

Research Article

RNA secondary structure analysis for abiotic stress resistant and housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa*

Syed Qasim Shah*, Muhammad Younas Khan Barozai, Muhammad Din, Iftekhar Ahmed Baloch and Humaira Abdul Wahid

Department of Botany, University of Balochistan, Quetta, Pakistan

*Corresponding author's email: Syed_qshah@yahoo.com

Citation

Syed Qasim Shah, Muhammad Younas Khan Barozai, Muhammad Din, Iftekhar Ahmed Baloch and Humaira Abdul Wahid. RNA secondary structure analysis for abiotic stress resistant and housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa*. Pure and Applied Biology. Vol. 5, Issue 3, pp476-482.

<http://dx.doi.org/10.19045/bspab.2016.50062>

Received: 25/04/2016

Revised: 20/05/2016

Accepted: 26/05/2016

Online First: 28/05/2016

Abstract

RNA is a single stranded nucleic acid type, produced by DNA via a process of transcription. RNA tend to form secondary structure by pairing the complementary base pairs within it. The RNA secondary structure is an essential parameter to understand its functions and expression into a protein. This research is based to understand the link between RNA secondary structure and abiotic stresses in two leading organisms viz; *Arabidopsis thaliana* and *Oryza sativa*. For this purpose the RNA secondary structures analyses for abiotic stress resistant and housekeeping genes were done in *A. thaliana* and *O. sativa*. Total 138 abiotic stress resistant and 21 housekeeping genes were selected through literature survey and subjected to RNA secondary structure tool. RNA secondary structure were analyzed and characterized in terms of minimum free energy, number of stems and loops. The 80% of abiotic stress and 67% of housekeeping genes show significant difference in the minimum free energy (MFE) while there were no significant variation in the number of stems and loops. The open reading frame analysis, as an additional parameter, of the selected genes were also done, that revealed the preference of +1 frame by most of the genes, i.e., 42% of abiotic stress and 43% of housekeeping genes. RNA secondary structure analysis help us to engineer stress resistant genes with stable RNA structures.

Key words: Abiotic stresses; Minimum free energy; Open reading frame

Introduction

The abiotic stresses such as cold, drought, osmotic, heat, salt and oxidative stress affects and reduce the growth and productivity yield of the plants. Plants also give response to these stress at the level of cell, molecule and at biochemical stage, that enable them to survive. At molecular level, stresses triggers the change in the expression of various genes in plants to

combat and cope from it [1]. *A. thaliana* because of its simple structure is modal plant in molecular genetics, therefore a wide study of stress tolerant genes is conducted [2]. About 45% of all stress tolerant reported genes are from *A. thaliana*, while nearly 50% of all stress tolerated genes have been characterized using *A. thaliana* as the transgenic species [3]. Seki *et al.* [2], reported 7000 gene

responding to cold, drought and salinity in *A. thaliana*. Stress tolerant genes are also studied by Vandembroucke [3], in *A. thaliana* and reported the role of different families of transcription factors (TFs) in abiotic stresses. Therefore, a total of 138 well characterized and annotated genes from *A. thaliana* were studied in this research. Similarly *O. sativa* is the most studied monocot plant at molecular level whose genome is sequenced in 2000 [4]. Hence, the combination of *A. thaliana* and *O. sativa* as a research organism would be great to elucidate the molecular mechanisms in plants.

The gene expression and function has a significant relationship with its transcript secondary structure [5]. The secondary structure of a nucleic acid molecule refers to the base pairing interactions within a single molecule or set of interacting molecules and can be represented as a list of bases which pair up in a nucleic acid molecule. The secondary structures of biological DNA's and RNA's tend to be different: biological DNA mostly exists as fully base paired double helices, while biological RNA is single stranded and often forms complicated base-pairing interactions due to its increased ability to form hydrogen bonds stemming from the extra hydroxyl group in the ribose sugar [6]. Secondary structure of RNAs for biotic stress resistant and housekeeping genes in *A. thaliana* and *O. sativa* is reported in terms of minimum free energy, number of loop and stem regions [5].

The current study is based on RNA secondary structure analyses of abiotic stress and housekeeping genes in *A. thaliana* and *O. sativa*. Main theme of this study is to find the relationship among the RNAs secondary structures of all abiotic stress and housekeeping genes. This would help us to better understand and manage the plant for abiotic stresses at molecular level.

