In silico identification, phylogenetic analysis and protein modeling of EREBP-1 genes of Phaseolus vulgaris, Arabidopsis thaliana and Cucumis sativus

A K Mishra*, Parush Chawla, Chetna Tyagi, Bharti Pandey, Ohika Chakraborty, Amrrender Kumar and A K Jain
Agricultural Knowledge Management Unit, Indian Agricultural Research Institute, New Delhi 110012, India

Received 29 November 2014; revised 28 May 2015; accepted 19 June 2015

Transcription factors (TFs) are proteins that bind to DNA and activate or repress gene expression at the transcriptional level. Ethylene-Responsive Element Binding Protein (EREBP) represents one such class of DNA binding proteins/TFs. In the present study, EREBP-1 gene of Glycine max was selected to find out similar genes in Phaseolus vulgaris, Arabidopsis thaliana and Cucumis sativus dicot plants by taking their protein sequences. A combined phylogenetic tree was constructed, which yielded four EREBP subgroups. From each subgroup, one protein sequence was evaluated, which had highest Bits (score) and lowest E-value. Consequently, two EREBPs of P. vulgaris and two of C. sativus were chosen from the four subgroups of the phylogenetic tree. Four EREBP structures were then modelled using ab initio based approach and were energy minimized. They were validated on the basis of Ramachandran plot analysis, where residues were in most favored region ranging from 75.80-87.70%. Further, intrinsically disordered regions were found out representing the ethylene-responsive nucleotide binding regions. All of the four modelled EREBP structures have been reported in this study with elaborate analysis and stability evaluation using a number of computational techniques.

Keywords: EREBP, phylogenetic analysis, protein modeling, transcription factor

Introduction

Transcription factors (TFs) are types of proteins that affect many biological processes, such as, growth, development and cell division, and responses to biotic and abiotic stresses. Plants have developed such a mechanism because of their inability to escape predation or environmental changes. Biotic stresses (fungal, bacterial as well as viral challenges) and abiotic stresses (cold, drought, wounding & salinity), which are sensed through a complex signal transduction network, result in biochemical, physiological and gene expression changes. TFs bind to DNA and either activate or repress gene expression at the transcriptional level. They mainly contain a DNA binding domain and a transcriptional activation domain; the DNA binding domain recognizes the target DNA sequences, while the transcriptional activation domain initiates transcription.

Ethylene is an endogenous and most intensely studied plant hormone that influences many aspects of plant growth and development, such as, germination, senescence and fruit ripening. The ethylene responsive element binding proteins (EREBPs) are the DNA binding proteins that exhibit no sequence homology with other known transcription factors or DNA binding proteins like bZIP and zinc finger families, because most of the known plant-derived bZIP proteins share DNA binding specificity for motifs that contain the core sequence ACGT and have a highly conserved basic region that is predicted to be the DNA binding domain. The DNA binding domain in EREBPs are found in both dicots and monocots, suggesting that it is evolutionary conserved in plants, which shows that it is a new class of DNA binding proteins relevant for the regulation of transcription. Conserved EREBP domains bind to the GCC box, which is an 11 bp sequence (TAAGAGCCGCC), an ethylene responsive element ERE or promoter element found in many pathogenesis related (PR) genes. The GCC box is necessary and sufficient for ethylene (ET) regulation of plant genes. EREBP-1 consists of 60 to 70 amino acids and is involved in DNA binding. The members of EREBP family may play wider role in responses to various stresses. In the present study, EREBP-1 of Glycine max, which is a 202 amino acid long, was chosen as template to find out similar genes in
Arabidopsis thaliana (dicot), Cucumis sativus (dicot) and Phaseolus vulgaris (dicot) plants. Therefore, identification, phylogenetic analysis and protein modeling of EREBP-1 genes in the above species were carried out.

Materials and Methods

Data Retrieval

G. max EREBP-1 [genpept acc. no.: AAM45475] was taken as the template sequence (202 amino acids) from NCBI. In the next step, the protein sequences of P. vulgaris (14.5 Mb) (phytozome.net/phaseolus.php), Arabidopsis thaliana (19.3 Mb) (phytozome.net/Arabidopsis.php) and C. sativus (12.9 Mb) (phytozome.net/cucumis.php) were downloaded from Phytozome v9.1.22. BLASTP was carried out taking G. max EREBP-1(AAM45475) protein sequence as query and the complete proteome of P. vulgaris, A. thaliana, and C. sativus, taking Bits (score) ≥ 100 and E-value ≤ 20 as a criteria to find significant blast hits.

