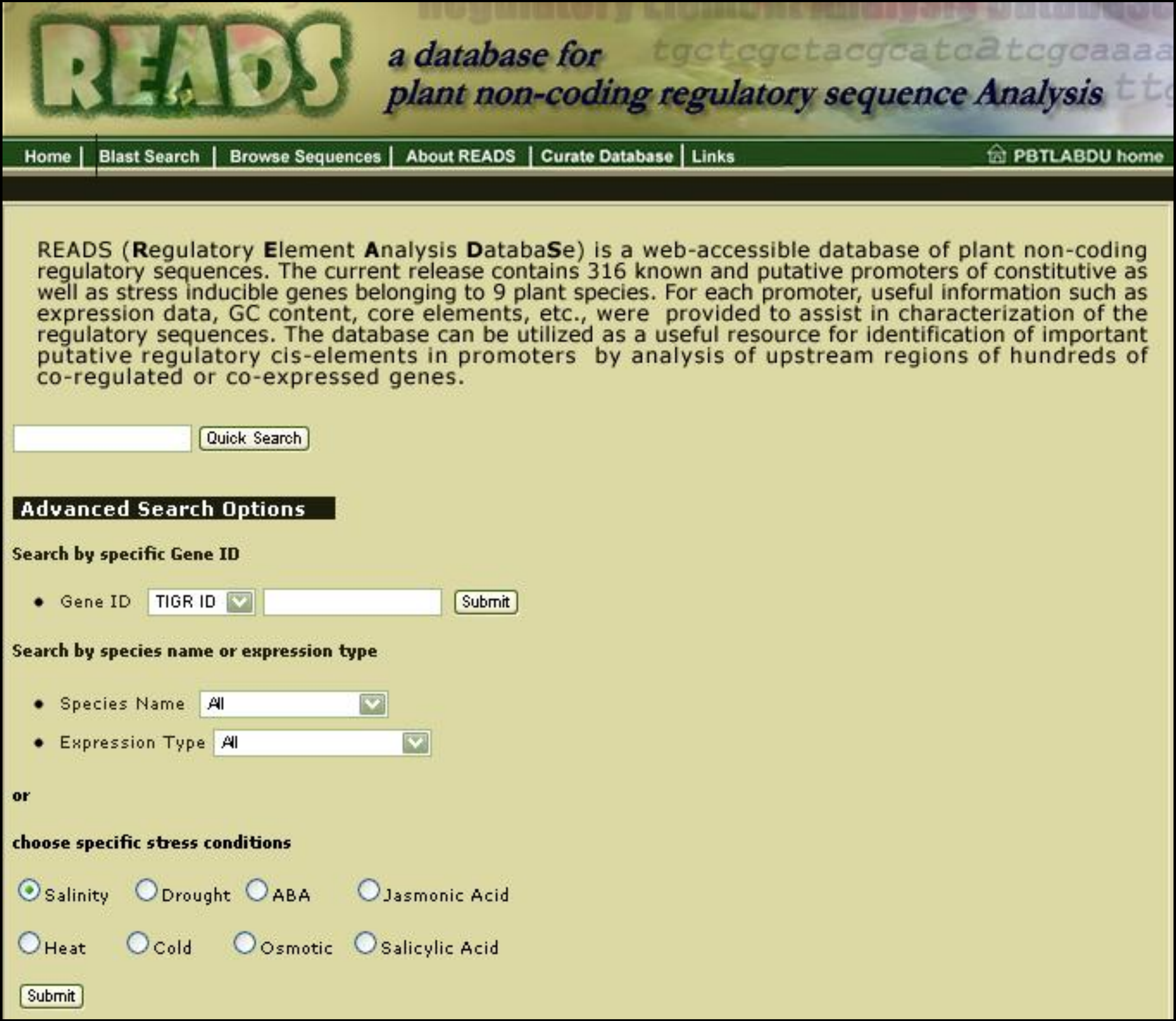
Category	Number
Total plant species	9
Monocots	6
Dicots	3
Halophytes	3
Total sequences	316
Constitutive promoters	218
Stress-inducible promoters	98

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

Website Architechture of READS

- www.pbtlabdu.net/READS/
- MYSQL based relational database
- Web server (Apache) and Perl modules
- FTP server for bulk data exchange
- Analysis program: NCBI BLAST
- Boolean logic in the quick and advanced search options



a database for
plant non-coding regulatory sequence Analysis

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.

