INDEPENDENT BIOINFORMATICS PROJECT

Comparative and Phylogenetic Analysis of DREB2B-Like Proteins Across Selected Plant Species

Protein-sequence comparison, multiple sequence alignment and phylogenetic reconstruction

Prepared by	Suneel Kumar
Academic background	BSc Plant Breeding and Genetics
Institution	Sindh Agriculture University, Tandojam
Bioinformatics tools	NCBI BLASTP · MUSCLE · MEGA 12
Dataset	21 protein records from selected plant species

Abstract

This project examines 21 DREB2B-like protein records from selected plant species. Twenty records were collected from the BLASTP subject results, together with the *Triticum aestivum* query sequence (BFU48973.1). The protein sequences were aligned using MUSCLE, and their relationships were inspected in a Neighbor-Joining tree displayed in MEGA 12. The report documents the supplied BLASTP statistics, the visible multiple-sequence alignment and the phylogenetic tree. The interpretation is limited to patterns that can be seen in the supplied results and screenshots.

Keywords: DREB2B-like proteins; protein sequence; BLASTP; MUSCLE; multiple sequence alignment; Neighbor-Joining; MEGA 12; phylogenetic analysis.

Project scope

The work is an in-silico comparison based on protein sequence records. It does not include laboratory experiments, gene-expression measurements or functional validation. The term “DREB2B-like” is retained to describe the records as labelled in the supplied BLASTP results.

Workflow at a glance

01	02	03	04
Sequence search	BLASTP results	MUSCLE alignment	MEGA 12 tree
NCBI protein records	Similarity statistics	21 protein sequences	Neighbor-Joining display

1. Background and aims

DREB proteins are plant transcription factors associated with responses to environmental conditions such as dehydration and other abiotic stresses. Comparing related protein sequences provides a way to examine sequence similarity and the groupings produced by a phylogenetic analysis. The practical aims of this project were to identify related protein records, compare their BLASTP statistics, align the sequences and inspect the resulting tree.

2. Materials and methods

Sequence dataset

The dataset contains 21 protein records: 20 BLASTP subject records and one *Triticum aestivum* query sequence, accession BFU48973.1 (354 amino acids, as listed in the supplied table). Accession numbers, organism names and BLASTP statistics are reproduced in Table 1.

Similarity search and record selection

The project used NCBI BLASTP output to identify protein records similar to the DREB2B-like query. The table reports maximum score, total score, query coverage, E-value, percentage identity and sequence length as provided by the user. These values have not been independently re-queried for this report.

Multiple sequence alignment

The 21 protein sequences were aligned with MUSCLE. Figure 1 reproduces the supplied MEGA 12 Alignment Explorer screenshot. It is a screenshot of the visible alignment region, not a complete machine-readable alignment export.

Phylogenetic analysis

A Neighbor-Joining tree was displayed in MEGA 12. Figure 2 reproduces the supplied Tree Explorer screenshot, including the visible branch-support values and the 0.10 scale bar.

Information that must be confirmed

The supplied screenshots do not show the bootstrap replicate count, amino-acid substitution model or gap-treatment setting. These parameters are therefore not stated as facts in this report. They should be checked in the original MEGA 12 analysis and added to the methods before formal academic submission.

Reproducibility note

For a reviewer to reproduce the analysis, retain the original FASTA file, MUSCLE alignment, MEGA project file and an exported high-resolution tree alongside this PDF. The same accession set should be used consistently across all files.

3. BLASTP results

Table 1. Protein records and BLASTP statistics supplied for this project. Sequence lengths are reported in amino acids.

