

## Article

# An Evidence Theory and Fuzzy Logic Combined Approach for the Prediction of Potential ARF-Regulated Genes in Quinoa

Nesrine Sghaier <sup>1,2,3</sup> , Jemaa Essemine <sup>2</sup> , Rayda Ben Ayed <sup>4,5</sup> , Mustapha Gorai <sup>6</sup>, Riadh Ben Marzoug <sup>7</sup>, Ahmed Rebai <sup>7</sup> and Mingnan Qu <sup>1,2,\*</sup> 

<sup>1</sup> National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China

<sup>2</sup> CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

<sup>3</sup> Laboratory of Advanced Technology and Intelligent Systems, National Engineering School of Sousse, Sousse 4023, Tunisia

<sup>4</sup> Department of Agronomy and Plant Biotechnology, National Institute of Agronomy of Tunisia (INAT), 43 Avenue Charles Nicolle, 1082 El Mahrajène, University of Carthage-Tunis, Tunis 1082, Tunisia

<sup>5</sup> Laboratory of Extremophile Plants, Centre of Biotechnology of Borj-Cédria, B.P. 901, Hammam Lif 2050, Tunisia

<sup>6</sup> Higher Institute of Applied Biology Medenine, University of Gabes, Medenine 4119, Tunisia

<sup>7</sup> Laboratory of Molecular and Cellular Screening Processes, Sfax Biotechnology Center, B.P. 1177, Sfax 3018, Tunisia

\* Correspondence: qumingnan@picb.ac.cn

**Abstract:** Quinoa constitutes among the tolerant plants to the challenging and harmful abiotic environmental factors. Quinoa was selected as among the model crops destined for bio-saline agriculture that could contribute to the staple food security for an ever-growing worldwide population under various climate change scenarios. The auxin response factors (ARFs) constitute the main contributors in the plant adaptation to severe environmental conditions. Thus, the determination of the ARF-binding sites represents the major step that could provide promising insights helping in plant breeding programs and improving agronomic traits. Hence, determining the ARF-binding sites is a challenging task, particularly in species with large genome sizes. In this report, we present a data fusion approach based on Dempster–Shafer evidence theory and fuzzy set theory to predict the ARF-binding sites. We then performed an “In-silico” identification of the ARF-binding sites in *Chenopodium quinoa*. The characterization of some known pathways implicated in the auxin signaling in other higher plants confirms our prediction reliability. Furthermore, several pathways with no or little available information about their functions were identified to play important roles in the adaptation of quinoa to environmental conditions. The predictive auxin response genes associated with the detected ARF-binding sites may certainly help to explore the biological roles of some unknown genes newly identified in quinoa.

**Keywords:** data fusion; machine learning; evidence theory; ARF-binding sites; *Chenopodium quinoa*



**Citation:** Sghaier, N.; Essemine, J.; Ayed, R.B.; Gorai, M.; Ben Marzoug, R.; Rebai, A.; Qu, M. An Evidence Theory and Fuzzy Logic Combined Approach for the Prediction of Potential ARF-Regulated Genes in Quinoa. *Plants* **2023**, *12*, 71. <https://doi.org/10.3390/plants12010071>

Academic Editor: Pengcheng Fu

Received: 15 November 2022

Accepted: 26 November 2022

Published: 23 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Natural systems, food security, and agricultural production have been adversely affected by devastating environmental changes [1]. Thus, the enrichment of our knowledge of plant systems will provide effective solutions and leading strategies for future plant yield improvement and breeding programs [2]. Particularly, quinoa constitutes among the tolerant plants to the challenging and harmful abiotic environmental factors [3]. Thanks to its nutritional characteristics and tolerance capacity to various environmental stress conditions, quinoa has become an attractive laboratory material for scientists and researchers worldwide. Hence, quinoa was selected as among the model crops destined for biosaline agriculture that could contribute to the staple food security for an ever-growing worldwide population under various climate change scenarios [4]. In these regards, the Food and

Agricultural Organization of the United Nations (FAO) declared 2013 as the international quinoa year to shed light on this crop as it is an essential staple food rich in proteins and fibres which needs more attention to be paid and further research activities to be performed to well understand and better decipher its valuable potential in fighting starvation problems to ensure and consolidate a promising nutritional value for the human beings [5].

The severe environmental conditions and the climate changes variability accentuate the effects of numerous stresses on plants [1]. So far, to cope with the surrounding continuously changing environment, plants respond by significant rearrangements at the transcriptomic level and modulation in the expression level of a large number of stress-related genes. Thus, plant hormones have been reported as well to be involved in plant adaptation to different biotic and abiotic stress factors [6,7]. Notably, auxin plays a critical and pivotal role in improving plant tolerance by controlling the expression of many stresses' responsive genes. Auxin signaling involves the activation or repression of gene expression by a specific class of ARF proteins that binds to the ARF-binding sites known as Auxin Response Elements, AuxREs [8,9]. The ARF-binding sites are the main contributors to the auxin response diversity. Thus, the knowledge and determination of the ARF-binding sites represent a key and significant step to understanding well and accurately determining the molecular basis of the auxin action, which could provide insights helping in plant breeding programs and thereby in the amelioration of certain agronomic traits. Therefore, inferring the presence of ARF-binding sites in the regulatory regions is essential both for functional and evolutionary analyses [9]. Thus, determining the ARF-binding sites constitutes a challenging task, particularly in species with large genome sizes.

Afterwards, a promising and powerful approach was used to determine the genome-wide ARF-binding site's location. This approach has been extensively approved to be achieved through various experimental techniques, including ChIP-Chip [10,11], ChIP-Seq [12,13], and ChIP-Pet [14]. However, these experiments are time-consuming and require huge financial resources and support; additionally, their given results remain relative and depend on the conditions adopted and/or being used during the experimental procedure [15]. Nevertheless, a complementary and alternative approach mainly based on computational methods has recently emerged, which allows fast and efficient identification of the ARF-binding sites [16,17]. Hence, the computational prediction of the binding sites remains a pivotal goal in bioinformatics with great priority and interest.

Most of the computational methods have used the TGTC-containing consensus core sequences as a tool to detect the ARF-binding sites [9,18]. Furthermore, other employed methods are mainly based on the position weight matrices (PWMs) describing the sequence preference for the transcription factor [19]. Unfortunately, the predictions using merely the consensus motifs or PWMs often generate a large number of false positives. This makes the detection of the binding site spurious sometimes and inaccurate. Accordingly, some unavoidable challenges and substantial limitations should be considered and considered since there are often several false positive events that could be generated and/or detected. In addition, many variants of DNA-binding sequences exist and can be recognized by the ARF TFs. Therefore, a number of computational approaches have been proposed to overcome these limitations and constraints to improve the prediction of the ARF-binding sites [20].

In the same trend, several recent reports have used various machine learning approaches, including the support vector machine, SVM [21], random forest [22,23], deep learning [24], and convolutional neural network, CNN [25]. These methods employ mostly different lines of evidence for the prediction of the ARF-binding site, such as sequence conservation, gene ontology (GO), and location of the binding sites [20]. Some other methods have used the distance and number of the nearest histone modifications [26]. Other groups based on DNA use three-dimensional (3D) structural information to describe the binding specificities [27,28]. Nevertheless, certain models could predict the binding sites based on the ChIP-Seq data once available [26]. Many other methods relied on gene expression data to predict the binding sites [28–30].

With the increasing number of evidence sources for ARF-binding events, the adoption of computational methods for integrating these various data sources can further improve the prediction of TF binding [31]. Recently, substantial progress has been made and yielded new valuable insights concerning the ARF-binding activity. Thus, the ARF-binding site varies sequences, repeat numbers, spacing, and orientation, contributing to the binding specificity and affinities of the different ARF family members [32–36]. Hence, a combination of this genomic information with other data sets, such as gene expression, may improve the prediction of the ARF-binding events. Plants respond to different environmental stresses by regulating stress-responsive gene expression [37–39].

In this study, we present a data fusion approach-based DS evidence theory and fuzzy set theory [40]. Thus, we combined the predictive data extracted from two techniques frequently used in the detection of the binding sites. These two are the detection of overrepresented motifs and the linear discriminant analysis (LDA). From each method, we extracted several features and combined them with an orthogonal sum of the DST rule. The specific features of the ARF-binding sites are also integrated to further improve their prediction accuracy and reliability. Then, we evaluated our predictions using ChIP-Seq data from *Zea mays*. Subsequently, we performed an “In-silico” identification of the ARF-binding sites in *Chenopodium quinoa*. Quinoa (*Chenopodium quinoa* Willd.) is a halophytic pseudo-cereal crop that originated from the Andean region of South America [41]. It is an allotetraploid ( $2n = 4x = 36$ ) with an estimated genome size of approximately 1.5 Gbp [42]. It is a more nutritious grain than any other major cereal [43–45].

## 2. Results

### 2.1. Modelling Approach

In this study, we perform a data fusion approach based on Dempster–Shafer theory and fuzzy set theory to predict ARF-binding sites. Thus, we combined different extracted features.

To meet our requirements, two hypotheses were considered herein, which are the following: “the motif can be an ARF-binding site” or “the motif can’t be an ARF-binding site”.

In terms of the DS evidence theory, we are located in the case where the frame of discernment is constructed with two single hypotheses, H1 and H2, then a single composite one (H3), unifying the two other hypotheses as mentioned in the following formula:  $H3 = H1 \cup H2$  (union of H1 and H2), where H3 represents the ignorance indeed. The modelling of our method follows six major steps to represent the confidence in the detection of the ARF-binding site.

