Journal of Cell and Tissue Research Vol. 7 (2) 1147-1154 (2007)
ISSN: 0973- 0028 *(Available online at www.tcrjournals.com)* **Original Article**

PRIMER DESIGNING FOR MICROBIAL ENDO-1, 4-b-XYLANASE GENE

GARG, S., SOHANI, N., PUNDHIR, S. AND ANIL KUMAR ?

School of Biotechnology, Devi Ahilya University, Khandwa Road, Indore 452017, India.
E mail: ak_sbt@yahoo.com

Received: April 5, 2006; Accepted: May 4, 2007

Abstract: Primer designing for endo-1,4-b- xylanase gene is done using Primer 3 software. Nineteen sequences for endo-1,4-b- xylanase were retrieved from NCBI's refseq database through Entrez. On multiple sequence alignment of the sequences, three conserved regions were identified. All the combinations of 18 to 24 nucleotides in these regions were selected as potential primers. The properties of these potential primers were analyzed using Premier Biosoft's NetPrimer tool. One forward (5'end) and one backward (3'end) primers having 50 to 60% GC contents, 54 to 58°C Tm and absence of secondary structures were finalized. Specificity of the primers was validated by carrying out the local alignment against the NCBI's nr database through BLAST. All the alignments showed significant alignment to xylanase validating the specificity of the primers.

Key words: Xylanase, Primer design, Primer 3, NetPrimer, BLAST

INTRODUCTION

In the last 10 -15 years, the computer has become an essential companion for cell and molecular biologists. Bioinformatics is an emerging scientific discipline that uses information technology to organize, analyze, and distribute biological information in order to answer complex biological questions. Bioinformatics is an interdisciplinary research area that may be broadly defined as the interface between biological and computational sciences. Bioinformatics has become an essential tool not only for basic research but also for applied research in biotechnology and biomedical science. It involves the solution of complex biological problems using computational tools and systems. It also includes the collection, organization, storage and retrieval of biological information from databases, selection of oligonucleotide primers for polymerase chain reaction (PCR), oligo- hybridization and DNA sequencing. Proper primer design is actually one of the most important factors/ steps in successful DNA sequencing. Optimal primer sequence and appropriate primer concentration are essential for maximal specificity and efficiency of Polymerase Chain Reaction (PCR) amplification. A poorly designed primer can result in little or no product due to non-specific amplification and/or primer-dimer formation that may become competitive enough to suppress product formation. There are several online tools available that are devoted to serve molecular biologist for effective PCR primer design.

Xylan is the second most abundant biopolymer after cellulose and the major hemicellulosic polysaccharide found in the plant cell wall [1]. Xylan is a heteropolymer with backbone of β-1,4-D-xylanopyranosyl residues and branches of neutral or uronic monosaccharides and oligosaccharides [2]. Endo-1, 4-β-xylanase (E.C 3.2.1.8) acts on β-1,4 xylan and cleaves β-1,4 glycosidic linkage randomly [3]. The products are xylose, xylobiose and xylo-oligosaccharides. These products are useful feedstock for food and fine chemicals [4]. It is of industrial importance and is used in paper manufacturing to degrade xylan to bleach paper pulp, increasing the brightness of pulp, improving the digestibility of animal feed and for clarification of fruit juices [5]. Use of xylanase avoids the use of chemical processes that are very expensive and cause pollution

[6-8]. Bajpai et al. [9] showed that chemical extraction of lignin from pulp may be improved by treatment with xylanases. The enzyme has been mostly isolated from microbial and fungal sources [3, 6, 10-13].

Mostly xylanases have been produced using solid state fermentation processes. Singh et al. [14] studied properties of xylanase from *Thermomyces lanuginosus*. Gene coding for xylanase has been cloned and sequenced from number of microbial and fungal sources [14]. Jiang et al. [15] characterized a thermophilic xylanase from *Thermomyces lanuginosus* and discussed its application in bread making. Collins et al. [16] reviewed microbial xylanases. Querido et al. [13] partially purified and characterized xylanase produced by *Penicillium expansum*. However, no report could be found in literature dealing with plant xylanases. Now a days, xylanase gene is used as a reporter gene in plant systems [www.pi.csiro.au/XylanaseAssays/index.htm, 17]. Considering the importance of xylanase enzyme, in the present study, we have designed primers for xylanase gene. Validation of the designed primers has been done by local alignment against the NCBI's nr database through BLAST.