No.	Protein record	Accession	Organism	Max score	Total score	Query cover	E-value	Identity	Length (aa)
1	DREB2B-like isoform X2	XP_044973707.1	Hordeum vulgare	518	518	100%	0.0	73.97%	368
2	DREB2B-like	XP_044953123.1	Hordeum vulgare	342	342	62%	2e-113	80.00%	300
3	DREB2B-like isoform X2	XP_062180805.1	Phragmites australis	317	317	100%	7e-103	53.19%	343
4	DREB2B-like isoform X4	XP_021302434.1	Sorghum bicolor	300	300	100%	2e-96	50.00%	335
5	DREB2B-like isoform X3	XP_062231128.1	Phragmites australis	301	301	89%	3e-96	55.82%	391
6	DREB2B-like	RLN00548.1	Panicum miliaceum	293	293	100%	1e-93	53.48%	330
7	DREB2B isoform X2	XP_004962409.1	Setaria italica	293	293	100%	1e-93	51.89%	332
8	DREB2B-like isoform X4	XP_015693218.1	Oryza brachyantha	276	276	89%	2e-86	53.68%	377
9	DREB2B-like isoform X2	XP_015693216.1	Oryza brachyantha	277	277	94%	3e-86	52.47%	423
10	DREB2B isoform 3	NP_001389443.1	Oryza sativa Japonica	269	269	100%	3e-83	47.93%	419
11	DREB2B	OEL15561.1	Dichanthelium oligosanthes	261	261	100%	5e-81	51.99%	340

The query sequence is shown in row 21 of the continuation table. Query-only fields are retained as “Query” because no subject-comparison statistics apply to that row.

3. BLASTP results (continued)

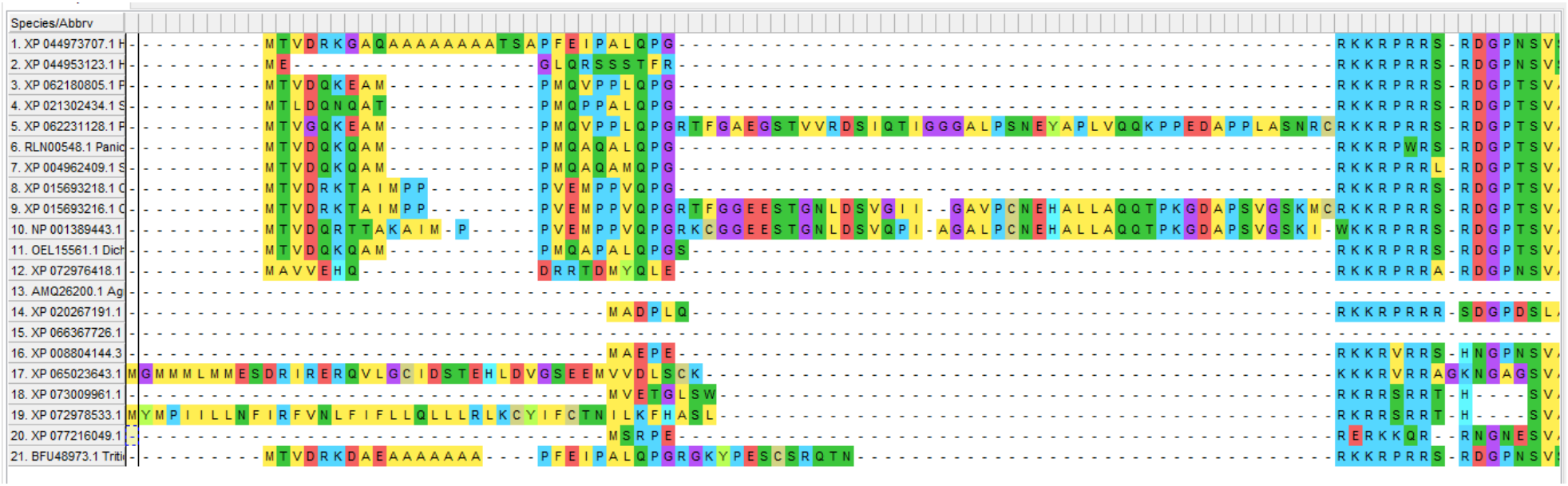
No.	Protein record	Accession	Organism	Max score	Total score	Query cover	E-value	Identity	Length (aa)
12	DREB2B-like	XP_072976418.1	Typha angustifolia	236	236	70%	1e-71	54.37%	314
13	DREB2B, partial	AMQ26200.1	Agrostis stolonifera	225	225	56%	5e-69	62.09%	206
14	DREB2B-like	XP_020267191.1	Asparagus officinalis	207	207	76%	4e-61	45.90%	274
15	DREB2B-like isoform X2	XP_066367726.1	Miscanthus floridulus	202	202	75%	3e-59	50.35%	257
16	DREB2B-like isoform X1	XP_008804144.3	Phoenix dactylifera	193	193	70%	5e-55	53.20%	322
17	DREB2B-like isoform X3	XP_065023643.1	Musa acuminata AAA Group	194	194	81%	1e-54	42.53%	372
18	DREB2B-like isoform X1	XP_073009961.1	Typha latifolia	186	186	70%	1e-52	46.83%	280
19	DREB2B-like isoform X1	XP_072978533.1	Typha angustifolia	186	186	70%	2e-52	46.40%	315
20	DREB2B-like	XP_077216049.1	Tasmannia lanceolata	184	184	69%	4e-52	43.51%	271
21	DREB2B transcript variant X2	BFU48973.1	Triticum aestivum (Query)	Query	Query	Query	Query	Query	354