Multiple Sequence Alignment and Phylogenetic Analysis

All top-ranking sequences of the three species were saved in a single fasta file. MEGA6.0623 software was used for the multiple sequence alignment (MSA) of all these sequences using CLUSTALW24. The phylogenetic tree was constructed using Neighbor-Joining(NJ)25 method with the parameters, such as, Poisson correlation26, pair wise deletion and bootstrap27 (1000 replicates; random seed).

In silico 3D Protein Modeling and Energy Minimization

Online BLASTP28 was performed with percentage identity of blast protein hits < 35 and very low query coverage. Due to lack of suitable structural template for EREBP, ab initio based modeling of EREBP using I-TASSER, a web based full chain protein prediction server, was used. Using I-TASSER, a comprehensive protocol was followed for protein modeling.

To check the quality of the modelled protein, certain scores like C-score and TM-score were calculated for ranking. The C-score is defined by:

\[
C \text{ score} = \ln \left( \frac{M}{M_{tot}} \cdot \frac{1}{\langle RMSD \rangle} \prod_{i=1}^{L} z(i) \right)
\]

Where M is the multiplicity of structures in the SPICKER cluster; \(M_{tot}\) is the total number of the I-TASSER structure decoys used in the clustering; RMSD is the average RMSD of the decoys to the cluster centroid; \(z(i)\) is the highest Z score (the energy to mean in the unit of standard deviation) of the templates by the \(i^{th}\) PPA threading program and \(Z_{d}(i)\) is a program-specified Z-score cut-off for distinguishing between good and bad templates, i.e., \(Z_{d}(1) = 7.0, Z_{d}(2) = 8.5, Z_{d}(3) = 8.0, Z_{d}(4) = 10.5\).

The TM score is defined to measure the topological similarity of two proteins:

\[
TM \text{ score} = \frac{1}{L} \sum_{i=1}^{L} \frac{1}{1 + \left( \frac{d_i}{d_0^i} \right)^2}
\]

Where \(d_i\) is the distance of the \(i^{th}\) pair of residues between two structures after an optimal superposition, \(d_0 = 1.24 L - 15 - 1.8\) and \(L\) is the protein length. TM score lies in between 0 to 1 with higher values indicating better models. Statistically, a TM-score ≤ 0.17 corresponds to a similarity between two randomly selected structures from the PDB library, while a TM-score > 0.5 corresponds approximately to two structures of the similar topology.

The energy minimization of the structures was done by the YASARA Energy Minimization Server30 and comparison of the energy and score of the structures before and after energy minimization was done as shown in Table 1. YASARA force field is mainly based on knowledge based potentials. It has been incorporated to calculate highly informative knowledge-based energies, which result in the most accurate force fields for structure prediction and refinement. This force field combined the AMBER force field equation with “multi-dimensional knowledge-based torsional potentials” to maximize the accuracy. To ensure that all forces responsible for

---

<table>
<thead>
<tr>
<th>No.</th>
<th>Acc. no.</th>
<th>Start energy (kJ/mol)</th>
<th>End energy (kJ/mol)</th>
<th>Start score</th>
<th>End score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phvul.006G179800.1</td>
<td>-008012.0</td>
<td>-072549.90</td>
<td>-6.04</td>
<td>-2.94</td>
</tr>
<tr>
<td>2</td>
<td>Cucsa.129320.1</td>
<td>000391.80</td>
<td>-107088.90</td>
<td>-5.70</td>
<td>-2.29</td>
</tr>
<tr>
<td>3</td>
<td>Cucsa.129630.1</td>
<td>049688.10</td>
<td>-089436.0</td>
<td>-6.84</td>
<td>-3.85</td>
</tr>
<tr>
<td>4</td>
<td>Phvul.008G046400.1</td>
<td>-022113.10</td>
<td>-106665.7</td>
<td>-4.28</td>
<td>-1.70</td>
</tr>
</tbody>
</table>
the experimentally observed structure should be considered, minimizations were done in crystal space, using complete unit cells. As a result, one obtained a force field that had stable energy minima as close as possible to native structures. This was essentially equivalent to a force field that moved models closer to the native structures during a simulation. All the energy minimized 3D EREBP structures were saved with their .pdb extension.

