

Research Article

Dempster-Shafer Theory for the Prediction of Auxin-Response Elements (AuxREs) in Plant Genomes

Nesrine Sghaier ^{1,2}, Rayda Ben Ayed ¹, Riadh Ben Marzoug,¹ and Ahmed Rebai ¹

¹Laboratory of Molecular and Cellular Screening Processes, Center of Biotechnology of Sfax, BP 1177, Sfax 3018, Tunisia

²Faculty of Sciences of Gabes, University of Gabes, City Riadh Zerig, 6072 Gabes, Tunisia

Correspondence should be addressed to Nesrine Sghaier; nesrine.sghaier@ymail.com

Received 17 May 2018; Accepted 15 October 2018; Published 1 November 2018

Guest Editor: Mostafa Abdelrahman

Copyright © 2018 Nesrine Sghaier et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Auxin is a major regulator of plant growth and development; its action involves transcriptional activation. The identification of Auxin-response element (AuxRE) is one of the most important issues to understand the Auxin regulation of gene expression. Over the past few years, a large number of motif identification tools have been developed. Despite these considerable efforts provided by computational biologists, building reliable models to predict regulatory elements has still been a difficult challenge. In this context, we propose in this work a data fusion approach for the prediction of AuxRE. Our method is based on the combined use of Dempster-Shafer evidence theory and fuzzy theory. To evaluate our model, we have scanning the DORNROSCHE promoter by our model. All proven AuxRE present in the promoter has been detected. At the 0.9 threshold we have no false positive. The comparison of the results of our model and some previous motifs finding tools shows that our model can predict AuxRE more successfully than the other tools and produce less false positive. The comparison of the results before and after combination shows the importance of Dempster-Shafer combination in the decrease of false positive and to improve the reliability of prediction. For an overall evaluation we have chosen to present the performance of our approach in comparison with other methods. In fact, the results indicated that the data fusion method has the highest degree of sensitivity (Sn) and Positive Predictive Value (PPV).

1. Introduction

Plants are genetically very diverse group and are playing a vital role in nutrition and livelihood in particular for rural and tribal masses for employment and income generation. In response to various developmental conditions and severe environmental changes by regulating gene expression. Transcription is at the core of physiological and developmental processes that requires well-coordinated players. Auxin is a major regulator of plant growth and development that plays important roles during all the stages of plant life and their action involves transcriptional activation. This phytohormone controls multiple fundamental aspects of the plant development [1] and environmental responses such as apical dominance [2], root development [3], phototropism, and gravitropism [4]. Also, Auxin is crucially involved in cell division, cell elongation, and cell differentiation [5]. The action of these plant hormone centres on the activation of early-response genes [6] and microarray studies has identified a large number

of early Auxin-response genes [7]. Many players are implicated in the transcriptional mechanism in the regulation of Auxin target gene expression. Auxin-response element (AuxRE) is a key element which is necessary in this process. The first and second reactions involve recognition of this specific element which contains the core sequence TGTCTC [8].

The identification of AuxREs is one of the most important issues to understand the Auxin regulation of gene expression at the genome level. Cis-regulatory elements can be elucidated by experimental technologies in vitro such as ChIP-chip [9], ChIP-seq [10, 11], and ChIP-PET [12]. However, using laboratory techniques is laborious and the process requires significant time and resources [13]. This is why many computational methods have been developed to allow fast and efficient identification of hormone receptor regulatory elements [14, 15]. Computational prediction of TFBS motifs remains a central goal in bioinformatics and intensive efforts have been dedicated to identifying putative cis-regulatory elements.

Several algorithms have been developed for the detection of consensus sequences. They can be categorized into two main strategies [16, 17]: enumeration of short words (counting and comparing oligonucleotide frequencies) [18, 19] and probabilistic methods [20, 21]. Usually, motif finding tool identifies short DNA sequence ‘motifs’ that are statistically overrepresented in regulatory regions (promoters) [21, 22]. A statistically overrepresented motif signify a motif that occurs more often than one would expect by chance [16]. Many computational approaches have been applied such as heuristic, greedy [23], and stochastic algorithms, some others used, expectation maximization (EM) [24], Gibbs Sampling algorithms [25], Hidden Markov model (HMM) [13], Bayesian network [26], Genetic algorithms (GA) [25], and others [16].

A pattern can be represented as a consensus sequence or a position weight matrix (PWM) [46]. PWMs are frequently applied for transcription factor binding site prediction [23, 47]. It describes the probability to find the nucleotides A,C,G,T on each position of a motif [48]. Searching pattern for matches with a PWM is more accurate than consensus string matching, but it also produces a large number of false positives [49, 50]. Other methods use localized distribution as a supplementary criterion to detect functional elements [51]. Over the past few years, a large number of motif identification tools have been developed, to name a few, MAPPER [52], AlignACE [21], MEME [53], Weeder [54], MotifSampler [55], and GAME [56]. Because of this diversity of algorithms and programs available, many studies present a comprehensive review of motifs predictors that provide comparison and guidance to researchers such as Stormo [48], Das and Dai [16], and Tompa et al. [57]. These studies show that despite these considerable efforts provided by computational biologists, building reliable models to predict regulatory elements was always a challenge in task. Stormo and Zhao [57] suggested that the majority of the current approaches are not accurate or complete and it is necessary to find more accurate prediction methods with higher specificity and sensitivity. So a new bioinformatics framework is required. Tompa et al. [57] recommended the use of a few complementary tools and follow up the top motifs by combining information from different predictions. Hu et al. [58] discussed the limitations of motif discovery algorithms and developed a new one, named, EMD, which is more significant for shorter input sequences [59].

In this context, we propose in this work a data fusion approach for the prediction of Auxin-response elements. Our method is based on the combined use of Dempster-Shafer (DS) evidence theory and fuzzy sets. It consists of modelling detection uncertainty and fusing the features using DS combination rule.

2. Material and Methods

2.1. Training Set (Data Collection). A training set of 64 experimentally verified that hormone response elements were collected from published data (Table 1). Whole genome dataset and upstream sequences of *Arabidopsis thaliana* were downloaded from TAIR (<http://arabidopsis.org/>).

Position weight matrix used for comparison tools was obtained from Ponomarenko and Ponomarenko [60]. Linear discriminant analysis was performed using SPSS (v. 16.0, Statistical Package for the Social Sciences, Chicago, IL, USA).

Microarray data of the primary response to Auxin in *Arabidopsis* was taken from Genevestigator database (<https://genevestigator.com/gv/>) [61]. Response in seedlings was selected: 1 μ M IAA for 1 h [62].

2.2. Implementation of the Algorithm. The main algorithm was implemented under the R environment language. All measurements were performed on a single CPU Intel Core i3 computer running at 2.8 GHz, with 6 GB main memory. The source code is available upon request.

2.3. Some Fundamentals of Dempster-Shafer Theory. The Dempster-Shafer (DS) evidence theory is a mathematical theory originated from the earlier works of Arthur P. Dempster in 1967 [63, 64] and extended by Glenn Shafer in 1976 [65]. DS theory can be considered as a generalization of Bayesian probability theory which uses the notions of imprecise, uncertain, and incomplete information. It has been applied in various domains such as medical diagnosis, image processing, and expert systems [66, 67]. DS theory can be used to combine information from different sources. DS theory uses ‘belief’ rather than probability. ‘Belief’ function is used to represent the uncertainty of the hypothesis. In DS theory, there is a finite set of N elements called the frame of discernment $\Theta = \{H_1, H_2, \dots, H_N\}$. It is a set of mutually exclusive and exhaustive propositions.

