

# Genome-wide identification and annotation of the Nuclear-factor YA gene family in cassava (*Manihot esculenta* Crantz)

Duc Ha Chu<sup>1\*</sup>, Thi Thuy Tam Do<sup>1,2</sup>, Xuan Dac Le<sup>3</sup>, Thi Ly Thu Pham<sup>1</sup>

<sup>1</sup>Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences

<sup>2</sup>University of Science and Technology of Hanoi

<sup>3</sup>Institute of Tropical Ecology, Vietnam-Russia Tropical Center

Received 5 May 2017; accepted 6 September 2017

## Abstract:

Nuclear factor Y (*NF-Y*) is a transcription factor which plays an important role in the regulation of various developmental processes and stress responses in plants. By using various bioinformatics tools, the identification and analyses of the *NF-YA* subunit of cassava (*Manihot esculenta* Crantz) have been attempted in this study. A total of 12 members of the *NF-YA* gene family were identified in the cassava genome. They were located on the 18 cassava chromosomes with different frequencies. Several initial structural analyses of the *NF-YA* family were also performed. Among them, the typical gene organization of the *MeNF-YA* gene family contained 5 exons/4 introns. Interestingly, the conserved region of *NF-YA* was characterized by the interaction of *NF-YB/C* domain and the DNA binding domain. This study provided information on *NF-Y* in plants.

**Keywords:** cassava, gene, in silico, *NF-YA*, transcription factor.

**Classification number:** 3.1

## Introduction

*NF-Y* (Nuclear factor Y) is known as one of the most important transcription factor groups in all eukaryotes. This family has evidentially played the key roles in the regulation of diverse genes [1]. *NF-Y* has three subunits (*NF-YA*, *NF-YB*, and *NF-YC*), which are connected with a range of biological processes, from the signalling pathways to stress responses in plants. Thus, it would be essential to study these subunits in order to expand our knowledge on plant's responses to adverse biotic/abiotic stresses.

To date, the *NF-Y* gene family has been found and characterized in many plant species such as rice (*Oryza sativa*) [2], canola (*Canola napus*) [3, 4],

soybean (*Glycine max*) [5], and foxtail millet (*Setaria italica*) [6]. Recently, the family has also been recorded in tomato (*Solanum lycopersicum*) [7], grape (*Vitis vinifera*) [8], and sorghum (*Sorghum bicolor*) [9]. Many *NF-YA* genes were reported to function in biological processes, especially in stress response in plants. For example, *Arabidopsis thaliana* transgenic plants overexpressing *AtNF-YA5* have shown a reduction of leaf water loss and a better resistance to drought stress than the wild-type plants, thus revealing that the *AtNF-YA5* might function in drought resistance through transcriptional and posttranscriptional regulatory mechanisms [10]. Additionally, *Arabidopsis AtNF-YA3* and *AtNF-YA8* were also found as redundant genes

required in early embryogenesis of plants [11]. In soybean, overexpression of *GmNF-YA3* conferred the reduction of leaf water loss and enhanced drought tolerance in transgenic *Arabidopsis* plants [12].

In this study, the *NF-YA* gene family in cassava (*Manihot esculenta*) was identified and annotated. The identifier, which was the chromosomal location of each gene encoding *NF-YA* subunit, was provided based on various available databases. Gene organization of *NF-YA* gene family in cassava was also analyzed by using bioinformatics approaches. Finally, protein features and conserved domains of *NF-YA* subunits were involved.

## Materials and methods

### Materials

The cassava genome database of "AM560-2" cultivar [13] is available in Phytozome v12.0 [14].

### Methods

*Identification and annotation of genes encoding NF-YA in cassava genome:* Members of *NF-Y* family in cassava from the Phytozome v12.0 [14] were identified. Their identifiers and chromosomal locations were then confirmed by blasting (BLASTP) against the cassava genome database [13] in NCBI server.

*Analysis of gene structure of NF-YA genes:* The genomic sequence and CDS (coding DNA sequence) of each

\*Corresponding author: Email: hachuamsr@yahoo.com

member of *NF-YA* genes were obtained from the cassava genome database [13] in the Phytozome v12.0 [14]. The GSDS (Gene Structure Display Server) v2.0 was used to analyze the exon/intron organization of *MeNF-Y* genes [15].