Validation of Energy Minimized 3D Protein Structures
The quality of backbone conformation of the model was assessed by PROCHECK based on Ramachandran plot for reliability. The observed Psi-Phi pairs were evaluated to be present in most favoured regions, in additional allowed regions, in generously allowed regions and in disallowed regions.

Disorder Prediction
The disorder prediction for EREBPs was done using a web-based server named IUPred\(^3\). It is a novel algorithm presented by the IUPred server for predicting intrinsically unstructured/disordered regions from amino acid sequences by estimating their total pair wise inter-residue interaction energy, assuming that IUP sequences do not fold due to their inability to form sufficient stabilizing inter-residue interactions. In this method, the pairwise interaction energies of a few selected globular proteins can be estimated using a quadratic expression set for each residue in terms of whether it contributes to the ordered or disordered regions. This contribution depends upon its own chemical composition as well as its sequential interaction partners as a whole.

Results and Discussion
Three plants were used for this study, namely, \(A.\) \(thaliana\), \(P.\) \(vulgaris\) and \(C.\) \(sativus\). The advantages of the study of genome analysis of \(A.\) \(thaliana\) include a short generation time, small size, and large number of offspring. That is why it is a popular model organism used in Plant Biology and Genetics due to the relatively short life cycle, small genome of approx 125 mega base-pairs (Mbps) and has five chromosomes. \(P.\) \(vulgaris\), commonly called as common bean is a crop of major importance for human consumption in the developing parts of the world. The size of its genome is approx 587 Mbps and has 11 chromosomes. It has approx 27,197 protein coding genes\(^18\). The third species \(C.\) \(sativus\), commonly called as cucumber, is an economically important cultivated plant as well as a model system for studies on sex determination and plant vascular biology as well. \(G.\) \(max\), commonly called as soybean, is one of the most important crop plants for seed protein and oil content, and also known for its capacity to fix atmospheric nitrogen through symbiosis with soil-borne microorganisms.

Data Retrieval
The protein sequences of \(P.\) \(vulgaris\) (14.5 Mb), \(A.\) \(thaliana\) (19.3 Mb) and \(C.\) \(sativus\) (12.9 Mb) were downloaded. The total number of genes present in the directory was 31,638, 35,386 and 30,364 for \(P.\) \(vulgaris\), \(A.\) \(thaliana\) and \(C.\) \(sativus\), respectively. \(G.\) \(max\) EREBP-1 [genpept acc. no.: AAM45475], comprising of 202 amino acids, was taken as the query sequence from NCBI database. BLASTP was performed between \(G.\) \(max\) EREBP-1(AAM45475) protein sequence and the protein sequences of \(P.\) \(vulgaris/A.\) \(thaliana/C.\) \(sativus\). As a result, 17 blast hits were found in \(P.\) \(vulgaris\), while 13 and 9 were observed in \(A.\) \(thaliana\) and \(C.\) \(sativus\), respectively, taking Bits (score) \(\geq 100\) and E-value \(\leq 20\) as a criteria for significant blast hits.

Multiple Sequence Alignment and the Phylogenetic Analysis
Fig. 1 depicts the phylogenetic tree representing four subgroups. From each subgroup, the sequence which had the lowest E-value and highest Bits (score) was evaluated. The E-value stands for expectation value and it indicates the probability by which resulting alignments from a database search are caused by chance. It provides information about the likelihood by which the given sequences solely match by chance. The lower the E-value, less likely the match is a result of random chance and, therefore, the more significant that match is. The E-value is related to the P-value used to assess significance of single pairwise alignment. BLAST compares a query sequence against all database sequences, and so the E-value is determined by the following formula:

\[
E = m \times n \times P
\]

Where \(m\) is the total number of residues in a database, \(n\) is the number of residues in the query sequence and \(P\) is the probability by which an HSP (high-scoring segment pair) alignment is a result of random chance\(^30\). To find the best alignment score, here similarity based method was used. In this method, a bit score is assigned to an alignment according to the formula given below, based on the number of matched pairs, penalty imposed for gaps of a particular length and the number of gaps. Among all
the alignments the one having the maximum score is considered to be the best one.

\[ S = x - \sum w_k z_k \]

With \( S \) = similarity; \( x \) = no. of matched pairs; \( w \) = penalty for gaps of length \( k \); \( z \) = no. of gaps of length \( k \); \( nt \) = nucleotide; and \( w_1 \) = represents the gap length of 1st nt; therefore, the gap length of nucleotides is represented as the subscript number of \( w \).