Advanced Search Options

Search by specific Gene ID

- Gene ID: TIGR ID

Search by species name or expression type

- Species Name: All
- Expression Type: All

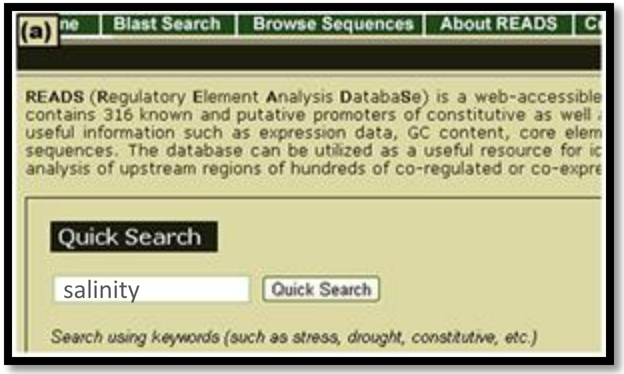
or

choose specific stress conditions

● Salinity ○ Drought ○ ABA ○ Jasmonic Acid

○ Heat ○ Cold ○ Osmotic ○ Salicylic Acid

Implementation of READS

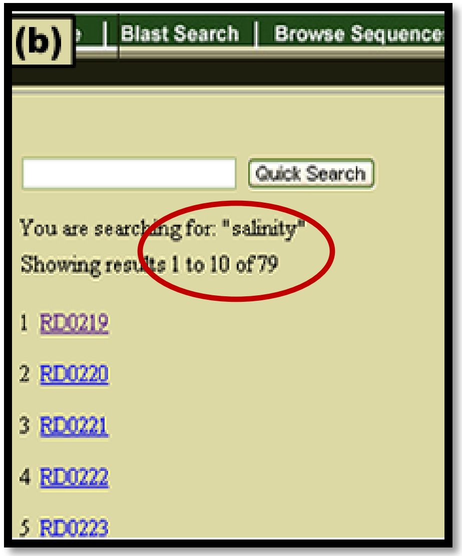


READS (Regulatory Element Analysis DatabaSe) is a web-accessible
contains 316 known and putative promoters of constitutive as well
useful information such as expression data, GC content, core elem
sequences. The database can be utilized as a useful resource for ic
analysis of upstream regions of hundreds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

Quick Search

salinity

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



You are searching for: "salinity"

