

Designing of Degenerate Primers-Based Polymerase Chain Reaction (PCR) for Amplification of WD40 Repeat-Containing Proteins Using Local Allignment Search Method

A. Jolodar*

Department of Basic Sciences, Biochemistry and Molecular Biology Section, Faculty of Veterinary Medicine, Shahid Chamran University of Ahvaz, Ahvaz, Islamic Republic of Iran

Received: 29 December 2018 / Revised: 1 May 2019 / Accepted: 26 May 2019

Abstract

Degenerate primers-based polymerase chain reaction (PCR) are commonly used for isolation of unidentified gene sequences in related organisms. For designing the degenerate primers, we propose the use of local alignment search method for searching the preserved areas to design acceptable primers. To test this method, a WD40 repeat-containing domain protein from *Beauveria bassiana* was used as reference sequence to find out a group of homologous gene sequences similar to some closely related to *Verticillium dahliae*. The retrieved sequences from Genbank amino acid sequence databases were used for designing a degenerate primer pairs. The allignments of the selected sequences revealed two highly conserve motif regions of amino acids that enabled us to design a degenerate primers. Forward and reverse degenerate primers for amplification of the related WD40 repeat-containing domain gene were designed from the conserved regions of the coded sequences from some other fungi. By using two rounds of Touchdown Semi-Nested RT- PCR, a partial 249 bp PCR products were isolated and sequenced. The results showed that the amplified fragment between two motifs from *V. dahliae* with the reference protein has 66.3% similarities at the amino acid level. However, this value when the entire of both sequences were compared is 81.9%. These data suggested that homologues transcripts encoded to WD40 repeat-containing protein can be isolated from *V. dahliae* using degenerate primers from the common motif regions through a local allignment search method.

Keywords: Degenerate primers; Fungi; WD40 repeat-containing proteins; *Verticillium dahliae*; PCR.

Introduction

Degenerate primers-based polymerase chain reaction (PCR) are commonly used for isolation of unidentified

gene sequences in related organisms. Generally, the more distant those related organisms, the more difficult it can be to design primers. The standard techniques for designing degenerate primers include several computer-based programs such as Gene Fisher [1], CODEHOP [2]

* Corresponding author: Tel: +986133330073; Fax: +986133360807; Email: jolodara@scu.ac.ir

or PrimaClade [3]. These methods demands a relatively high degree of conservation between all the aligned sequences in multiple global alignments method. Using this kind of algorithms did not reveal obvious preserved regions. As a result, degenerate primer designing, was not achievable. Alternatively, we suggest the employ of local alignment technique to make primers from those sequences with low global similarity. On the basis of this technique, the multiple local alignments in place of the global are performed. This technique was used the Multiple Expectation-Maximization for Motif Elicitation (MEME) algorithm which was originally appropriated to look for preserved patterns in protein sequences.

Many proteins containing the WD40 repeat sequence have been identified in all prokaryotes and eukaryotes [4]. It was revealed that WD40 gene functions are involved in the fiber regulation in cotton [5] and Maize [6]. Many WD40 families have shown sex-related expression during metamorphosis in the Silkworm, *Bombyx mori* [7]. CDC4, a yeast protein is essential for several biological processes at several steps in the cell cycle [8], a neurogenic gene product in *Drosophila* [9], PRP4, PRP17, proteins associated with splicing processing in yeast [10], RACK1, a WD40 repeat family of proteins of *Dictyostelium discoideum* [11] and also *Verticillium dahliae* [12], MAKI 1, an crucial yeast membrane related protein [13] and expression profiling of WD40 genes in foxtail millet were identified [14].

In this study, a homologues transcripts encoded to N-terminus part of WD40 repeat-containing protein is isolated from *Verticillium dahliae*, a soil borne fungus causing wilt in many crops [15]. We employed the local alignment search method to design degenerate primers from the preserved motif areas for the sequences associated to this gene.

Materials and Methods

Sequences obtained from GenBank database

Sequences associated to the WD40 repeat-containing proteins were obtained from the GenBank database, after an extensive study related to target genes in several fungi species. The assemblies from different species included in the family of ascomyces were retrieved, to make a database of proteins associated with *Verticillium*. The maximum approved e-value in this study was $8e-07$; sequence scores less than this value were rejected.

