J Bioinform Syst Biol 2020; 3 (4): 092-104 DOI: 10.26502/jbsb.5107017

Research Article

Exploring the Intrinsic Disorder of the WRKY Transcription Factor Family in the Cereals

Mouna Choura[1*], Ahmed Rebaï[2], Faiçal Brini[1]

[1]Biotechnology and Plant Improvement Laboratory, Center of Biotechnology of Sfax, University of Sfax, Route Sidi Mansour Km 6, P.O.Box 1177, 3018, Sfax, Tunisia

[2]Laboratory of Molecular and Cellular Screening Processes, Center of Biotechnology of Sfax, University of Sfax, Route Sidi Mansour Km 6, P.O.Box 1177, 3018, Sfax, Tunisia

[*]Corresponding Authors: Mouna Choura, Biotechnology and Plant Improvement Laboratory, Center of Biotechnology of Sfax, University of Sfax, Route Sidi Mansour Km 6, P.O.Box 1177, 3018, Sfax, Tunisia, Tel: 216 74 871816, Fax: 216 74 875818; E-mail: mouna.choura@cbs.rnrt.tn

Received: 09 October 2020; **Accepted:** 05 November 2020; **Published:** 16 December 2020

Citation: Mouna Choura, Ahmed Rebaï, Faiçal Brini. Exploring the Intrinsic Disorder of the WRKY Transcription Factor Family in the Cereals. Journal of Bioinformatics and Systems Biology 3 (2020): 092-104.

Abstract

The WRKY transcription factors superfamily are involved in diverse biological processes in plants including responses to biotic and abiotic stresses. It is also known that WRKYs exert many of its biological functions via interaction with other proteins of the signalling pathway. To understand the potential roles of intrinsic disorder in the functioning of WRKYs, we investigated the intrinsic disorder distribution of WRKYs in *O.sativa, S.bicolor, Z.mays, T.aestivum, H.vulgare and A.thaliana* and their protein-protein interactions. The analysis revealed that their intrinsic disorder distributions are similar. The data show that hub proteins have a higher disorder content in all species. These findings indicate that the peculiarities of the intrinsic disorder of WRKYs are evolutionary conserved, and therefore could be of crucial for their biological activities. The multifunctionality of WRKYs suggests that these proteins might utilize intrinsic disorder for their various functions.

Keywords: Cereals; Intrinsic disorder; Interactions; Multifunctionality; WRKY

1. Introduction

Intrinsically disordered proteins (IDPs) lack a defined tertiary structure, yet they play key biological roles. The IDPs are over represented in signalling and transcription regulatory networks which involve hubs having several protein partners [1, 2].

Transcription factors are modular proteins that contain at least one DNA-binding domain, recognize and bind to specific DNA sequences, involved in transcription regulation. The analysis reported that the AT-hooks and basic regions of DNA-binding domains of transcription factors exhibit high level of disorder [3].

According to [4], seventy-eight different families of transcription factors including MYB, AP2-EREBP, bHLH, MADS, C2H2, NAC, HB, WRKY, bZIP and C3H contain different numbers of intrinsically disordered regions (IDRs).

The present paper focused on WRKY transcription factors family. WRKY transcription factors are characterized by their highly conserved DNA-binding domain (DBD), called the WRKY domain.

The WRKY domain is characterised by the conserved amino acid sequence WRKYGQK at its N-terminal end, with a zinc-finger-like motif. The WRKY domain is found in one or two copies in a superfamily of plant transcription factors involved in the regulation of various physiological processes, including pathogen defence, response to salt stress, response to cold, response to wounding, salicylic response and the biosynthesis of secondary metabolites. The WRKY domain is required to DNA binding. WRKY TFs are structurally classified into three main groups (I, II and III), and also multiple subgroups (e.g. IIa, IIb and IIc, etc.) based on the number of WRKY domains [5].

Here, we aim to investigate the long disorder in the WRKY transcription factors in *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum*, *H.vulgare and A.thaliana*.

2. Methods

2.1 Data sources

The dataset was retrieved from the plant transcription factor database Plant TFDB 5.0 [6].

2.2 Protein disorder prediction

The prediction of intrinsic protein disorder was carried out by IUPred2A web server (http://iupred.enzim.hu/) [7]. This server disordered region from amino acid sequence based on pairwise energy content. There are three different prediction types: long disorder, short disorder, and structured domains. Here, we used long disorder prediction type in which at least 30 consecutive disordered residues are considered.

2.3 Protein binding region prediction

The binding regions prediction was used by ANCHOR tool (http://anchor.enzim.hu/) [8], which is based on the IUPred program mentioned above. Anchor predicts binding regions located in disordered proteins from the amino acid sequence.

The software IUPred and ANCHOR were requested from the authors and were compiled and executed locally.

2.4 Protein-protein interaction networks and topological analysis

The protein-protein interactions of WRKYs in *O.sativa, S.bicolor, Z.mays, T.aestivum, H.vulgare and A.thaliana* were obtained from STRING version 11.0 database [9].