Information sources can distribute mass values on subsets of the frame of discernment. A numerical measure of uncertainty, termed basic probability masses, may be assigned to sets of hypotheses as well as individual hypotheses.

The mass functions verify the following constraints:

$$\begin{aligned} 0 &\leq m(A_i) \leq 1 \\ m(\emptyset) &= 0 \\ \sum_{A_i \in 2^\Theta} m(A_i) &= 1 \end{aligned} \quad (1)$$

where A_i designates a simple hypothesis H_i or composite hypotheses (union of simple hypotheses), $A_i = 2^\Theta$.

If we consider two mass distributions m_1 and m_2 from two different information sources, m_1 and m_2 can be combined with Dempster’s orthogonal rule, and a new distribution $m = m_1 \oplus m_2$ is calculated in the following manner:

$$m(A_i) = (1 - K)^{-1} \sum_{A_p \cap A_q = A_i} m_1(A_p) m_2(A_q) \quad (2)$$

where

$$K = \sum_{A_p \cap A_q = \emptyset} m_1(A_p) m_2(A_q) \quad (3)$$

K is the conflict between the two sources.

TABLE I: Datasets.

Cis-regulatory elements	abbreviations	Numbers	References
Auxin-response element	AuxRE	16	[27–37]
ABA response element	ABRE	12	[38–40]
TATA Box	TATA Box	16	[41]
Ypatch	Ypatch	11	[41]
drought-responsive element	DRE	9	[38, 39]

Dempster-Shafer uses ‘belief’ rather than probability. Belief function is used to represent the uncertainty of the hypothesis.

To evaluate the uncertainty of the hypothesis, two functions can be calculated from a mass distribution: the belief function (Bel) and the plausibility function (Pls). Belief and plausibility functions can be considered as lower and upper estimations of probabilities.

$$\begin{aligned} \text{Bel}(A_i) &= \sum_{A_j \subseteq A_i} m(A_j) \\ \text{Pls}(A_i) &= \sum_{A_j \cap A_i = \emptyset} m(A_j) \end{aligned} \quad (4)$$

$\text{Bel}(A) = 0$ represents lack of evidence about A.

3. Results and Discussion

3.1. Modelling Uncertainty of AuxRE Detection. The objective of our study is detection of AuxRE. We applied a data fusion approach which consists of a combination of predictions coming from two techniques commonly used in pattern finding: overrepresented motifs and linear discriminant analysis. The idea is to extract, for each method, some features (parameters) and combine these parameters using the Dempster-Shafer (DS) rule, called orthogonal sum. We have applied our model to the *Arabidopsis thaliana* genome. The *Arabidopsis* genome sequence was obtained from TAIR [68].

Two hypotheses are involved: “this motif is an AuxRE”: “this motif is not an AuxRE” (i.e., not a motif or a motif other

than AuxRE). In terms of the Dempster-Shafer evidence theory, we are in the case where the frame of discernment is constructed of two single hypotheses H1 and H2 and one composite hypothesis H3= H1 U H2 (union of H1 and H2). H3 represents in fact the ignorance.

The modelling process is proceeding with six major steps (Figure 1):

- (i) Step 1: extraction of parameters
- (ii) Step 2: construction of learning graphs
- (iii) Step 3: determination of confidence regions
- (iv) Step 4: modelling the doubt on the hypotheses
- (v) Step 5: fuzzification of the learning graphs
- (vi) Step 6: data fusion methodology

3.1.1. Extraction of Parameters. From the first method (detection of overrepresented motifs), we have prepared four parameters which are position P, significance score Sc, occurrence O, and density D. The position was located from the ATG. Significance score obtained from Weeder algorithm [54]. The occurrence represents the total number of a validated motif sequence in the whole genome of *Arabidopsis thaliana*. We have considered the density as the rate of a validated AuxRE motif sequence in promoter (-1000 bp) of response gene of Auxin. To prepare density, we have extracted the 2-fold Auxin-response gene from the microarray data.

$$D = \frac{\text{Number of a validated motif sequence in the promoters of 2 folds Auxin response gene}}{\text{Total number of a validated motif sequence in all the promoters of arabidopsis genes}} \quad (5)$$

We used the Z-curve parameters [69] and the GC% as potentially discriminative parameters and we performed a linear discriminant analysis. The Z-curve is a unique three-dimensional curve representation of a DNA sequence. We used three Z-curve parameters which are

$$\begin{aligned} x1 &= (a1 + g1) - (c1 + t1) \\ y1 &= (a1 + c1) - (g1 + t1) \\ z1 &= (a1 + t1) - (g1 + c1) \end{aligned} \quad (6)$$

3.1.2. Construction of Learning Graphs. In the following sections, two methods will be presented that use the available data on a positive and a negative training set to construct a discriminative prediction model. A training set of 64 experimentally proven hormone response elements were collected from published data.

Method 1: Overrepresented Motifs. First, the validated motifs are studied in feature spaces which make the interpretation of the link between the selected features (P, SC, O, and D) and the type of motifs straightforward. We chose to study separately knowledge from position P and significance score