RNA isolation and cDNA synthesis

Total RNAs from fungus were prepared using the RNX plus solution (CinnaGen, Iran) based on the

company's protocols. The extracted total RNA was calculated by absorbance at 260 nm and utilized directly or kept at -80°C . Briefly, 12 μl (2 μg each) of fungal total RNA was incubated with 0.5 μg of ModT (modified oligodT) (5'-GGGTCTAGAGCTCGAGTCACTTTTTTTTTTTTTTTTTT-3') primer at 70°C , for 10 min. The reaction was located on ice before adding 1 μl RNasin (CinnaGen, Iran), 1 μl dNTP mixture (120 mM of each nucleotide), 2.5 μl of 5 X enzyme buffer and 1 μl (200 U) of Moloney Murine Leukemia Virus (M-MuLV) reverse transcriptase (CinnaGen, Iran). The reaction was incubated at 42°C for 1 h before inactivation at 70°C . Reverse transcriptase was excluded in negative control tubes.

Cloning of the Touchdown Semi-Nested RT-PCR products

Degenerate primers were designed on the basis of two highly conserved regions using local alignment search method. For locating the forward and reverse primers, these two conserved regions should be reasonably close together. On the basis of the preserved sequences, two degenerate primers were made. Primer sequences were: forward primer WD-F (5'-GCGgaattcAAYYTNGCNGAYGA), related to N-terminal region of protein (NLADE) and reverse primer WD-R (5'-GCGggatccTCNSWNCRCRTA), related to C-terminal region of protein (YDGSE). (N, R, S, W and Y represent A/G/C/T, A/G, G/C, A/T and C/T, respectively). To make easy the directional cloning, Eco RI and *Bam*HI were introduced to the 5'ends of the forward and reverse primers, respectively [16]. Those restriction sites were shown in lower cases and underlined. As the cleavage sites were located at the ends of the primers, and some restriction endonucleases may digest poorly at this situations, some bases (gcg) were introduced to the 5'-ends to guarantee cutting the amplified fragments. The degenerate primers are designed as a pool of all the possible combinations of nucleotides that could code for a given amino acid sequence. All primers are cautiously analyzed for their specifications. Semi-Nested RT-PCR reactions were done for cDNA by use of touchdown PCR (TD-PCR) conditions. TD-PCR reactions were performed the following conditions: $94^{\circ}\text{C} \times 3$ min for one cycle; $94^{\circ}\text{C} \times 30\text{s}$, $42^{\circ}\text{C} \times 50\text{s}$, $72^{\circ}\text{C} \times 1$ min for 15 cycles by increasing 1°C per cycle; $94^{\circ}\text{C} \times 30\text{s}$, $57^{\circ}\text{C} \times 50\text{s}$, $72^{\circ}\text{C} \times 40\text{s}$ for 20 cycles. PCR products of the first round of amplification were used as template for the second round of amplification. The PCR conditions for second rounds were 30 cycles with denaturation at 94°C (30s), annealing at 57°C (50s) and extension at 72°C (40s)

with a initial denaturation at 95°C (3 min) and final extention at 72°C (5 min). PCR reaction was performed using cDNA from *V. dahlia* as template with the mix: Tris-HCl 20mM, KCl 50mM, MgCl₂ 1.4mM, dNTPs 0.2mM of each, 0.5 U Taq Polimerase, each primer at 0.4 μM, and 5 μl of cDNA. Analytical agarose gel containing 50 ug/ml of safe stain was prepared. PCR products were fractionated on 1% agarose gel electrophoresis and detected by UV transilluminatore. The fragments were inserted into the pET32 plasmid (Invitrogen, Hamburg, Germany) and propagated in *Escherichia coli* TOP10. The selected colonies were cultured on LB-agar supplemented with 100 μg/ml of ampicillin. The positive colonies were selected and cultured overnight in liquid LBA medium. Plasmid DNA was purified using QIA mini prep kit (QIAGEN, Hilden, Germany).

Sequencing and DNA analysis

For sequencing the resulting amplicons, the cloned bands were purified using Qiagen miniprep kit and then sequenced. Sequences were edited and analyzed using BioEdit software. DNA sequences were translated to amino acids to facilitate alignment and aligned using Clustal_W [17]. In order to search for sequence similarity, BLAST program from the NCBI website was used (ncbi.nlm.nih.gov). DNA sequences were analyzed using the BLASTX program. The translated DNA sequences were analyzed using the EXPASY tool (ca.expasy.org/tools/dna.html).