The topological and statistical parameters of networks have been analysed using Cytoscape (version 3.5.1) [10].

2.5 Gene Ontology analysis

The gene ontology (GO) analyses were performed by agriGOv2.0 tools. It is specifically focused on gene ontology (GO) enrichment analyses of plant and agricultural species [11].

3. Results

3.1 Overall disorder content

The overall disorder contents of WRKYs in *O.sativa, S.bicolor, Z.mays, T.aestivum, H.vulgare* and *A.thaliana* are listed in Table1. These data clearly show similar mean content of disorder and of disordered binding regions between WRKYs analysed in this study. By grouping proteins according to the percentage of predicted disorder of their sequence, we found that the higher disorder content (>30%) of WRKY proteins is within the interval [30-60%] (Figure 1(A)) and (Supplemental file1).

Prediction of disordered binding regions (DBRs) showed that the average number of DBRs per protein was also similar among studied species (Table1). When proteins were grouped according to intervals of DBR residues content, we found higher number of WRKYs in the [5-10) interval excepting *T.aestivum* (Figure 1 (B)) and (Supplemental file 1).

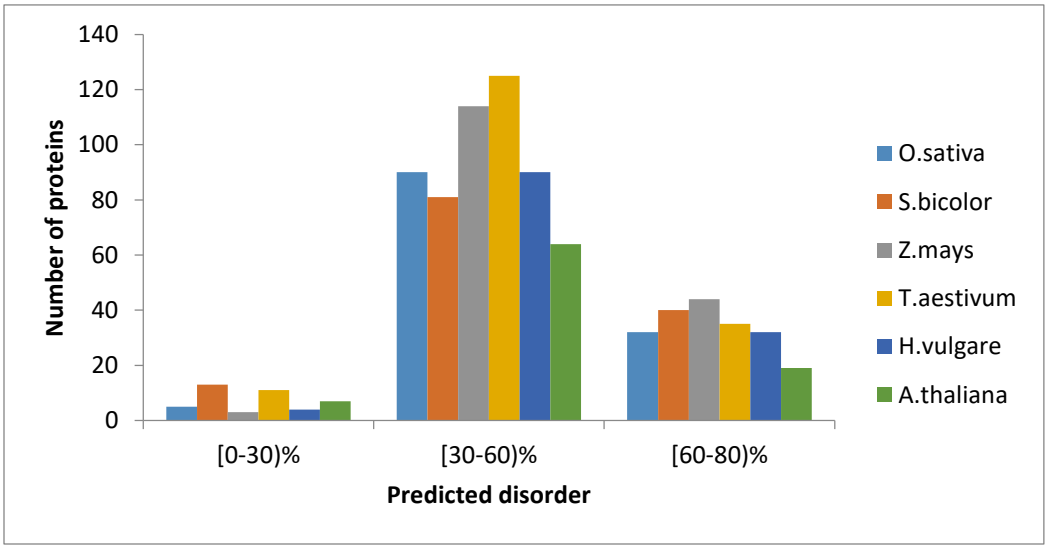
Table1: Summary of intrinsic disorder metrics for *O.sativa, S.bicolor, Z.mays, T.aestivum, H.vulgare* and *A.thaliana*.

Organism	*O.sativa*	*S.bicolor*	*Z.mays*	*T.aestivum*	*H.vulgare*	*A.thaliana*
Mean content of disorder	0.52	0.51	0.54	0.5	0.52	0.49
DBR*	11.2	11	10.5	9.46	9.44	10.4

*Disordered binding regions

We found that the mean of disorder is significantly correlated with the DBR content for *A.thaliana, O.sativa, S.bicolor,Z.mays, T. aestivum* and *H.vulgare* (*Pearson correlation, r = 0.73*, 0.6,0.57, 0.65, 0.65,0.43 respectively).

(A)



(B)