*Multiple alignments and phylogenetic analysis of MeNF-Y proteins:* The protein sequence of each member of *NF-YA* subunits was obtained from the Phytozome v12.0 [14]. The MEGA (Molecular Evolutionary Genetics Analysis) software v7.0 [16] was utilized for multiple alignments of *MeNF-YA* proteins. The parameters of sequence alignments were composed of a gap open penalty of 10 and a gap extension penalty of 0.2. An unrooted phylogenetic tree of all full-length *NF-YA* proteins was constructed with the Neighbor Joining Method as previously studied [17].

*Analysis of protein features of NF-YA subunit:* The general information, including the isoelectric point (pI) and molecular weight (mW), was collected through the ExPASy tool [18]. The subcellular localization of proteins was predicted via the TargetP v1.1 web-based tool [19, 20].

**Results and discussions**

**Genome-wide identification of the NF-YA gene family in the cassava genome**

In order to identify the *NF-YA* family in cassava, a comprehensive search of all proteins containing typical *NF-YA* conserved domain [1] was performed against the family in cassava from the Phytozome v12.0 [14]. As a result, a total of 12 members of the *NF-YA* family were found in the cassava genome (E-value < 1 × 10<sup>-6</sup>). The gene annotation and nomenclature of *NF-YA* gene family were harvested by searching against the NCBI database (Bioproject: PRJNA86123) (Table 1).

The *NF-YA* subunit found in cassava genome was also encoded by a gene belonging to a multigene family

**Table 1. Annotation of NF-Y gene family in cassava genome.**

#	Gene name	Transcript name <sup>1,2</sup>	Alias name <sup>1</sup>	Locus name <sup>2</sup>
1	<i>MeNF-YA1</i>	Manes.04G142600.1	cassava4.1_014256m.g.v4.1	OAY53185
2	<i>MeNF-YA2</i>	Manes.06G163900.1	cassava4.1_010819m.g.v4.1	OAY48517
3	<i>MeNF-YA3</i>	Manes.06G054900.1	cassava4.1_016099m.g.v4.1	OAY47137
4	<i>MeNF-YA4</i>	Manes.07G006600.1	cassava4.1_010627m.g.v4.1	OAY44802
5	<i>MeNF-YA5</i>	Manes.08G034700.1	cassava4.1_011620m.g.v4.1	OAY43008
6	<i>MeNF-YA6</i>	Manes.09G025200.1	cassava4.1_012382m.g.v4.1	OAY40472
7	<i>MeNF-YA7</i>	Manes.09G044200.1	cassava4.1_012637m.g.v4.1	OAY40725
8	<i>MeNF-YA8</i>	Manes.10G141400.1	cassava4.1_011264m.g.v4.1	OAY40005
9	<i>MeNF-YA9</i>	Manes.11G022300.1	cassava4.1_013364m.g.v4.1	OAY36452
10	<i>MeNF-YA10</i>	Manes.14G003100.1	cassava4.1_007505m.g.v4.1	OAY30095
11	<i>MeNF-YA11</i>	Manes.14G123000.1	cassava4.1_017907m.g.v4.1	OAY31569
12	<i>MeNF-YA12</i>	Manes.16G097900.1	cassava4.1_011576m.g.v4.1	OAY27078

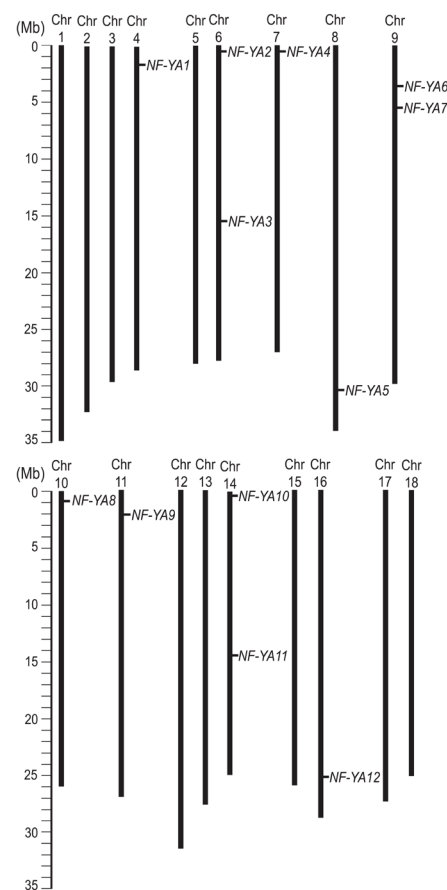
Information obtained from <sup>1</sup>Phytozome v12.0 and <sup>2</sup>NCBI databases.