Results

Designing of degenerate primers

Schematic illustration of the local alignment search method used in this study is shown in Figure 1. After the literature review, a homologous sequence of WD40

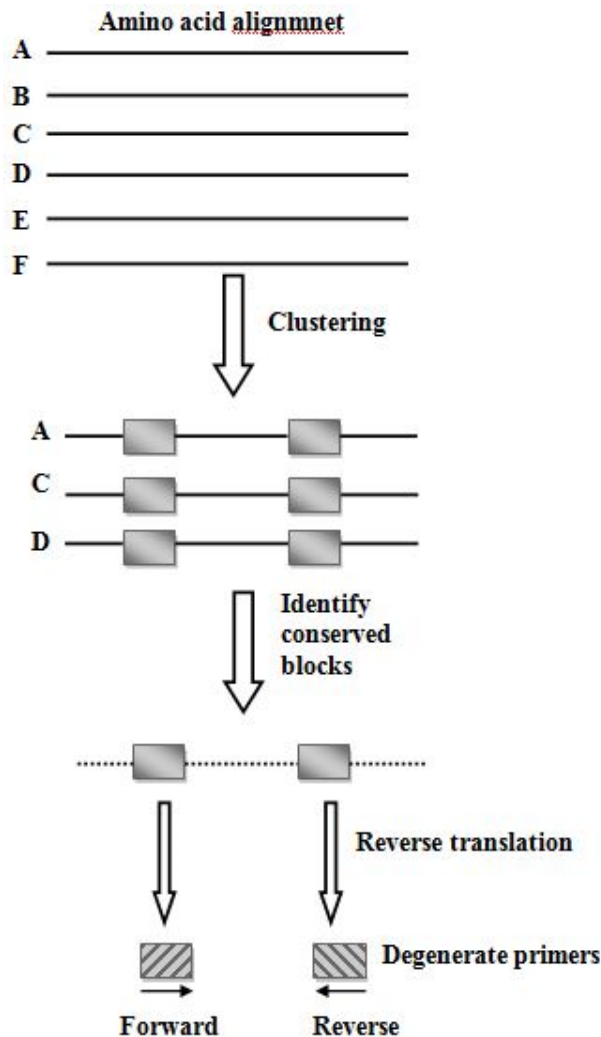


Figure 1. Schematic illustration of the local alignment search method used in this study.

domain-containing protein from *Beauveria bassiana* (Bb; XP_008600572) was determined and selected as reference sequence. For finding out the homologues genes in relative species of *V. dahliae*, we performed

HomoloGene program in the NCBI database. The reference sequence was used as template for searching the NCBI GenBank database by employing BLASTP program. By comparing the reference sequence with the