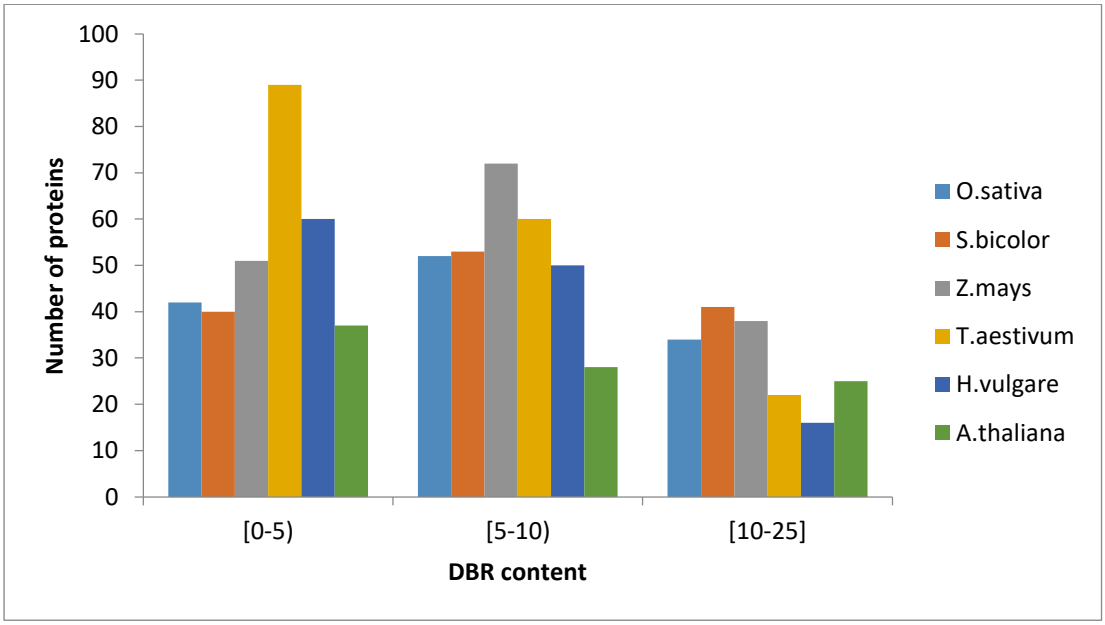


Figure 1: Fraction of proteins with different degrees of predicted disorder (A) and disordered binding regions (B) in *O.sativa, S.bicolor, Z.mays, T.aestivum, H.vulgare and A.thaliana*.

3.2 Location of DBRs within proteins

The locations of DBRs within proteins are shown in Figure 2 and (Supplemental file1). We found that DBRs are in N-terminal and C-terminal regions flanking the WRKY domain, and inter-WRKY domains. We also noted short disorder in the WRKY domain itself. This suggests that the disorder content increase with the number of WRKY domains and the pattern of the zinc-finger motif. In fact, members of Group 1 typically contain two WRKY

domains, while most proteins with one WRKY domain belong to Group 2. Group 3 proteins have a single WRKY domain with a zinc-finger motif.

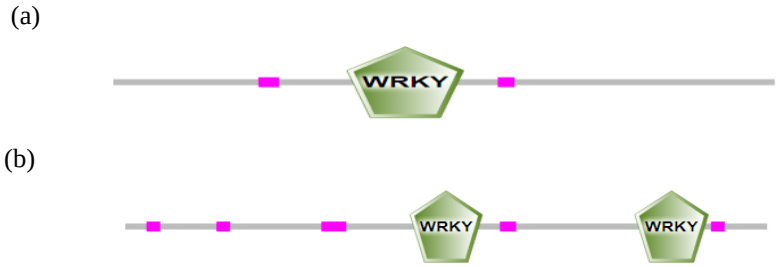


Figure 2: Schematic representation of disorder regions (in pink) flanking WRKY domain (in green) (a) Protein with one WRKY domain, (b) protein with two WRKY domains.

3.3 Gene ontology (GO) analysis

The enrichment analysis of WRKY proteins revealed similar GO categories among the studied species. The cellular components encompassed the nucleus, intracellular membrane-bounded organelle, organelle and cell. Their molecular functions are mainly related to sequence-specific DNA binding, transcription factor activity and binding. The biological processes mediated by the WRKYs included regulation of transcription, regulation of cellular metabolic process and biological process, regulation of defense response and gene expression. The Gene Ontology categories are listed by organism in supplemental file 2.

3.4 WRKY interaction networks

Based on genomic context, high-throughput experiments, co-expression and text mining, protein-protein networks related to WRKYs in *O.sativa* (Figure 3), *S.bicolor* (Figure 4), *Z.mays* (Figure 5), *H.vulgare* (Figure 6), *T.aestivum* (Figure 7) *and A.thaliana* (Figure 8) were constructed. We then analyzed the topological features of these networks. These parameters are listed in table 2 and in supplemental file 3.

Table2: Network parameters calculated for each species

Parameters	*O.sativa*	*Z.mays*	*S.bicolor*	*H.vulgare*	*T.aestivum*	*A.thaliana*
number of nodes	85	122	96	71	162	72
number of edges	93	93	66	36	543	87
Avg node degree	2.14	1.52	1.38	1.01	6.7	2.42
Avg.clust coeff	0.346	0.267	0.337	0.295	0.373	0.314

The highly ranked nodes, namely hubs, are identified by combining degree distribution and betweenness centrality measures. In all studied species, we have noticed that these hubs have high disorder content and disorder binding

regions (DBRs) allowing them high flexibility and interaction with multiple partners, of which participate in different functions.