Materials and methods

Identification of abiotic stress resistant genes & housekeeping genes

Total 138 well annotated and characterized abiotic stress and 21 housekeeping genes were identified through literature survey in *A. thaliana* [2 & 3]. These 159 genes were used as an initial data for downstream analyses.

Retrieval of sequences

The sequences of 159 abiotic stress and housekeeping genes for *A. thaliana* were retrieved from the Gene bank (NCBI). The FASTA format of these sequences was saved and used as query to find their homologue in *O. sativa*. For this purpose all the *A. thaliana* retrieved FASTA sequences were blast against *O. sativa* via BLASTn [7]. The *O. sativa* sequences homologue to *A. thaliana* were also retrieved and saved.

RNA secondary structure prediction and analyses

The RNA secondary structures prediction and analyses were done for all 159 abiotic stress resistant and housekeeping genes transcripts of *A. thaliana* and *O. sativa*. For this purpose, the transcripts sequences in FASTA format were submitted to Zuker folding algorithm, MFOLD (version2.3)[8], publicly available at <http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi>. to produce the folded structure. The Mfold parameters were set similar as adopted earlier [9]. All the secondary structures of the abiotic and housekeeping genes were selected on the basis of minimum free energy (MFE) and saved. These secondary structures were also subjected to manual analyses for number of loops and stems.

Open reading frame (ORF) analysis

All the selected 159 abiotic stress and housekeeping genes were also studied and characterized for the open reading frame (ORF). This would help us to better understand the link between transcripts and their translated proteins. For this purpose, the 159 abiotic stress resistant and housekeeping genes were analyzed through, available at NCBI (www.ncbi.nlm.nih.gov/gorf/gorf.html) [10]. The longest frame produced by ORF

finder were selected and subjected to physical analysis.

Results and discussion

Characterization and analyses of RNA secondary structures

The RNA secondary structures of abiotic stress resistant and housekeeping genes in *A. thaliana* and *O. sativa* were analysed and characterized by Mfold algorithm. Some of these secondary structures are shown in Figure-1. The RNA secondary structures of 159 abiotic stress and housekeeping genes belonging to *A. thaliana* and *O. sativa* were characterized in terms of the number of loops, stems and minimum free energy (MFE).

MFE is an important parameter to find the stability of the RNA secondary structure. The MFE of abiotic stress and housekeeping genes are evaluated as shown in (Table 1). The range of difference in MFE of abiotic stress resistant genes in *A. thaliana* and *O. sativa* are $-0.7 \text{ KcalMole}^{-1}$ to $-287 \text{ KcalMole}^{-1}$ while in case of housekeeping genes are $-8.6 \text{ KcalMole}^{-1}$ to

$-128 \text{ KcalMole}^{-1}$. The average value of difference in MFE of abiotic stress resistant and housekeeping genes in *A. thaliana* and *O. sativa* are $-144 \text{ KcalMole}^{-1}$ and $-68 \text{ KcalMole}^{-1}$ respectively. The 110 abiotic stress resistant and 14 housekeeping genes showed the values greater than their average values respectively, while 28 and 6 genes showed values less than their respective average values correspondingly. All the 159 genes are classified into four different categories on the basis of their MFEs. These are category-I; -1 to -72, II; -73 to -144, III; -145 to -216 and IV; -217 to $-288 \text{ KcalMole}^{-1}$. The maximum number of abiotic stress resistant and housekeeping genes are lies in the category-I, i.e. 60 and 15 respectively followed by category-II (50 and 6 respectively), category-III (16 and 0 respectively) and category-IV (04 and 0 respectively). Similar findings are also reported earlier for the characterization of secondary structures of viral biotic stress and housekeeping genes [5].



Figure 1. Secondary structures of few genes

Table 1. Characterization of abiotic stress resistant and housekeeping genes in *A. thaliana* and *O. sativa* in terms of differences in minimum free energy (MFE) of RNA secondary structures