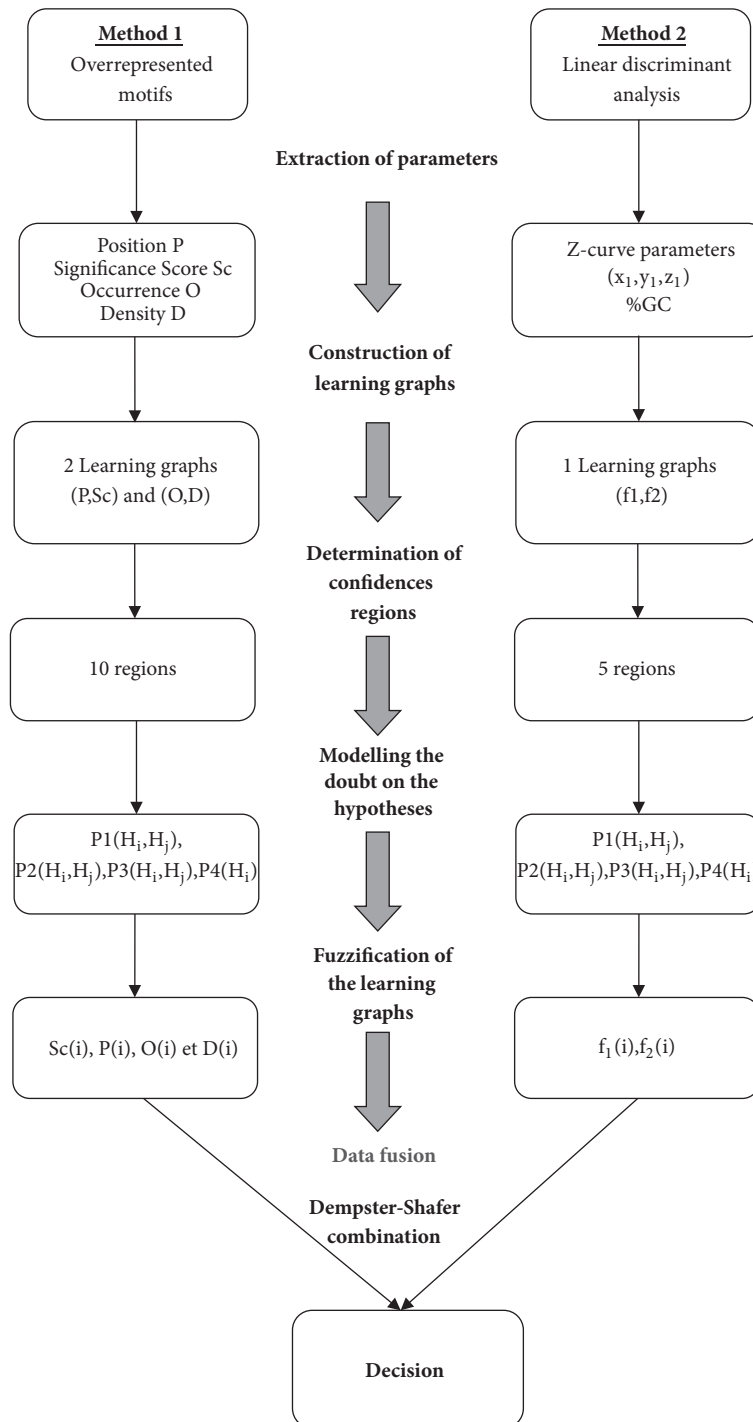


FIGURE 1: Modelling approach.

Sc and those provided by occurrence and density in order to separate as much as possible AuxRE from other types of cis-regulatory elements. Two learning graphs have been created (Figures 2 and 3). Figure 2 represents the distribution of validated motifs according to their parameters position P and significance score Sc. We distinguish, at the bottom of the graph, a region containing only AuxRE; the other part

of the graph corresponds to an area of uncertainty which contains all types of motifs. This figure shows that only AuxREs are located relatively far from the translational start site (start codon). However, it is not a discriminative parameter, as many AuxREs were found in -500 bp upstream regions. Therefore, we have decided to study two other parameters (occurrence and density) in order to improve the

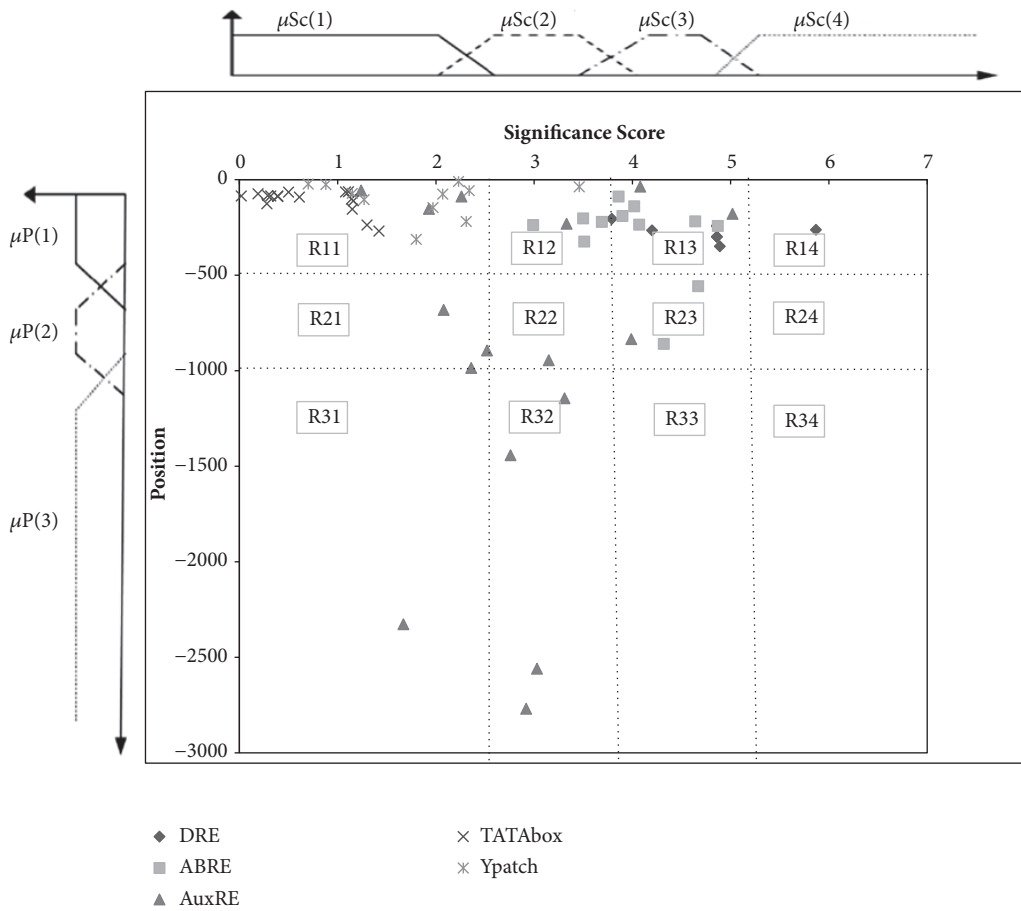


FIGURE 2: Learning graph 1: distribution of different type of motifs in significance score/position feature space.

classification and try to differentiate AuxREs, especially those found in the mixed region shown in Figure 2.

Figure 3 illustrates the classification of training cis-elements based on two parameters: the occurrence of the patterns in the -1000 bp upstream regions and the density.

Method 2: Linear Discriminant Analysis. For the linear discriminant analysis, we have used the Z-curve parameter and the % GC. Figure 4 shows the first two discriminant functions which allow a good discrimination of AuXRE from other motifs except Ypatch. The first discriminant function explains 59.6% of variability and has the highest correlation with GC% (-0.88) and Z1 (0.85) while the second function (32% of variability) is correlated to X1 (0.75).

3.1.3. Confidence Regions. All the previous graphs do not allow a clear discrimination of AuxRE from other motifs. Each graph can be subdivided in several ways into different regions that will be enriched in one or few motifs. Here, we have chosen to partition the graph into five confidence regions shown in the Figures 1, 2, and 3 based on the percentage of AuxRE that belong to this region. The graph partition is given in Figures 1, 2, and 3 and Tables 2, 3, and 4.

3.1.4. Modelling the Doubt on the Hypotheses. In order to make the graph partition an automatic process we attributed a confidence level to any unknown detected motif that would be located on the graph.

For that purpose, we define a gradual doubt through a set of four propositions:

- (i) P1(H_i,H_j): total ignorance
- (ii) P2(H_i,H_j): low preference for the H_i hypothesis but high doubt between H_i and H_j
- (iii) P3(H_i,H_j): strong preference for the H_i hypothesis but low doubt between H_i and H_j
- (iv) P4(H_i): total confidence in the H_i hypothesis, no doubt

Next, these propositions are translated in terms of masses as detailed in Table 5. The preference level for a hypothesis from P1 to P4 is gradually represented by a mass value, respectively, equal to 0, 0.33, 0.67, and 1 [66]. Likewise, the gradual doubt between hypotheses is modelled by a mass value. In case of total doubt, the mass value affected equals 0. On the other hand, the mass value assigned to the total confidence is equal to 1.

