HNUE JOURNAL OF SCIENCE DOI: 10.18173/2354-1059.2021-0057
Natural Sciences 2021, Volume 66, Issue 3, pp. 161-169
This paper is available online at http://stdb.hnue.edu.vn

# GENOME-WIDE IDENTIFICATION AND COMPUTATIONAL CHARACTERIZATION OF THE NUCLEAR FACTOR-YC SUB-UNITS IN GRAIN AMARANTH (*Amaranthus hypochondriacus*)

Tran Thi Thanh Huyen[1], La Viet Hong[2], Le Thi Ngoc Quynh[3], Pham Chau Thuy[4], Ha Thi Quyen[4] and Chu Duc Ha[4]
[1]*Faculty of Biology, Hanoi National University of Education*
[2]*Institute of Science and Application, Hanoi Pedagogical University 2*
[3]*Department of Biotechnology, Thuyloi University*
[4]*Faculty of Agricultural Technology, University of Engineering Technology, VNU*

**Abstract**. Nuclear factor-Y (NF-Y) has been known as one of the plant-specific transcription factors that play key roles in numerous biological processes during the growth and development of plant species. In this study, a comprehensive analysis of NF-YC sub-units in grain amaranth (*Amaranthus hypochondriacus*) was carried out based on the bioinformatics approaches. Firstly, a total of five members of the NF-YC sub-units was reported in the grain amaranth. Its structural analyses revealed that the NF-YC sub-units were variable in physic-chemical properties, like protein sizes, molecular masses, isoelectric point, instability index, and grand average of hydropathy. Of our interest, the expression profiles of genes encoding NF-YC sub-units in various tissues/organs during the growth and development of grain amaranth. We found that three genes, including *AhNF-YC01*, *AhNF-YC04*, and *AhNF-YC05* were highly expressed in leaf, root, floral, immature seed, and stem tissues. Interestingly, *AhNF-YC05* was exclusively expressed in leaf and stem tissues. Taken together, our study could provide a solid understanding for further functional characterization of genes encoding NF-YC sub-units in grain amaranth.

***Keywords***. nuclear factor-Y, grain amaranth, expression, structure, identification.

## 1. Introduction

Grain amaranths (*Amaranthus hypochondriacus*) have been known as edible C4 dicots. Containing a high accumulation of lysine, grain amaranths could be used as a highly

Received September 15, 2021. Revised October 18, 2021. Accepted October 25, 2021.
Contact Tran Thi Thanh Huyen, e-mail address: tranthanhhuyen@hnue.edu.vn

nutritional food for daily uses and raw materials for feed production [1, 2]. Since grain amaranth plants are grown in many areas, adverse environmental conditions are the major factors that dramatically affect their growth and development. It is thought that transcription factors (TFs) play critical roles in numerous biological processes, like growth, development, and responses to adverse environmental stresses. Among them, Nuclear factor-Y (NF-Y) was one of the major TF families in whole monocotic and dichotic plant species [3].

Generally, this TF family is composed of three sub-units, namely NF-YA, NF-YB, and NF-YC, of which NF-YC consisted of three conserved sub-domains, like two short regions related to the NF-YA interaction and a sub-domain involved in the dimerization with the NF-YB sub-unit [3, 4]. Many studies reported that NF-Y TFs, especially NF-YC members, participated in the regulation of various biological processes in plants [5]. Up till now, great efforts have been made in order identify and characterize the NF-Y TFs in various important crops, like *Triticum aestivum* [6], *Oryza sativa* [7], *Zea mays* [8], *Phaseolus vulgaris* [9], *Cicer arietinum* [10], *Glycine max* [11] and *Solanum lycopersicum* [12]. Unfortunately, the information of NF-Y TFs, particularly NF-YC sub-units in grain amaranth have been still lacked, even its assembly has been reported for several years [13]. Recently, two remaining sub-units, NF-YA, and NF-YB were reported in grain amaranth [14, 15].

In the present study, we carried out a genome-wide identification, annotation, and characterization of the NF-YC sub-units in grain amaranths. Particularly, all putative members of NF-YC groups were identified in the grain amaranth assembly. We next characterized their general features, like gene structure and physic-chemical parameters of proteins. Finally, we investigated the expression profiles of genes encoding NF-YC sub-units in various tissues/organs during the growth and development of grain amaranth plants.