Showing results 1 to 10 of 79

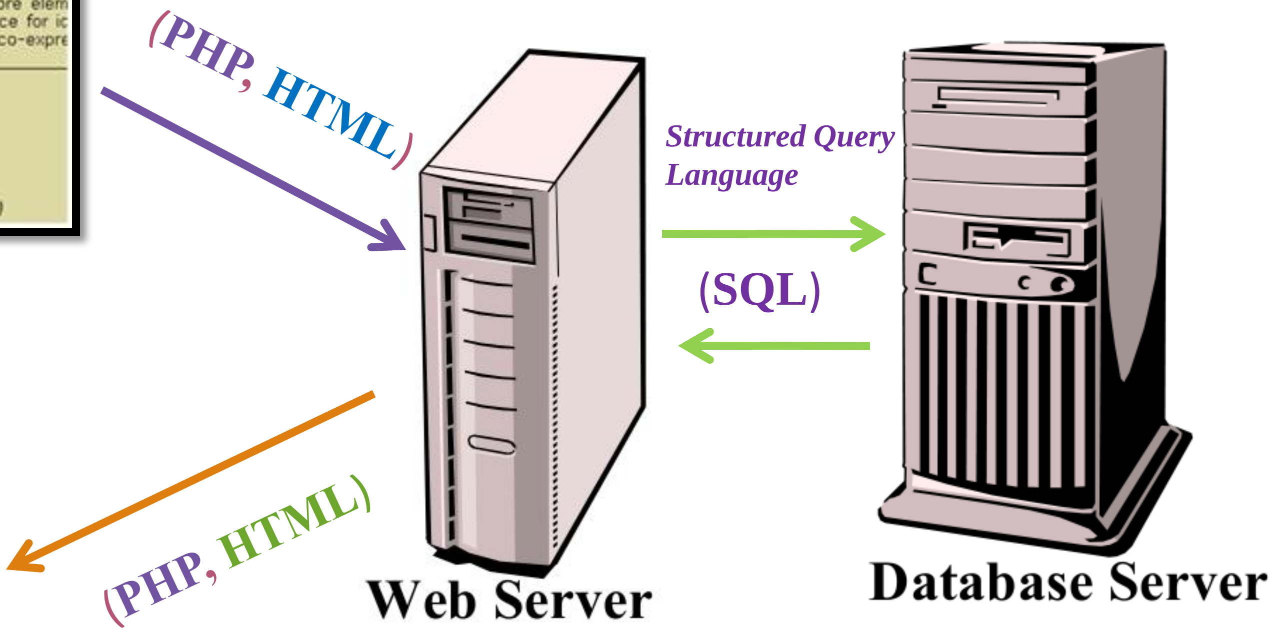
1 RD0219

2 RD0220

3 RD0221

4 RD0222