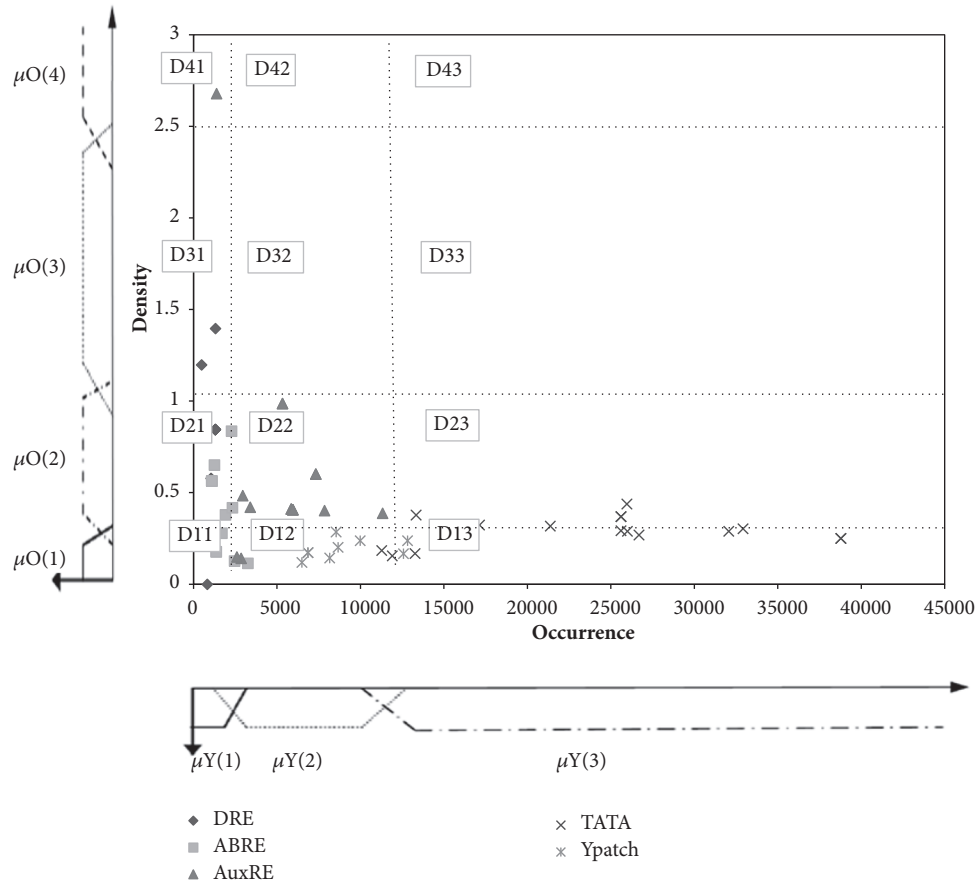


FIGURE 3: Learning graph 2: distribution of different type of motifs in occurrence/density feature space.

TABLE 2: Proportion of false positive and true positive in the regions of the significance score/position feature space and associated propositions.

Region R_{ij}	% AuxRE	% non AuxRE	Proposition
Z1: R11	7	93	P4(H2)
Z2: R12	25	75	P2(H2)
Z3: R13	0	100	P4(H2)
Z4: R14, R24, R34	0	100	P4(H2)
Z5: R23	80	20	P3(H1)
Z6: R21, R31, R32, R33, R22	100	0	P4(H1)

Finally, a proposition is assigned to each region from the previous analyses on percentages of AuxRE and other motifs in each region. The link between the percentages and the related proposition are presented in Tables 2, 3, and 4.

3.1.5. Fuzzification of the Learning Graphs. In the previous section we used discrete representation to define regions, which is not very objective because it can allocate confidence significantly different, for two near motifs from either side of boundaries. Moreover, the boundaries between regions are not well defined, and the transition from one region of the graph to another is not abrupt but a smooth one. Thus, In order to have a fuzzy, gradual continuous transition, we introduce the fuzzy logic theory. Therefore, we define

fuzzy sets for each measured feature to predict its membership degrees to different possible feature families. For the parameter significance score four sets were defined (small, average, high, and very high). For the parameter position, three sets were described (core, proximal, and distal). For the parameters occurrence and density, three sets were defined (small, average, and high) for each of them.

3.1.6. Data Fusion Methodology. The process of data fusion consists of fusing a number of learning graphs based on the definition of the so-called masses.

For each detected motifs, three masses are calculated, corresponding to the three learning graphs. They are given, respectively, by

TABLE 4: Proportion of false positive and true positive in the regions of the f1/f2 feature space and associated propositions.

Region Q_{ij}	% AuxRE	% non AuxRE	Proposition
Z11: Q11, Q21, R12	0	100	P4(H2)
Z12: Q22	70	30	P3(H1)
Z13: Q13, Q23, Q33	0	100	P4(H2)
Z14: Q31, Q32	10	90	P3(H2)
Z15: Q41, Q42, Q43	0	100	P4(H2)

TABLE 5: Association of propositions with mass values.

Proposition	m(H1) (AuxRE)	m(H2) (Pas AuxRE)	m(H1 U H2) (ignorance)
P1(H1,H2)	0	0	1
P2(H1,H2)	0,33	0	0,67
P3(H1,H2)	0,67	0	0,33
P4(H1)	1	0	0
P2(H2,H1)	0	0,33	0,67
P3(H2,H1)	0	0,67	0,33
P4(H2)	0	1	0

The final mass function is then calculated by fusing the two masses $m_1(O \in S)$ and $m_{R_{ij}}(O \in S/f1\&f2)$; the orthogonal sum of Dempster is

$$m_{fusion}(O \in S) = m_1(O \in S) \oplus m_{R_{ij}}(O \in S/f1\&f2) \quad (9)$$

3.2. Scan of the Auxin Responsive DRN Promoter. DORNRÖSCHEN (DRN) promoter is one of the most studied Auxin responsive promoters which have an essential role in Auxin transport and perception in the *Arabidopsis* embryogenesis [70]. Two AuxREs that are not used in training have been experimentally identified in this promoter. To verify the reliability of the prediction, we tested our method to the DRN promoter. At a threshold of 0.9, the scanning of the DRN promoter by the model has detected the two validated AuxREs and at the same time we have not detected a false positive. Among 1200 motifs, we considered the two proven AuxREs as a true positive and the others as false positives (Figure 5).

3.3. Comparison between Method 1, Method 2, and Fusion. In order to study the influence of the data fusion by Dempster-Shafer combination, we have presented in Figure 6 the ration between true and false positive before and after combination. Figure 6 shows that, based on method 1 and method 2 separately, we have a large number of false positives. Their percentage exceeds 90% in both cases. After combination, it appears that the number of false positive significantly decreases to the point of cancelled when the credibility value equals 0.9. The reliability of detection is improved by data fusion. In parallel, the comparing of Tree ROC curves as shown in Figure 7 confirms the higher predictive reliability of the model after fusion compared with that based on only one method, when we scan DRN promoter.