## 2. Content

### 2.1. Materials and methods

**** Materials***

The recent assembly of grain amaranth (AhG2s cultivar), including genome, proteome, and transcriptome was explored from Phytozome [16] and NCBI [13]. The well-characterized NF-YC sub-units from *C. arietinum* downloaded from the previous study [10] were used for our computational analyses.

**** Methods***

- *Identification and annotation of NF-YC*

The well-characterized NF-YC sub-units from *C. arietinum* [10] were used as templates for BLAST searches against the current assembly of grain amaranth [13]. All reliable results (E-value ≥ 1e-10) were collected to validate the presence of the NF-YC-specific domain [4] in the Pfam [17] as previously described [14, 15]. The annotation of

each putative NF-YC sub-unit, including genomic DNA sequence (gDNA), coding DNA sequence (CDS), full-length protein sequence, and identifiers were then explored for further analyses.

- *Characterization of protein features of NF-YC*

The full-length protein sequence of each NF-YC was used as a query to analyze in the Expasy Protparam [18]. Briefly, five features, like protein size (amino acid residues, aa), molecular weight (kilo Dalton, kDa), isoelectric point (pI), instability index (II), and grand average of hydropathy (GRAVY) were obtained from the Expasy Protparam [18] as previously described [10]. Among them, pI values of less than 7 and more than 7 indicated acidity and base ranges, respectively, while II values of less than 40 and more than 40 indicated stability and instability natures, respectively. The GRAVY values of less than 0 and more than 0 indicated hydrophilic and hydrophobic, respectively.

- *Construction of phylogenetic tree of NF-YC*

The full-length protein sequences of all NF-YC sub-units were used to construct an unrooted phylogenetic tree by the MEGA [19] as previously [10]. Particularly, all proteins were firstly aligned by the MEGA with default parameters to generate a phylogenetic tree by the Neighbor-Joining method [10]. The phylogenetic tree was then exported into Adobe Illustrator for visualization.

- *Motif compositions of NF-YC*

The full-length protein sequences of all NF-YC sub-units were searched against the MEME tool [20] to identify the motif enrichments. The length of each potential motif was set from 6 to 50 aa as the default parameters. The prediction was then arranged based on the order of occurrence in the phylogenetic tree.

- *Expression analysis of genes encoding NF-YC*

The RNA-Seq dataset was obtained from the Phytozome [16] as previously described [14, 15]. Particularly, the transcriptome atlas in seven tissues, including root, stem, leaf, floral, maturing seed, immature seed, and green cotyledon under the normal condition was explored [13]. The expression profiles of genes encoding NF-YC sub-units were represented by the $\log_2$ FPKM (Fragments per kilo base million) values and visualized by R script as the previous method [14, 15].

## 2.2. Results and discussion

### 2.2.1. Genome-wide identification and annotation of the NF-YC sub-units in grain amaranth

In order to identify all putative members in the NF-YC sub-units, intensive searches have been carried out based on the available seed sequences [10]. The conserved NF-YC-specific domain was then confirmed by validating the full-length protein sequence in the Pfam [17]. As a result, a total of five putative members of the

NF-YC sub-units has been identified in grain amaranths (Table 1). Based on the order of TF identifiers, gene names, from *AhNF-YC01* to *05* were listed in Table 1.

***Table 1. Summary of the NF-YC sub-units in grain amaranths***

| No. | Gene name | Gene ID | Transcript ID | TF ID |
|---|---|---|---|---|
| 1 | *AhNF-YC01* | AH020120 | AH020120-RA | AHYPO_005743-RA |
| 2 | *AhNF-YC02* | AH000360 | AH000360-RA | AHYPO_006083-RA |
| 3 | *AhNF-YC03* | AH001329 | AH001329-RA | AHYPO_008475-RA |
| 4 | *AhNF-YC04* | AH010440 | AH010440-RA | AHYPO_010955-RA |
| 5 | *AhNF-YC05* | AH002982 | AH002982-RA | AHYPO_020681-RA |