Data note

The protein names, accession numbers and numerical values in Table 1 have been retained from the supplied BLASTP table. They have not been independently checked against current NCBI records. AMQ26200.1 is labelled as a partial protein in the supplied data and has a reported length of 206 amino acids.

4. Multiple sequence alignment

Figure 1. MUSCLE protein alignment displayed in MEGA 12. The screenshot shows the visible alignment region for the 21 sequence records.



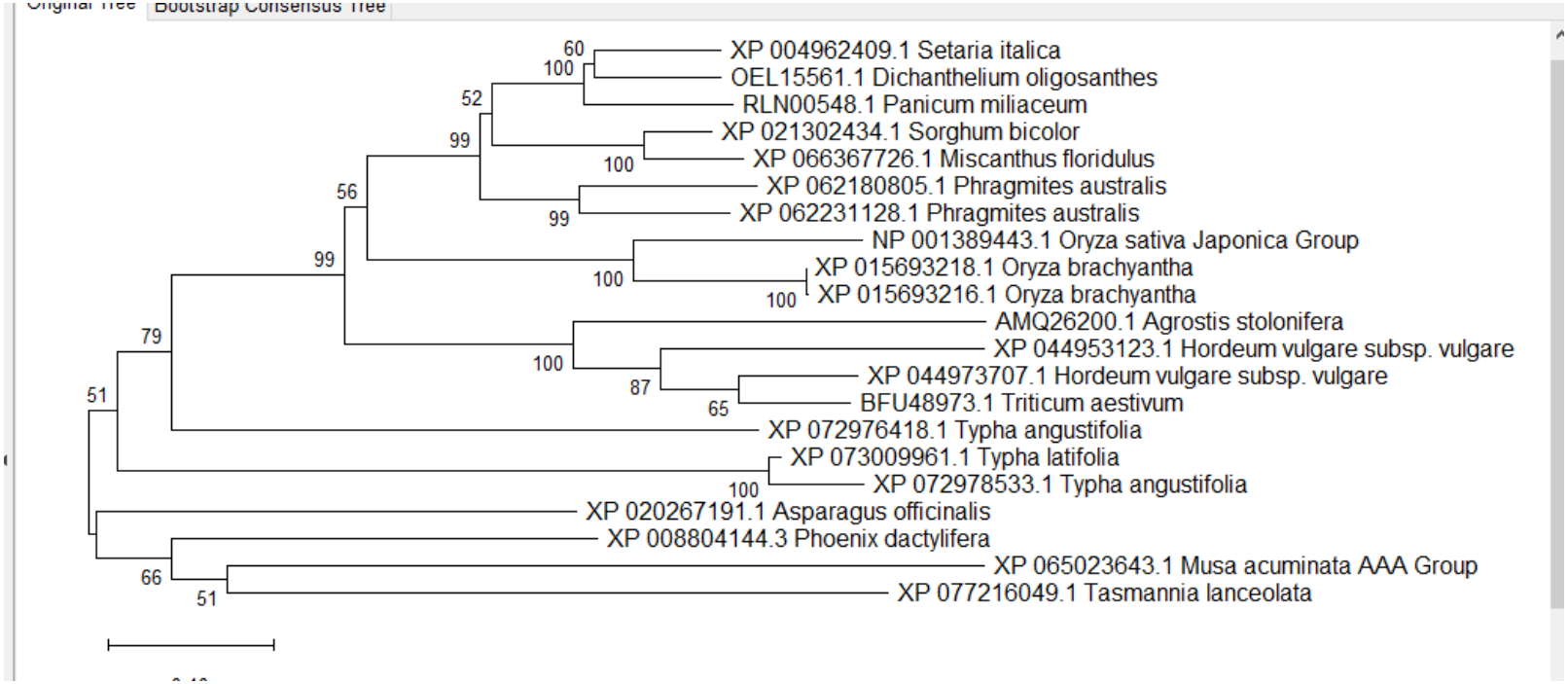
The figure is a screenshot of the displayed Alignment Explorer region, rather than a complete export of every alignment position. The original alignment file should be retained with the project data so that the full alignment can be reviewed.