3.4. Scan of DRN Promoter by Other Methods. To evaluate our method, we have scanned the DRN promoter by previous

tools: Consensus [71], MEME [20], Gibbs Sampler [25], MDScan [72], and Weeder [54]. On the analysis platform MELINA II [73], the result indicates that the four motifs finding tools do not detect any AuxRE. These basic tools are unable to detect specific hormone responsive elements, but they detect cis-elements in general. We have also compared our model to the PWM method. PWM detects the two AuxREs but in return it produces a high frequency of false positive predictions. In fact, four false positives have been detected at a threshold equal to 0.9. For example, PWM detects the motif TTGTCAA as an AuxRE with a score equal to 0.93 because this motif sequence is similar to the AuxRE sequence and, on the other hand, the PWM is based only on the composition. Conversely, this motif was not detected with our method since the prediction depends on several parameters. Likewise, the Plant Promoter Database (PPDB) has not detected these two validated AuxRE present in the DRN promoter. In this database, cis-regulatory elements are identified by the Local Distribution of Short Sequences (LDSS) and a prediction method based on microarray data methods (RARf-based approach)[74].

3.5. Scan of RD29B Promoter. The promoter of RD29B gene contains no AuxRE according to the literature. Several studies have shown the presence of other types of cis-regulatory elements such as ABA and DRE. The scan of this promoter by our model did not detect any false positives.

3.6. Validation of the Results. Because of the limited number of confirmed Auxin responsive elements, there is not enough data to divide it into training and validation sets. So, we have performed the Gold Standard [75] test to evaluate our model. A library of random DNA sequences (100 sequences) was generated using Unipro UGENE software version 1.26.1. (<http://ugene.unipro.ru/>) [76]. A set of 14 AuxRE was prepared. In each randomly generated DNA sequence only one AuxRE from preparing set was inserted at a random

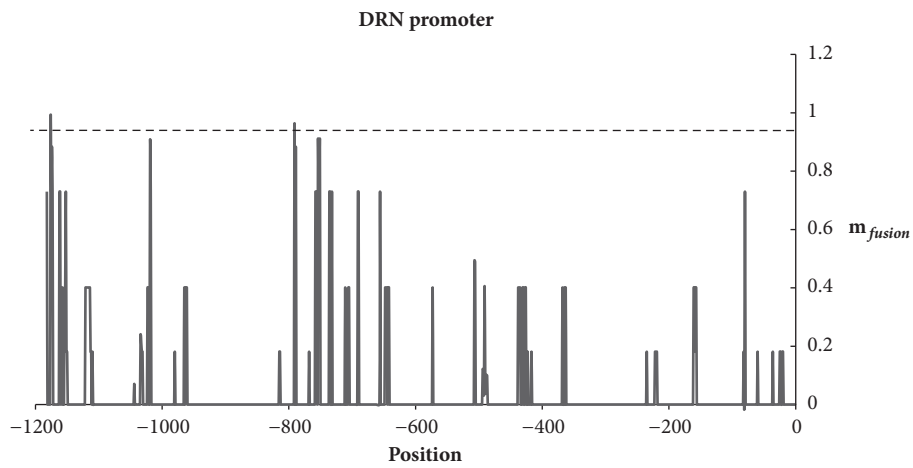


FIGURE 5: Scanning of the DRN promoter by the data fusion method.

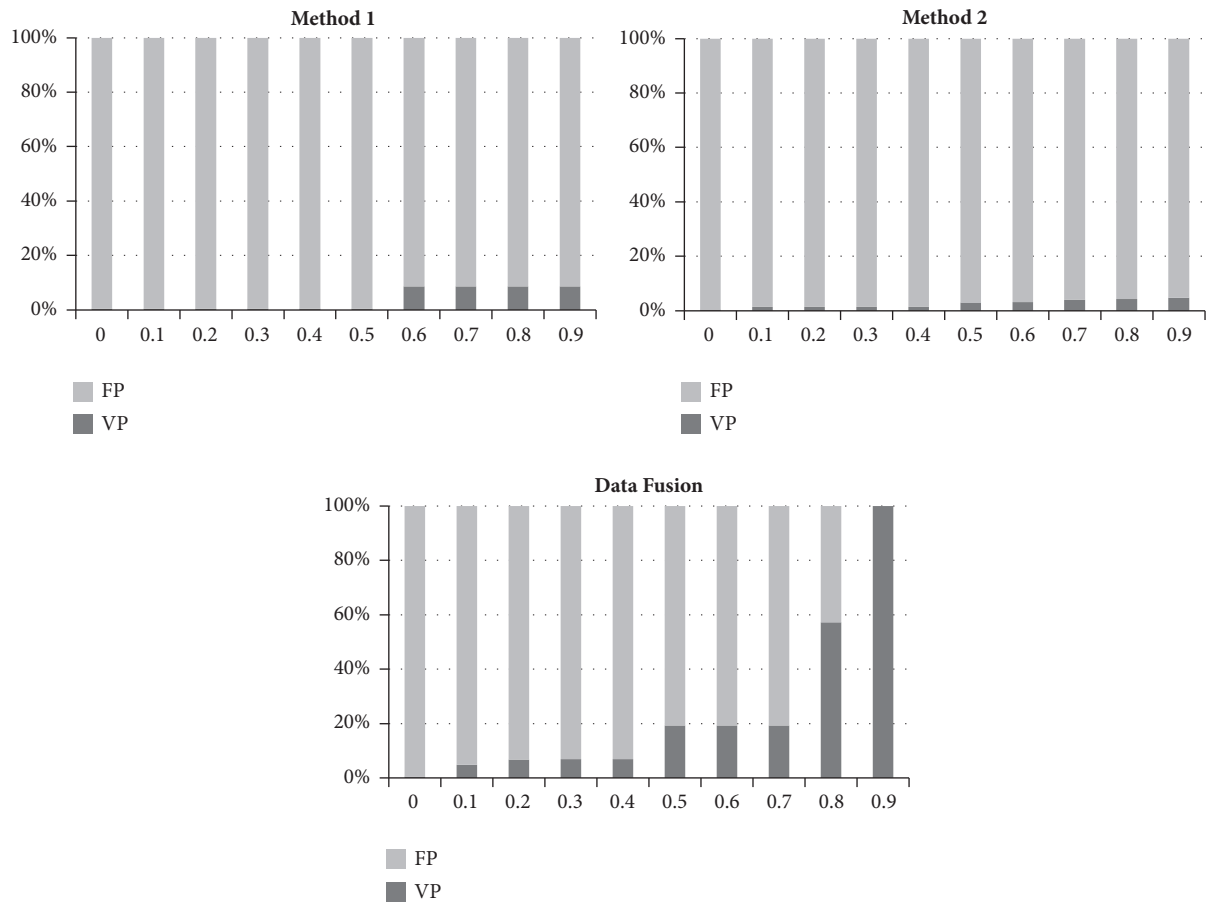


FIGURE 6: Evolution of the positive and false detection as a function of credibility obtained before and after data fusion.