*TF: Transcription factor, ID: Identifier.*

As compared with previous studies, the amount of NF-YC sub-units in higher plant species was highly variable. Particularly, the number of NF-YC sub-units were recorded to range from five (in grain amaranths), seven (in *P. vulgaris*) [9], 11 (in *C. arietinum*) [10], 12 (in *O. sativa*) [7], 14 (in *T. aestivum*), 15 (in *G. max*) [11], 18 (in *Z. mays*) [8], to 20 (in *S. lycopersicum*) [12] (Table 2). Recently, information on the NF-YA and NF-YB sub-units in grain amaranths was also reported. Six and 16 members of NF-YA and NF-YB sub-units were identified and characterized in grain amaranths based on the computational approaches [14, 15]. The low amount of members of NF-Y TF in grain amaranths could be explained that the estimated genome size of grain amaranths (~ 256 Mb) was much less than that in other plant species [13].

***Table 2. Summary of the NF-YC sub-units in several plant species***

| No. | Plant species | NF-YC sub-units | References |
|---|---|---|---|
| 1 | *Amaranthus hypochondriacus* | 5 | Our study |
| 2 | *Phaseolus vulgaris* | 7 | [9] |
| 3 | *Cicer arietinum* | 11 | [10] |
| 4 | *Oryza sativa* | 12 | [7] |
| 5 | *Triticum aestivum* | 14 | [6] |
| 6 | *Glycine max* | 15 | [11] |
| 7 | *Zea mays* | 18 | [8] |
| 8 | *Solanum lycopersicum* | 20 | [12] |

### 2.2.2. Structural analysis of the NF-YC sub-units in grain amaranth

In order to characterize the properties of NF-YC sub-units in grain amaranths, the full-length protein sequence was analyzed in the Expasy Protparam [18] to assess several general parameters, like protein size molecular weight, pI and II values, and GRAVY. The analysis of the physic-chemical properties of the five putative AhNF-YC was provided in Table 3.

***Table 3. Characterization of NF-YC sub-units in grain amaranths***

| No. | NF-YC sub-units | Size | mW | pI | II | GRAVY |
|---|---|---|---|---|---|---|
| 1 | AhNF-YC01 | 233 | 26.57 | 5.90 | 56.07 | -0.74 |
| 2 | AhNF-YC02 | 255 | 27.66 | 5.60 | 59.68 | -0.51 |
| 3 | AhNF-YC03 | 133 | 14.70 | 5.91 | 40.19 | -0.32 |
| 4 | AhNF-YC04 | 255 | 27.93 | 5.95 | 56.19 | -0.47 |
| 5 | AhNF-YC05 | 146 | 16.44 | 7.96 | 43.12 | -0.58 |

*Size: Amino acid residues, mW: Molecular weight (kilo Dalton), pI: Isoelectric point, II: Instability index, GRAVY: Grand average of hydropathy*

Our results showed that the sizes of AhNF-YC sub-units were varied from 133 (AhNF-YC03) to 255 aa residues (AhNF-YC02 and 04). Next, the molecular weights of AhNF-YC sub-units ranged from 14.70 (AhNF-YC03) to 27.93 kDa (AhNF-YC04). The pI scores of AhNF-YC sub-units were mostly in the acidic range, from 5.60 (AhNF-YC02) to 5.95 (AhNF-YC04), except for AhNF-YC05 (pI = 7.96) (Table 3). Interestingly, the II and GRAVY values of all AhNF-YC sub-units were more than 40 and less than 0, respectively. These findings suggested that NF-YC sub-units in grain amaranths were unstable in the test tube condition and hydrophilic.

Previously, the physic-chemical properties of the NF-YC sub-units in other plant species were also investigated. For example, the length of NF-YC sub-units from *Z. mays* was varied from 127 to 439 aa residues [8], while the CaNF-YC proteins from *C. arietinum* ranged from 114 to 357 aa residues in size, 12.60 to 40.52 kDa in mass [10]. It has been also demonstrated that the II scores of all NF-YC sub-units in other plant species, such as *C. arietinum*, *S. lycopersicum,* and *T. aestivum* were more than 40, indicating that these sub-units were unstable [6, 10, 12]. Recently, the features of NF-YA and NF-YB sub-units from grain amaranth were also discussed, which could together provide comprehensive information on the NF-Y TFs in this important crop. Particularly, NF-YA and NF-YB sub-units were reported to range from 230 to 337 and 51 to 328 aa residues in sizes, respectively [14, 15]. The majority of NF-YA and NF-YB sub-units were in the acidic range, while all NF-YA sub-units were also unstable under the test tube condition [14, 15].