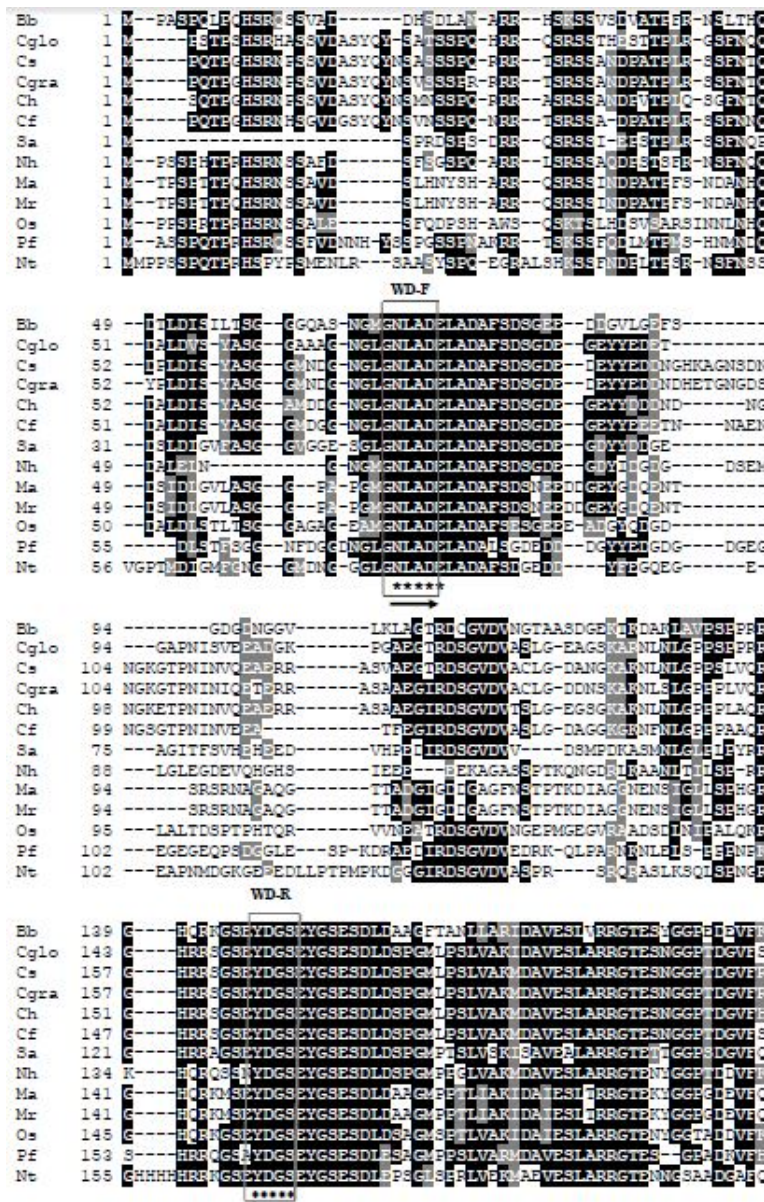


Figure 2. Multiple sequence alignment of WD40 domain containing protein from *Beauveria bassiana* (Bb) with other fungi proteins. The aa sequence of *B. bassiana* (Bb) was aligned with the hypothetical protein of *Colletotrichum gloeosporioides* (Cglo; XP_007280055), hypothetical protein of *Colletotrichum sublineola* (Cs; KDN69486), hypothetical protein of *Colletotrichum graminicola* (Cgra; EFQ26146), hypothetical protein of *Colletotrichum higginsianum* (Ch; CCF37269), hypothetical protein of *Colletotrichum fioriniae* (Cf; XP_007599508), Uncharacterized protein of *Scedosporium apiospermum* (Sa; KEZ46718), predicted protein of *Nectria haematococca* (Nh; XP_003047540), hypothetical protein of *Metarhizium anisopliae* (Ma; XP_007819849), hypothetical protein of *Metarhizium robertsii* (Mr; EXV03610), hypothetical protein of *Ophiocordyceps sinensis* (Os; EQK97708), hypothetical protein of *Pestalotiopsis fici* (Pf; XP_007837810) and hypothetical protein of *Neurospora tetrasperma* (Nt; EGO60803). The aa are represented by one-letter symbols. Shading designates identity (black) or conservative substitutions (grey) associated with the target sequence (Bb). Gaps were designated by dashes. Two protein blocks showed in boxes. Amino acid of motifs are showed by stars. Primers sites indicated by arrows below the sequence.