Categories of differences in MFE Kcalmol ⁻¹	No. of Abiotic Stress Genes	Type of Abiotic Stress Resistant Genes	No. of Housekeeping Genes	Type of Housekeeping Genes
(-1) - (-72)	60	MBF1c, MYB60, CaMBP25, SRK2C, TPS1, RZ-1a, SRL1, SOS1, PCS1, GI-3, GPX3, SOD/MSD1, VTC1, VTC2, UVH6, PLDdelta, DHN (RAB18), DJA2, ProDH, DHS, FTA, SDIR1, ABI1, MKK9, NDPK2, RGS1, PP7, AHK2, AHK3, TOR, CIPK23, CPK23, EIN2/ORE3, CTR1, GLI1, ABF2/AREB1, ABF3, ANAC002, ANAC055/NAC3, HD2C, HSF1, MYB2, MYC2, SZF1, SHN1, ZAT12, DREB2C, ALS3, MRP5, MTP11, PDR12, PDR8, PIP1;4, ZIF1, SBP1, ESK1, CHYB, TMAC2, PHYA (ars4ars5), RCD1	15	PP2AA3 (AT1G13320), helicase (AT1G58050), PPR gene (AT1G62930), SAND family (AT2G28390), polypyrimidine track-binding protein (AT3G01150), expressed protein (AT4G26410), UBC9 (AT4G27960), expressed protein (AT4G33380), TIP41-like (AT4G34270), expressed protein (AT5G12240), F-box protein (AT5G15710), UBQ10 (AT4G05320), ACT (AT3G18780), TUB (AT5G62690), EF-1 α (AT1G07920),
(-73) - (-144)	50	GOLS2, AVP1, NHX1, tAPX, HSP17.6a, NCED3/STO1, GLB1, BCB, HSFA2, RHL41/ZAT12, AAO, ALDH7B4, APX1, FRO2, SOD/CSD2, PARP2, PLDalpha1, CYP707A3, ERD15, F9E10.5, DJA3, LOS4/CRYOPHITE, STRS1, GRP2, PMSR4, HAB1, GSK1, AHK1/ATHK1, PP2CA, RAB7, CBL9, ABF4, ABO1/ELO2, ANAC019, ANAC072/RD26, CBF1 / DREB1B, CBF3 / DREB1A, CBF4/DREB1D, CPL1/FRY2, HSF3, MBF1a, STZ/ZAT10, BOR1, PIP2;5, ALX8, ORE1, TRG1, SPS, HAL3A, PCR1,	06	expressed protein (AT2G32170), ubiquitin transferase (AT3G53090), mitosis protein YLS8 (AT5G08290), clathrin adaptor complex subunit (AT5G46630), GAPDH (AT1G13340), UFO (AT1G30950),
(-145) - (-216)	24	AOX1a, RCI3, HSP101, DREB2A, CBL1, ALDH3I3, LTL1, ABA2/GIN, COR15a, DHN (COR47), STRS2, ORE9/M, AX2, MKK2, ABI3, CBF2 / DREB1C, SZF2, XERICO, HRD/HARDY, NF-YB1, MTP3, SULTR1;2, GPP2, CESA8/IRX1/LEW2,	00	None
(-217) - (-288)	04	ICE1, LEA5, ZAT7, ATM3,	00	None

Table 2. ORF- Frame preferred by abiotic stress resistant and housekeeping genes of *A. thaliana*