Next, the motif compositions in the NF-YC sub-units in grain amaranths were also investigated in order to assess the conserved structure of these proteins. The foundation of these conserved motifs was then provided in Figure 1. As a result, a total of three conserved motifs was found in the full-length sequences of NF-YC sub-units in grain amaranths (Figure 1A). Among them, motif 1 was localized in all AhNF-YC proteins, while motif 2 and 3 were noted to distribution on four (AhNF-YC01, 02, 03, and 04) and three (AhNF-YC01, 02, and 04) members of NF-YC sub-units in grain amaranth (Figure 1A). The detailed sequences of the three motifs were described in Figure 1B.

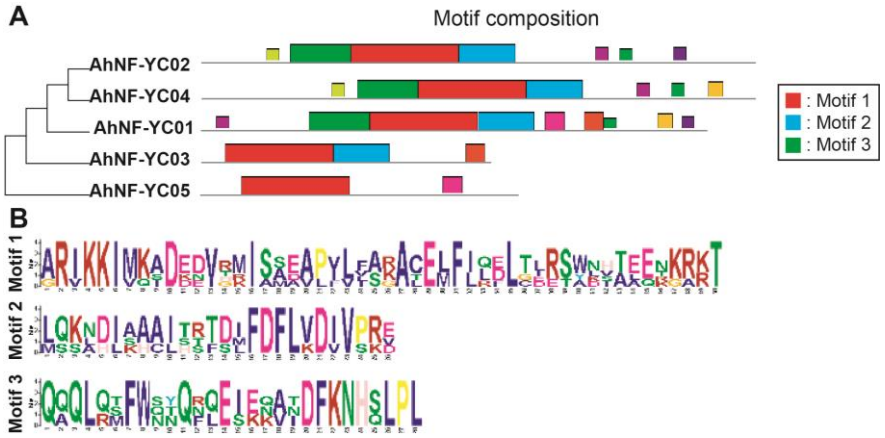


***Figure 1. Motif compositions in the NF-YC sub-units in grain amaranths and their corresponding sequences. (A) Indicates the motif positions corresponding to the phylogenic tree of AhNF-YC amino acids; (B) The consensus motif sequences***

It is thought that these motifs might act as functional domains of the NF-YC sub-units, like interacting with NF-YA (motif 2 and 3) and NF-YB sub-units (motif 1) as previously reported [3, 4]. Recently, the NF-YC sub-units from *C. arietinum* were recorded to contain three core regions, including two short domains that participated in the NF-YA interaction and an 'NF-YB interaction' domain that participated in the dimerization with NF-YB sub-units [10]. In tomato, the NF-YC sub-units contained a highly conserved ~80-aa-domain. This domain has been shown to harbor three sub-domains that act for DNA binding and interactions between NF-Y sub-units [12]. In this study, only three members, including AhNF-YC01, 02, and 04 were described to contain three motifs, indicating that these molecules might involve in the construction of NF-Y TF complexes.

### 2.2.3. Expression profiles of genes encoding the NF-YC sub-units in grain amaranth

In order to analyze the expression profiles of genes encoding NF-YC sub-units in grain amaranth, the recent transcriptome dataset [13] was applied to generate the $\log_2$ FPKM values. The obtained data was then described by R and provided in Figure 2. As a result, five genes encoding NF-YC sub-units in grain amaranth were exhibited different expression levels in various tissues/organs (Figure 2).

Based on the $\log_2$ FPKM values, we found that all genes encoding NF-YC sub-units were not expressed (under the detection) in maturing seed and green cotyledon (Figure 2). Two genes, *AhNF-YC02* and *03* were not expressed or were expressed at low levels on vegetative plant parts include leaf, floral, maturing seed, immature seed, stem, and green cotyledon. Next, three genes, including *AhNF-YC01*, *04*, and *05* were noted to be highly expressed in five tissues (leaf, root, floral, immature seed, and stem) (Figure 2). Among them, *AhNF-YC05* was noted to be exclusively expressed in leaf and stem, suggesting that this gene may play important roles in the biological processes in these tissues.