NCBI database, a total of 12 sequences including, hypothetical protein of *Colletotrichum gloeosporioides* (Cglo; XP_007280055), hypothetical protein of *Colletotrichum sublineola* (Cs; KDN69486), hypothetical protein of *Colletotrichum graminicola* (Cgra; EFQ26146), hypothetical protein of *Colletotrichum higginsianum* (Ch; CCF37269), hypothetical protein of *Colletotrichum fioriniae* (Cf; XP_007599508), Uncharacterized protein of *Scedosporium apiospermum* (Sa; KEZ46718), predicted protein of *Nectria haematococca* (Nh; XP_003047540), hypothetical protein of *Metarhizium anisopliae* (Ma; XP_007819849), hypothetical protein of *Metarhizium robertsii* (Mr; EXV03610), hypothetical protein of *Ophiocordyceps sinensis* (Os; EQK97708), hypothetical protein of *Pestalotiopsis fici* (Pf; XP_007837810) and hypothetical protein of *Neurospora tetrasperma* (Nt; EGO60803) were retrieved. In Figure 2, at least two preserved areas in the alignment part were identified. The conserved regions have at least 5 aa long. In addition to this, primers designed based on these regions, require to pay attention to several points. This includes a sufficient annealing temperature, a proper GC-content value, locating non-degenerate bases at the end of primers and low degeneracy fold [18]. Based on these criteria, the multiple alignments using sequence data of reference sequence shows two blocks of highly preserved protein sequences N-L-A-D-E (for forward primer) and Y-D-G-S-E (for reverse primer). Degenerate primers were chosen from regions of amino acid positions around 69-73 for the forward primers and around 147-151 for the reverse primers (Fig. 2). The next task is to determine the degeneracy fold of each primer. Looking at the amino acid code, some amino acids are coded for by more triplet codon possibilities than others. The degeneracy of a sequence is the number of different codon that it corresponded to. Those two blocks are translated using IUPAC coding system. The fold of degeneracy for each primer is estimated by multiplying the degenerate values of different amino acids using the IUPAC coding system. The fold of degeneracy for both primers was calculated 192. For increasing the PCR efficiencies, a sequence tails were added to the degenerate primers on the 5' ends. This increased the melting point and also facilitating the cloning procedure.

Touchdown Semi-Nested PCR amplification

Starting with 0.5 g of fungi sample, 4 µg of total RNA were extracted and cDNA was synthesized. PCR reactions were performed using the cDNA of *V. dahliae*. Originally, we started with 25 ng of the synthesizing cDNA as template. A primer pair in the first normal

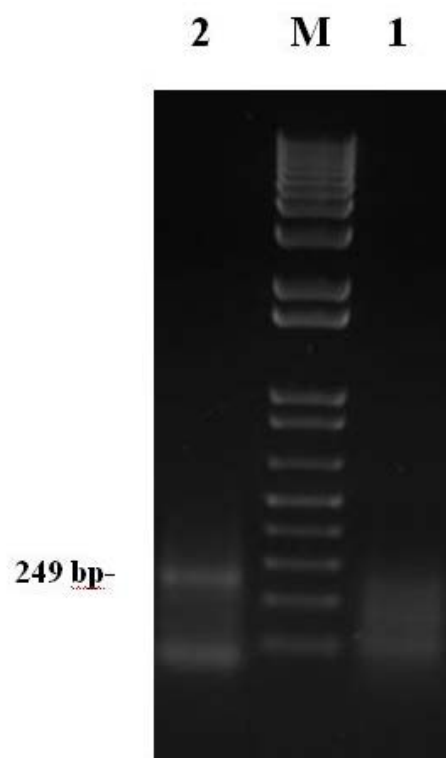


Figure 3 Agarose gel electrophoresis (1%) of WD40 repeat-containing DNA sequences. PCR products from *V. dahliae*. Lane 1 is the first round of Touchdown Semi-nested RT-PCR amplification products, lane 2 is the second round of PCR amplification products and lane M is DNA size marker. Each lane was loaded with 8 [

PCR was tested first with an annealing temperature of 58°C. But, the primers demonstrated no positive products. It means that all amplified products have low expression levels. Since amplification was not observed under these conditions, the cDNA concentration was elevated. Again, we showed no positive results. When more DNA did not improve the results, we decided to perform a TD-PCR technique. In order to facilitate Touchdown Semi-Nested RT-PCR, cDNA was produced using the modified oligo (dT) (Mod-T). This technique let us the amplification of the right transcripts in a two-round PCR protocol. For performing the first round of amplification, ModT-R primer (5'-CCCAGATCTCGAGCTCAGTG) was designed. This was complemented to the 5'-end tail of the modified oligodT (ModT) primer. The first round of Touchdown PCR (TD-PCR) was done by ModT-R and WD-F primers and the original cDNA as template. The second round of PCR was done by using WD-F and WD-R primers and the diluted first PCR (one-tenth) as templates (Fig. 3). It has been observed that the TD-

	10	20	30	40	50	60	70	80
Bb	NLADELADAFSDSGEEDDGVLFGEFSGDGDNG--GVLKLAGTRDCGVVNGTAASDGEKTKDAKLAVSPPPRRGHQRKGSEYDGSE							
Vd	NLADELADAFSDSGEEGE-YYEEDTLDPEQRTIEGREGIDGIRDSDGVDV-AEEVTTGV-TKPTNLNVSPVVRKGHARSQSGYDGSE							
	10	20	30	40	50	60	70	80

Figure 4. Alignments of the translated sequences between two conserved motifs from *B. bassiana* (Bb) with *V. dahliae* (Vd). Gaps were designated by dashes. Double dots represent identities and single dots represent conservative substitutions between *B. bassiana* and *V. dahliae* sequences.