MATERIALS AND METHODS

Web-based resources for primer design: There are numerous web-based resources for PCR primer design. Though most are freely available, they are of variable quality and not well maintained. Therefore, web-based resources often result in missing links and web sites that have been useful previously, may not be functional at a later date. There are number of criteria viz., primer length, Tm, GC contents, 3'-end sequence, dimer formation, false priming, specificity, degenerate primers, hairpin loop that need to be established in the design of primers [18,19].

Software in primer design : The use of software in biological applications has given a new dimension to the field of bioinformatics. Many different programs for the design of primers are now available. Free ware software are available on the internet and many Universities have established servers where a user can log on and perform free analyses of proteins and nucleic acid sequences. There are number of stand –alone programs as well as complex integrated networked versions of the commercial software available. These software packages may be for complete DNA and protein analysis, secondary structure predictions, primer design, molecular modeling, development of cloning strategies, plasmid drawing or restriction endonuclease analyses. Many companies all over the world are engaged in biosoftware development. Some scientists have also developed algorithms and computer programs for various purposes of primer design.

Primer design: Nineteen sequences of Endo-1,4-β-xylanase were taken from GenBank database. Multiple sequence alignment was done using Bioinformatics tool ClustalX. After alignment, the conserved region was taken and primers were designed using software Primer3.

Primer3 software: It is a software developed by Rozen and Skaletsky [20]. It is freely available on Internet (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi). This software is provided by the Whitehead Institute "as is" and any express or implied warranties, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose are disclaimed. The conserved region of the sequence was inserted in the box of the GUI of Primer3. The parameters were adjusted as follows: Tm mini. = 54°C, max. = 58°C, Primer size mini. = 18, max.= 24, GC% mini.= 50, max.= 60.

Gene runner software: Gene runner is a simple and convenient program for sequence analysis. It may be downloaded from the web site http://www.generunner.com. Sometimes one has to try two or three times if it does not work. Gene runner for Windows is now free. Gene runner now has protein motif database searching abilities. One may visit to NCBI web site and download Genbank flat files for searching. One can search dbEST or GenBank for secondary protein structure by protein motifs in all 6 translation frames using Motif runner.

NetPrimer: NetPrimer combines the latest primer design algorithms with a web-based interface allowing the user to analyze primers over the Internet (http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html). All primers are analyzed for melting temperature using the nearest neighbor thermodynamic theory to ensure accurate Tm prediction. Primers are analyzed for all secondary structures including hairpins, self-dimers, and cross-dimers in primer pairs. This ensures the availability of the primer for the reaction as well as minimizing

the formation of primer dimer. The program eases quantitation of primers by calculating primer molecular weight and optical activity. To facilitate the selection of an optimal primer, each primer is given a rating based on the stability of its secondary structures. A comprehensive analysis report can be printed for individual primers or primer pairs.

ClustalX software: ClustalX is a new windows interface for the ClustalW multiple sequence alignment program [21]. It provides an integrated environment for performing multiple sequence and profile alignments for DNA and proteins and analyzing the results. The sequence alignment is displayed in a window on the screen. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

BLAST: The **Basic Local Alignment Search Tool (BLAST)** finds regions of local similarity between sequences [22]. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