5 RD0223



(PHP, HTML)

Structured Query Language

(SQL)

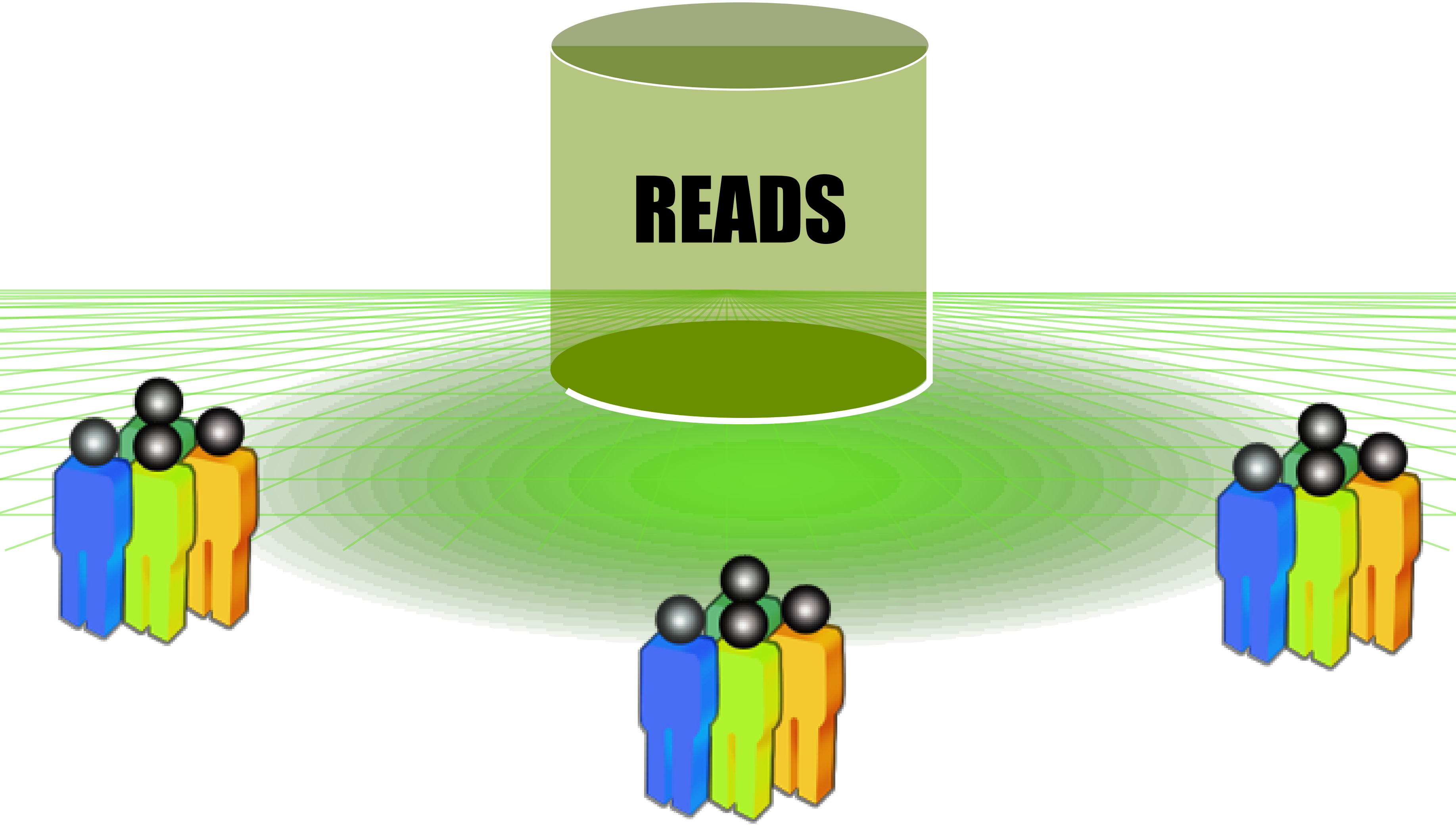
(PHP, HTML)