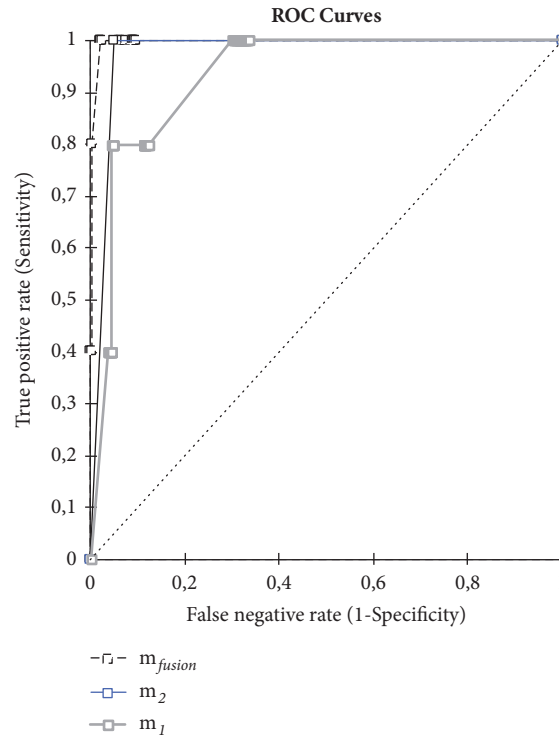


FIGURE 7: ROC curves before and after data fusion (scan of DRN promoter).

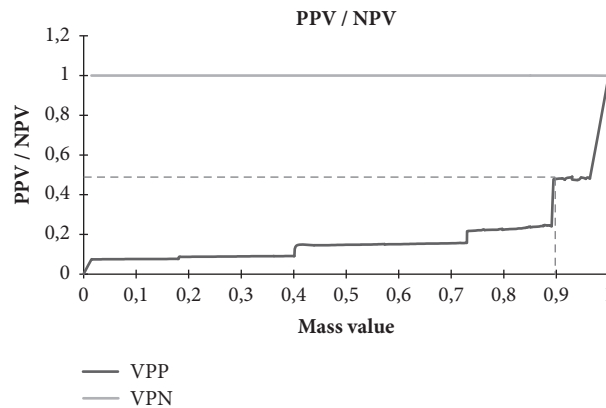


FIGURE 8: Variation of Positive Predictive Value (PPV).

position using SeqKit toolkit [77]. A TSV file which contains a list of the sequences of inserted AuxRE and their positions of insertion was generated using csvtk (<https://github.com/shenwei356/csvtk>).

In the next step, to further investigate the prediction performance and to choose the optimum cutoff, we applied our prediction method and we look at the variation of Positive Predictive Value (PPV). The results showed that we achieve maximal PPV for a cutoff value of 0.9 (Figure 8).

For an overall evaluation we have chosen to present the performance of our approach in comparison with other methods. The chosen methods are the five individual TFBS prediction tools evaluated by Jayaram et al. [78].

We do this by first summing true/false positives and negatives, and then statistical parameters were calculated in order to illustrate the best predictive approach. Table 6 presented the obtained results. Our method is based on the joint using of Dempster-Shafer (DS) evidence theory and fuzzy sets and has the high degree of sensitivity (S_n) and Positive Predictive Value (PPV) with a value of 79 and 48.17, respectively, compared to the best previous methods. Even the Youden index (YI) and the X2 test parameters generated higher value than the other reference tools. Moreover, Table 6 shows that our approach (Data fusion) followed by the Clover computer program implemented by Frith et al. [42] are the best performing transcription factor binding sites

TABLE 6: Comparison between our method and the other published methods.

	<i>Sn</i>	<i>Sp</i>	<i>PPV</i>	<i>NPV</i>	<i>FPR</i>	<i>FNR</i>	<i>YI</i>	<i>QCY</i>	X_2	<i>Ref</i>
Data fusion	79	99.91	48.17	99.98	51.83	0.02	0.79	1	36490.3	
Clover	69	99.92	47.9	99.97	52.08	0.03	0.69	1	31702.8	[42]
Matrix-Scan	51	99.94	46.36	99.95	53.64	0.05	0.51	1	22661.6	[43]
Patser	63.64	99.92	43.45	99.96	56.55	0.04	0.64	1	27285.9	[43]
FIMO	22	99.97	42.31	99.92	57.69	0.08	0.22	1	8911.9	[44]
PoSSuMsearch	56.41	83.84	40.74	90.71	59.26	9.29	0.4	0.74	30	[45]

Average sensitivities (*Sn*). Specificity (*Sp*). Positive Predictive Value (*PPV*). Negative Predictive Value (*NPV*). False Positive Rate (*FPR*). False Negative Rate (*FNR*). Youden index (*YI*). Q coefficient of Yule (*QCY*) and X_2 test value (X_2). The best-performing tools. Data fusion and Clover are highlighted in bold.

(TFBS) prediction tools for individual sites. On the other side, Table 6 shows that the Find Individual Motif Occurrences (FIMO) method described by Grant et al. [44] has the worst sensitivity ($Sn=22$) on all the six presented tools. Besides, position specific scoring matrices (PoSSuMsearch) developed by Beckstette et al. [45] and FIMO tool have lower Positive Predictive Value (PPV) than the other previous methods, with a value of 40.74 and 42.31, respectively.

Our method strikes a good balance between sensitivity and PPV.

4. Conclusion

In this study, we applied a data fusion approach for the prediction of Auxin-response elements. Our method is based on the combined use of Dempster-Shafer (DS) evidence theory and fuzzy theory. We have tested our model to the DRN promoter and we have compared the prediction to previous tools. The results show that false positives are significantly decreased.

Data Availability

All the data used in this manuscript are included within the article and will be freely accessible upon its publication in BioMed Research International.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Tunisian Ministry of Higher Education and Scientific Research.

References

- [1] B. Möller and D. Weijers, "Auxin control of embryo patterning," *Cold Spring Harbor Perspectives in Biology*, vol. 1, no. 5, p. a001545, 2009.
- [2] O. Leyser, "The fall and rise of apical dominance," *Current Opinion in Genetics & Development*, vol. 15, no. 4, pp. 468–471, 2005.
- [3] T. Bennett and B. Scheres, "Root development—two meristems for the price of one?" *Current Topics in Developmental Biology*, vol. 91, no. C, pp. 67–102, 2010.
- [4] G. K. Muday, "Auxins and tropisms," *Journal of Plant Growth Regulation*, vol. 20, no. 3, pp. 226–243, 2001.
- [5] Z. Ding and J. Friml, "Auxin regulates distal stem cell differentiation in Arabidopsis roots," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 26, pp. 12046–12051, 2010.
- [6] S. Abel and A. Theologis, "Early genes and auxin action," *Plant Physiology*, vol. 111, no. 1, pp. 9–17, 1996.
- [7] J. L. Nemhauser, F. Hong, and J. Chory, "Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses," *Cell*, vol. 126, no. 3, pp. 467–475, 2006.
- [8] T. Ulmasov, Liu Zhan-Bin, G. Hagen, and T. J. Guilfoyle, "Composite structure of auxin response elements," *The Plant Cell*, vol. 7, no. 10, pp. 1611–1623, 1995.
- [9] A. S. Weinmann and P. J. Farnham, "Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation," *Methods*, vol. 26, no. 1, pp. 37–47, 2002.
- [10] G. Robertson, M. Hirst, M. Bainbridge et al., "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing," *Nature Methods*, vol. 4, no. 8, pp. 651–657, 2007.
- [11] A. Barski, S. Cuddapah, K. Cui et al., "High-resolution profiling of histone methylations in the human genome," *Cell*, vol. 129, no. 4, pp. 823–837, 2007.
- [12] Y. Loh, Q. Wu, J. Chew et al., "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells," *Nature Genetics*, vol. 38, no. 4, pp. 431–440, 2006.
- [13] A. Sandelin and W. W. Wasserman, "Prediction of nuclear hormone receptor response elements," *Molecular Endocrinology*, vol. 19, no. 3, pp. 595–606, 2005.
- [14] B. Lenhard, A. Sandelin, L. Mendoza, P. Engström, N. Jareborg, and W. W. Wasserman, "Identification of conserved regulatory elements by comparative genome analysis," *Journal of Biology*, vol. 2, no. 2, article no. 13, 2003.
- [15] A. Brazma, I. Jonassen, J. Vilo, and E. Ukkonen, "Predicting gene regulatory elements in silico on a genomic scale," *Genome Research*, vol. 8, no. 11, pp. 1202–1215, 1998.
- [16] M. K. Das and H.-K. Dai, "A survey of DNA motif finding algorithms," *BMC Bioinformatics*, vol. 8, supplement 7, article S21, 2007.
- [17] I. W. Davis, C. Benninger, P. N. Benfey, and T. Elich, "Powers: Position-sensitive motif discovery," *PLoS ONE*, vol. 7, no. 7, 2012.
- [18] C. Linhart, Y. Halperin, and R. Shamir, "Transcription factor and microRNA motif discovery: The Amadeus platform and a