as observed in other higher plants' genomes [1]. In comparison with recent annotated dicot species, a total of 21 *GmNF-YA* genes were identified in soybean [5], while 10 *NF-YA* genes were computationally predicted in tomato [7]. More recently, the genome-wide identification of eight *NF-YA* genes has been reported in grape [8].

The chromosomal locations of 12 *NF-YA* genes were identified based on the cassava genome database [13]. As manually illustrated in Fig. 1, these 12 members of *MeNF-YA* genes were mapped on the 18 cassava chromosomes with different frequencies. Among them, chromosomes 6, 9, and 14 contained two *MeNF-YA* genes, whereas only one *MeNF-Y* gene was distributed on each of the chromosomes 4, 7, 8, 10, 11, and 16 (Fig. 1).

**Analysis of the structure of MeNF-YA genes**

To analyze the structures of *NF-YA* genes in cassava, the genomic sequence and CDS of each *NF-YA* member were obtained from the cassava genome [13]. They were then used as query sequences



**Fig. 1. Chromosomal distributions of MeNF-YA genes in cassava genome.**

in the GSDS web-based tool to explore the structures of *NF-YA* genes in cassava (Table 2).

As provided in Table 2, the genomic regions of *NF-YA* genes had a variable length of from 4042 (*MeNF-YA4*, Manes.07G006600.1) to 16084 nucleotides (*MeNF-YA11*, Manes.14G123000.1). Previously, the genomic length of a gene was evidentially associated with the transcription level of this gene [21]. Hence, it would be proposed that all *MeNF-YA* genes were highly expressed in the cells, thus they might function in various biological processes and stress response in cassava plants.

Interestingly, the CDS of *NF-YA* genes varied from 645 (*MeNF-YA11*) to 1065 nucleotides (*MeNF-YA10*, Manes.14G003100.1) (Table 2). The structures of *MeNF-YA* genes commonly consisted of 5 exons/4 introns. Only *MeNF-YA6* (Manes.09G025200.1) had 4 exons/3 introns (Fig. 2). Our results clearly indicated that *NF-YA* gene family was completely conserved in cassava as well as in other higher plant species [1]. Furthermore, the introns in the CDS region of a gene might cause the structural diversity and complexity. Consequently, the presences of introns in *MeNF-YA* genes might be directly related to the evolution of *NF-YA* gene family in cassava.

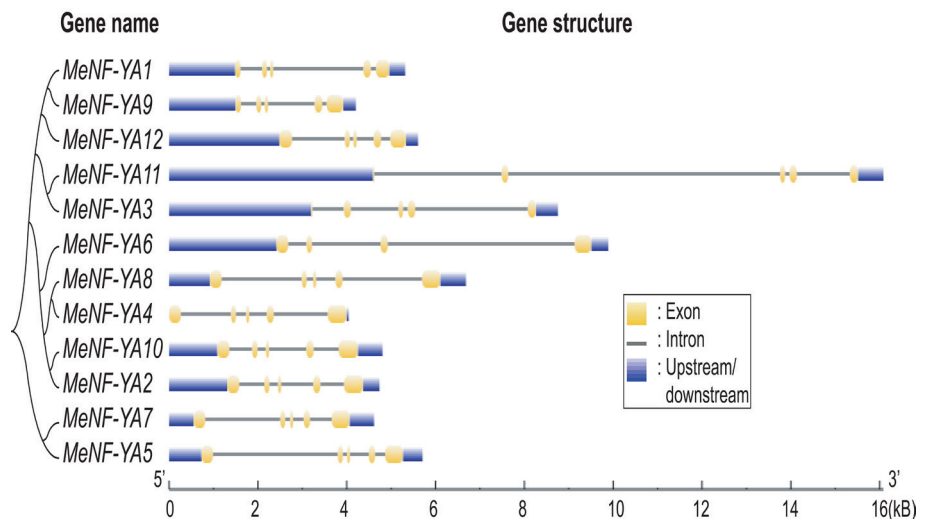
#### Analysis of protein features of *MeNF-YA*