PCR thermal condition can considerably change the achievement rate of degenerate PCR. The TD-PCR starts with a low annealing temperature (42°C) for 15 cycles, followed by 20 cycles at a high annealing temperature (57°C). The low annealing situations permit the short conserved primer regions to hybridize to their complementary strand. After the second cycle of amplification, the 5'-ends of each primer integrated into the amplified products and will provide as a template for following rounds of amplification. By changing to a high annealing temperature situations, specificity can be increased. However, multiple bands were obtained at the first round of PCR. PCR products of the second round of PCR were utilized for cloning and sequencing. Sequence comparison was done with GenBank database using the BLAST program from NCBI Genbank. The primers designed were amplified the putative WD40 repeat gene fragment of 249 bp from the cDNA template in the second round of PCR. This is in concordance with the expected amplicons predicted by manual calculation. It is interesting to note that the first round of amplification gave multiple bands (almost obtained smear products with no more than 0.3Kb in size). However, in the second round of amplification a major band of 249 bp was observed (Fig. 2). The PCR products were cloned into a vector and then constructs were transformed into *Escherichia coli* Top-10F'. For screening the transformed cells, the blue/white bacterial colony method was performed and right DNA clones were purified from the positive bacterial colonies. Clone analysis by PCR was done to confirm the positive transformants.

Comparison of amino acid sequences

Three clones containing product inserts of 249 bp were sequenced after plasmid DNA extraction. After deleting of 5' and 3' vector sequence, multiple alignments were done. The sequence data from all three selected clones were exactly the same. The cDNA sequence was obtained and translated into all possible amino acid sequences and used in BLASX searches. Comparisons between the amplified sequence with the database showed 99% amino acid identity with sequence from *V. dahliae* (EGY23143). When this

sequence from *V. dahlia* was compared with the exact counterpart region of the reference sequence (*Beauveria bassiana* ; XP_008600572), the translated sequence between two conserved motifs shared 52.3 identity and 66.3% similarities at the amino acid level. However, more homology with the comparison of the entire two sequences were found with 60.8 % identities and 81.9% similarities.

Discussion

By synthesizing the degenerate primers, the preserved sequence genes from the related organisms can be isolated. This let us to identify unknown sequences and novel members of gene families. Therefore, applying this technique can be used for sequences those have very few known homologues. It also can be used for many of the reported transcripts with unknown molecular functions. To do this, we designed degenerate primers from several known sequences associated with WD40 domain in different fungi species, for obtaining their related target genes in the genus *Verticillium*.

Creating primers for degenerate PCR needs following several stages. Firstly, the homologous sequences for which degenerate primers are to be created are retrieved from Genbank database. Secondly, a group of homologous Wd40 repeat-containing proteins is to calculate a sequence local alignment of the genes at the protein level. A multiple alignment algorithm was performed for a cluster of less preserved homologous gene sequences associated with WD40 repeat-containing gene. Primers created using these conserved sections are analyzed for several possessions. These contain GC content, melting point and self complementarity. As a final stage, those primers that have all the specifications are selected to amplify the related sequences.

Careful design of a TD-PCR technique was an essential factor for successful amplification using degenerate primers. In this technique, the temperature selected for the annealing step is initially set 5°C lower than the calculated melting point of the primers. Annealing under conditions of low stringency allows the formation of more primer-template hybrids. In

subsequent cycles, the annealing temperature is gradually increased by a small amount so that by the end of the PCR, the annealing temperature is 5°C above the calculated melting point of the primers. This is favors the formation of perfect primer-template hybrids [19].

We made efforts to design degenerate primer from a group of homolog sequences using local multiple search methos to find out two block regions of the highly conserved sequences. Despite of having low sequence similarity outside those two conserved blocks, the sequences between two primers were succesfully amplified.