**In silico Protein Modeling and Energy Minimization**

**Sequence Selection**

For further analysis, 4 EREBP sequences were selected from each subgroup of the finely constructed combined phylogenetic tree with the highest Bits (score) and lowest E-value of protein sequences. The four sequences were as follows: Phvul.006G179800.1|PACid:27154361, Cucsa.129320.1|PACid:16962264, Cucsa.129630.1|PACid:16962315, and Phvul.008G046400.1|PACid:27154361. BLASTp was performed with these selected EREBP sequences to find a template for protein modelling but, in each case, blast hits were covering very low percentage of the whole query coverage and with \( \leq 35\% \) identity. So, \textit{ab initio} based approach was preferably used for the analysis. The comparative modelling of EREBP protein sequence was performed using a scratch based approach implemented in I-TASSER. In each case, a set of 5 models of target protein was constructed based on the 10 best templates. The resulting 3D models of each subgroup [here each subgroup was represented by a single amino acid sequence of an EREBP based on highest Bits (score) and lowest E-value] were sorted according to the C-score.

The tertiary structure prediction was performed by I-TASSER server using the best aligned template by the BLAST program against the Protein Data Bank. Templates that were chosen are 3gccA, the ERF-1 of \textit{A. thaliana} for Phvul006g179800, Cucsa129320 and Cucsa129630 models, and 3gv2A (Capsid protein...
p24 + ccmK + homolog 4 of Human immunodeficiency virus type 1) for Phvul.008g046400.1. The template was selected to model and analyze 3D structure because a high level of sequence identity should guarantee a more accurate alignment between the target sequence and template whose structure is known. Out of five generated models of the target sequence, the best ones were chosen owing to the criteria of good alignment with the chosen template and C-score, TM score and RMSD values were calculated as shown in Table 2.

The C-score refers to the confidence in the quality of the predicted structure and is based on threading template alignments and convergence parameters involved in the simulation. It lies between −5 to 2, where a higher C-score signifies the absolute quality of the model and vice versa. The Phvul006g179800 EREBP model has a C-score of −3.68, Cucsa129320 model a C-score of −3.90, and Cucsa129630 a C-score of −2.76 and Phvul008g046400 a C-score of −1.91. To measure the structural similarity between two structures consequent in determining the quality of the predicted model, TM score and RMSD were calculated. However, owing to the error-prone state of the RMSD calculations, we relied on the TM score, in which the smaller distances between structures are weighted stronger than the larger distances and thus making it free of local errors. A TM value of more than 0.5 ensures a model of correct topology and a value of less than 0.17 refers to random similarity and hence should be discarded. The Phvul006g179800 EREBP model has a TM score of 0.31±0.10, Cucsa129320 model a TM score of 0.29±0.09, Cucsa129630 a TM score of 0.40±0.13 and Phvul008g046400 a TM score of 0.49±0.15. Based on the C score and TM score for all the models, it can be understood that the predicted models are of good quality.

**Energy Minimization**

Using YASARA energy minimization server, each protein model was input in the server that resulted in .sce files. The energy minimized protein models show little deviation from the original structures. The loop regions are the most affected and show larger shifts than the secondary structural regions. Energy minimization leads to the attainment of native-like conformation for almost all protein models predicted by computational methods and thus becomes a necessary exercise in such experiments. The energy minimized structures are provided in Fig. 2.