General features of *MeNF-YA* members of cassava were also figured out by analyzing the protein sequence of each member obtained from the Phytozome v12.0 [14] in the ExPASy tool [18]. The lengths of *MeNF-YA* proteins in cassava ranged from 212 (*MeNF-YA3*) to 354 amino acids (*MeNF-YA10*). The mW values of *NF-YA* family also reached from 23.34 (*MeNF-YA3*) to 38.30 kDa (*MeNF-YA2*)

**Table 2. The structures of *NF-YA* genes in cassava.**

#	Gene name	Chromosomal location	Genomic length	CDS length
1	<i>MeNF-YA1</i>	Chr04R:26895916..26901232	5317	780
2	<i>MeNF-YA2</i>	Chr06R:26629560..26634297	4738	1053
3	<i>MeNF-YA3</i>	Chr06F:15795526..15804280	8755	639
4	<i>MeNF-YA4</i>	Chr07F:807124..811165	4042	1029
5	<i>MeNF-YA5</i>	Chr08R:3126251..3131957	5707	993
6	<i>MeNF-YA6</i>	Chr09F:3779466..3789354	9889	930
7	<i>MeNF-YA7</i>	Chr09F:5927016..5931630	4615	996
8	<i>MeNF-YA8</i>	Chr10R:25259719..25266401	6683	1020
9	<i>MeNF-YA9</i>	Chr11F:2064501..2068703	4203	852
10	<i>MeNF-YA10</i>	Chr14F:401483..406287	4805	1065
11	<i>MeNF-YA11</i>	Chr14R:10921305..10937388	16084	645
12	<i>MeNF-YA12</i>	Chr16F:25396313..25401918	5606	996

Information was obtained from the Phytozome v12.0; Chr: Chromosome; F: Forward; R: Reverse; Genomic and CDS length were measured by nucleotides.



**Fig. 2. Gene structure of *NF-YA* family in cassava.**

(Table 3). Previously, eight members of *NF-YA* subunit were also identified in sorghum (*Sorghum bicolor*). Among them, SbNF-YA2 (ABXC01000113.1) was found to be the smallest member (90 amino acids, 10.21 kDa), whereas the size of SbNF-YA3 was 305 amino acids and 33.37 kDa [9].

Additionally, a majority of *MeNF-YA* proteins were the basic proteins,

from 8.53 (*MeNF-YA2*) to 9.61 (*MeNF-YA5*). The pI of four remaining *NF-YA* members approximately reached 7, thus indicating that they were likely neutral proteins (Table 3). As mentioned above, all *NF-YA* members in sorghum were also shifted towards basicity [9]. It is understood that the pI value of a protein was directly linked with its subcellular localization. Here, it was observed

Table 3. General features of NF-YA proteins in cassava.

#	Gene name	Transcript name	Protein length	pI	mW
1	MeNF-YA1	Manes.04G142600.1	259	8.97	28.83
2	MeNF-YA2	Manes.06G163900.1	350	8.53	38.30
3	MeNF-YA3	Manes.06G054900.1	212	7.16	23.34
4	MeNF-YA4	Manes.07G006600.1	342	8.76	37.32
5	MeNF-YA5	Manes.08G034700.1	330	9.61	36.03
6	MeNF-YA6	Manes.09G025200.1	309	8.77	34.78
7	MeNF-YA7	Manes.09G044200.1	331	9.13	36.30
8	MeNF-YA8	Manes.10G141400.1	339	8.99	36.56
9	MeNF-YA9	Manes.11G022300.1	283	6.95	31.11
10	MeNF-YA10	Manes.14G003100.1	354	6.67	38.05
11	MeNF-YA11	Manes.14G123000.1	214	6.75	23.44
12	MeNF-YA12	Manes.16G097900.1	331	9.01	36.37

Data were obtained from the ExPasy tool; Protein length (amino acid); pI: Isoelectric point; mW: Molecular weight (kDa).