Step 1: Features determination.

Step 2: Construction of feature space.

Step 3: Determination of confidence regions.

Step 4: Modelling hypotheses.

Step 5: Fuzzification.

Step 6: Dempster–Shafer combination.

#### 2.1.1. Features Determination

The first step consists of extracting the different features that can be selected from each studied method. From the first method (detection of overrepresented motifs), we have extracted four features which are position (P), significance score (Sc), occurrence (O), and density (D). The position represents the initial feature that must be determined for each motif of the database that we enclosed. Thus, several regulatory elements have been identified in the 5'UTR regions [46,47]. In fact, we have chosen to use the position relative to the start codon ATG. However, the significance score (Sc) has been calculated using the Weeder algorithm as previously performed by [48]. The occurrence indicates the total number of the detected motifs (core sequences) in a whole genome. The density feature has been defined as the ARF-binding sites rate in the promoter of auxin responsive

genes. In this context, we have selected the auxin response genes with twofold changes (FC) expression from the microarray data [49].

For the second method, an LDA has been performed using Z-curve features [50] and the GC% as potentially the discriminative features. Hence, the LDA represents among the most important supervised linear dimensional reduction techniques [51]. The Z-curve is a unique 3D representative curve of a DNA sequence (Equations (1)–(3)). Notably, the three Z-curve features used here are:

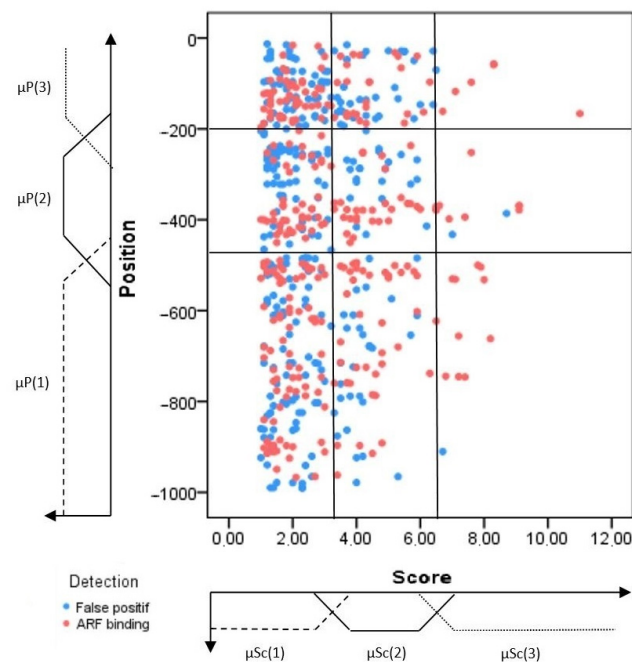
$$x1 = (a1 + g1) - (c1 + t1) \quad (1)$$

$$y1 = (a1 + c1) - (g1 + t1) \quad (2)$$

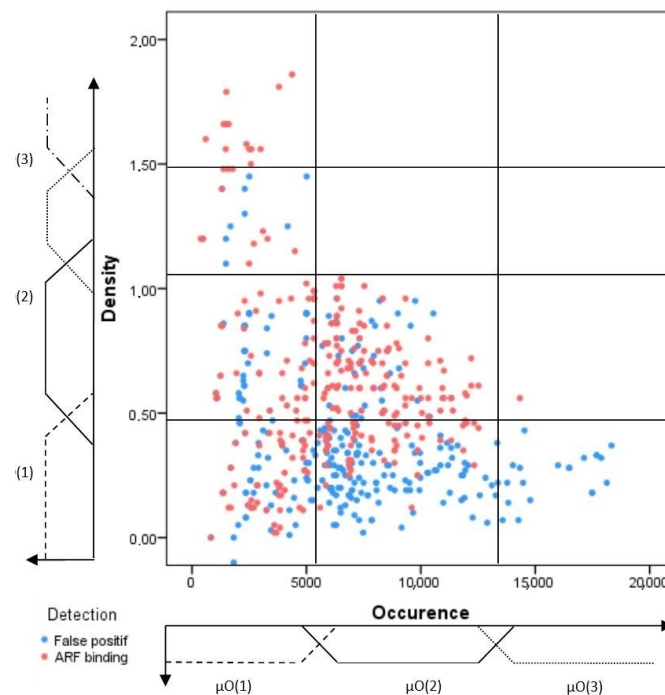
$$z1 = (a1 + t1) - (g1 + c1) \quad (3)$$

### 2.1.2. Construction of Feature Space

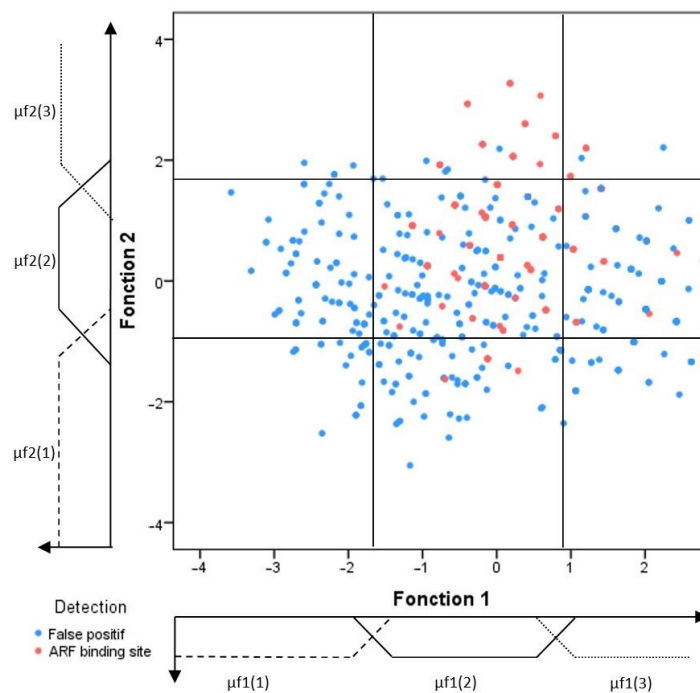
In the following section, the training motifs have been divided into positive and negative training sets to construct a discriminative prediction model. These motifs are studied in the feature space, which helps to investigate the link between the selected features and the type of the considered motifs. Thereby, three learning graphs have been created (Figures 1–3). These learning graphs represent the distribution of training motifs according to their calculated features. We have chosen to study independently the knowledge and insights acquired from, on one hand, the position and significance score (Figure 1) and, on the other hand, those provided by the Occurrence and Density (Figure 2) in order to distinguish as much as possible, the ARF-binding sites from the false positive's events. For the LDA, we have used the Z-curve feature and the GC%. Function 1 and function 2 are the first and the second discriminant functions, respectively. The first function maximizes the differences between groups on that function. The second function maximizes differences in that function. So, the third learning graph depicts the distribution of the different types of motifs in function 1 versus function 2 feature space (Figure 3).



**Figure 1.** Feature space of the position and significance score representing the training sets distribution of the ARF-binding sites (red color) and false positives (blue color). Three fuzzy sets and their corresponding membership degree functions ( $\mu_{Sc(i)}$  and  $P(j)$ ). The latter parameters were defined for each feature (position and score) and found to yield nine regions. The boundaries of the different sets were determined through a learning step as much as possible to define the discriminative regions.



**Figure 2.** Feature space of the occurrence and density defining the training sets distribution of the ARF-binding sites (red color) and false positives (blue color). Three fuzzy sets and their corresponding membership degree functions ( $\mu O(i)$  and  $D(j)$ ).  $\mu O(i)$  and  $D(j)$  were defined for the occurrence and density, respectively, yielding twelve regions. The boundaries of the different sets were determined through a learning step as much as we could to find the discriminative regions.



**Figure 3.** Feature space of the two first discriminative functions linear discriminant analysis of the represented training sets distribution of the ARF-binding sites (red color) and false positives (blue color). Three fuzzy sets and their corresponding membership degree functions ( $\mu F1(i)$  and  $F2(j)$ ) were defined for each feature, yielding nine regions. The different sets of boundaries were determined through a learning step as much as possible to delimit discriminative regions.



The first graph corresponds to the area of uncertainty that contains all types of regulatory elements. Thus, they are not discriminative features, as many ARF-binding sites were found in a very common region. Therefore, the interpretation of feature space relative to the two other features (occurrence and density) improves the TF classification and helps segregate or differentiate the ARF-binding sites, especially those found in the common region.

### 2.1.3. Confidence Regions

Generally, the constructed learning graphs did not provide clear discrimination of the ARF-binding sites from the false positive hits. Indeed, for the sake of clarity, each graph was sub-divided into different areas called confidence regions that would be enriched by the ARF-binding sites. Each confidence region was defined according to the percentage of ARF-binding sites included. Mostly, this percentage varies from one region to another, and the graph partitioning was as illustrated in Figures 1–3.

### 2.1.4. Modelling Hypotheses

In order to perform automated detection processing for the ARF-binding sites, a confidence level should be automatically assigned for each unknown detected motif that can be placed on the graph. To achieve that, we have defined a gradual doubt through a set of four propositions:

- P1(H<sub>i</sub>, H<sub>j</sub>): Total ignorance
- P2(H<sub>i</sub>, H<sub>j</sub>): Low preference for the H<sub>i</sub> hypothesis but high doubt between H<sub>i</sub> and H<sub>j</sub>
- P3(H<sub>i</sub>, H<sub>j</sub>): Strong preference for the H<sub>i</sub> hypothesis but low doubt between H<sub>i</sub> and H<sub>j</sub>
- P4(H<sub>i</sub>): Total confidence in the H<sub>i</sub> hypothesis, no doubt

Table 1 shows the seven cases representing the various preference degrees that enable expressing the hesitation between the two hypotheses (H1 and H2).