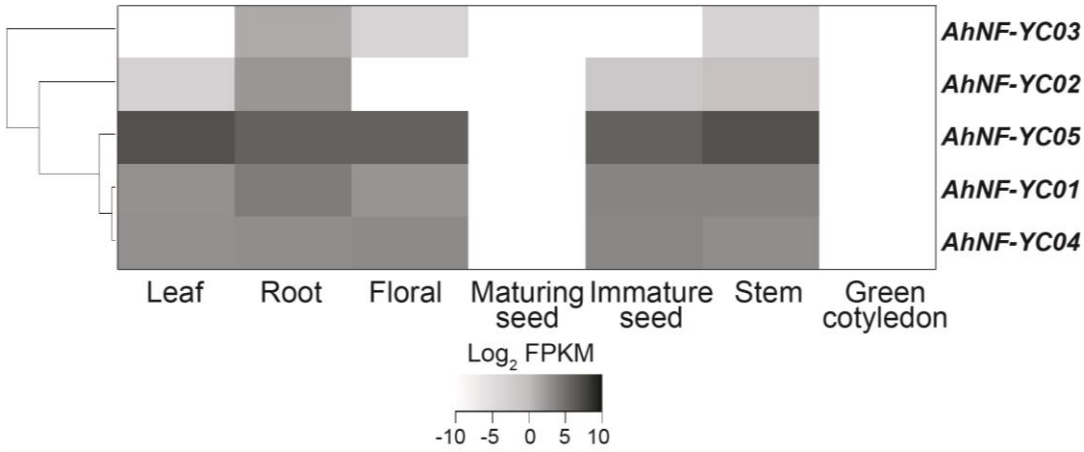


***Figure 2. Expression profiles of genes encoding NF-YC sub-units in different tissues in grain amaranth. The color bar represents the normalized FPKM values, black represents high expression level, gray represents low expression level, and white represents no expression***

Previously, the expression profiles of genes encoding NF-YA and NF-YB sub-units were also reported [14, 15]. Among genes encoding NF-YA sub-units, *AHYPO_07754-RA* was highly expressed in flower only, while *AHYPO_002483-RA* and *AHYPO_003114-RA* gene expression were shown at high levels in immature seed, flower, mature seed, and green cotyledon [15]. Additionally, four genes encode NF-YB sub-units, like *AHYPO_001076-RA*, *AHYPO_003409-RA*, *AHYPO_013452-RA*, and *AHYPO_014999-RA* were found to highly express in all seven tissues. Furthermore, several genes encoding NF-YC sub-units from *A. thaliana* and *O. sativa* were demonstrated to control the flowering time in plants [5].

## 3. Conclusions

Five members of NF-YC sub-units have been identified and characterized in grain amaranths. The number of members of NF-YC sub-units was variable between plant species.

Our analysis indicated that the AhNF-YC family was ranged from 133 to 255 aa residues and 14.70 to 27.93 kDa in sizes and weights, respectively. All members of NF-YC sub-units were found to be hydrophilic and unstable.

Three genes were highly expressed in leaf, root, floral, immature seed, and stem tissues. *AhNF-YC05* was noted as the leaf- and stem-specific gene.

In further studies, the expression patterns of the genes encoding NF-YC sub-units under adverse environmental conditions will be carried out to propose a candidate gene for functional characterization.

## REFERENCES

[1] Caselato-Sousa V. M., Amaya-Farfán J., 2012. State of knowledge on amaranth grain: a comprehensive review. *J. Food Sci*., Vol. 77, Iss. 4, pp. R93-104.

[2] Oteri M., Gresta F., Costale A., Lo Presti V., Meineri G., Chiofalo B., 2021. Amaranthus hypochondriacus L. as a sustainable source of nutrients and bioactive compounds for animal feeding. *Antioxidants*, Vol. 10, Iss. 6, pp. 876.

[3] Laloum T., De Mita S., Gamas P., Baudin M., Niebel A., 2013. CCAAT-box binding transcription factors in plants: Y so many? Trends Plant. Sci, Vol. 18, Iss. 3, pp. 157-166.