RESULTS AND DISCUSSION

On searching for endo-1, 4-β-xylanase in Genbank database through Entrez, total 74 records were found. In order to refine the results, "3.2.1.8[EC/RN number] AND srcdb_refseq[prop]" was fed in the Entrez search Input box, resulting in 19 records. The 3.2.1.8 is the Enzyme Commission number for Endo-1,4-β-xylanase and srcdb_refseq is used to restrict the search results from NCBI RefSeq [23] database only. The database was accessed from the website, http://www.ncbi.nlm.nih.gov/RefSeq. The Reference Sequence (RefSeq) database aims to provide a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products, for major research organisms. Nucleotide sequences corresponding to 19 records were retrieved from RefSeq. The resulting nineteen sequences belong to diverse group of microorganisms namely *Saccharophagus*, *Xanthomonas*,*Pseudoalteromonas*, *Pseudomonas*, *Synechococcus*, *Streptomyces*, *Thermobifida*, *Rhizobium* and *Rhodopirellula*. On Multiple Sequence Alignment (MSA) of nine nucleotide sequences of endo-1,4-β-xylanase (one nucleotide sequence from each organism) using ClustalX, no significant alignment was found and each sequence was highly divergent from the other sequence as evident by the Phylogenetic tree shown in Fig. 1. Phylogenetic tree is constructed using NJPLOT program in ClustalX software. NJPLOT program uses the Neighbor-joining method to calculate the distance between each sequence pair in MSA and develops a distance matrix that is used for building Phylogenetic tree.

Table 1: Various properties of four forward and four backward primers analyzed using NetPrimer.

Name of the primer	Sequence	Hairpin	Dimer	Tm	GC%
F1	GGGACGTTGTCAACGAAGT	1(-1.7)	1(-8.85)	55.17	52.6
F2	GGACGTTGTCAACGAAGTC	1(-1.7)	1(-8.85)	52.85	52.6
F3	TGGGACGTTGTCAACGAAGT	1(-1.7)	1(-8.85)	58.44	50
F4	GGGACGTTGTCAACGAAGTC	1(-1.7)	1(-8.85)	57.5	55
R1	GTCGATGGGAACACCTTG	0	1(-6.76)	53.23	55.56
R2	CGTCGATGGGAACACCTT	0	1(-6.76)	55.81	55.56
R3	CCGTCGATGGGAACACCT	0	1(-6.76)	58.27	61.11
R4	CCCGTCGATGGGAACACC	1(-1.84)	1(-6.76)	61.17	66.67

However, on searching Interpro [24] for any conserved domain in endo-1,4-β-xylanase, glycosyl hydrolase domain was found and endo-1,4-β-xylanase was found to belong to glycosyl hydrolase family 10. The database was accessed from the website, http://www.ebi.ac.uk/interpro. Protein sequences corresponding to 19 records of endo-1,4-β-xylanase were retrieved and proteins predicted as putative endo-1,4-β-xylanase were removed from the dataset resulting in 11 protein sequences. Multiple Sequence alignment of 11 protein sequences was carried out using ClustalX in order to identify conserved region corresponding to the glycosyl hydrolase domain. Protein sequences were chosen against nucleotide sequences because according to equation "E(match by chance)=$\log_e(mn)$", the longest expected run of matches between two random nucleotide sequences and protein sequences by chance is 6.65 and 3.08 respectively, where m and n is the length of two sequence. Thus, protein sequences have less probability of showing significant alignment by chance.

On MSA, three conserved regions were found as shown in Fig. 2, Fig. 3 and Fig. 4, respectively. Nucleotide sequences corresponding to each conserved region in MSA of protein sequence was determined from the corresponding nucleotide