**Table 1.** Seven possibilities for the various preference degrees expressing the hesitation between the two hypotheses, H1 and H2.

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

Thereafter, all propositions are interpreted in a numerical form for which the information sources can provide evidence by using elementary mass values. The preference levels from P1 to P4 are modeled by a gradual mass value and are equal to 0, 0.33, 0.67, and 1, respectively [52]. P4 represents the hypothesis with total confidence and no doubt that the detected motif is an ARF-binding site. In the absence of doubt, the mass value assigned is equal to 1. The mass value corresponds to the total doubt equals zero. The transformation of doubt into a quantitative mass value was provided in detail in Table 1. Afterwards, a proposition was assigned to each confidence region from the previous analysis of the learning graphs based on the percentages of ARF-binding sites already existing in each region.

### 2.1.5. Fuzzification

In the previous sections, we applied a discrete representation to define the regions of the learning graphs. This representation assigns different classifications for the close motifs from both the boundaries sides. However, the boundaries between regions were not

well defined, and the transition from one region to another remains tricky and ambiguous. Thus, to further describe this fuzziness aspect in the learning graphs when using the mass functions, we suggest incorporating the theory of fuzzy logic. Accordingly, a gradual, continuous, and smooth transition between regions can be achieved using the membership function concept. Therefore, we have defined the fuzzy sets for each measured feature. For instance, for the feature significance score ( $Sc$ ), four distinct sets (small, average, high, and very high) were defined and recognized. Thus, a smooth transition from one region to another can merely occur. During the learning stage, such a detected motif could be weighted by its membership degrees to different fuzzy sets and characterized by a mass value according to the doubt level for the hypothesis related to its corresponding region.

For each detected motif, three masses were calculated according to Equations (4)–(6) below, corresponding to the three learning graphs (Figures 1–3). They are given, respectively, as follows:

$$m(O \in S/Sc\&P) = \sum_{i=1, j=1}^{i=3, j=3} \mu_{Sc(i)}(x) \times \mu_{p(j)}(y) \times m_{R_{ij}}(O \in S/Sc\&P) \quad (4)$$

$$m(O \in S/O\&D) = \sum_{i=1, j=1}^{i=3, j=4} \mu_{O(i)}(x) \times \mu_{D(j)}(y) \times m_{R_{ij}}(O \in S/O\&D) \quad (5)$$

$$m(O \in S/f1\&f2) = \sum_{i=1, j=1}^{i=3, j=3} \mu_{f1(i)}(x) \times \mu_{f2(j)}(y) \times m_{R_{ij}}(O \in S/f1\&f2) \quad (6)$$

where  $S$  represents any sub-set of the hypotheses,  $m_{R_{ij}}(O \in S/Sc\&P)$ ,  $m_{R_{ij}}(O \in S/O\&D)$ , and  $m_{R_{ij}}(O \in S/f1\&f2)$  designates the mass corresponding to the region  $R_{ij}$  of the significance score/position graph, Occurrence/Density graph, and  $f1/f2$  graph, respectively.

#### 2.1.6. Dempster–Shafer Combination

Subsequently, the data fusion step consists of a combination of the confidence levels deduced from the two methods of detection. Firstly, we must combine the two masses of method 1 (Equations (4) and (5)). Thus, the mass function was obtained by fusing the two masses from the two learning graphs of method 1 by using the orthogonal sum of DS evidence:

$$m_1(O \in S) = m(O \in S/Sc\&P) \oplus m(O \in S/O\&D) \quad (7)$$

Then, the final mass function ( $m_{fusion}$ ) (Equation (8)) was defined by fusing the two masses  $m_1(O \in S)$  (Equation (7)) and  $m(O \in S/f1\&f2)$  (Equation (6)) as shown in the following equation:

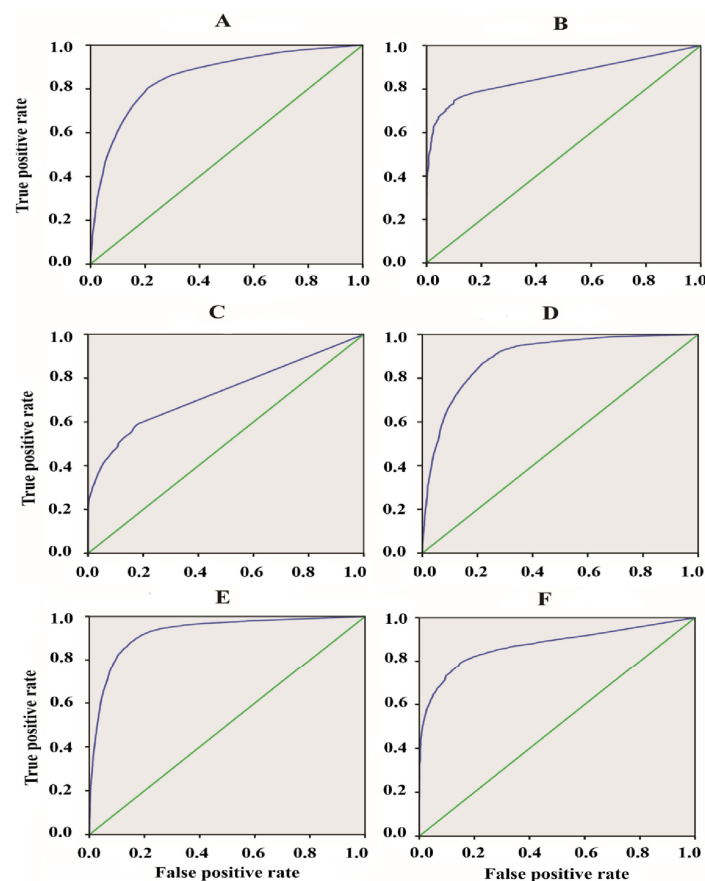
$$m_{fusion}(O \in S) = m_1(O \in S) \oplus m(O \in S/f1\&f2) \quad (8)$$

Thus, this approach integrates many specific features of the ARF-binding sites, such as sequence conservation, cut counts in a 200 bp window around the site, location relative to the transcription starting sites, and motif orientation and spacing. The number of motifs repeats and their spacing represents a very characteristic and major feature for the ARF-binding sites. In this regard, Galli and co-workers have demonstrated that the ARFs bind more frequently with high affinity to the sites containing multiple TGTC core sequences in *Zea mays* [32]. The same authors found that most of the peaks containing two or more TGTCs (55–86%) and/or the peaks with higher TGTC motif numbers showed stronger peak signal intensity [32]. Conversely, the randomly selected genomic regions contain a much lower percentage of fragments with two or more TGTC motifs and a much higher proportion of instances with zero or only a single TGTC core sequence [32]. Additionally, motif orientation and spacing are both important features in the ARF-specific binding sites. Overall, 90% of total peaks reveal less than 50 intervening nucleotides for all orientations, while 29–46% showed spacing of fewer than 20 nucleotides [32,34]. These findings are like previous studies performed on *Arabidopsis thaliana* [53,54]. The integration of these important features reduces the prediction space and the potential of false positive rates, thereby promoting prediction reliability and precision.

## 2.2. Evaluation of Data Fusion Approach on the Experimental Data

In order to assess the predictive accuracy of our methodology, we constructed gold-standard datasets for six ARFs which are ARF4, ARF13, ARF14, ARF18, ARF35, and ARF39. The ChIP-Seq data from *Zea mays* was used to evaluate and/or test the performance of the data fusion method for identifying the ARF-binding sites. These ARFs members were chosen because they are available in a narrow peak format with peak summit values. Therefore, we used the middle 100 bases of each peak to ensure including enough sequence length in identifying the ARF-binding sites while minimizing the false detections. All ChIP-Seq peaks of the ARFs are taken (considered) as positive binding events. The positive sequences correspond to 50 bases from each side of the maximum signal for each ChIP-Seq peak. The control set contains randomly generated, non-overlapping peaks harboring the same mean peak width as the positive ones (peaks).

The receiver operation characteristic curve (ROC) can be performed by plotting the true positive rate against the false positive one (rate) at different thresholds. Thus, we mainly considered the area under the ROC curve (AUC) to assess the aggregated classification performance. Figure 4 and Table 2 display the ROC curves and AUC values, respectively, for the six evaluated ARFs members. The data fusion-based algorithm could discriminate the ChIP-Seq peaks from the control sequences for the whole evaluated ARFs to some extent (degree), as evidenced by the fact that the AUC scores of all the ARFs members surpassed the random expectation of 0.5.