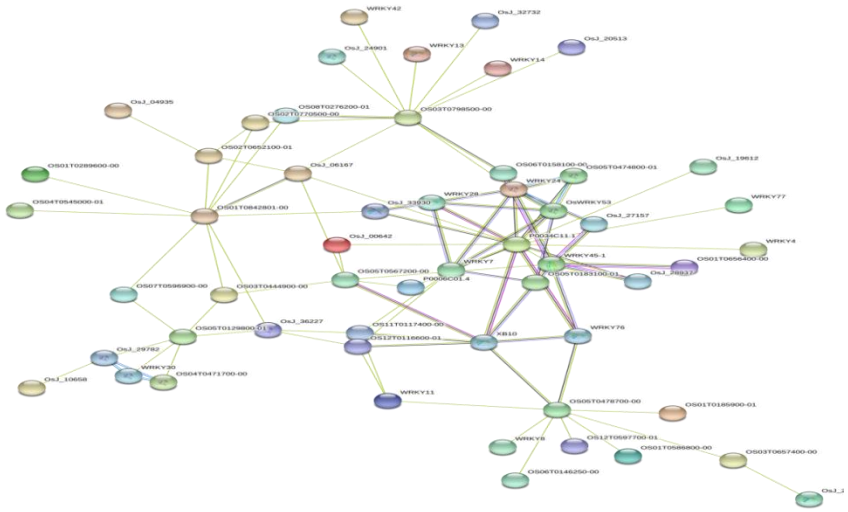


Figure 3: Protein-Protein interaction networks of WRKYs in *O.sativa*.

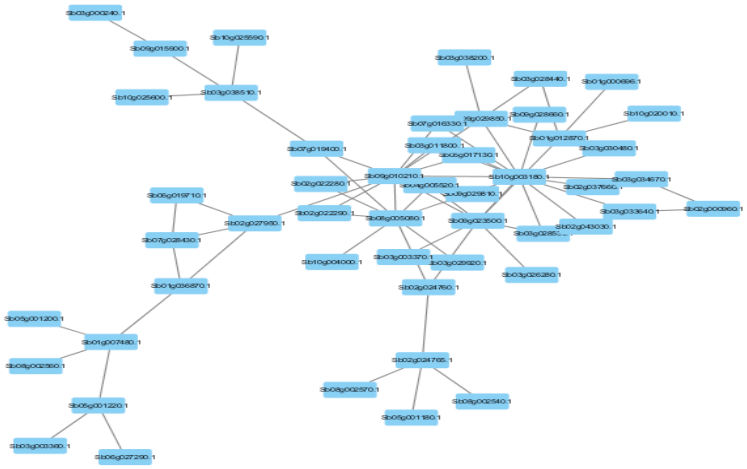


Figure 4: Protein-Protein interaction networks of WRKYs in *S.bicolor*.

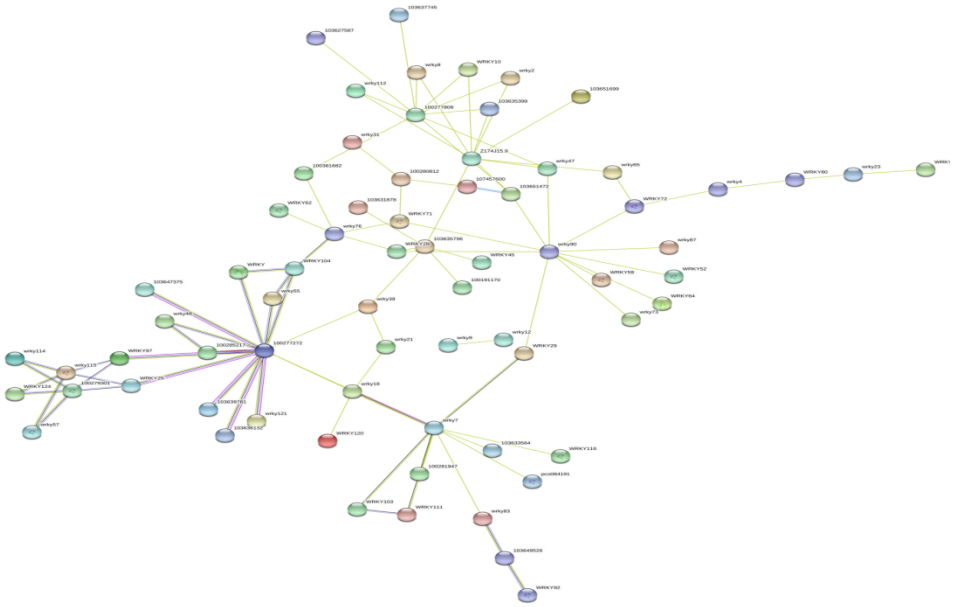


Figure 5: Protein-Protein interaction networks of WRKYs in *Z.mays*.