Interpretation

A multiple sequence alignment places amino-acid residues into corresponding positions so that conserved and variable regions can be compared. The screenshot documents the alignment display; a detailed conservation analysis would require examination of the full alignment file.

5. Phylogenetic analysis

Figure 2. Neighbor-Joining tree for the 21 protein records, as displayed in MEGA 12. Visible branch-support values and the 0.10 scale bar are reproduced from the supplied screenshot.



The dataset includes taxa outside the grasses, including *Asparagus officinalis*, *Phoenix dactylifera*, *Musa acuminata* and *Tasmannia lanceolata*. The title and description therefore refer to selected plant species rather than grasses alone.

Tree-reading note

In a phylogenetic tree, nearby branching patterns indicate the relationships inferred from the aligned sequences and the selected analysis method. The scale bar represents the branch-length scale shown in the MEGA display. Branch-support values indicate how consistently a grouping was recovered under the resampling procedure used; the number of replicates is not visible in the supplied screenshot.

6. Results and interpretation

The supplied tree shows several close groupings. The two *Phragmites australis* records (XP_062180805.1 and XP_062231128.1) form a pair with a displayed support value of 99. The two *Oryza brachyantha* records (XP_015693218.1 and XP_015693216.1) form a pair with a displayed value of 100; the *Oryza sativa* record appears in the same broader cluster.

The two *Hordeum vulgare* records (XP_044953123.1 and XP_044973707.1) appear as a pair with a displayed support value of 87. The *Typha latifolia* and *Typha angustifolia* records form a pair with a displayed value of 100. *Setaria italica* and *Dichanthelium oligosanthes* also appear together with a displayed value of 100. These values are reported as they appear in the supplied tree screenshot.

The *Agrostis stolonifera* record (AMQ26200.1) is identified as partial in the supplied BLASTP table and has a reported length of 206 amino acids. Because it covers a shorter sequence, its placement should be interpreted with care when compared with the full-length records.

These observations describe the supplied tree; they do not establish that the proteins have identical functions or experimentally confirmed biological roles. A stronger interpretation would include conserved-domain analysis, sequence-quality checks and confirmation of the tree settings.

7. Limitations and recommended verification

- Confirm the bootstrap replicate count, amino-acid substitution model and gap-treatment setting in the original MEGA 12 project.
- Verify that the same 21 accession numbers occur in the FASTA file, alignment and tree.
- Check sequence completeness and review the partial AMQ26200.1 record.
- Export a high-resolution tree with readable accession labels.
- Preserve the FASTA sequences, alignment and MEGA project file with the report.

8. Conclusion

This project assembled 21 DREB2B-like protein records for BLASTP comparison, MUSCLE alignment and Neighbor-Joining phylogenetic analysis in MEGA 12. The supplied tree displays several supported groupings among the selected plant records. The work documents a practical bioinformatics workflow and provides a basis for further sequence and domain analysis. The interpretation remains limited by the analysis settings that were not visible in the supplied screenshots.

9. Software and database resources

NCBI Protein Database — https://www.ncbi.nlm.nih.gov/protein/
NCBI BLAST — https://blast.ncbi.nlm.nih.gov/Blast.cgi
MEGA 12 — https://www.megasoftware.net/
MUSCLE — https://www.drive5.com/muscle/

Prepared as an independent student portfolio project by Suneel Kumar. No claim of peer review or journal publication is made.