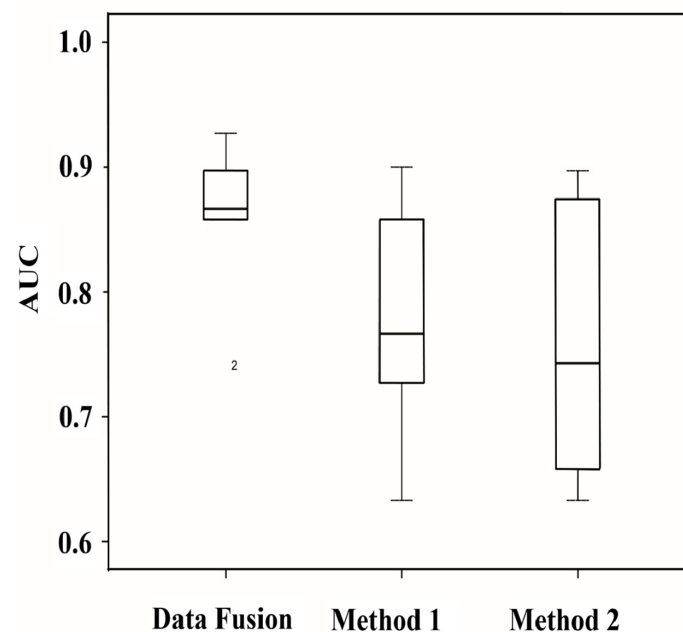
**Figure 4.** The receiver operating characteristic (ROC) curves of ARF-binding sites predicted using our data fusion method. The ARF members (4, 13, 14, 18, 35, and 39) were depicted by the alphabetic letters from (A–F), respectively. The true positive rate was evaluated as follows:  $TPR = TP / (TP + FN)$  and the false positive one  $FPR = TN / (TN + FP)$ . The reference lines are displayed in green color for all the ARFs members.



**Table 2.** Area under the receiver operator characteristic (ROC) curve (AUC) and average Area Under the Precision-Recall curve (AUPR) for six evaluated ARFs: ARF4, ARF14, ARF35, ARF39, ARF13, and ARF18.

ARF	ARF4	ARF14	ARF35	ARF39	ARF13	ARF18
AUC	0.859	0.733	0.927	0.874	0.858	0.897
AUPR	Data fusion 0.91		Matrix scan 0.8		Fimo 0.78	

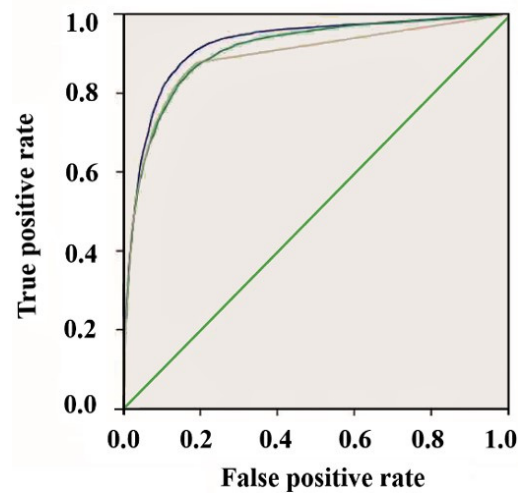
In order to investigate the influence of the combination by data fusion approach, we presented in the boxplots of Figure 5 the average AUC for the 6 studied ARFs members before and after combination using specific features such as the number of motifs repeats and their spacing. Notably, the comparison of the AUC values obviously shows that the data fusion method greatly outperforms the other prediction methods (Figure 5). This reveals the utility of our proposed methodology for the recognition of the ARF-binding sites. Hence, our results depicted in Figure 5 clearly show that using the combination of specific features substantially reduced the number of false positives. Thus, the average AUC of our used method ranged from 0.85 to 0.93, and this reflects the high efficiency of this method and the low rate of spurious combination events generated during our implemented method.



**Figure 5.** Boxplot of the average AUC of 6 different ARFs members studied herein before and after combination using specific features of ARF-binding. Data fusion method is a combination of method\_1 and method\_2. Method\_1 represents the prediction using overrepresented motifs, and method\_2 corresponds to the prediction based on linear discriminant analysis (LDA).

### 2.3. Comparison of Our Approach to Other Methods

In order to evaluate the performance of the data fusion approach for identifying the ARF-binding sites, we compared our methodology with other TFs site prediction software, such as Fimo, as well as Matrix scan. For example, we obtained ChIP-Seq data for the ARF39 and compared the performances and reliabilities of the programs to detect the true positive ARF-binding sites using ROC curves (Figure 6). The area under the curve (AUC) was calculated for each program as well as area under precision recall curve (AUPR). The AUPR evaluates the classification performance in terms of precision and recall.



**Figure 6.** The accuracy of Data fusion methodology (blue curve) was compared to Fimo (green curve) and Matrix scan (yellow curve) using ROC curves for the ARF39 binding sites. Higher curve (close to the top left corner) represents the ROC curve of our method corresponding to the model with better ARF-binding sites prediction quality.

In comparison to the evaluated tools using ARF39 ChIP-Seq data, our method, represented by the higher curve close to the top left corner, shows the best fit (high efficiency) to predict with high accuracy the true positive ARF sites (Figure 6; blue color curve). Our model also achieves higher overall AUC than all the other previously tested methods (Fimo, Matrix scan). As well, the AUPR comparison (Table 2) clearly indicates that our model performs well in terms of the area under the precision-recall curve. This reveals that our model exhibits better performance for the AUPR if compared to Fimo and Matrix scan.

To evaluate our model, we have compared our prediction to previously published detected auxin binding sites in quinoa. A valuable recent study by Yu et al. identified four auxin-responsive elements in the promoter of AUR62002523, AUR62002810, AUR62004953, and AUR62004956 genes. Our analysis also identified with a high score these AuxRE as ARF-binding sites. The high scores are explained by the fact that the studied genes by Yu et al. are Auxin Response Protein (AUX/IAA) since our model takes into consideration the genes' expression levels in response to auxin.

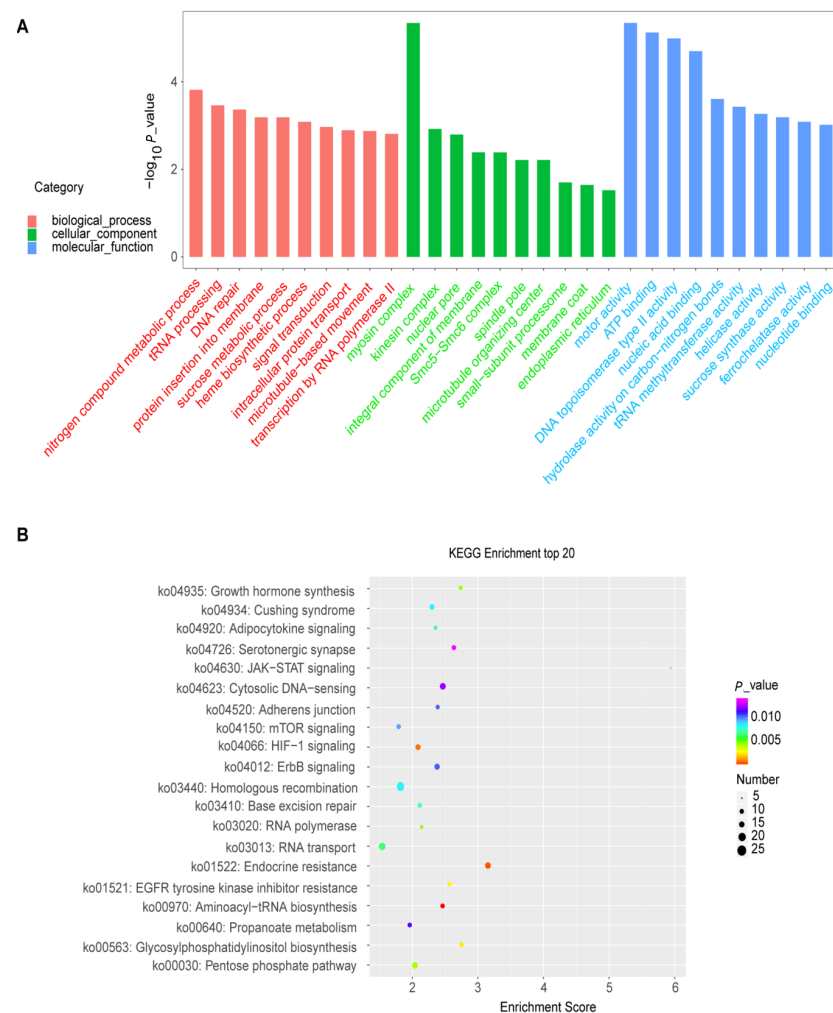
Another recent study by Zhu et al. identified several cis-acting elements associated with auxin in the promoter region of the quinoa SRS genes. The authors highlight the importance of Cis-acting elements in plant defense against various biotic and abiotic stresses. Auxin response elements have been particularly detected in the promoters of three CqSRS gene family members, which are CqSRS 4 (AUR62007636), CqSRS 5 (AUR62007664), and CqSRS 8 (AUR62016794). These AuxRe are likewise identified as ARF-binding sites by our model. All studied cis elements are detected par data fusion model. In addition, some other motifs are also identified as ARF-binding sites. As an example, we detect a second reverse motif at the  $-354$  bp position of the AUR62004956 gene. The spacing between the detected motifs is about thirty bp [55,56].