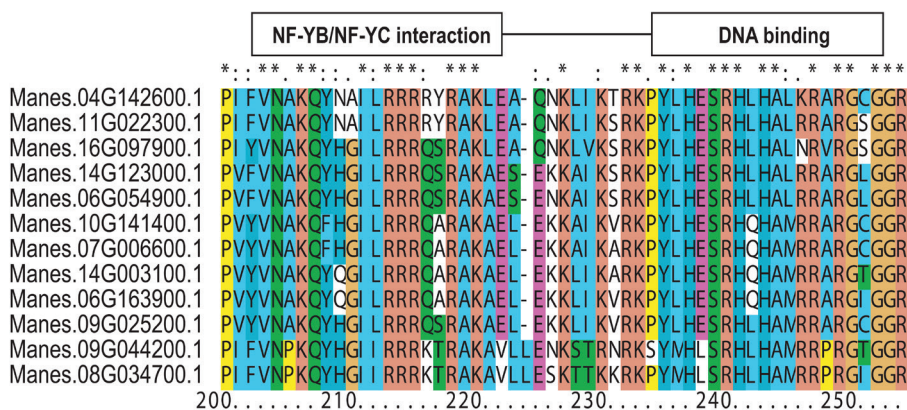


Fig. 3. The conserved domain of NF-YA subunit in cassava.

that basic MeNF-YA proteins seemed to belong to an integral membrane proteome.

The conserved domain of the NF-YA family in cassava was analyzed by using the MEGA software [16]. As shown in Fig. 3, the MeNF-YA proteins in cassava could be characterized by two conserved

regions, including a protein interaction and DNA binding domains. A twenty-amino-acid-domain could be bound to the combined surface of NF-YB/NF-YC complex [22] that was clearly observed in the alignment of MeNF-YA proteins. Interestingly, most of yeast and mammals functionally required amino

acids [23, 24] that were also obviously found in NF-YA family in cassava. These findings highlighted that the NF-YA family was completely conserved during the evolution.

Conclusions

A total of 12 members of the NF-YA gene family have been found in the cassava genome. The identified MeNF-YA genes were distributed on the 18 cassava chromosomes with different frequencies. The analysis of gene structure showed that the genomic regions of the MeNF-YA genes ranged from 4042 to 16084 nucleotides, while the CDS varied from 645 to 1065 nucleotides. The most common motif of NF-YA genes in cassava was 5 exons/4 introns.

Most of MeNF-YA members were basic proteins. This strongly suggested that they belonged to the integral membrane proteome in the cells. In addition, the MeNF-YA proteins could be recognized by two conserved regions, including NF-YB/NF-YC interaction and DNA binding domains.

This research provided an initial description of the NF-YA gene family in cassava plants. In further studies, the expression profiles of these identified MeNF-YA genes under various conditions should be analyzed.

REFERENCES

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

[2] T. Thirumurugan, Y. Ito, T. Kubo, A. Serizawa, N. Kurata (2008), "Identification, characterization and interaction of HAP family genes in rice", *Mol. Genet. Genomics*, **279**(3), pp.279-289.

[3] L. Xu, Z. Lin, Q. Tao, M. Liang, G. Zhao, X. Yin, R. Fu (2014), "Multiple NUCLEAR FACTOR Y transcription factors respond to abiotic stress in *Brassica napus* L.", *PLoS One*, **9**(10), p.e111354.

[4] M. Liang, X. Yin, Z. Lin, Q. Zheng, G. Liu, G. Zhao (2014), "Identification and characterization of NF-Y transcription factor families in canola (*Brassica napus* L.)", *Planta*, **239**(1), pp.107-126.