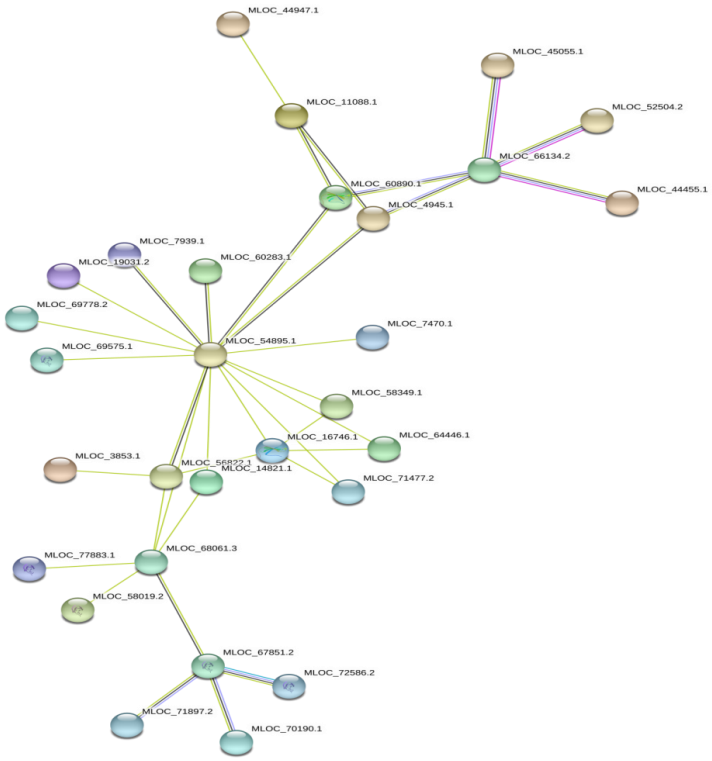


Figure 6: Protein-Protein interaction networks of WRKYs in *H.vulgare*.

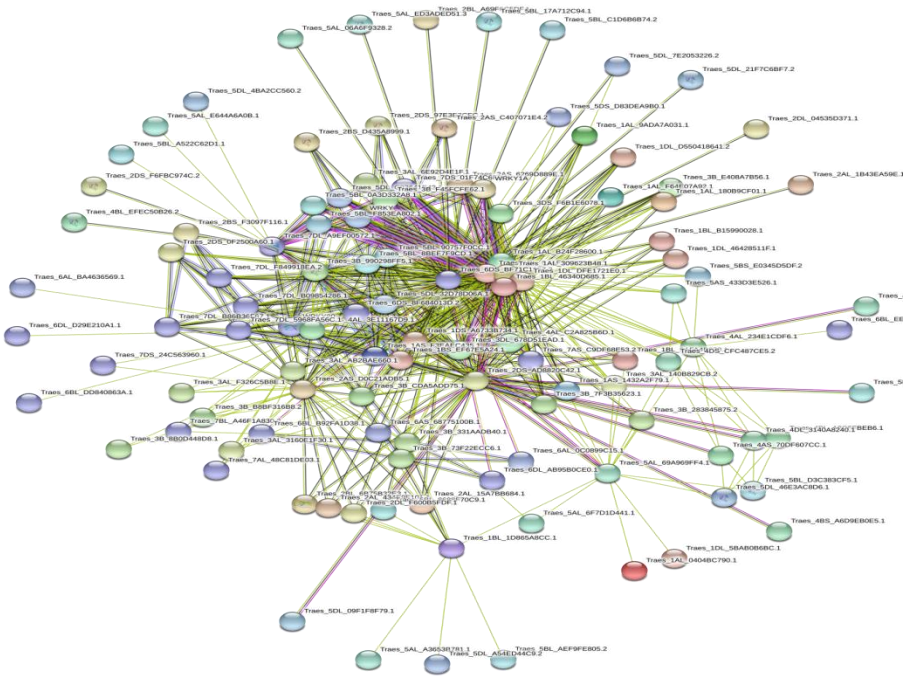


Figure 7: Protein-Protein interaction networks of WRKYs in *T.aestivum*.