Web Server

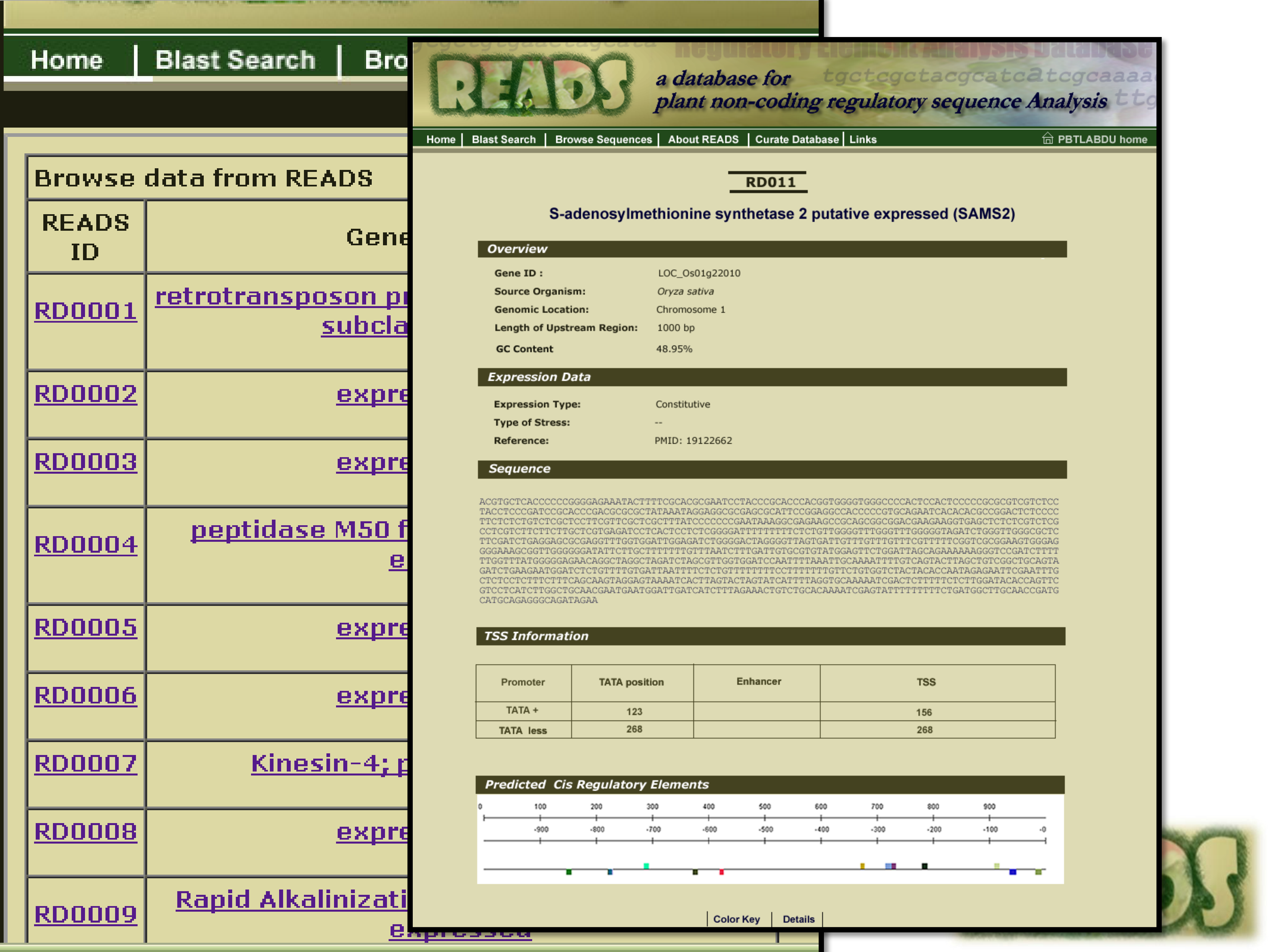
Database Server

(Apache)