- compendium of metazoan target sets," *Genome Research*, vol. 18, no. 7, pp. 1180–1189, 2008.
- [19] S. Georgiev, A. P. Boyle, K. Jayasurya, X. Ding, S. Mukherjee, and U. Ohler, "Evidence-ranked motif identification," *Genome Biology*, vol. 11, no. 2, article R19, 2010.
- [20] T. L. Bailey and C. Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers," vol. 2, pp. 28–36, 1994.
- [21] J. D. Hughes, P. W. Estep, S. Tavazoie, and G. M. Church, "Computational identification of Cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*," *Journal of Molecular Biology*, vol. 296, no. 5, pp. 1205–1214, 2000.
- [22] J. Van Helden, B. André, and J. Collado-Vides, "Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies," *Journal of Molecular Biology*, vol. 281, no. 5, pp. 827–842, 1998.
- [23] G. Z. Hertz, G. W. Hartzell, and G. D. Stormo, "Identification of consensus patterns in unaligned DNA sequences known to be functionally related," *Bioinformatics*, vol. 6, no. 2, pp. 81–92, 1990.
- [24] C. E. Lawrence and A. A. Reilly, "An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences," *Proteins: Structure, Function, and Bioinformatics*, vol. 7, no. 1, pp. 41–51, 1990.
- [25] C. E. Lawrence, S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton, "Detecting subtle sequence signals: a gibbs sampling strategy for multiple alignment," *Science*, vol. 262, no. 5131, pp. 208–214, 1993.
- [26] R. Siddharthan, E. D. Siggia, and E. Van Nsmwegea, "Phylo-Gibbs: A gibbs sampling motif finder that incorporates phylogeny that incorporates phylogeny," *PLoS Computational Biology*, vol. 1, no. 7, pp. 0534–0556, 2005.
- [27] Y. Okushima, I. Mitina, H. L. Quach, and A. Theologis, "AUXIN RESPONSE FACTOR 2 (ARF2): A pleiotropic developmental regulator," *The Plant Journal*, vol. 43, no. 1, pp. 29–46, 2005.
- [28] I. Ismail, "Function and Regulation of Xylem Cysteine Protease 1 and Xylem Cysteine Protease 2 in *Arabidopsis*," 2003, <http://scholar.lib.vt.edu/theses/available/etd-08152004-231624>.
- [29] T. J. Donner, I. Sherr, and E. Scarpella, "Regulation of prepro-cambial cell state acquisition by auxin signaling in *Arabidopsis* leaves," *Development*, vol. 136, no. 19, pp. 3235–3246, 2009.
- [30] E. Scacchi, P. Salinas, B. Gujas et al., "Spatio-temporal sequence of cross-regulatory events in root meristem growth," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 52, pp. 22734–22739, 2010.
- [31] Z. Zhao, S. U. Andersen, K. Ljung et al., "Hormonal control of the shoot stem-cell niche," *Nature*, vol. 465, no. 7301, pp. 1089–1092, 2010.
- [32] C. L. Walcher and J. L. Nemhauser, "Bipartite promoter element required for auxin response," *Plant Physiology*, vol. 158, no. 1, pp. 273–282, 2012.
- [33] A. Hirota, T. Kato, H. Fukaki, M. Aida, and M. Tasaka, "The auxin-regulated AP2/EREBP gene PUCHI is required for morphogenesis in the early lateral root primordium of *Arabidopsis*," *The Plant Cell*, vol. 19, no. 7, pp. 2156–2168, 2007.
- [34] K. W. Berendzen, C. Weiste, D. Wanke, J. Kilian, K. Harter, and W. Dröge-Laser, "Bioinformatic cis-element analyses performed in *Arabidopsis* and rice disclose bZIP- and MYB-related binding sites as potential AuxRE-coupling elements in auxin-mediated transcription," *BMC Plant Biology*, vol. 12, article no. 125, 2012.
- [35] C. Zhu and S. E. Perry, "Control of expression and autoregulation of AGL15, a member of the MADS-box family," *The Plant Journal*, vol. 41, no. 4, pp. 583–594, 2005.
- [36] A. Schlereth, B. Möller, W. Liu et al., "MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor," *Nature*, vol. 464, no. 7290, pp. 913–916, 2010.
- [37] Z. J. Cheng, L. Wang, W. Sun et al., "Pattern of auxin and cytokinin responses for shoot meristem induction results from the regulation of cytokinin biosynthesis by AUXIN RESPONSE FACTOR3," *Plant Physiology*, vol. 161, no. 1, pp. 240–251, 2013.
- [38] K. Yamaguchi-Shinozaki and K. Shinozaki, "A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress," *The Plant Cell*, vol. 6, no. 2, pp. 251–264, 1994.
- [39] K. Nakashima, Y. Fujita, K. Katsura et al., "Transcriptional regulation of ABI3- and ABA-responsive genes including *RD29B* and *RD29A* in seeds, germinating embryos, and seedlings of *Arabidopsis*," *Plant Molecular Biology*, vol. 60, no. 1, pp. 51–68, 2006.
- [40] M. Denekamp and S. C. Smeeckens, "Integration of wounding and osmotic stress signals determines the expression of the AtMYB102 transcription factor gene," *Plant Physiology*, vol. 132, no. 3, pp. 1415–1423, 2003.
- [41] A. Hieno, H. A. Naznin, M. Hyakumachi et al., "Ppddb: Plant promoter database version 3.0," *Nucleic Acids Research*, vol. 42, no. 1, pp. D1188–D1192, 2014.
- [42] M. C. Frith, Y. Fu, L. Yu, J.-F. Chen, U. Hansen, and Z. Weng, "Detection of functional DNA motifs via statistical over-representation," *Nucleic Acids Research*, vol. 32, no. 4, pp. 1372–1381, 2004.
- [43] J.-V. Turatsinze, M. Thomas-Chollier, M. Defrance, and J. van Helden, "Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules," *Nature Protocols*, vol. 3, no. 10, pp. 1578–1588, 2008.
- [44] C. Grant, T. Bailey, and W. Noble, "Scanning for occurrences of a given motif," *Bioinformatics*, vol. 27, no. 2, pp. 1017–1018, 2011.
- [45] M. Beckstette, R. Homann, R. Giegerich, and S. Kurtz, "Fast index based algorithms and software for matching position specific scoring matrices," *BMC Bioinformatics*, vol. 7, article no. 389, 2006.
- [46] M. R. Barnes, "Bioinformatics Challenges for the Geneticist," *Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data: Second Edition*, pp. 1–16, 2007.
- [47] G. D. Stormo, T. D. Schneider, L. Gold, and A. Ehrenfeucht, "Use of the 'perceptron' algorithm to distinguish translational initiation sites in *E. coli*," *Nucleic Acids Research*, vol. 10, no. 9, pp. 2997–3011, 1982.
- [48] G. D. Stormo, "DNA binding sites: representation and discovery," *Bioinformatics*, vol. 16, no. 1, pp. 16–23, 2000.
- [49] G. Cuellar-Partida, F. A. Buske, R. C. McLeay, T. Whittington, W. S. Noble, and T. L. Bailey, "Epigenetic priors for identifying active transcription factor binding sites," *Bioinformatics*, vol. 28, no. 1, pp. 56–62, 2012.
- [50] H. Lähdesmäki, A. G. Rust, and I. Shmulevich, "Probabilistic inference of transcription factor binding from multiple data sources," *PLoS ONE*, vol. 3, no. 3, 2008.
- [51] Y. Y. Yamamoto, H. Ichida, M. Matsui et al., "Identification of plant promoter constituents by analysis of local distribution of short sequences," *BMC Genomics*, vol. 8, article no. 67, 2007.