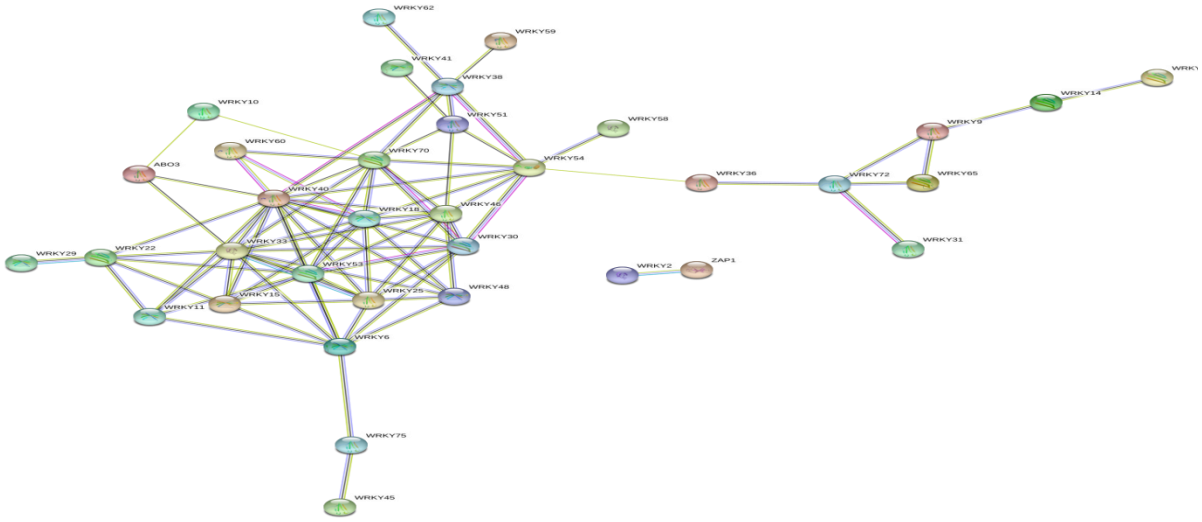


Figure 8: Protein-Protein interaction networks of WRKYs in *A.thaliana*.

3.5 Evolutionary conservation of intrinsic disorder of WRKYs in cereals

Comparative analysis of WRKY DNA-binding domain and flanking regions by sequence alignment of some WRKYs in *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum*, *H.vulgare* and *A.thaliana* showed some interesting insights (Figure 9).

We noticed that in spite of the strong evolutionary conservation of the WRKY DNA-binding domain, the flanking sequences of proteins are highly divergent, which might reflect their different functions.

```
Q9LY00      WRK70_ARATH       70  EGSQNAS-------------C--DNDGKFEDSGDSRKRLGPVKGKR----GCYKRKKRS-      109
A0A3B5Y5P3  A0A3B5Y5P3_WHEAT   3  SGAAPPPPPESGAGSS---------SGVGREETKGK--G----SARGR-G---SRKASRP       43
Q6QHD1      WRK71_ORYSJ      125  DSPPPPPPPPHPHAAPHHMHVMPGAAAAGYADQTECTS-GEPCKRIR---EEC-KPKISKL      179
C5XCU5      C5XCU5_SORBI     114  ----PPSNS--SDGNANAKAEPGD--------------HAAVESALSDEGTCRRIKVTRV      153
B2KJ87      B2KJ87_HORVV     150  TAGRNPSPPLAAAATGGFAISV--NVGPGRDQAECTSVHEPCNSKRVRADECKASRISKL      207
Q32SG4      WRKY1_MAIZE      270  SGSQTASTPELGL--VQRRRCAGREDGTG----RCA-TGSRC--------HCSKKRKLRI      314
                                                                                          :

Q9LY00      WRK70_ARATH      110  -------ETCTTESTILEDAFSWRKYGQKEILNAKFPRSYFRCTHKYTQGCKATKQVQKV      162
A0A3B5Y5P3  A0A3B5Y5P3_WHEAT  44  ----RFAFQTKSENDVLDDGYRWRKYGQKAVKNSAFPRSYYRCTH---HTCNVKKQVQRL       96
Q6QHD1      WRK71_ORYSJ      180  ----YVHADPSDLSLVVKDGYQWRKYGQKVTKDNPCPRAYFRCSF--APACPVKKKVQRS      233
C5XCU5      C5XCU5_SORBI     154  ----CTRIDPADATLTVKDGYQWRKYGQKVTRDNPSPRAYFRCAY--APSCPVKKKVQRS      207
B2KJ87      B2KJ87_HORVV     208  ----YVHADPSDLSLVVKDGYQWRKYGQKVTKDNPCPRAYFRCSF--APSCQVKKKVQRS      261
Q32SG4      WRKY1_MAIZE      315  RRSIKVPAISNKVADIPADEFSWRKYGQKPIKGSPHPRGYYKCSS--VRGCPARKHVERC      372
                                              * : *******   .   **.*::*:     * . *:*::

Q9LY00      WRK70_ARATH      163  ELEPKMFSITYIGNHTCNTNAETPKSKTCDHHDEIFMDSEDHKSPSLSTSMKEEDNPHRH      222
A0A3B5Y5P3  A0A3B5Y5P3_WHEAT  97  AKDTSIVVTTYEGVHNHPCEKLMEALN---------------------------PILRQLQ      130
Q6QHD1      WRK71_ORYSJ      234  AEDNTILVATYEGEHNHGQPPPPLQSAAQNSDGSGKSAGKPPHAPAAAPP--APVVPHRQ      291
C5XCU5      C5XCU5_SORBI     208  AEDSSLLVATYEGEHNHPSPTRAGELP---SSASATASGPVPCSISINSS--GPTITLDL      262
B2KJ87      B2KJ87_HORVV     262  AEDKTVLVATYDGDHNHAPPPKHQGSGGRKS-------G-DAAPVRVSPP--APVLVQQQ      311
Q32SG4      WRKY1_MAIZE      373  VDDPSMLIVTYEGDHNHNRVLAQPA-----------------------------------      397
                                    : .:.  ** * *.

Q9LY00      WRK70_ARATH      223  HGSSTENDLSLVWPEMVFEEDYHHQASYVNGKTSTSIDVLGSQDLMVFGGGGDFEFSENE      282
A0A3B5Y5P3  A0A3B5Y5P3_WHEAT 131  FLSQL---------------------------------------------------------      135
Q6QHD1      WRK71_ORYSJ      292  HEPVVVNGE---------------QQAAAAS--EMIRRNLAEQMAMTLTR--DPSFKAAL      332
C5XCU5      C5XCU5_SORBI     263  TKNGGGGGVRVLDAA--EAPDLKKLCQEIAS--PDFRTALVEQMARSLTS--DSKFTHAL      316
B2KJ87      B2KJ87_HORVV     312  REQEAS---------------------TADQ--VADRKKLVEQMAATLTR--DPGFKAAL      346
Q32SG4      WRKY1_MAIZE      398  ---------------------------------------------------------      397
```