#### 2.4. Functional Annotation of ARF-Binding Sites in *Chenopodium quinoa*

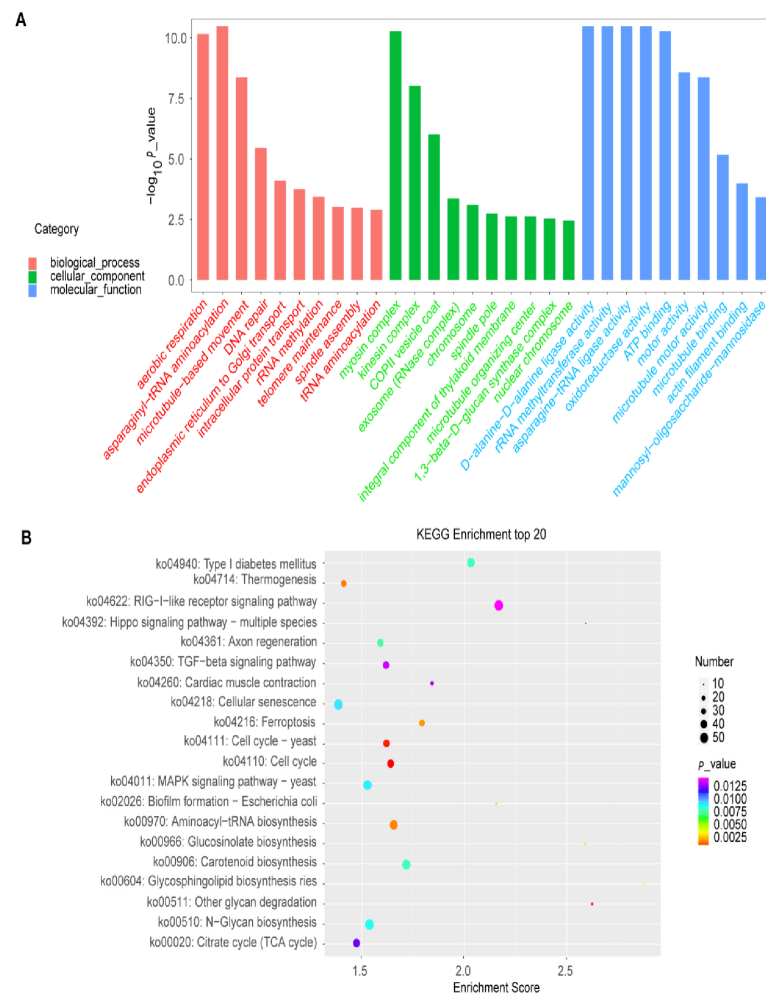
To further explore the gene function and gain more insights into the biological pathways implicated in the auxin response in quinoa, we conducted the GO enrichment and KEGG analysis to illustrate the auxin-responsive genes harboring potential ARF-binding sites in their promoters' regions (Table S1). To unravel the major auxin response processes, three top biological pathways from the enriched analysis were selected for further investigation. In terms of GO enrichment, we found that the proteins showing significantly enriched expression were involved in three top pathways of the biological process of ARF4 (nitrogen compound metabolic process, tRNA processing, and DNA repair;

Figure 7A; Table S2), ARF13 (IMP salvage, (1→3)-beta-D-glucan biosynthetic process and endoplasmic reticulum to Golgi vesicle-mediated transport; Figure S1A; Table S2) and ARF14 (telomere maintenance, DNA repair, and rRNA methylation; Figure S2A; Table S2), ARF18 (regulation of DNA replication, DNA damage checkpoint, and chromosome organization; Figure S3A; Table S2), ARF35 (regulation of DNA replication, DNA damage checkpoint, and chromosome organization; Figure S4A; Table S2) and ARF39 (aerobic respiration, asparaginyl-tRNA amino-acylation, and microtubule-based movement; Figure 8A; Table S2). For the other pathways, complete and detailed descriptions are given in Table S2.

The KEGG analysis reveals that various metabolic pathways were enriched regarding the up-regulated genes by elevated CO<sub>2</sub>. These pathways include starch and sucrose metabolism, Fatty acid biosynthesis, Nitrogen metabolism, Seleno-compound metabolism, Amino-acyl-tRNA biosynthesis, and Carotenoid biosynthesis (Figures 7B, 8B and S1B–S4B; Table S3). For the other pathways, complete and detailed descriptions are given in Table S3.



**Figure 7.** GO and KEGG analyses performed on the ARF4. (A), GO analysis showing the top 20 enriched pathways for the biological process, cellular component, and molecular function. (B), KEGG analysis displaying the top 20 enriched metabolic pathways based on the ARF4 gate.



**Figure 8.** GO and KEGG analyses performed on the ARF39. (A), GO analysis displaying the top 20 enriched pathways for the biological process, cellular component, and molecular function. (B), KEGG analysis displaying the top 20 enriched metabolic pathways based on the ARF39 gate.

### 3. Discussion

In this study, we developed a new algorithm for the prediction of ARF-binding sites by combining a set of genomic features extracted from two basic methods (overrepresented motifs and LDA). To achieve that, we attempted to extract the features from the existing ARFs ChIP-Seq data based on the two chosen methods, and we built up the model by applying a data fusion approach. The supervised learning model step constitutes an important and pivotal task of the work. We then extensively evaluated the transferability using this algorithm and found that the learned model was well accomplished in predicting the ARF-binding sites with high accuracy and reliability.

Quinoa (*Chenopodium quinoa* Willd.) is a pseudo cereal of the Amaranthaceae family, which originates from the Andean region and can adapt to different edaphic and climatic conditions. Both quinoa seeds and leaves are edible parts of the plant; however, the seeds are considered most in terms of economic and scientific aspects. It is a seed crop with high nutritional value since seeds are rich in proteins, lipids, fiber, vitamins, and minerals and have a remarkable balance of essential amino acids. Moreover, due to the absence of gluten, quinoa constitutes a suitable diet for celiac patients or gluten-related disorders [57]. Thus, the nutritional value of quinoa seeds has been reported to meet, and even exceed, that recommended by the World Health Organization, WHO [58]. Furthermore, the quinoa plant is resistant to various environmental stresses, including cold [59], salt [60], and

drought [61]. Very likely, for these reasons, quinoa has been called since a while the “golden grain” [57].

In the same trend, the GO enrichment reveals that the various pathways were enriched regarding the list of genes containing the predictive ARF-binding sites in their promoters. Some pathways are known to be involved in the auxin signaling and response in other plants, such as *Arabidopsis* [62], rice [63], and tea [64]. These results corroborate our prediction reliability based on the data fusion approach and potentially suggest the implication of these pathways in auxin response in *Chenopodium quinoa* as well. For instance, the myosin complex pathway was highly enriched in the following ARFs: ARF4, ARF13, and ARF39 (Figure 7, Figure 8 and Figure S1A). In line with our findings, a recent work also suggested that the myosin XIs could play a significant role in the auxin regulation network in *Arabidopsis thaliana* [62]. Hence, the myosin XIs were found to be involved in mediating and orchestrating (in a concerting manner) the root organogenesis via their effects on the polar distribution of auxin responses and on the cell division process [62].

Besides, another interesting significantly enriched pathway was related to the tRNA amino-acylation and processing. In line with our findings, recently, Chen and co-workers have found that the tRNA modification plays an essential role in auxin signaling in rice plants in response to a moderately high temperature of about 35 °C [63]. Our results confirm and agree with this finding and suggest the implication of tRNA processing in the auxin signalling output in quinoa under heat-stress conditions. Furthermore, the GO enrichment exhibits that the nitrogen compound metabolic process was enriched regarding the genes containing ARF4 binding sites in their promoters (Figure 7A). In the same trend, a recently published study reported that the lateral root formation could be induced by a low nitrogen (N) concentration via auxin biosynthesis and accumulation in tea plants [64]. In this regard, our results strongly also suggested the implication of the ARF4 family in this process (lateral root formation under low N) in quinoa.

Some other pathways were highly enriched in the predicted auxin-responsive genes but have no or less information about their functional annotations and relation to auxin response. Thus, our results proposed that these pathways play crucial or prominent roles in the auxin signaling and adaptation of quinoa to changing environmental conditions. Among these pathways, the kinesin complex was significantly enriched in the list of all the studied ARFs response genes (ARF4, ARF13, ARF14, ARF18, ARF35, and ARF39) (Figure 7A, Figure 8A and Figures S1A–S4A). This suggests a potentially prominent role of kinesin in the auxin signaling process. The (1→3)-beta-D-glucan biosynthetic process was also enriched in the ARF13 gene list. This (1→3)-beta-D-glucan is highly represented in monocotyledons and known to be involved in plants' response to oxidative and heat stresses [65–67]. Our analyses suggest thus the involvement of this pathway in quinoa adaptation to oxidative and heat stresses by implying the ARF13 auxin signaling (Figure S1A). Hence, various pathways were enriched regarding the list of genes containing the predictive ARF-binding sites in their promoters, including telomere maintenance, DNA repair, asparaginyl-tRNA amino-acylation, and chloroplast movements (Table S2). These different pathways were extensively reported to be implicated in the *Chenopodium quinoa* tolerance to adverse environmental factors, such as salinity [3,68,69], drought [70], or their combined effect [71,72]. Eventually, this finding may help scientists to better understand and identify the crucial molecular mechanisms of the auxin action and to further uncover how the auxin signaling pathway could involve, in the quinoa plant, against different abiotic stresses such as salinity since it (quinoa) has earlier been widely used as a model crop for understanding the salt-tolerance in halophytes [4]. In addition, the predictive auxin response genes with the detected ARF-binding sites could help explore the molecular functions and the biological roles of some unknown genes in quinoa since its genome has still not yet sequenced till date, which makes it tough and delicate to decipher its deep genomic functions with precision. One of the main advantages of Dempster–Shafer theory (DST) is that we can utilize it to generate a degree of belief by taking all the evidence into account. This evidence can be obtained from different sources. Merge several types of information

specific to ARF-binding sites to reduce false positives significantly. However, it still needs to be improved by introducing other parameters like digital genomic footprinting or DNase-I hypersensitivity score [73]. Further directions will be the development of algorithms by fusion more specific and recently investigated to predict ARF-binding sites.