**Validation of Energy Minimized 3D Protein Structures**

The quality of backbone conformation of the model was assessed by PROCHECK in Ramachandran plot for reliability. PROCHECK checks the stereochemical quality of a protein by analyzing residue by residue geometry and overall structural geometry. Before the energy minimization, the observed Psi-Phi pairs had 72.00-86.60% residues in most favoured regions, 10.70-21.10% of residues in additional

<table>
<thead>
<tr>
<th>No.</th>
<th>Model</th>
<th>C-score</th>
<th>TM value</th>
<th>RMSD calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phvul.006g179800.1</td>
<td>PACid:27165358</td>
<td>-3.68</td>
<td>0.31±0.10</td>
</tr>
<tr>
<td>2</td>
<td>Cucsa.129320.1</td>
<td>PACid:16962264</td>
<td>-3.90</td>
<td>0.29±0.09</td>
</tr>
<tr>
<td>3</td>
<td>Cucsa.129630.1</td>
<td>PACid:16962315</td>
<td>-2.76</td>
<td>0.40±0.13</td>
</tr>
<tr>
<td>4</td>
<td>Phvul.008g046400.1</td>
<td>PACid:27154361</td>
<td>-1.91</td>
<td>0.49±0.15</td>
</tr>
</tbody>
</table>

Fig. 2 (A-D)—Energy minimized 3D structures of four selected EREBPs-1 proteins, representing four subgroups of the newly constructed combined phylogenetic tree, viz., Phvul.006g179800.11 PACid:27165358 (A), Cucsa.129320.1|PACid:16962264 (B), Cucsa.129630.1|PACid:16962315 (C), and Phvul.008g046400.11 PACid:27154361 (D). [I-TASSER software was used for 3D modelling, while YASARA energy minimization server was used for energy minimization. Orange colour represents for α-helix, green colour for β-sheets and turquoise colour for random coils.]
allowed regions, 0.10-0.430% residues in disallowed regions as reported in Table 3. After energy minimization, the observed Psi-Phi pairs had 75.80-87.70% residues in most favoured regions, 9.60-19.90% of residues in additional allowed regions, 0.10-0.30% residues in generously allowed regions and 0.10-0.30% residues in disallowed regions (Table 4; Fig. 3).

The red regions correspond to conformations where there are no steric clashes, i.e., these are the most favoured regions, namely, the α-helical and β-sheet conformations. The yellow areas show the allowed regions if slightly shorter van der Waals radii are used in the calculation, i.e., the atoms are allowed to come a little closer together. This brings out an additional region, which corresponds to the left-handed α-helix. Disallowed regions generally involve steric hindrance between the side chain C methylene group and main chain atoms.

The results clearly indicate the stability of the conformations of modelled proteins. Energy minimization adjusts the conformation so as to reach the minimum energy state which is most stable. A high percentage of residues falling in the allowed regions of the Ramachandran plot signify that the phi-psi angle values are well within the limits according to native protein conformations. These regions correspond to the secondary structures of a protein like α-helices and β-sheets. With very less number of residues in the disallowed region of Ramachandran plot after energy minimization, it signifies that a more stable conformation was attained.

**Disorder Prediction**

Intrinsically unstructured/disordered proteins and domains (IUPs) lack a well-defined 3D structure under native conditions. The assumption is that IUP sequences do not fold due to their inability to form sufficient stabilizing inter-residue interactions. In this, the web-server takes a single amino acid sequence as an input and calculates the pairwise energy profile along the sequence. Then, the energy values are transformed into a probability score ranging from

<table>
<thead>
<tr>
<th>No.</th>
<th>Acc. no.</th>
<th>% residues in most favoured region</th>
<th>% residues in additional allowed region</th>
<th>% residues in generously allowed region</th>
<th>% residues in disallowed region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phvul.006G179800.1</td>
<td>PACid:27165358</td>
<td>72.00</td>
<td>21.10</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>Cucsa.129320.1</td>
<td>PACid:16962264</td>
<td>85.00</td>
<td>11.00</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>Cucsa.129630.1</td>
<td>PACid:16962315</td>
<td>75.80</td>
<td>14.20</td>
<td>0.70</td>
</tr>
<tr>
<td>4</td>
<td>Phvul.008G046400.1</td>
<td>PACid:27154361</td>
<td>86.60</td>
<td>10.70</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>Acc. no.</th>
<th>% residues in most favoured region</th>
<th>% residues in additional allowed region</th>
<th>% residues in generally allowed region</th>
<th>% residues in disallowed region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phvul.006G179800.1</td>
<td>PACid:27165358</td>
<td>75.80</td>
<td>18.60</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>Cucsa.129320.1</td>
<td>PACid:16962264</td>
<td>84.00</td>
<td>11.50</td>
<td>0.30</td>
</tr>
<tr>
<td>3</td>
<td>Cucsa.129630.1</td>
<td>PACid:16962315</td>
<td>75.80</td>
<td>19.90</td>
<td>0.20</td>
</tr>
<tr>
<td>4</td>
<td>Phvul.008G046400.1</td>
<td>PACid:27154361</td>
<td>87.70</td>
<td>09.60</td>
<td>0.10</td>
</tr>
</tbody>
</table>
0 (complete order) to 1 (complete disorder). The residues can be considered as disordered with a score above 0.5. It resulted in a disorder prediction graph along with the disorder probability for each residue position. Intrinsically unstructured/disordered regions were 43.58% in Phvul.006G179800.1, 34.47% in Cucsa.129320.1, 13.5% in Cucsa.129630.1 and 30.91% in case of Phvul.008G046400.1 as shown in Table 5. The residue position of the intrinsically unstructured/disordered regions were 1-80 and 140-179 in Phvul.006G179800.1; 34-69, 138-199, and 232-235 in Cucsa.129320.1; 132-143 and 190-252 in Cucsa.129630.1; and 1-60, 133-207 in Phvul.008G046400.1. The graphical representation of disorder prediction results have been provided in Fig. 4. These intrinsically disordered regions signify the presence of active regions of a protein molecule. In the case of EREBP proteins, the regions involved