Frame Preferred	No. of Abiotic Stress Genes	Type of Abiotic Stress Resistant Genes	No. of Housekeeping Genes	Type of Housekeeping Genes
+1	58	ICE1, AOX1a, RCI3, HSP101, SOS1, BCB, HSFA2, CBL1, PCS1, ALDH3I3, ALDH7B4, FRO2, GI-3, PARP2, LTL1, COR15a, DHN (COR47), F9E10.5, DJA2, ProDH, GRP2, SDIR1, ORE9/MAX2, PMSR4, HAB1, AHK1/ATHK1, MKK2, PP2CA, RGS1, AHK2, AHK3, CIPK23, CBL9, CTR1, GLI1, ABF4, ABO1/ELO2, ANAC002, ANAC055/NAC3, CBF4/DREB1D, CPL1/FRY2, HSF1, MBF1a, MYC2, SZF2, XERICO, HRD/HARDY, NF-YB1, ATM3, MTP3, PIP1;4, PIP2;5, SULTR1;2, ORE1, TRG1, CHYB, CESA8/IRX1/LEW2, PCR1,	09	Helicase (AT1G58050), PPR (AT1G62930), SAND family (AT2G28390), expressed protein (AT2G32170), UBC9 (AT4G27960), TIP41-like (AT4G34270), UBQ10 (AT4G05320), GAPDH (AT1G13340), UFO (AT1G30950)
+2	42	MBF1c, CaMBP25, RZ-1a, LEA5, GLB1, APX1, GPX3, SOD/MSD1, SOD/CSD2, PLDalfa1, PLDdelta, ABA2/GIN, CYP707A3, DHN (LTI29/ERD10+LTI30), DHS, STRS1, GSK1, MKK9, RAB7, PP7, TOR, CPK23, EIN2/ORE3, ABF2/AREB1, ABF3, ABI3, ANAC072/RD26, CBF1 /DREB1B, CBF3 /DREB1A, HD2C, MYB2, STZ/ ZAT10, ZAT7, DREB2C, ALS3, PDR12, SBP1, ALX8, HAL3A, PHYA (ars4ars5), RCD1,	06	Expressed protein (AT4G26410), Expressed protein (AT4G33380), mitosis protein YLS8 (AT5G08290), clathrin adaptor complex subunit (AT5G46630), ACT (AT3G18780), TUB (AT5G62690),
+3	38	MYB60, SRK2C, GOLS2, TPS1, AVP1, NHX1, tAPX, HSP17.6a, NCED3/STO1, SRL1, DREB2A, RHL41/ZAT12, AAO, VTC1, VTC2, UVH6, DHN (RAB18), ERD15, DJA3, LOS4/CRYOPHITE, STRS2, ABII, NDPK2, ANAC019, CBF2 /DREB1C, HSF3, SZF1, SHN1, ZAT12, BOR1, MRP5, MTP11, PDR8, ZIF1, ESK1, GPP2, SPS, TMAC2,	06	PP2AA3 (AT1G13320), polypyrimidine track-binding protein (AT3G01150), ubiquitin transferase (AT3G53090), Expressed protein (AT5G12240), F-box protein (AT5G15710), EF-1 α (AT1G07920),

Table 3. ORF- Frame preferred by abiotic stress resistant and housekeeping genes of *O. sativa*

Frame Preferred	No. of Abiotic Stress Genes	Type of Abiotic Stress Resistant Genes	No. of Housekeeping Genes	Type of Housekeeping Genes
+1	63	MBF1c, ICE1, SRK2C, AVP1, NHX1, tAPX, RCI3, HSP101, RZ-1a, SOS1, BCB, DREB2A, RHL41/ZAT12, CBL1, APX1, GI-3, PARP2, PLDdelta, LTL1, ABA2/GIN, CYP707A3, DHN (LTI29/ERD10+LTI30), DHN (RAB18), DHN (COR47), ERD15, DJA2, DJA3, ProDH, GRP2, FTA, PMSR4, HAB1, MKK2, RGS1, CIPK23, CBL9, CTR1, GLI1, ABF2/AREB1, ABF3, ABO1/ELO2, CBF2 /DREB1C, CBF3 /DREB1A, CBF4/DREB1D, HSF1, HSF3, MBF1a, MYC2, SZF1, ZAT7, HRD/HARDY, ALS3, ATM3, BOR1, PDR8, SULTR1:2, ZIF1, SBP1, ALX8, HAL3A, PHYA (ars4ars5), RCD1,	05	ubiquitin transferase (AT3G53090), Expressed protein (AT4G26410), UBC9 (AT4G27960), TIP41-like (AT4G34270), TUB (AT5G62690),
+2	35	MYB60, GOLS2, TPS1, SRL1, GLB1, HSFA2, ALDH3I3, ALDH7B4, FRO2, VTC2, UVH6, PLDalfa1, COR15a, F9E10.5, STRS2, ORE9/MAX2, MKK9, PP7, AHK2, CPK23, EIN2/ORE3, ANAC019, ANAC055/NAC3, MYB2, SZF2, XERICO, ZAT12, NF-YB1, MRP5, MTP11, PIP2:5, GPP2, ORE1, TRG1, PCR1,	06	PP2AA3 (AT1G13320), Helicase (AT1G58050), SAND family (AT2G28390), polypyrimidine track-binding protein (AT3G01150), mitosis protein YLS8 (AT5G08290), UBQ10 (AT4G05320),
+3	31	CaMBP25, NCED3/STO1, PCS1, SOD/MSD1, VTC1, SOD/CSD2, DHS, LOS4/CRYOPHITE, SDIR1, AB11, GSK1, AHK1/ATHK1, NDPK2, PP2CA, RAB7, AHK3, ABF4, ABI3, ANAC002, ANAC072/RD26, CBF1 /DREB1B, CPL1/FRY2, HD2C, STZ/ ZAT10, DREB2C, MTP3, PIP1:4, SPS, CHYB, CESA8/IRX1/LEW2, TMAC2,	06	PPR (AT1G62930), expressed protein (AT2G32170), Expressed protein (AT5G12240), clathrin adaptor complex subunit (AT5G46630), ACT (AT3G18780), EF-1 α (AT1G07920),
-1	03	AOX1a, GPX3, STRS1,	01	Expressed protein (AT4G33380),
-2	01	LEA5,	03	F-box protein (AT5G15710), GAPDH (AT1G13340), UFO (AT1G30950),
-3	05	HSP17.6a, AAO, SHN1, PDR12, ESK1,	00	None