Acknowledgement

This study was financially supported by research grant from the Vice President of the Research Affairs Office at the Shahid Chamran University of Ahvaz, Ahvaz- Iran.

References

- Giegerich R., Meyer F., Schleiermacher C. (1996) GeneFisher-software support for the detection of postulated genes. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **4**: 68-77(1996).
- Rose T.M., Henikoff J.G., Henikoff S. CODEHOP (Consensus-Degenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res.* **31**(13): 3763-6 (2003).
- Gadberry M.D., Malcomber S.T., Doust A.N., Kellogg E.A. Primaclade-a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics.* **21**(7): 1263-4 (2005).
- Hu X.J., Li T., Wang Y., Xiong Y., Wu. X.H., Zhang D.L. Prokaryotic and Highly-Repetitive WD40 Proteins: A Systematic Study. *Sci Rep.* **7**(1): 10585. (2017).
- Salih H., Gong W., Mkulama M., Du X. Genome-wide characterization, identification, and expression analysis of the WD40 protein family in cotton. *Genome.* **61**(7):539-547. (2018).
- Mascheretti I., Battaglia R., Mainieri D., Altana A., Lauria M., Rossi V. The WD40-Repeat Proteins NFC101 and NFC102 Regulate Different Aspects of Maize Development through Chromatin Modification. *Plant Cell.* **25**(2):404-20. (2013).
- He S., Tong X., Han M., Hu H., Dai F. Genome-Wide Identification and Characterization of WD40 Protein Genes in the Silkworm, *Bombyx mori*. *Int.J. Mol. Sci.* **19**(2): e527(2018).
- Csizmok V, Orlicky S, Cheng J, Song J, Bah A, Delgoshai N. An allosteric conduit facilitates dynamic multisite substrate recognition by the SCF^{Cdc4} ubiquitin ligase. *Nat. Commun.* **8**: 13943 (2017).
- Guillemot F., Billault A., Auffray C. Physical linkage of a guanine nucleotide-binding protein-related gene to the chicken major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* **86**: 4594-4598 (1989).
- Bordonne R., Banroques J., Abelson J., Guthrie C. Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP-snRNP interactions, and pre-mRNA splicing in vivo. *Genes Dev.* **4**(7):1185-1196 (1990).
- Omosigho N.N., Swaminathan K., Plomann M., Müller-Taubenberger A., Noegel A.A., Riyahi T.Y. The Dictyostelium discoideum RACK1 orthologue has roles in growth and development. *Cell Commun. Signal.* **12**: 37 (2014).
- Yuan L., Su, Y., Zhou S., Feng Y., Guo W., Wang X. A RACK1-like protein regulates hyphal morphogenesis, root entry and in vivo virulence in *Verticillium dahliae*. *Fungal Genet. Biol.* **99**:52-61 (2017).
- Icho, T., Wickner, R.B. The MAK11 protein is essential for cell growth and replication of M double-stranded RNA and is apparently a membrane-associated protein. *J Biol Chem.* **263** (3): 1467- 1475(1988).
- Mishra A.K., Muthamilarasan M., Khan Y., Parida S.K., Prasad M. Genome-wide investigation and expression analyses of WD40 protein family in the model plant foxtail millet (*Setaria italica* L.). *PloS. One.* **23**: 9(1) e86852 (2014).
- Neumann M.J., Dobinson K.F. Sequence tag analysis of gene expression during pathogenic growth and microsclerotia development in the vascular wilt pathogen *Verticillium dahliae*. *Fungal Genetics and Biology.* **38**: 54-62 (2003).
- Hou J. Design of endonuclease restriction sites into primers for PCR cloning. *Bioinformatics.* **18**(12): 1690-1 (2002).
- Thompson J.D., Higgins D.G., and Gibson T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680 (1994). DOI: doi.org/10.1093/nar/22.22.4673. PMID: 7984417.
- Kampke T., Kieninger M. Mecklenburg M. Efficient primer design algorithms. *Bioinformatics.* **17**(3): 214-25 (2001).
- Scalon M.C, da Silva T.F, Aquino L.C, Carneiro F.T., Lima M.G., Lemos M.D., Paludo G.R. Touchdown polymerase chain reaction detection of polycystic kidney disease and laboratory findings in different cat populations. *J Vet Diagn Invest.* **26**(4): 542-546 (2014).