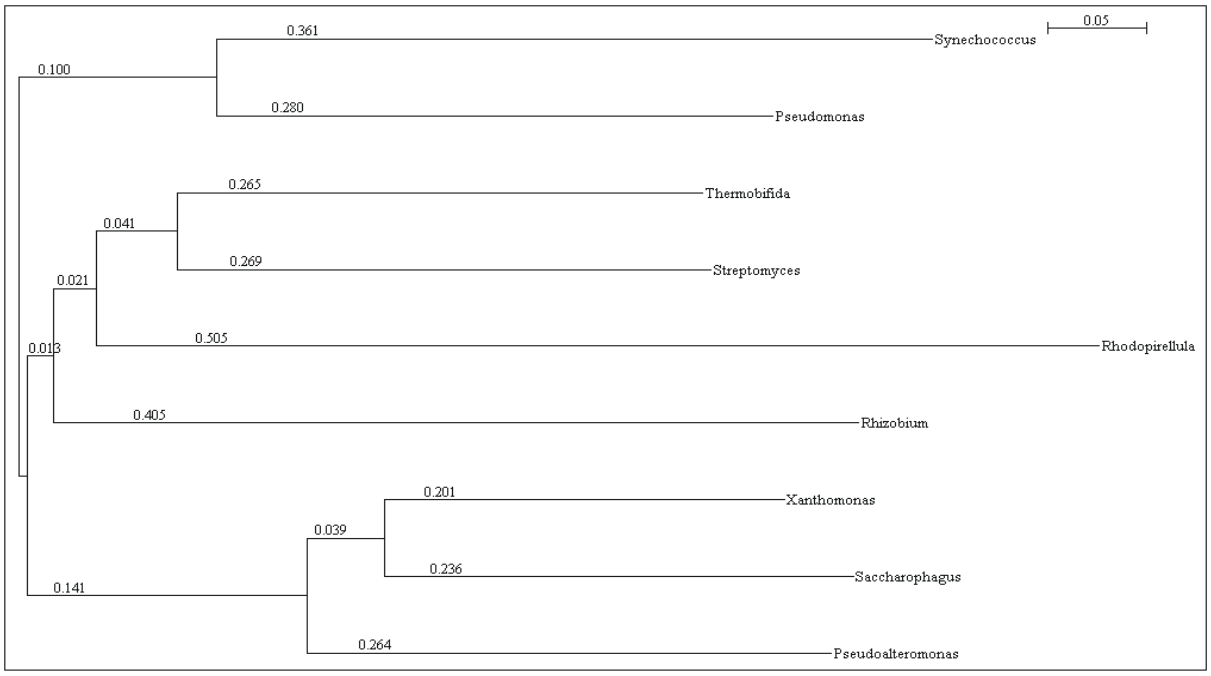

Fig. 1: Highly divergent endo-1,4-β-xylanase. Phylogenetic Tree for nine sequences of endo-1,4-β-xylanase from various microorganisms is shown. There is high divergence between all sequences as evident from the branch lengths.