Open reading frame (ORF) analysis

The open reading frame is another additional parameter analysed for 159 abiotic stress and housekeeping genes. The 58 out of 138 abiotic stress genes preferred +1 frame, 42 genes preferred +2 frame and 38 genes preferred +3 frame in *A. thaliana* as shown in table 2, while in *O. sativa* 63, 35 and 31 genes are observed to prefer +1, +2 and +3 frame respectively. Three genes in *O. sativa* preferred -1 frame, one gene -2 and 5 genes preferred -3 frame as shown in table 3, whereas, no single *A. thaliana* gene is observed in -1, -2 and -3 frames. This indicates a little variation in the ORF preference between *A. thaliana* and *O. sativa*, and could serve a point to analyse in detail at individual gene level. The role of ORF in the expression of gene is reported in various organisms [11].

Housekeeping genes also showed the remarkable results i.e., 9 genes out of 21 preferred +1 frame, 6 genes preferred +2 and +3 frame for each in *A. thaliana*. Five genes preferred +1 frame, 6 genes preferred +2 and +3 frame by each in *O. sativa*. Four genes in *O. sativa* also preferred -1 and -2 frame.

Conclusion

This research is based on the relationship of secondary structure folding for abiotic stress resistant and housekeeping genes in *A. thaliana* and *O. sativa*. The concluded results showed that there are variation on the basis of RNA secondary structures and ORFs for abiotic stress resistant and housekeeping genes in *A. thaliana* and *O. sativa*. These findings would be helpful to understand and manage the dicot and monocot plants under abiotic stresses.

Authors' contributions

Conceived and designed the experiments: MYK Barozai, Performed the experiments: SQ Shah, Analyzed the data: M Din, IA Baloch & HA Wahid, Wrote the paper: MYK Barozai & SQ Shah.

References

1. Bartels D & Sunkar R (2005). Drought and salt tolerance in plants. *Crit Rev Plant Sci* 24: 23–58.
2. Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A & Sakurai T (2002). Monitoring the expression profiles of ca 7000 *Arabidopsis* genes under drought, cold and high salinity stresses using a full-length cDNA microarray. *Plant J* 31: 279–292.
3. Vandembroucke (2008). Role for Hydrogen Peroxide during Abiotic and Biotic Stress Signaling in Plants. *Thesis submitted in partial fulfillment of the requirements for Ph.D., Biotechnology Academic year 2007-2008*.
4. Eckardt NA (2000). Sequencing the rice genome. *The Plant Cell* 12(11): 2011-2017.
5. Barozai MYK, Shah SQ, Din M & Muhammad R (2014). Codon Usage Bias and RNA Secondary Structures Analysis for Virus Resistant Genes in *A. thaliana* and *O. sativa*. *Pure Appl. Bio* 3(2): 81-91.
6. Krieger M, Scott Matsudaira MP, Lodish PT, Darnell HF, Lawrence JE, Kaiser Z & Berk C.A (2004). *Structure of Nucleic Acids*. Molecular cell biology. New York: W.H. Freeman & CO. ISBN 0-7167-4366-3. <http://www.ncbi.nlm.nih.gov/books/NBK21514/>.
7. Altschul, SF, Gish W, Miller W, Myers EW & Lipman DJ (1990). Basic local alignment search tool. *J of Mol Bio* 215: 403-410.
8. Zuker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res* 31: 3406-3415.
9. Barozai MYK (2012). In-silico identification of microRNAs and their targets in fiber and oil producing plant Flax (*Linum usitatissimum* L.) *Pak J Bot* 44(4): 1357-1362.
10. Rombel IT, Sykes KF, Rayner S & Johnston SA (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282(1): 33-41.
11. Kastenmayer JP, Chu NiL A, Kitchen LE & Au WC (2006). Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res* 16: 3.