Figure 9: Multiple alignment of WRKY domain and flanking sequences of some members from *A.thaliana*, *T.aestivum,O.sativa*, *S.bicolor*, *H.vulgare* and *Z.mays* performed by clustal O [19]. Domain and DNA-binding sequences are highlighted in yellow and purple respectively.

To understand whether the intrinsic disorder is evolutionary conserved among studied species, we evaluated the evolutionary relationships between few AtWRKYs (from Group I, Group II and Group III) and their orthologs in *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum*, *H.vulgare* (Figure10).

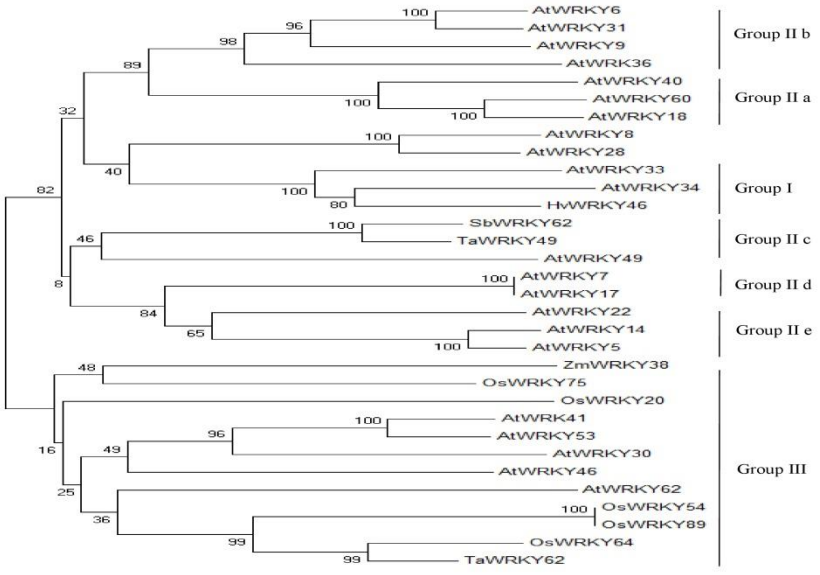


Figure 10: The phylogenetic tree of some WRKY proteins in the studied species.

Ta, Triticum aestivum: At, Arabidopsis thaliana: Hv, Hordeum vulgare, Os, Oryza sativa, Zm, Zea mays, Sb, Sorghum bicolor. The phylogenetic tree was constructed with MEGA5.2 [20] using a bootstrap test of phylogeny with a minimum evolution test and a parameter of 1000 replications.

The results indicated that WRKY family members within the same group, having similar functions, have also similar disorder content and DBR numbers among the species (Table 3).

Table3: Some WRKYs used in the phylogenetic analysis.

Sequence ID	Description	Group	Disorder(%)	DBR	Position*
AT5G43290.1	AtWRKY49	IIc	0.46	1	114-170
Traes_3AL_3160E1F30.1	TaWRKY49	IIc	0.49	3	152-207
MLOC_74884.1	HvWRKY46	I	0.73	9	189-244 362-419
AT2G38470.1	AtWRKY33	I	0.65	16	184-239 361-418
AT2G24570.1	AtWRKY17	IId	0.49	8	242-300
AT4G24240.1	AtWRKY7	IId	0.59	9	280-338
AT1G30650.1	AtWRKY14	IIe	0.54	10	216-247
AT4G01250.1	AtWRKY22	IIe	0.49	7	128-186
LOC_Os08g29660.1	OsWRKY75	III	0.47	5	129-191
AT4G23810.1	AtWRKY53	III	0.43	3	157-219
Traes_5AL_69A969FF4.1	TaWRKY62	III	0.46	4	81-147
GRMZM2G005207_P01	ZmWRKY38	III	0.47	6	110-176

*WRKY domain position

4. Discussion

This study has aimed to explore the protein disorder in the WRKYs in *O.sativa, S.bicolor, Z.mays, T.aestivum, H.vulgare and A.thaliana*. We found that these proteins have similar disorder content, DBR numbers and biological processes among these species. Similar conclusions were reached by performing the disorder analysis on the complete proteomes of these species [12,13]. The results have also shown that the WRKY domain has regions of conserved disorder as well as the flanking disordered sequences. This indicates that disorder tendencies are kept in these proteins, suggesting that their function depends on disorder such as in protein-protein/DNA interaction. By examining the protein-protein interaction networks of the WRKYs in these species, we found that these networks are highly connected by hubs, having a higher disorder content, particularly important for flexible binding with many interaction partners allowing thus multiple functions. Hereafter, some examples are discussed.