## 4. Material and Methods

### 4.1. Training Set

A set of validated ARF-binding sites was collected from the published data online and ChIP-Seq data for six of the *Zea mays* ARF-binding sites regions from the gene expression omnibus (GEO) under the accession number GSE111857 [32,34].

The data were available in narrow peak format with peak max values. We extracted the binding peaks for each dataset with a length of 100 bp centering on the summit of the originally called binding peaks. We divided our dataset into training and test sets. The whole genomes dataset and upstream sequences of *Zea mays* and *Chenopodium quinoa* were downloaded from phytozome ([www.phytozome.org](http://www.phytozome.org) (accessed on 18 September 2019)) and used for all our analysis below.

A Linear discriminant analysis (LDA) was performed using SPSS (v. 16.0, statistical package for the social sciences, Chicago, IL, USA). The microarray data of the primary response to auxin in Arabidopsis was taken from the Genevestigator database (<https://genevestigator.com/gv/> (accessed on 18 November 2019)) [74].

### 4.2. Algorithm Implementation

The main algorithm was implemented under the R software environment language version R-3.5.3. All computations were performed on a single CPU Intel Core i7 computer running at 2.8 GHz, with 8 GB main memory. The source code is available upon request.

In order to assess the predictive accuracy of our methodology, we constructed gold-standard datasets for six ARFs which are ARF4, ARF13, ARF14, ARF18, ARF35, and ARF39.

### 4.3. Evaluation of Data Fusion Approach

The ChIP-Seq data from *Zea mays* was used to evaluate and test the performance of the data fusion method for identifying the ARF-binding sites. The gold-standard data set was constructed based on ChIP-Seq data of the given ARFs currently under investigation. These ARFs members were chosen because they are available in a narrow peak format with peak summit values. Therefore, we used the middle 100 bases of each peak to ensure including enough sequence length in identifying the ARF-binding sites while minimizing the false detections. All ChIP-Seq peaks of the ARFs are considered positive binding events. The positive sequences correspond to 50 bases from each side of the maximum signal for each ChIP-Seq peak. The control set contains randomly generated, non-overlapping peaks harboring the same mean peak width as the positive ones.

Thus, we mainly considered the area under the ROC curve (AUC) to estimate the aggregated classification performance. The receiver operation characteristic curve (ROC) and average Area Under the Precision-Recall curve (AUPR) can be performed by plotting the true positive rate against the false positive one at different thresholds.

### 4.4. Functional Annotation

To further explore the gene function and gain more insights about the biological pathways implicated in the auxin response in quinoa, we conducted the GO enrichment and KEGG analysis to illustrate the auxin-responsive genes containing potential ARF-binding sites in their promoters' regions.

A multi-omics data analysis tool, OmicsBean (<http://www.omicsbean.cn> (accessed on 1 July 2020)), dedicated to integrating the gene ontology (GO) enrichment and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, was employed to investigate the obtained gens lists (Table S2). A  $p$ -value  $< 0.05$  (Fisher's exact test) was considered as the threshold to determine the significant enrichments for the GO and KEGG pathways.



## 5. Conclusions

In this article, we present a new data fusion approach based on DS evidence theory and fuzzy set theory to predict the ARF-binding sites in quinoa. To achieve that, we developed a new algorithm for the prediction of ARF-binding sites by combining a set of genomic features extracted from two basic methods (overrepresented motifs and LDA). Thus, we found that the learned model was well accomplished in predicting with high accuracy and reliability these ARFs. Afterwards, we performed an “In-silico” identification of the ARF-binding sites in *Chenopodium quinoa*. The GO enrichment corroborates that various biological pathways were enriched regarding the list of genes containing the predictive ARF-binding sites in their promoters. These pathways were identified to play important roles in the auxin signaling and adaptation of quinoa to severe environmental conditions.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12010071/s1>, Figure S1: GO and KEGG analyses performed on the ARF13. A, GO analysis displaying the top 20 enriched pathways for the biological process, cellular component and the molecular function. B, KEGG analysis displaying the top 20 enriched metabolic pathways based on the ARF13 gate; Figure S2: GO and KEGG analyses performed on the ARF14. A, GO analysis displaying the top 20 enriched pathways for the biological process, cellular component and the molecular function. B, KEGG analysis displaying the top 20 enriched metabolic pathways based on the ARF14 gate; Figure S3: GO and KEGG analyses performed on the ARF18. A, GO analysis displaying the top 20 enriched pathways for the biological process, cellular component and the molecular function. B, KEGG analysis displaying the top 20 enriched metabolic pathways based on the ARF18 gate; Figure S4: GO and KEGG analyses performed on the ARF35. A, GO analysis displaying the top 20 enriched pathways for the biological process, cellular component and the molecular function. B, KEGG analysis displaying the top 20 enriched metabolic pathways based on the ARF35 gate; Table S1: List of putative auxin response genes containing ARF-binding sites in their promoters regions predicted by data fusion approach; Table S2: Enriched biological pathways based on gene ontology (GO) analysis in *Chenopodium quinoa*; Table S3: Enriched biological pathways based on Kyoto encyclopedia of genes and genomes (KEGG) analysis in *Chenopodium quinoa*.

**Author Contributions:** Conceptualization, N.S. and J.E.; methodology, N.S.; software, R.B.M.; validation, J.E., M.Q. and A.R.; formal analysis, M.G.; investigation, N.S.; resources, N.S.; data curation, N.S.; writing—original draft preparation, N.S.; writing—review and editing, J.E.; visualization, R.B.A.; supervision, M.Q.; project administration, M.Q.; funding acquisition, M.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper is partially funded by the National Natural Science Foundation of China (32170245; 32260447) and the project of Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2022-04).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data used to support the findings of this study are included within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

ARF auxin-responsive transcription factor; AUC, area under receiver operation characteristic curve; AUPR, area under precision-recall curve; AuxRE, auxin response element; ChIP-Chip, chromatin immuno-precipitation followed by microarray hybridization; ChIP-PET, chromatin immuno-precipitation with the paired-end tags; ChIP-Seq, chromatin immuno-precipitation sequencing; CNN, convolutional neural network; DNA, deoxyribo-nucleic-acid; DST, Dempster-Shafer theory; FAO, food and agricultural organization; GEO, gene expression omnibus; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; LDA, linear discriminant analysis; PWM, position weight matrices; ROC, receiver operating characteristic; SVM, support vector machine; TF, transcription factor.