Information we can get from READS



READS



Browse data from READS

READS ID	Gene
RD0001	retrotransposon p subcla
RD0002	expre
RD0003	expre
RD0004	peptidase M50 f e
RD0005	expre
RD0006	expre
RD0007	Kinesin-4; p
RD0008	expre
RD0009	Rapid Alkalinizati expressed

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

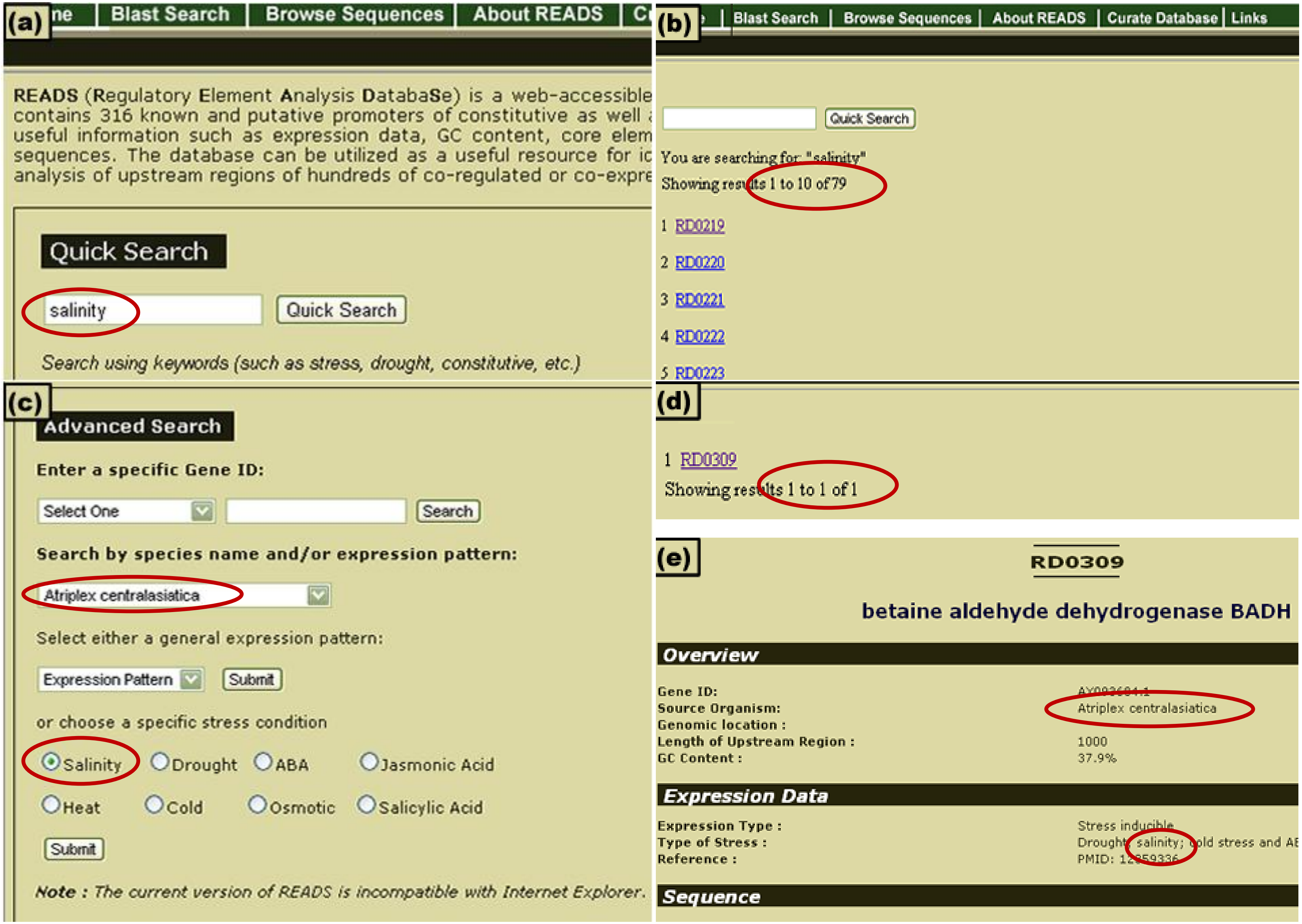
Sequence

```
ACGTGCTCACCCCCCGGGGAGAAATACTTTTCGCACGCGAATCCTACCCGCACCCACGGTGGGGTGGGCCCCACTCCACTCCCCCGCGCGTCGTCTCC
TACCTCCCGATCCGCACCCGACGCGCGCTATAAATAGGAGGCGCGAGCGCATTCCGGAGGCCACCCCCGTGCAGAATCACACACGCCGGACTCTCCCC
TTCTCTCTGTCTCGCTCCTTCGTTCGCTCGCTTTATCCCCCCCGAATAAAGGCGAGAAGCCGCAGCGGCGGACGAAGAAGGTGAGCTCTCTCGTCTCG
CCTCGTCTTCTTCTTGCTCGTGAGATCCTCACTCCTCTCGGGGATTTTTTTTTCTCTGTTGGGGTTTGGGTTTGGGGGTAGATCTGGGTTGGGCGCTC
TTCGATCTGAGGAGCGCGAGGTTTGGTGGATTGGAGATCTGGGGACTAGGGGTTAGTGATTGTTTGTTTGTTTCGTTTTTCGGTCGCGGAAGTGGGAG
GGGAAAGCGGTTGGGGGGATATTCTTGCTTTTTTTGTTTAATCTTTGATTGTGCGTGTATGGAGTTCTGGATTAGCAGAAAAAAGGGTCCGATCTTTT
TTGGTTTATGGGGGAGAACAGGCTAGGCTAGATCTAGCGTTGGTGGATCCAATTTTAAATTGCAAAATTTTGTCAGTACTTAGCTGTCGGCTGCAGTA
GATCTGAAGAATGGATCTCTGTTTTGTGATTAATTTTCTCTGTTTTTTTTCCTTTTTTTGTTCTGTGGTCTACTACACCAATAGAGAATTCGAATTTG
CTCTCCTCTTTCTTTCAGCAAGTAGGAGTAAAATCACTTAGTACTAGTATCATTTTAGGTGCAAAAATCGACTCTTTTTCTCTTGGATACACCAGTTC
GTCCTCATCTTGGCTGCAACGAATGAATGGATTGATCATCTTTAGAAACTGTCTGCACAAAATCGAGTATTTTTTTTTCTGATGGCTTGCAACCGATG
CATGCAGAGGGCAGATAGAA
```