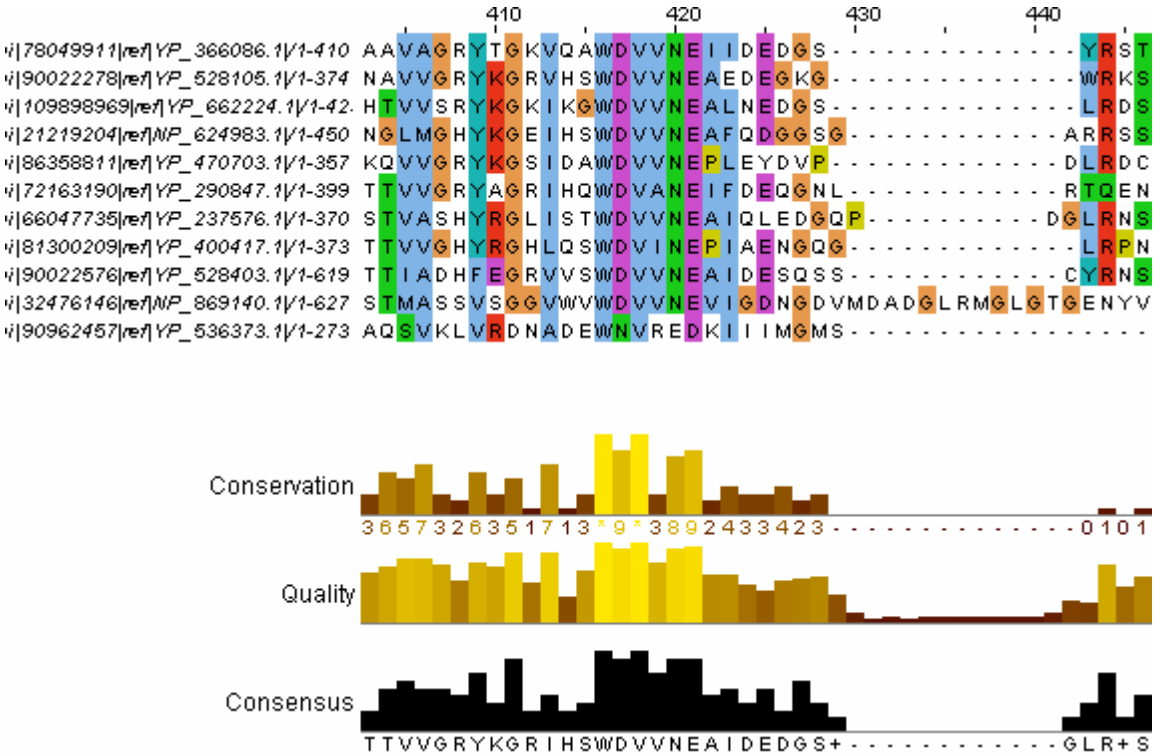

Fig. 2: First Conserved domain in endo-1,4-β-xylanase. Despite high divergence in endo-1,4-β-xylanase. A highly conserved region is found. Blue box indicates the position of conserved region. This is the conserved domain 'First' in our study. Also shown are the Conservation, Quality and Consensus for the aligned region.

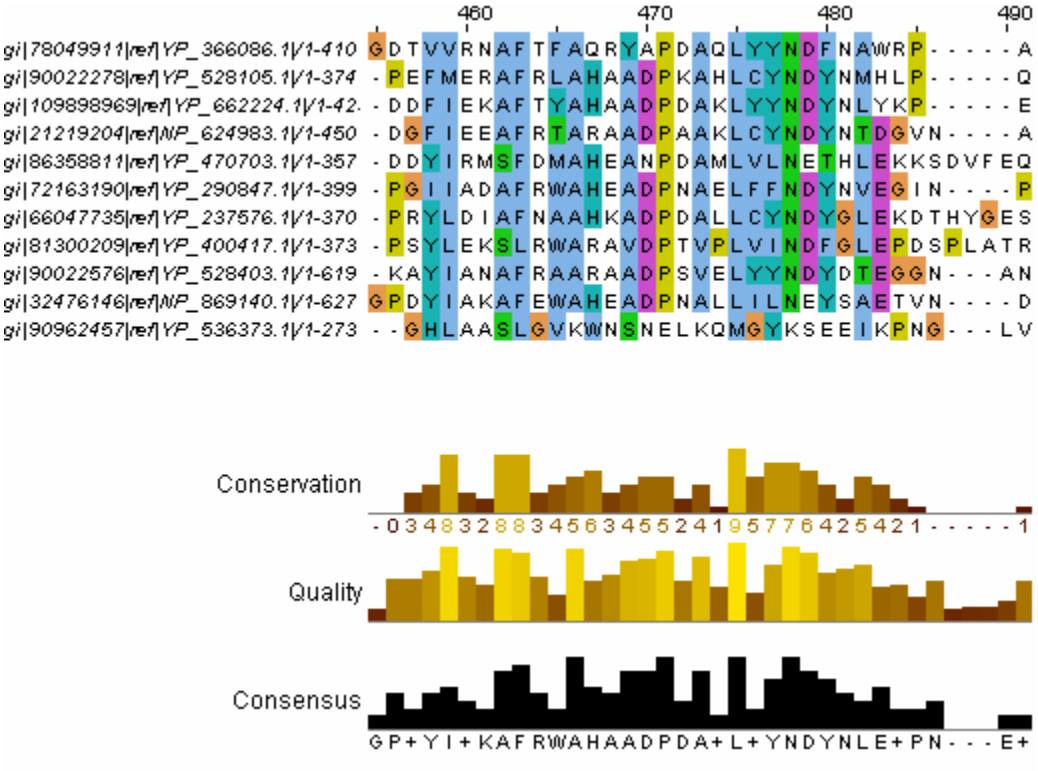

Fig. 3: Second Conserved domain in endo-1,4-β-xylanase. A highly conserved region in endo-1,4-β-xylanase. Blue box indicates the position of conserved region. This is the conserved domain 'Second' in our study. Also shown are the Conservation, Quality and Consensus for the aligned region.