- [5] Truyen N. Quach, Hanh T.M. Nguyen, Babu Valliyodan, Trupti Joshi, Dong Xu, Henry T. Nguyen (2015), "Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response", *Mol. Genet. Genomics*, **290**(3), pp.1095-1115.
- [6] Z.J. Feng, G.H. He, W.J. Zheng, P.P. Lu, M. Chen, Y. Gong, Y.Z. Ma, Z.S. Xu (2015), "Foxtail millet NF-Y families: Genome-wide survey and evolution analyses identified two functional genes important in abiotic stresses", *Front. Plant Sci.*, **6**, doi: 10.3389/fpls.2015.01142. eCollection 2015.
- [7] S. Li, K. Li, Z. Ju, D. Cao, D. Fu, H. Zhu, B. Zhu, Y. Luo (2016), "Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening", *BMC Genomics*, **17**, doi: 10.1186/s12864-015-2334-2.
- [8] C. Ren, Z. Zhang, Y. Wang, S. Li, Z. Liang (2016), "Genome-wide identification and characterization of the NF-Y gene family in grape (*Vitis vinifera* L.)", *BMC Genomics*, **17**, doi: 10.1186/s12864-016-2989-3.
- [9] N. Malviya, P. Jaiswal, D. Yadav (2016), "Genome-wide characterization of Nuclear Factor Y (NF-Y) gene family of sorghum [*Sorghum bicolor* (L.) Moench]: a bioinformatics approach", *Physiol. Mol. Biol. Plants*, **22**(1), pp.33-49.
- [10] W.X. Li, Y. Oono, J. Zhu, X.J. He, J.M. Wu, K. Iida, X.Y. Lu, X. Cui, H. Jin, J.K. Zhu (2008), "The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance", *Plant Cell*, **20**(8), pp.2238-2251.
- [11] M. Fornari, V. Calvenzani, S. Masiero, C. Tonelli, K. Petroni (2013), "The *Arabidopsis* NF-YA3 and NF-YA8 genes are functionally redundant and are required in early embryogenesis", *PLoS One*, **8**(11), p.e82043, doi: 10.1371/journal.pone.0082043. eCollection 2013.
- [12] Z. Ni, Z. Hu, Q. Jiang, H. Zhang (2013), "*GmNFYA3*, a target gene of miR169, is a positive regulator of plant tolerance to drought stress", *Plant Mol. Biol.*, **82**(1-2), pp.113-129.
- [13] J.V. Bredeson, J.B. Lyons, S.E. Prochnik, G.A. Wu, et al. (2016), "Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity", *Nat. Biotechnol.*, **34**(5), pp.562-570.
- [14] D.M. Goodstein, S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D.S. Rokhsar (2012), "Phytozome: A comparative platform for green plant genomics", *Nucleic Acids Res.*, **40**, pp.D1178-1186.
- [15] B. Hu, J. Jin, A.Y. Guo, H. Zhang, J. Luo, G. Gao (2015), "GSDS 2.0: An upgraded gene feature visualization server", *Bioinformatics*, **31**(8), pp.1296-1297.
- [16] S. Kumar, G. Stecher, K. Tamura (2016), "MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets", *Mol. Biol. Evol.*, **33**(7), pp.1870-1874.
- [17] C.V. Ha, M.N. Esfahani, Y. Watanabe, U.T. Tran, S. Sulieman, K. Mochida, D.V. Nguyen, L.S. Tran (2014), "Genome-wide identification and expression analysis of the CaNAC family members in chickpea during development, dehydration and ABA treatments", *PLoS One*, **9**(12), p.e114107.
- [18] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch (2003), "ExPASy: The proteomics server for in-depth protein knowledge and analysis", *Nucleic Acids Res.*, **31**(13), pp.3784-3788.
- [19] O. Emanuelsson, S. Brunak, G. Von Heijne, H. Nielsen (2007), "Locating proteins in the cell using TargetP, SignalP and related tools", *Nat. Protoc.*, **2**(4), pp.953-971.
- [20] O. Emanuelsson, H. Nielsen, S. Brunak, G. Von Heijne (2000), "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence", *J. Mol. Biol.*, **300**(4), pp.1005-1016.
- [21] H.N. Lim, Y. Lee, R. Hussein (2011), "Fundamental relationship between operon organization and gene expression", *Proc. Natl. Acad. Sci. USA*, **108**(26), pp.10626-10631.
- [22] D. Hackenberg, Y. Wu, A. Voigt, R. Adams, P. Schramm, B. Grimm (2012), "Studies on differential nuclear translocation mechanism and assembly of the three subunits of the *Arabidopsis thaliana* transcription factor NF-Y", *Mol. Plant*, **5**(4), pp.876-888.
- [23] S.N. Maity, S. Sinha, E.C. Ruteshouser, B. De Crombrughe (1992), "Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor", *J. Biol. Chem.*, **267**(23), pp.16574-16580.
- [24] Y. Xing, J.D. Fikes, L. Guarente (1993), "Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain", *EMBO J.*, **12**(12), pp.4647-4655.