TSS Information

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

Predicted Cis Regulatory Elements



READS (**R**egulatory **E**lement **A**nalysis **D**ataba**S**e) 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

salinity

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

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

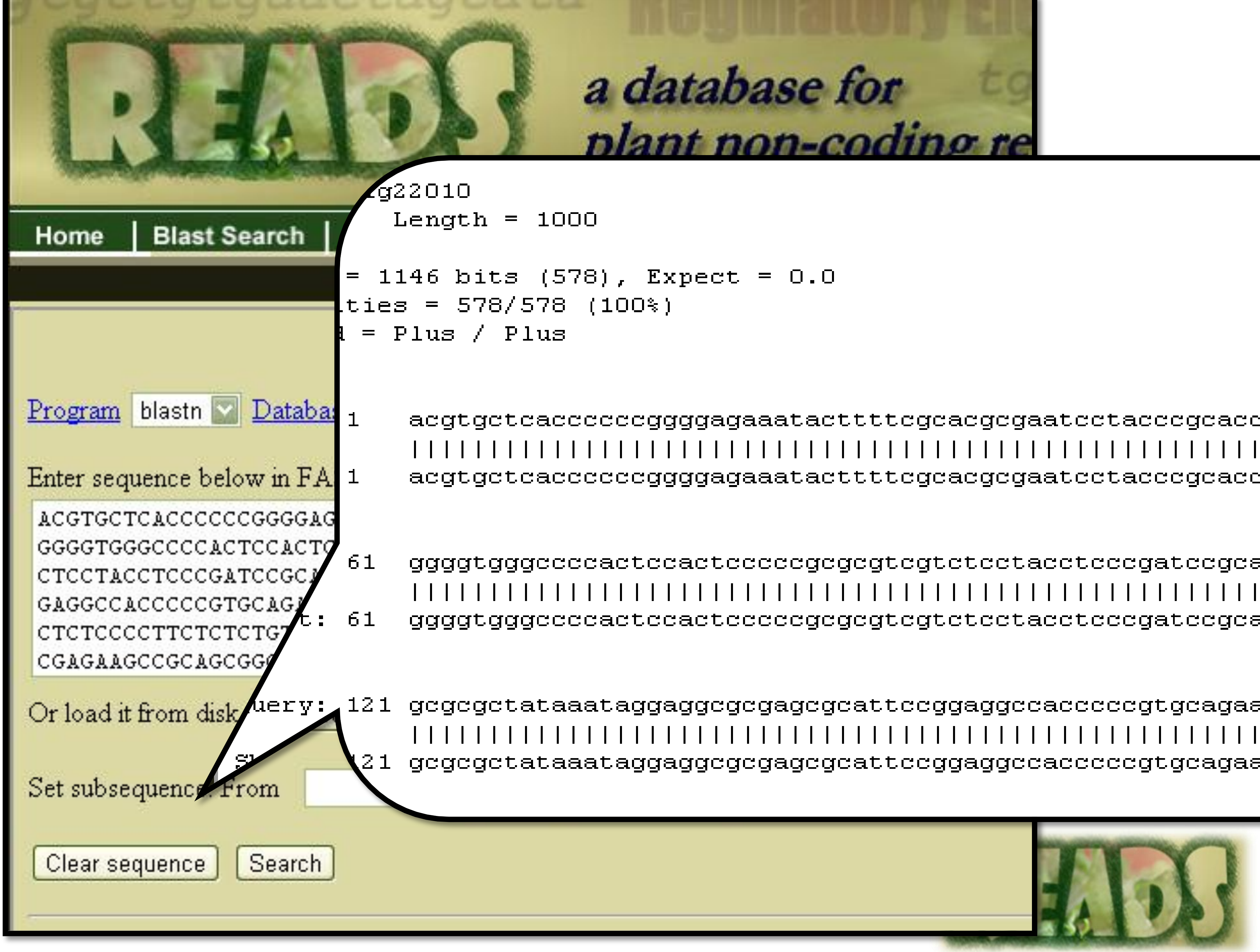
Showing results 1 to 1 of 1

RD0309

betaine aldehyde dehydrogenase BADH

Gene ID: AY093604.1
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 A
Reference : PMID: 12859336