## References

- Raza, A.; Razzaq, A.; Mehmood, S.S.; Zou, X.; Zhang, X.; Lv, Y.; Xu, J. Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. *Plants* **2019**, *8*, 34. [CrossRef]
- Pareek, A.; Dhankher, O.P.; Foyer, C.H. Mitigating the impact of climate change on plant productivity and ecosystem sustainability. *J. Exp. Bot.* **2020**, *71*, 451–456. [CrossRef]
- Manaa, A.; Goussi, R.; Derbali, W.; Cantamessa, S.; Abdelly, C.; Barbato, R. Salinity tolerance of quinoa (*Chenopodium quinoa* Willd) as assessed by chloroplast ultrastructure and photosynthetic performance. *Environ. Exp. Bot.* **2019**, *162*, 103–114. [CrossRef]
- Ruiz, K.B.; Biondi, S.; Oses, R.; Acuña-Rodríguez, I.S.; Antognoni, F.; Martínez-Mosqueira, E.A.; Coulibaly, A.; CanahuaMurillo, A.; Pinto, M.; Zurita-Silva, A.; et al. Quinoa biodiversity and sustainability for food security under climate change. A review. *Agron. Sustain. Dev.* **2014**, *34*, 349–359. [CrossRef]
- Bazile, D.; Bertero, D.; Nieto, C. (Eds.) *State of the Art Report on Quinoa around the World in 2013*; FAO: Roma, Italy, 2015; 589p, Available online: [https://www.academia.edu/21134586/Bazile\\_D\\_ed\\_Bertero\\_H\\_D\\_ed\\_Nieto\\_C\\_ed\\_2015\\_State\\_of\\_the\\_art\\_report\\_on\\_quinoa\\_around\\_the\\_world\\_in\\_2013\\_Roma\\_Italy\\_FAO\\_589\\_p](https://www.academia.edu/21134586/Bazile_D_ed_Bertero_H_D_ed_Nieto_C_ed_2015_State_of_the_art_report_on_quinoa_around_the_world_in_2013_Roma_Italy_FAO_589_p) (accessed on 17 September 2020).
- Xiong, L.; Ishitani, M.; Zhu, J.-K. Interaction of Osmotic Stress, Temperature, and Abscisic Acid in the Regulation of Gene Expression in Arabidopsis. *Plant Physiol.* **1999**, *119*, 205–212. [CrossRef]
- Shinozaki, K.; Yamaguchi-Shinozaki, K. Gene Expression and Signal Transduction in Water-Stress Response. *Plant Physiol.* **1997**, *115*, 327–334. [CrossRef]
- Pierre-Jerome, E.; Moss, B.L.; Lanctot, A.; Hageman, A.; Nemhauser, J.L. Functional analysis of molecular interactions in synthetic auxin response circuits. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11354–11359. [CrossRef]
- Stigliani, A.; Martin-Arevalillo, R.; Lucas, J.; Bessy, A.; Vinos-Poyo, T.; Mironova, V.; Vernoux, T.; Dumas, R.; Parcy, F. Capturing Auxin Response Factors Syntax Using DNA Binding Models. *Mol. Plant.* **2019**, *12*, 822–832. [CrossRef]
- Horak, C.E.; Snyder, M. A genomic approach for identifying transcription factor binding sites. *Methods Enzymol.* **2002**, *350*, 469–483. [CrossRef]
- Weinmann, A.S.; Farnham, P.J. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* **2002**, *26*, 37–47. [CrossRef]
- Robertson, G.; Hirst, M.; Bainbridge, M.; Bilenky, M.; Zhao, Y.; Zeng, T.; Euskirchen, G.; Bernier, B.; Varhol, R.; Delaney, A.; et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **2007**, *4*, 651–657. [CrossRef]
- Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **2007**, *129*, 823–837. [CrossRef]
- Loh, Y.-H.; Wu, Q.; Chew, J.-L.; Vega, V.B.; Zhang, W.; Chen, X.; Bourque, G.; George, J.; Leong, B.; Liu, J.; et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **2006**, *38*, 431–440. [CrossRef]
- Sandelin, A.; Wasserman, W.W. Prediction of nuclear hormone receptor response elements. *Mol. Endocrinol.* **2005**, *19*, 595–606. [CrossRef]
- Lenhard, B.; Sandelin, A.; Mendoza, L.; Engström, P.; Jareborg, N.; Wasserman, W.W. Identification of conserved regulatory elements by comparative genome analysis. *J. Biol.* **2003**, *2*, 13. [CrossRef]
- Brazma, A.; Jonassen, I.; Vilo, J.; Ukkonen, E. Predicting gene regulatory elements in silico on a genomic scale. *Genome Res.* **1998**, *8*, 1202–1215. Available online: <http://www.ncbi.nlm.nih.gov/pubmed/9847082> (accessed on 20 July 2016). [CrossRef]
- Zemlyanskaya, E.V.; Wiebe, D.S.; Omelyanchuk, N.A.; Levitsky, V.G.; Mironova, V.V. Meta-analysis of transcriptome data identified TGTCNN motif variants associated with the response to plant hormone auxin in *Arabidopsis thaliana* L. *J. Bioinform. Comput. Biol.* **2016**, *14*, 1641009. [CrossRef]
- Stormo, G.D.; Schneider, T.D.; Gold, L.; Ehrenfeucht, A. Use of the “Perceptron” algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res.* **1982**, *10*, 2997–3011. Available online: <http://www.ncbi.nlm.nih.gov/pubmed/7048259> (accessed on 20 July 2016). [CrossRef]
- Nguyen, T.T.; Androulakis, I.P. Recent Advances in the Computational Discovery of Transcription Factor Binding Sites. *Algorithms* **2009**, *2*, 582–605. [CrossRef]
- Kumar, S.; Bucher, P. Predicting transcription factor site occupancy using DNA sequence intrinsic and cell-type specific chromatin features. *BMC Bioinform.* **2016**, *17*, 4. [CrossRef]
- Tsai, Z.T.Y.; Shiu, S.H.; Tsai, H.K. Contribution of Sequence Motif, Chromatin State, and DNA Structure Features to Predictive Models of Transcription Factor Binding in Yeast. *PLoS Comput. Biol.* **2015**, *11*, e1004418. [CrossRef] [PubMed]
- Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
- Lan, G.; Zhou, J.; Xu, R.; Lu, Q.; Wang, H. Cross-cell-type prediction of TF-binding site by integrating convolutional neural network and adversarial network. *Int. J. Mol. Sci.* **2019**, *20*, 3425. [CrossRef] [PubMed]
- Alipanahi, B.; Delong, A.; Weirauch, M.T.; Frey, B.J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **2015**, *33*, 831–838. [CrossRef]
- Qin, Q.; Feng, J. Imputation for transcription factor binding predictions based on deep learning. *PLoS Comput. Biol.* **2017**, *13*. [CrossRef]
- Zhou, T.; Shen, N.; Yang, L.; Abe, N.; Horton, J.; Mann, R.S.; Bussemaker, H.J.; Gordân, R.; Rohs, R. Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4654–4659. [CrossRef]

28. Liu, S.; Zibetti, C.; Wan, J.; Wang, G.; Blackshaw, S.; Qian, J. Assessing the model transferability for prediction of transcription factor binding sites based on chromatin accessibility. *BMC Bioinf.* **2017**, *18*, 355. [CrossRef]
29. Faith, J.J.; Hayete, B.; Thaden, J.T.; Mogno, I.; Wierzbowski, J.; Cottarel, G.; Kasif, S.; Collins, J.J.; Gardner, T.S. Large-Scale Mapping and Validation of *Escherichia coli* Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol.* **2007**, *5*, e8. [CrossRef]
30. Ramsey, S.A.; Klemm, S.L.; Zak, D.E.; Kennedy, K.A.; Thorsson, V.; Li, B.; Gilchrist, M.; Gold, E.S.; Johnson, C.D.; Litvak, V.; et al. Uncovering a Macrophage Transcriptional Program by Integrating Evidence from Motif Scanning and Expression Dynamics. *PLoS Comput. Biol.* **2008**, *4*, e1000021. [CrossRef]
31. Ernst, J.; Plasterer, H.L.; Simon, I.; Bar-Joseph, Z. Integrating multiple evidence sources to predict transcription factor binding in the human genome. *Genome Res.* **2010**, *20*, 526–536. [CrossRef]
32. Galli, M.; Khakhar, A.; Lu, Z.; Chen, Z.; Sen, S.; Joshi, T.; Nemhauser, J.L.; Schmitz, R.J.; Gallavotti, A. The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nat. Commun.* **2018**, *9*, 4526. [CrossRef]
33. Matthes, M.S.; Best, N.B.; Robil, J.M.; Malcomber, S.; Gallavotti, A.; McSteen, P. Auxin EvoDevo: Conservation and Diver-sification of Genes Regulating Auxin Biosynthesis, Transport, and Signaling. *Mol. Plant.* **2019**, *12*, 298–320. [CrossRef] [PubMed]
34. Ricci, W.A.; Lu, Z.; Ji, L.; Marand, A.P.; Ethridge, C.L.; Murphy, N.G.; Noshay, J.M.; Galli, M.; Mejía-Guerra, M.K.; Colomé Tatché, M.; et al. Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **2019**, *5*, 1237–1249. [CrossRef] [PubMed]
35. Freire-Rios, A.; Tanaka, K.; Crespo, I.; van der Wijk, E.; Sizentsova, Y.; Levitsky, V.; Lindhoud, S.; Fontana, M.; Hohlbein, J.; Boer, D.R.; et al. Architecture of DNA elements mediating ARF transcription factor binding and auxin-responsive gene expression in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 24557–24566. [CrossRef]
36. Kato, H.; Mutte, S.K.; Suzuki, H.; Crespo, I.; Das, S.; Radoeva, T.; Fontana, M.; Yoshitake, Y.; Hainiwa, E.; van den Berg, W.; et al. Design principles of a minimal auxin response system. *Nat. Plants* **2020**, *6*, 473–482. [CrossRef] [PubMed]
37. Liu, X.; Meng, P.; Yang, G.; Zhang, M.; Peng, S.; Zhai, M.Z. Genome-wide identification and transcript profiles of walnut heat stress transcription factor involved in abiotic stress. *BMC Genom.* **2020**, *21*, 1–13. [CrossRef]
38. Li, W.; Pang, S.; Lu, Z.; Jin, B. Function and Mechanism of WRKY Transcription Factors in Abiotic Stress Responses of Plants. *Plants* **2020**, *9*, 1515. [CrossRef]
39. Vriet, C.; Hennig, L.; Laloi, C. Stress-induced chromatin changes in plants: Of memories, metabolites and crop improvement. *Cell. Mol. Life Sci.* **2015**, *72*, 1261–1273. [CrossRef]
40. Feng, F. Generalized rough fuzzy sets based on soft sets. In Proceedings of the 2009 International Workshop on Intelligent Systems and Applications, Wuhan, China, 23–24 May 2009. [CrossRef]
41. Morales, A.; Zurita-Silva, A.; Maldonado, J.; Silva, H. Transcriptional responses of chilean quinoa (*Chenopodium quinoa* Willd.) under water deficit conditions uncovers ABA-independent expression patterns. *Front. Plant Sci.* **2017**, *8*, 216. [CrossRef]
42. Li, F.; Liu, J.; Guo, X.; Yin, L.; Zhang, H.; Wen, R. Genome-wide survey, characterization, and expression analysis of bZIP transcription factors in *Chenopodium quinoa*. *BMC Plant Biol.* **2020**, *20*, 405. [CrossRef]
43. Zurita-Silva, A.; Fuentes, F.; Zamora, P.; Jacobsen, S.E.; Schwember, A.R. Breeding quinoa (*Chenopodium quinoa* Willd.): Potential and perspectives. *Mol. Breed.* **2014**, *34*, 13–30. [CrossRef]
44. Graf, B.L.; Rojas-Silva, P.; Rojo, L.E.; Delatorre-Herrera, J.; Baldeón, M.E.; Raskin, I. Innovations in Health Value and Functional Food Development of Quinoa (*Chenopodium quinoa* Willd.). *Compr. Rev. Food Sci. Food Saf.* **2015**, *14*, 431–445. [CrossRef] [PubMed]
45. El-Shamy, M.A.; Alshaal, T.; Mohamed, H.H.; Rady, A.M.S.; Hafez, E.M.; Alsohim, A.S.; Abd El-Moneim, D. Quinoa Response to Application of Phosphogypsum and Plant Growth-Promoting Rhizobacteria under Water Stress Associated with Salt-Affected Soil. *Plants* **2022**, *11*, 872. [CrossRef] [PubMed]
46. Kutach, A.K.; Kadonaga, J.T. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol. Cell. Biol.* **2000**, *20*, 4754–4764. Available online: <http://www.ncbi.nlm.nih.gov/pubmed/10848601> (accessed on 18 September 2018). [CrossRef] [PubMed]
47. Lim, C.Y.; Santoso, B.; Boulay, T.; Dong, E.; Ohler, U.; Kadonaga, J.T. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* **2004**, *18*, 1606–1617. [CrossRef] [PubMed]
48. Pavesi, G.; Mereghetti, P.; Mauri, G.; Pesole, G. Weeder Web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **2004**, *32*, W199–W203. [CrossRef]
49. Goda, H.; Sasaki, E.; Akiyama, K.; Maruyama-Nakashita, A.; Nakabayashi, K.; Li, W.; Ogawa, M.; Yamauchi, Y.; Preston, J.; Aoki, K.; et al. The AtGenExpress hormone- and chemical-treatment data set: Experimental design, data evaluation, model data analysis, and data access. *Plant J.* **2008**, *55*, 526–542. [CrossRef]
50. Guo, F.-B.; Ou, H.-Y.; Zhang, C.-T. ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res.* **2003**, *31*, 1780–1789. Available online: <http://www.ncbi.nlm.nih.gov/pubmed/12626720> (accessed on 20 July 2016). [CrossRef]
51. Li, Z.; Nie, F.; Chang, X.; Yang, Y. Beyond trace ratio: Weighted harmonic mean of trace ratios for multiclass discriminant analysis. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2100–2110. [CrossRef]
52. Kaftandjian, V.; Zhu, Y.M.; Dupuis, O.; Babot, D. The Combined Use of the Evidence Theory and Fuzzy Logic for Improving Multimodal Nondestructive Testing Systems. *IEEE Trans. Instrum. Meas.* **2005**, *54*, 1968–1977. [CrossRef]