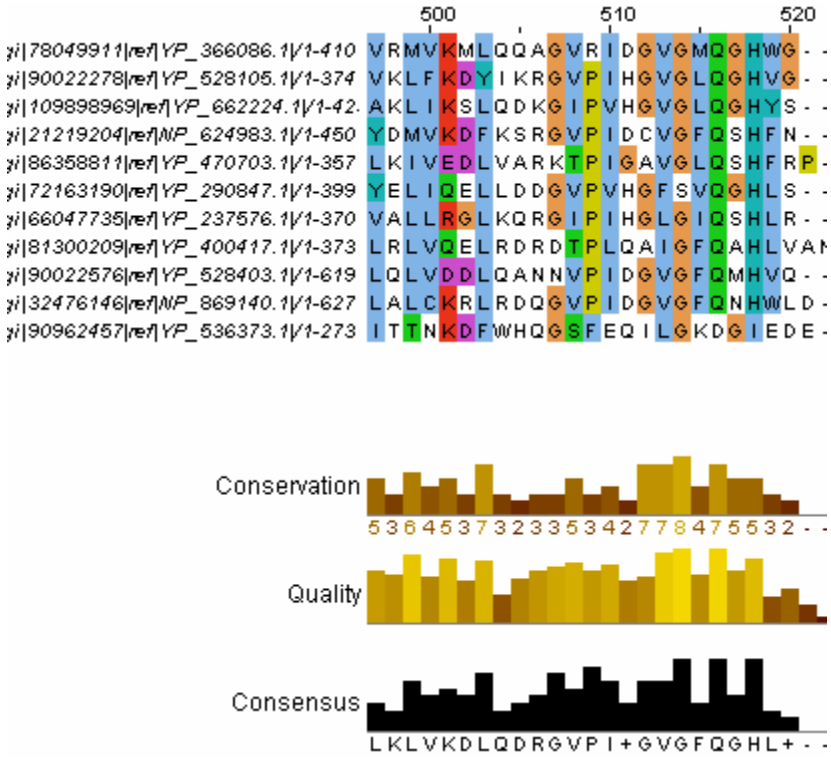

Fig. 4: Third Conserved domain in endo-1,4-β-xylanase. A highly conserved region in endo-1,4-β-xylanase. Blue box indicates the position of conserved region. This is the conserved domain 'Third' in our study. Also shown are the Conservation, Quality and Consensus for the aligned region.

```
Primer3 Output

--------------------------------------------------------------------------------

PRIMER PICKING RESULTS FOR gi|32470666:c5342036-5340153 Rhodopirellula baltica SH 1,
complete genome

No mispriming library specified
Using 1-based sequence positions

OLIGO            start  len      tm     gc%   any    3' seq
LEFT PRIMER        641   19   58.55   52.63  9.00  4.00 GGGACGTTGTCAACGAAGT
RIGHT PRIMER       916   18   59.47   55.56  4.00  0.00 CGTCGATGGGAACACCTT
SEQUENCE SIZE: 1884
INCLUDED REGION SIZE: 1884

PRODUCT SIZE: 276, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 2.00

  601 TCGACGATGGCGAGCAGTGTGTCCGGCGGTGTTTGGGTGTGGGACGTTGTCAACGAAGTC
                                                >>>>>>>>>>>>>>>>>>>

  661 ATTGGTGACAACGGCGACGTGATGGATGCCGATGGGCTTCGGATGGGGCTTGGTACCGGT


  721 GAAAACTATGTGCCCTACAAAGAGTACGCGGCGATGGGGCCGGACTACATCGCGAAAGCG


  781 TTTGAGTGGGCTCACGAAGCCGATCCGAACGCGTTGCTGATTCTCAACGAGTACTCCGCC


  841 GAAACGGTGAACGACAAATCGGACCGTTTGCTGGCGTTGTGCAAGCGACTGCGAGACCAA
                                                              <<

  901 GGTGTTCCCATCGACGGGGTTGGTTTTCAAAACCACTGGCTCGACCTTCGCTACGAACCC
      <<<<<<<<<<<<<<<<
```