[4] Myers Z. A., Holt B. F., 3rd, 2018. Nuclear Factor -Y: still complex after all these years? *Curr. Opin. Plant Biol.*, Vol. 45, pp. 96-102.

[5] Hwang Y. H., Kim S. K., Lee K. C., Chung Y. S., Lee J. H., Kim J. K., 2016. Functional conservation of rice OsNF-YB/YC and Arabidopsis AtNF-YB/YC proteins in the regulation of flowering time. *Plant Cell Rep.*, Vol. 35, Iss. 4, pp. 857-865.

[6] Stephenson T. J., McIntyre C. L., Collet C., Xue G. P., 2007. Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum. *Plant Mol. Biol*., Vol. 65, Iss. 1-2, pp. 77-92.

[7] Yang W., Lu Z., Xiong Y., Yao J., 2017. Genome-wide identification and co-expression network analysis of the OsNF-Y gene family in rice. *Crop J.*, Vol. 5, Iss. 1, pp. 21-31.

[8] Zhang Z., Li X., Zhang C., Zou H., Wu Z., 2016. Isolation, structural analysis, and expression characteristics of the maize nuclear factor Y gene families. *Biochem. Biophys. Res. Commun*, Vol. 478, Iss. 2, pp. 752-758.

[9] Ripodas C., Castaingts M., Clua J., Blanco F., Zanetti M. E., 2014. Annotation, phylogeny and expression analysis of the nuclear factor Y gene families in common bean (*Phaseolus vulgaris*). *Front Plant Sci*., Vol. 5, pp. 761.

[10] Chu H. D., Nguyen K. H., Watanabe Y., Le D. T., Pham T. L. T., Mochida K., Tran L. P., 2018. Identification, structural characterization and gene expression analysis of members of the Nuclear factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments. *Int. J. Mol. Sci.*, Vol. 19, Iss. 11, pp. 3290.

[11] Quach T. N., Nguyen H. T., Valliyodan B., Joshi T., Xu D., Nguyen H. T., 2015. Genome-wide expression analysis of soybean NF-Y genes reveals a potential

function in development and drought response. *Mol. Genet. Genomics*, Vol. 290, Iss. 3, pp. 1095-1115.

[12] Li S., Li K., Ju Z., Cao D., Fu D., Zhu H., Zhu B., Luo Y., 2016. Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. *BMC Genomics*, Vol. 17, Iss. 1, pp. 36.

[13] Sunil M., Hariharan A. K., Nayak S., Gupta S., Nambisan S. R., Gupta R. P., Panda B., Choudhary B., Srinivasan S., 2014. The draft genome and transcriptome of Amaranthus hypochondriacus: a C4 dicot producing high-lysine edible pseudo-cereal. *DNA Res.*, Vol. 21, Iss. 6, pp. 585-602.

[14] Quynh T. N. L., Ha D. C., 2020. Study on the structure and *in silico* expression analysis of genes encoding nuclear factor-YB in amaranth. *J. Vietnam Agri. Sci. Tech.*, Vol. 2, Iss. 111, pp. 31-35 (in Vietnamese).

[15] Quynh T. N. L., Ha D. C., 2021. Characterization of the conserved structure and expression analysis of the genes encoding YA subunits of the transcription factor Nuclear factor –Y in amaranth. *J. Vietnam Agri. Sci. Tech.*, Vol., 2, Iss. 123, pp. 40-45 (in Vietnamese).

[16] Goodstein D. M., Shu S., Howson R., Neupane R., Hayes R. D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D. S., 2012. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.*, Vol. 40, pp. D1178-D1186.

[17] Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar Gustavo A., Sonnhammer E. L. L., Tosatto S. C. E., Paladin L., Raj S., Richardson L. J., Finn R. D., Bateman A., 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res.*, Vol. 49, Iss. D1, pp. D412-D419.

[18] Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D., Bairoch A., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, Vol. 31, Iss. 13, pp. 3784-3788.

[19] Kumar S., Stecher G., Tamura K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, Vol. 33, Iss. 7, pp. 1870-1874.

[20] Bailey T. L., Johnson J., Grant C. E., Noble W. S., 2015. The Meme Suite. *Nucleic Acids Res.*, Vol. 43, Iss. W1, pp. W39-49.