53. Mironova, V.V.; Omelyanchuk, N.A.; Wiebe, D.S.; Levitsky, V.G. Computational analysis of auxin responsive elements in the *Arabidopsis thaliana* L. genome. *BMC Genom.* **2014**, *15*, S4. [[CrossRef](#)] [[PubMed](#)]
54. Cherenkov, P.; Novikova, D.; Omelyanchuk, N.; Levitsky, V.; Grosse, I.; Weijers, D.; Mironova, V. Diversity of cis-regulatory elements associated with auxin response in *Arabidopsis thaliana*. *J. Exp. Bot.* **2018**, *69*, 329–339. [[CrossRef](#)]
55. Yu, H.; Yuan, Y.; Wang, S.; Wu, G.; Xu, H.; Wei, J.; Ju, L.; Huang, Y.; Chen, H. Interspecies Evolution and Networks Investigation of the Auxin Response Protein (AUX/IAA) Family Reveals the Adaptation Mechanisms of Halophytes Crops in Nitrogen Starvation Agroecological Environments. *Agriculture* **2021**, *11*, 780. [[CrossRef](#)]
56. Zhu, X.; Wang, B.; Wang, X.; Wei, X. Genome-wide identification, structural analysis and expression profiles of short internodes related sequence gene family in quinoa. *Front Genet.* **2022**, *13*, 961925. [[CrossRef](#)] [[PubMed](#)]
57. Angeli, V.; Silva, P.M.; Massuela, D.C.; Khan, M.W.; Hamar, A.; Khajehi, F.; Graeff-Hönninger, S.; Piatti, C. Quinoa (*Chenopodium quinoa* Willd.): An overview of the potentials of the “golden grain” and socio-economic and environmental aspects of its cultivation and marketization. *Foods* **2020**, *9*, 216. [[CrossRef](#)] [[PubMed](#)]
58. Hirose, Y.; Fujita, T.; Ishii, T.; Ueno, N. Antioxidative properties and flavonoid composition of *Chenopodium quinoa* seeds cultivated in Japan. *Food Chem.* **2010**, *119*, 1300–1306. [[CrossRef](#)]
59. Fischer, S.; Wilckens, R.; Jara, J.; Aranda, M.; Valdivia, W.; Bustamante, L.; Graf, F.; Obal, I. Protein and antioxidant composition of quinoa (*Chenopodium quinoa* Willd.) sprout from seeds submitted to water stress, salinity and light conditions. *Ind. Crops Prod.* **2017**, *107*, 558–564. [[CrossRef](#)]
60. Gu, M. Transcriptome analysis and differential gene expression profiling of two contrasting quinoa genotypes in response to salt stress. *BMC Plant Biol.* **2020**, *20*, 568. [[CrossRef](#)]
61. Rafik, S.; Chaoui, M.; Assabban, Y.; Jazi, S.; Choukr-Allah, R.; el Gharouss, M.; Hirich, A. Quinoa value chain, adoption, and market assessment in Morocco. *Environ. Sci. Pollut. Res.* **2021**, *28*, 46692–46703. [[CrossRef](#)]
62. Abu-Abied, M.; Belausov, E.; Hagay, S.; Peremyslov, V.; Dolja, V.; Sadot, E. Myosin XI-K is involved in root organogenesis, polar auxin transport, and cell division. *J. Exp. Bot.* **2018**, *69*, 2869–2881. [[CrossRef](#)]
63. Chen, K.; Guo, T.; Li, X.M.; Zhang, Y.M.; Yang, Y.B.; Ye, W.W.; Dong, N.Q.; Shi, C.L.; Kan, Y.; Xiang, Y.H.; et al. Translational Regulation of Plant Response to High Temperature by a Dual-Function tRNAHis Guanylyltransferase in Rice. *Mol. Plant* **2019**, *12*, 1123–1142. [[CrossRef](#)] [[PubMed](#)]
64. Hu, S.; Zhang, M.; Yang, Y.; Xuan, W.; Zou, Z.; Arkorful, E.; Chen, Y.; Ma, Q.; Jeyaraj, A.; Chen, X.; et al. A novel insight into nitrogen and auxin signaling in lateral root formation in tea plant [*Camellia sinensis* (L.) O. Kuntze]. *BMC Plant Biol.* **2020**, *20*, 1–17. [[CrossRef](#)] [[PubMed](#)]
65. Majda, M.; Robert, S. The role of auxin in cell wall expansion. *Int. J. Mol. Sci.* **2018**, *19*, 951. [[CrossRef](#)]
66. Zeng, L.; Wang, Y.H.; Ai, C.X.; Zhang, J.S. Differential effects of  $\beta$ -glucan on oxidative stress, inflammation and copper transport in two intestinal regions of large yellow croaker *Larimichthys crocea* under acute copper stress. *Ecotoxicol. Environ. Saf.* **2018**, *165*, 78–87. [[CrossRef](#)]
67. Havrlentová, M.; Deáková, L.; Kraic, J.; Zofajová, A. Can  $\beta$ -D-Glucan Protect Oat Seeds against a Heat Stress? *Nov. Biotechnol. Chim.* **2016**, *15*, 107–113. [[CrossRef](#)]
68. Hariadi, Y.; Marandon, K.; Tian, Y.; Jacobsen, S.E.; Shabala, S. Ionic and osmotic relations in quinoa (*Chenopodium quinoa* Willd.) plants grown at various salinity levels. *J. Exp. Bot.* **2011**, *62*, 185–193. [[CrossRef](#)]
69. Shabala, S.; Hariadi, Y.; Jacobsen, S.E. Genotypic difference in salinity tolerance in quinoa is determined by differential control of xylem  $\text{Na}^+$  loading and stomatal density. *J. Plant Physiol.* **2013**, *170*, 906–914. [[CrossRef](#)] [[PubMed](#)]
70. Sun, Y.; Liu, F.; Bendevis, M.; Shabala, S.; Jacobsen, S.-E. Sensitivity of Two Quinoa (*Chenopodium quinoa* Willd.) Varieties to Progressive Drought Stress. *J. Agron. Crop Sci.* **2014**, *200*, 12–23. [[CrossRef](#)]
71. Razzaghi, F.; Ahmadi, S.H.; Jacobsen, S.-E.; Jensen, C.R.; Andersen, M.N. Effects of Salinity and Soil-Drying on Radiation Use Efficiency, Water Productivity and Yield of Quinoa (*Chenopodium quinoa* Willd.). *J. Agron. Crop Sci.* **2012**, *198*, 173–184. [[CrossRef](#)]
72. Razzaghi, F.; Ahmadi, S.H.; Adolf, V.I.; Jensen, C.R.; Jacobsen, S.-E.; Andersen, M.N. Water Relations and Transpiration of Quinoa (*Chenopodium quinoa* Willd.) Under Salinity and Soil Drying. *J. Agron. Crop Sci.* **2011**, *197*, 348–360. [[CrossRef](#)]
73. Rivièrè, Q.; Corso, M.; Ciortan, M.; Noël, G.; Verbruggen, N.; Defrance, M. Exploiting Genomic Features to Improve the Prediction of Transcription Factor-Binding Sites in Plants. *Plant Cell Physiol.* **2022**, *63*, 1457–1473. [[CrossRef](#)] [[PubMed](#)]
74. Zimmermann, P.; Hirsch-Hoffmann, M.; Hennig, L.; Gruissem, W. GENEVESTIGATOR. *Arabidopsis* Microarray Database and Analysis Toolbox. *Plant Physiol.* **2004**, *136*, 2621–2632. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.