Fig. 5: Primer3 output for the designed primers. Primer3 results validating the sensitivity of the designed primers for endo-1,4-β- xylanase.

```
> emb|BX294150.1|  Rhodopirellula baltica SH 1 complete genome; segment 18/24
Length=283050

 Features in this part of subject sequence:
   endo-1,4-beta-xylanase

 Score = 38.2 bits (19),  Expect = 0.023
 Identities = 19/19 (100%), Gaps = 0/19 (0%)
 Strand=Plus/Minus

Query  1       GGGACGTTGTCAACGAAGT  19
               ||||||||||||||||||
Sbjct  156996  GGGACGTTGTCAACGAAGT  156978

> dbj|AB098080.1|  Paenibacillus sp. W-61 xyn5 gene for xylanase 5, complete cds
Length=4129

 Score = 36.2 bits (18),  Expect = 0.091
 Identities = 18/18 (100%), Gaps = 0/18 (0%)
 Strand=Plus/Plus

Query  1     GGGACGTTGTCAACGAAG  18
              ||||||||||||||||||
Sbjct  1605  GGGACGTTGTCAACGAAG  1622

> gb|CP000075.1|  Pseudomonas syringae pv. syringae B728a, complete genome
Length=6093698

  Features in this part of subject sequence:
    Endo-1,4-beta-xylanase

  Score = 36.2 bits (18),  Expect = 0.091
  Identities = 18/18 (100%), Gaps = 0/18 (0%)
  Strand=Plus/Plus

Query  1        GGGACGTTGTCAACGAAG  18
                ||||||||||||||||||
Sbjct  5372314  GGGACGTTGTCAACGAAG  5372331
```

Fig. 6. BLAST results for designed primers. Primers showed significant alignment to endo-1,4-β-xylanase.

sequence and all combinations of 18 to 24 nucleotide (ideal length of Primer) long sequences were selected as potential primer. Nucleotide sequences corresponding to second conserved region, being highly rich in G+C content were not used for further work.

The various properties namely Hairpin loop, Primer dimer, Tm and GC% were calculated for all potential primers using Premier Biosoft's NetPrimer tool. The results of NetPrimer for four forward and four backward primers are shown in Table 1. As per results, one forward (F1) and one backward (R2) primers having 50 to 60 %GC contents, 54 to 58°C Tm and absence of secondary structures were finalized. Primer3 was used to check the product made by the primers, nucleotide sequence of endo-1,4-β-xylanase was fed in the box of input sequences of the GUI of Primer3. Minimum and maximum temperature, 54 ^{0}C and 58 ^{0}C respectively; minimum and maximum primer size 18 to 24 nucleotides respectively, and GC% minimum and maximum 50 and 60 respectively, were set in the appropriate box of the GUI of the software. Afterwards, pick primer button was clicked. Primer3 results showed the primer designed as valid primers as shown in Fig. 5. Specificity of the primers was validated by carrying out the local alignment against the NCBI's nr database through BLAST. The BLAST search was carried against the non-redundant (nr) database and Bacteria specified as the Organism. All the results showed significant alignment to xylanase thus validating the specificity of the primers as shown in Fig 6. The designed forward and backward primers may be used for PCR amplification of endo-1,4-β-xylanase gene using microbial genomic DNA.

ACKNOWLEDGEMENTS

The authors acknowledge the Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi for its facilities under the Bioinformatics Sub Centre and also under M.Sc. Biotechnology Program sanctioned to this School. SG also acknowledges the bioinformatics studentship awarded to her by the Bioinformatics Sub Centre.

REFERENCES

[1] Timell, T.E.: Wood Sci. Technol., 1: 45-70 (1967).

[2] Joseleau, J.P., Combat, J. and Ruel, K.: In: *Xylans and Xylanases* (Visser, J., Beldman, G. Van-Someren, M.A.K., and Voragen, A.G. J. eds.), Elsevier, Amsterdam, pp 1-15 (1992).