Fig. 4—Intrinsically disordered regions of four EREBPs: (A) Phvul.006G179800.1|PACid:27165358; (B) Cucsa.129320.1|PACid:16962264; (C) Cucsa.129630.1|PACid:16962315; and (D) Phvul.008G046400.1|PACid:27154361. [Horizontal axis represents for Residue position and the vertical axis for Disorder tendency. The graph above the value 0.5 represents the intrinsically disordered region of the protein.]
in nucleotide binding (as transcription factors) are predicted. These regions should not have a rigid conformation so that they can change form during interaction. The degree of disorder signifies this interaction ability. The models have large disordered regions except Cucsa.129630.1. The N-terminal of Phvul.006G179800.1 and Phvul.008G046400.1 seems to be majorly involved in interaction with a nucleotide chain, while the central parts of Cucsa sequences seem to drive this interaction.

Conclusion

We have reported here a comparative study on plant stress-responsive EREBP transcription factors. A number of EREBP protein sequences were selected to identify the phylogenetic relationship between them. Then a combined phylogenetic tree was constructed, which yielded four EREBP subgroups. From each subgroup, one protein sequence was evaluated which had the highest Bits (score) and lowest E-value. Consequently, two EREBPs of P. vulgaris and two of C. sativus were chosen from the four subgroups of the phylogenetic tree. Four EREBP structures were modelled using ab initio based approach and were energy minimized. They were validated on the basis of Ramachandran plot analysis where residues were in most favoured region ranging from 75.80-87.70%. Further, intrinsically disordered regions were found out representing the ethylene-responsive nucleotide binding regions. All of the four modelled EREBP structures have been reported in this study with elaborate analysis and stability evaluation using a number of computational techniques. Given the importance of knowledge of various transcription factors in understanding gene regulation in plants, especially those related to stress tolerance; this study was carried out to elucidate structures of EREBP proteins. Knowledge of 3D structure is a pre-requisite to a protein’s function along with its mode of interaction with other molecules.

Table 5—Percentage of intrinsically unstructured/disordered regions of four EREBPs.

<table>
<thead>
<tr>
<th>No.</th>
<th>Acc. no.</th>
<th>Percentage of intrinsically unstructured/disordered regions within a protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phvul.006G179800.1</td>
<td>43.58</td>
</tr>
<tr>
<td></td>
<td>PACid:7165358</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Cucsa.129320.1</td>
<td>34.47</td>
</tr>
<tr>
<td></td>
<td>PACid:16962264</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Cucsa.129630.1</td>
<td>13.50</td>
</tr>
<tr>
<td></td>
<td>PACid:16962315</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Phvul.008G046400.1</td>
<td>30.91</td>
</tr>
<tr>
<td></td>
<td>PACid:27154361</td>
<td></td>
</tr>
</tbody>
</table>

regions which bind to the nucleotides are functionally relevant and therefore add to the existing arsenal of stress-related knowledge.

References

17. TAIR, The Arabidopsis information resource. [https://www.arabidopsis.org/]
22 Phytozome v9.1: Home. [https://www.phytozome.com/]