In agreement with previous report, we found that *AtWRKY33* acts as a one of the hubs in the protein-protein interaction in the WRKYs of *A.thaliana* [14]. *AtWRKY33* (disorder content=0.65). Moreover, *AtWRKY33* is involved in the regulation of the defense pathways mediating responses to P. syringae and necrotrophic fungal pathogens [15]. It is also involved in response to salt stress and abscisic acid (ABA) signalling [16].

Previous study suggests that *OsWRKY6* (disorder content=0.5) positively regulates defense responses through activation of OsICS1 expression and *OsWRKY6* stabilization [17]. Furthermore, *HvWRKY46* (disorder content=0.68), denoted also SUSIBA2 has been demonstrated as a regulatory transcription factor in starch synthesis and involved in carbohydrate anabolism [18].

5. Conclusion

In this study, we have provided a comparative analysis of the protein disorder of WRKY transcription factors in *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum*, *H. vulgare*, and *A. thaliana*. Interestingly, we found similar and conserved disorder features such as the disorder content, the disorder binding regions (DBR) numbers and the Gene ontology including biological processes, molecular function and cellular components. The data provided here points towards the important role of the disorder of WRKYs in different biological processes notably in stress tolerance, deserving hence further investigation and experimental validation.

Acknowledgements

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

Conflicts of interest

The authors declare that they have no conflicts of interests.

References

1. Ward JJ, Sodhi JS, McGuffin LJ, et al. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of Molecular Biology 337 (2004): 635-645.
2. Dunker AK, Cortese MS, Romero P, et al. Flexible nets: the roles of intrinsic disorder in protein interaction networks. The FEBS Journal 272 (2005): 5129-548.
3. Liu J, Perumal NB, Oldfield CJ, et al. Intrinsic disorder in transcription factors. Biochemistry 45 (2006): 6873-6888.
4. Howton TC, Zhan YA, Sun Y, et al. Intrinsically disordered proteins: controlled chaos or random walk. International Journal of Plant Biology 6 (2015).
5. Eulgem T, Rushton PJ, Robatzek S, et al. The WRKY superfamily of plant transcription factors. Trends in Plant Science 5 (2000): 199-206.
6. Jin J, Tian F, Yang DC, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research 45 (2017): D1040.

7. Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Research 46 (2018): W329-W337.
8. Dosztányi Z, Mészáros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25 (2009): 2745-2746.
9. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45 (2017): D362–D368.
10. Su G, Morris JH, Demchak B, et al. Biological network exploration with Cytoscape 3. Current Protocols in Bioinformatics 47 (2014): 8-13.
11. Tian T, Liu Y, Yan H, et al. agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Research 45 (2017): W122-W129.
12. Choura M, Ebel C, Hanin M. Genomic analysis of intrinsically disordered proteins in cereals: From mining to meaning. Gene 714 (2019): 143984.
13. Choura M, Rebaï A, Hanin M. Proteome-wide analysis of protein disorder in Triticum aestivum and Hordeum vulgare. Computational Biology and Chemistry 84 (2020): 107138.
14. Choura M. Unraveling the WRKY transcription factors network in Arabidopsis thaliana by integrative approach. Network Biology 5 (2015): 55.
15. Birkenbihl RP, Diezel C, Somssich IE. Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. Plant Physiology 159 (2012): 266-285.
16. Jiang Y, Deyholos MK. Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. Plant Molecular Biology 69 (2009): 91-105.
17. Choi C, Hwang SH, Fang IR, et al. Molecular characterization of Oryza sativa WRKY 6, which binds to W-box-like element 1 of the Oryza sativa pathogenesis-related (PR) 10a promoter and confers reduced susceptibility to pathogens. New Phytologist 208 (2015): 846-859.
18. Sun C, Palmqvist S, Olsson H, et al. A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. The Plant Cell 15 (2003): 2076-2092.
19. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7 (2011): 539.
20. Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28 (2011): 2731-2739.

Supplementary Data

Supplemental file 1 Intrinsic disordered proteins in of WRKYs in *A.thaliana*, *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum* and *H.vulgare* predicted by IUPred. It contains the disorder means, the DBRs their positions and their lengths.

Supplemental file 2 Gene Ontology (GO) analysis of WRKYs in *A.thaliana*, *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum* and *H.vulgare* including biological process, molecular functions and cellular components.

Supplemental file 3 Topological features of WRKYs in *A.thaliana*, *O.sativa*, *S.bicolor*, *Z.mays*, *T.aestivum* and *H.vulgare*.