[3] Khandeparkar, R.D.S. and Bhosle, N.B.: Enzyme Microbial Technol., 39: 732-742 (2006).

[4] Gilbert, H.J. and Hazlewood, G.P.: J. Gen. Microbiol., 139: 187-194 (1993).

[5] Biely, P., Markovic, O. and Mislovicova, D.: Anal. Biochem., 144: 147-151 (1985).

[6] Ruiz-Arribas, A., Fernandez-Abaloz, J.M., Sanchez, P., Garda, A.L. and Santamaria, R.I. : Appl. Environ. Microbiol., 61: 2414-2419 (1995).

[7] Viikari, L.: FEMS Microbiol. Rev., 13: 335-350 (1994).

[8] Srinivasan, M.C. and Rele, V.M.: Ferment. Sci. Technol., 137-162 (1988).

[9] Bajpai, B., Bhardwaj, N.K., Bajpai, P.K. and Jauhari, M.B.: J. Biotechnol., 38: 1-6 (1994).

[10] Ghatora, S.K., Chadha, B.S., Badhan, A.K., Saini, H.S. and Bhat, M.K.: Bioresources, 1: 18-33 (2006).

[11] Bachmann, S.L. and McCarthy, A.J.: Appl. Environ. Microbiol., 57: 2121-2130 (1991).

[12] Yang, V.W., Zhuang, Z., Elegir, G. and Jeffries, T.W.: J. Indus. Microbiol. 15: 434-441 (1995).

[13] Querido, A.L.D.S., Coelho, J.L.C., Araujo, E.F.D. and Chaves-Alves, V.M.: Brazilian Arch. Biol. Technol., 49: 475-480 (2006).

[14] Singh, S., Madlala, A. M. and Prior, B.A.: FEMS Microbiol. Rev., 27: 3-16 (2003).

[15] Jiang, Z.Q., Yang, S.Q., Tan, S.S., Li, L.T. and Li, X.T.: Letters Appl. Microbiol., 41: 69-76 (2005).

[16] Collins, T., Gerday, C. and Feller, G.: FEMS Microbiol. Rev., 29: 3-23 (2005).

[17] Herbers, K., Wilke, I. and Sonnewald, U.: Biotechnology, 13: 63-66 (1995).

[18] Dieffenbach, C.W., Lowe, T.M.J. and Dveksler, G.S.: In: *PCR Primer, A Laboratory Manual* (Dieffenbach, C.W. and Deveksler, G.S. eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 133-155 (1995).

[19] Singh, V.K. and Kumar, A.: Mol. Biol. Today, 2: 27-32 (2001).

[20] Rozen, S. and Skatelsky, H.J.: In: *Bioinformatics Methods and Protocols* :*Methods in Molecular Biology* (Krawetz, S. and Misener, S. eds.), Humana Press, Totowa, New Jersey, pp 365-386 (2000).

[21] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G.: Nucleic Acids Research, 24: 4876-4882 (1997).

[22] Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman Nucleic Acids Res., 25: 3389-3402 (1997).

[23] Pruitt K.D., Tatusova T. and Maglott D.R.: Nucleic Acids Research, 33: D501-D504 (2003).

[24] Mulder N.J., Apweiler R., Attwood T.K., Bairoch A., Bateman A., Binns D., Bradley P., Bork P., Bucher, P., Cerutti L., Copley R., Courcelle E., Das U., Durbin R., Fleischmann W., Gough J., Haft D., Harte N., Hulo N.,

Kahn D., Kanapin A., Krestyaninova M., Lonsdale D., Lopez R., Letunic I., Madera M., Maslen J., McDowall J., Mitchell A., Nikolskaya A.N., Orchard S., Pagni M., Ponting C.P., Quevillon E., Selengut J., Sigrist C.J., Silventoinen V., Studholme D.J., Vaughan R., and Wu C.H.: Nucleic Acids Res. 33 (Database Issue): D201-205 (2005).