- [52] V. D. Marinescu, I. S. Kohane, and A. Riva, "MAPPER: A search engine for the computational identification of putative transcription factor binding sites in multiple genomes," *BMC Bioinformatics*, vol. 6, article no. 79, 2005.
- [53] T. L. Bailey and C. Elkan, "The value of prior knowledge in discovering motifs with MEME," *Proceedings / International Conference on Intelligent Systems for Molecular Biology; ISMB. International Conference on Intelligent Systems for Molecular Biology*, vol. 3, pp. 21–29, 1995.
- [54] G. Pavesi, P. Mereghetti, G. Mauri, and G. Pesole, "Weeder web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes," *Nucleic Acids Research*, vol. 32, pp. W199–W203, 2004.
- [55] G. Thijs, M. Lescot, K. Marchal et al., "A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling," *Bioinformatics*, vol. 17, no. 12, pp. 1113–1122, 2002.
- [56] Z. Wei and S. T. Jensen, "GAME: detecting cis-regulatory elements using a genetic algorithm," *Bioinformatics*, vol. 22, no. 13, pp. 1577–1584, 2006.
- [57] M. Tompa, N. Li, T. L. Bailey et al., "Assessing computational tools for the discovery of transcription factor binding sites," *Nature Biotechnology*, vol. 23, no. 1, pp. 137–144, 2005.
- [58] J. Hu, Y. D. Yang, and D. Kihara, "EMD: An ensemble algorithm for discovering regulatory motifs in DNA sequences," *BMC Bioinformatics*, vol. 7, article no. 342, 2006.
- [59] J. Hu, B. Li, and D. Kihara, "Limitations and potentials of current motif discovery algorithms," *Nucleic Acids Research*, vol. 33, no. 15, pp. 4899–4913, 2005.
- [60] P. M. Ponomarenko and M. P. Ponomarenko, "Sequence-based prediction of transcription upregulation by auxin in plants," *Journal of Bioinformatics and Computational Biology*, vol. 13, no. 1, Article ID 1540009, 2015.
- [61] P. Zimmermann, M. Hirsch-Hoffmann, L. Hennig, and W. Gruissem, "GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox," *Plant Physiology*, vol. 136, no. 1, pp. 2621–2632, 2004.
- [62] H. Goda, E. Sasaki, K. Akiyama et al., "The AtGenExpress hormone and chemical treatment data set: Experimental design, data evaluation, model data analysis and data access," *The Plant Journal*, vol. 55, no. 3, pp. 526–542, 2008.
- [63] A. P. Dempster, "New methods for reasoning towards posterior distributions based on sample data," *Annals of Mathematical Statistics*, vol. 37, pp. 355–374, 1966.
- [64] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.
- [65] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [66] V. Kaftandjian, Y. M. Zhu, O. Dupuis, and D. Babot, "The combined use of the evidence theory and fuzzy logic for improving multimodal nondestructive testing systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 5, pp. 1968–1977, 2005.
- [67] Y. M. Zhu, L. Bentabet, O. Dupuis, V. Kaftandjian, D. Babot, and M. Rombaut, "Automatic determination of mass functions in Dempster-Shafer theory using fuzzy c-means and spatial neighborhood information for image segmentation," *Optical Engineering*, vol. 41, no. 4, pp. 760–770, 2002.
- [68] E. Huala, A. W. Dickerman, M. Garcia-Hernandez et al., "The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant," *Nucleic Acids Research*, vol. 29, no. 1, pp. 102–105, 2001.
- [69] F.-B. Guo, H.-Y. Ou, and C.-T. Zhang, "ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes," *Nucleic Acids Research*, vol. 31, no. 6, pp. 1780–1789, 2003.
- [70] M. Cole, J. Chandler, D. Weijers, B. Jacobs, P. Comelli, and W. Werr, "DORNROESCHEN is a direct target of the auxin response factor MONOPTEROS in the Arabidopsis embryo," *Development*, vol. 136, no. 10, pp. 1643–1651, 2009.
- [71] G. D. Stormo and G. W. Hartzell III, "Identifying protein-binding sites from unaligned DNA fragments," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 4, pp. 1183–1187, 1989.
- [72] X. S. Liu, D. L. Brutlag, and J. S. Liu, "An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments," *Nature Biotechnology*, vol. 20, no. 8, pp. 835–839, 2002.
- [73] N. Poluliakh, T. Takagi, and K. Nakai, "MELINA: Motif extraction from promoter regions of potentially co-regulated genes," *Bioinformatics*, vol. 19, no. 3, pp. 423–424, 2003.
- [74] Y. Y. Yamamoto, Y. Yoshioka, M. Hyakumachi et al., "Prediction of transcriptional regulatory elements for plant hormone responses based on microarray data," *BMC Plant Biology*, vol. 11, article no. 39, 2011.
- [75] P. Rudd, "In search of the gold standard for compliance measurement," *JAMA Internal Medicine*, vol. 139, no. 6, pp. 627–628, 1979.
- [76] K. Okonechnikov, O. Golosova, M. Fursov et al., "Unipro UGENE: a unified bioinformatics toolkit," *Bioinformatics*, vol. 28, no. 8, pp. 1166–1167, 2012.
- [77] W. Shen, S. Le, Y. Li, and F. Hu, "SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation," *PLoS ONE*, vol. 11, no. 10, 2016.
- [78] N. Jayaram, D. Usvyat, and A. C. R. Martin, "Evaluating tools for transcription factor binding site prediction," *BMC Bioinformatics*, 2016.