Program blastn

Enter sequence below in FA

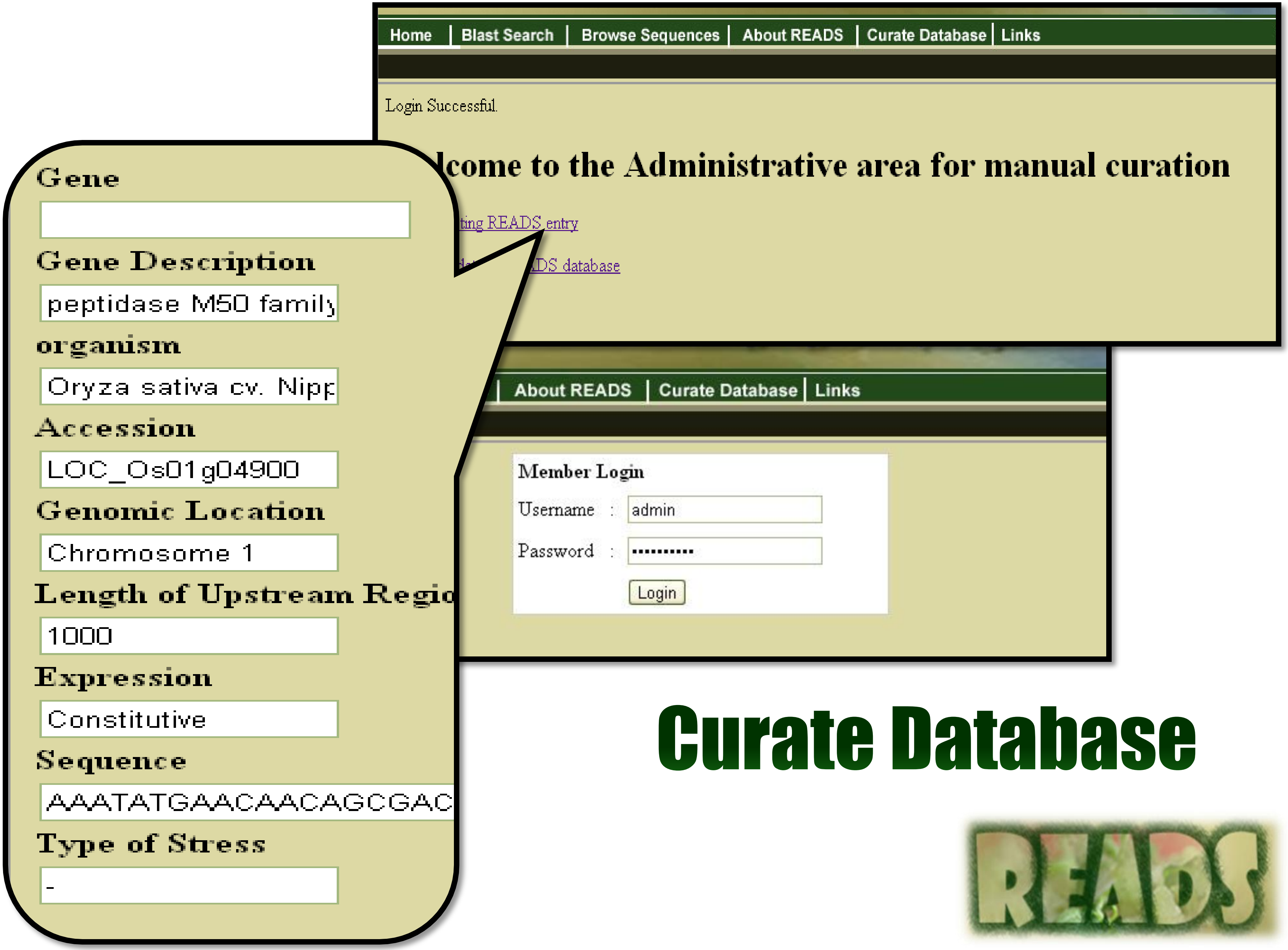
```
ACGTGCTCACCCCCCGGGGAG
GGGGTGGGCCCCACTCCACTC
CTCCTACCTCCCGATCCGCA
GAGGCCACCCCCGTGCAGA
CTCTCCCCTTCTCTCTGT
CGAGAAGCCGCAGCGG
```

Or load it from disk

Set subsequence From

```
g22010
  Length = 1000

= 1146 bits (578), Expect = 0.0
ties = 578/578 (100%)
d = Plus / Plus


     1   acgtgctcaccccccggggagaaatactttttcgcacgcgaatcctacccgcacc
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1   acgtgctcaccccccggggagaaatactttttcgcacgcgaatcctacccgcacc


    61   ggggtgggccccactccactcccccgcgcgtcgtctcctacctcccgatccgca
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
  t: 61  ggggtgggccccactccactcccccgcgcgtcgtctcctacctcccgatccgca


uery: 121 gcgcgctataaataggaggcgcgagcgcattccggaggccacccccgtgcagaa
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Sb  121 gcgcgctataaataggaggcgcgagcgcattccggaggccacccccgtgcagaa
```

Curate Database



Login Successful.

lcome to the Administrative area for manual curation

ting READS entry

DS database

Member Login

Username : admin

Password : ••••••••••

Gene

Gene Description

peptidase M50 famil

organism

Oryza sativa cv. Nipp

Accession

LOC_Os01g04900

Genomic Location

Chromosome 1

Length of Upstream Regio

1000

Expression

Constitutive

Sequence

AAATATGAACAACAGCGAC

Type of Stress

-

READS can be helpful for........

- Discriminative motif identification
- Select suitable promoters for transgene expression under specific condition or stage
- Primer design for isolation and cloning experiments

Future Endeavor to Update READS

- Addition of more sequences from diverse species, specially halophytes
- Integration of novel tools with database to aid promoter analysis

Credits

READS

Zeba I. Seraj

- ***Development & Website:***

  Sabrina M. Elias

- ***Curation:***

  Saima Shahid

  Sudip Biswas

Hosted at:

Plant Biotechnology Lab
University of Dhaka

Thank You