ISSN 2278–5469
EISSN 2278–5450

Discovery

In-silico analysis and identification of *Bordetella sp.* from domestic waste water

Geetha S[1☼], Brindha V[2]

1. Department of Biomedical Engineering, Bharath University, Seliayur-600073, Chennai, Tamil Nadu, India, E-mail: Geethag24@gmail.com
2. Department of Biotechnology, Hindustan College of Arts and Science, Chennai, Tamilnadu, India, E-mail: brindhav2@gmail.com

☼Corresponding author:
Geetha S, Department of Biomedical Engineering, Bharath University, Selaiyur-600073, Chennai, Tamil Nadu, India, E-mail: Geethag24@gmail.com

Publication History
Received: 24 April 2012
Accepted: 29 May 2012
Published: 1 July 2012

Citation
Geetha S, Brindha V. *In-silico* analysis and identification of *Bordetella sp.* from domestic waste water. *Discovery*, 2012, 1(1), 18-24

General Note
Article is recommended to print as color digital version in recycled paper.

ABSTRACT

Discharge of house hold waste water into surface water may present a variety of concerns such as high biochemical oxygen demand, significant nutrient inputs, high suspended solids, ecosystem disturbance and health hazard due to potential pathogens. House hold waste water consists of black water (urine and faeces) and grey water from the kitchen, bathroom and washing machine. Bacteria are the most important group of microorganisms found in this biological contact process. Water plays a very important role in supporting all forms of life. However if contaminated, it has great potential for transmitting a variety of diseases and illnesses. The

objective of the work is to study the molecular characterization by isolating DNA from the selected organism. The conserved region was amplified and the amplified product was sequenced based on 16s rRNA. The fasta format of the sequence was submitted in BLAST for the similarity searches. The functional region was predicted.

Keywords: Pertussis; *Bordetella parapertussis*; In-silico analysis; Bioinformatics.

Abbreviations: HMM - Hidden Markov model; BLAST - Basic Local Alignment Search Tool; ORF - Open reading frame.

1. INTRODUCTION

Bordetella parapertussis is a small Gram-negative bacterium of the genus *Bordetella* which is adapted to colonize the mammalian respiratory tract. Pertussis caused by *B. parapertussis* manifests with similar symptoms to *B. pertussis*. *Bordetella pertussis* is the only organism of major clinical significance in this genus, causing whooping cough in infants and young children (Ashok Gadgil, 1998). However, a closely related organism, *B. parapertussis* can also cause a milder form of bronchitis. It is an extremely small, strictly aerobic, Gram negative, non-motile cocobacillus (short rod). Compared to other *Bortdetella species, B. pertussis* does not grow on common laboratory media. *B. pertussis* can be distinguished from *B. parapertussis* in that *B. pertussis* is oxidase positive but urease negative, while *B. parapertussis* is oxidase negative and urease positive. *B. bronchosepticus* is positive for both enzymes. Symptoms are characteristic (Blackall, 1994). Laboratory diagnosis is made by obtaining a nasopharyngeal aspirate and primary culture in Bordet-Gengou medium or potato-glycerol-blood agar. Growth is inhibited by peptones, unsaturated fatty acids, sulphides, etc. found in ordinary media. The organism grows as small transparent hemolytic colonies on blood agar. It can be serologically distinguished from *B. parapertussis* and *B. bronchosepticus* (Boddinghaus et al., 1990).

>/tmp/readseq.in.1338 [Unknown form], 783 bases
acttcggtctggtggcgagtggcgaacgggtgagtaattgtatcggaacg
tgcccagtagcgggggataactacgcgaaagcgtggctaataccgcatac
gccctacgggggaaagcgggggaccttcgggcctcgcactattggagcgg
ccgatatcggattagctagttggtggggtaacggctcaccaaggcgacga
tccgtagctggtttgagaggacgaccagccacactgggactgagacacgg
cccagactcctacgggaggcagcagtggggaattttggacaatgggggca
accctgatccagccatcccgcgtgtgcgatgaaggccttcgggttgtaaa
gcacttttggcaggaaagaaacggcgccggctaatacctggcgctaatga
cggtacctgcagaataagcaccggctaactacgtgccagcagccgcggta
atacgtagggtgcaagcgttaatcggaattactgggcgtaaagcgtgcgc
aggcggttcggaaagaaagatgtgaaatcccagagcttaactttggaact
gcatttttaactaccgggctagagtgtgtcagagggaggtggaattccgc
gtgtagcagtgaaatgcgtagatatgcggaggaacaccgatggcgaaggc
agcctcctgggataacactgacgctcatgcacgaaagcgtggggagcaaa
caggattagataccctggtagtccacgccctaaacgatgtcaactagctg
ttggggccttcgggccttagtagcgcagctaac

Figure 1
FASTA format of the sequence

The GeneMark gene-prediction algorithm was developed in several steps. The first step was performed by a group at the Institute of Molecular Genetics in Moscow in 1986. It was demonstrated that markov chain models were useful tools for DNA sequence analysis and particularly for gene prediction. The GeneMark.hmm algorithm (Lukashin and Borodovsky, 1998) was designed to improve gene-prediction quality in terms of finding exact gene boundaries. GeneMark program identified a gene mainly as the open reading frame where the gene is residing. However, the 5′ boundary of the gene (the translation initiation codon associated with the protein amino terminus) might not be precisely predicted. GeneMark indicates several possible start codons and scores them. The underlying idea of GeneMark.hmm was to embed the GeneMark models for coding and noncoding regions into the naturally derived hidden Markov model (HMM) framework, with gene boundaries modeled as transitions between hidden states. BLAST (Basic Local Alignment Search Tool) is a web-based program that is able to align your search sequence to thousands of different sequences in a database (that you choose) and shows you a list of the top matches. This program can search through a database of thousands of entries in under a minute. BLAST performs its alignment by matching up each position of your search sequence to each position of the sequences in the database. For each position, BLAST gives a positive score if the nucleotides match. BLAST can also insert gaps when performing the alignment. Each gap inserted has a negative effect on the alignment score, but if enough nucleotides align as a result of the gap, this negative effect is overcome and the gap is accepted in the alignment (Malik, 2008). These scores are then used to calculate the alignment score in “bits” which is converted to the statistical E-value. A

high bit score correlates with a low E-value. The lower the E value, the more similar the sequence found in the data base is to your query sequence (Wu et al., 2001). The objective of this work is to predict the similarity search of the microbe, submit the sequence in gene bank, analyze the phylogenetic relationship between the species, predict the functional region of the sequence and predict the promoter site of the sequence for transcription.

2. MATERIALS AND METHODS

2.1. Sample Collection

Water samples are collected in a sterile containers from the residential (cooking and washing) waste water. It was then plated on the nutrient agar medium for the growth of microbes.

2.2. Identification of microbes

Characterizations of the selected isolates are confirmed by various physiological and biochemical test.

2.3. Sequence analysis

DNA was isolated from *Bordetella sp.* It was then amplified and sequenced based on 16s rRNA.

2.4. In–silico analysis

Sequence was submitted in BLAST for similarity searches. Evolutionary and Phylogenetic relationship was identified using MSA, Phylip and ATV software. Functional region of the sequence was identified using ORF finder. Coding region of the sequence was identified using Gene Mark. Promoter site region of the sequence was identified using promoter finder.

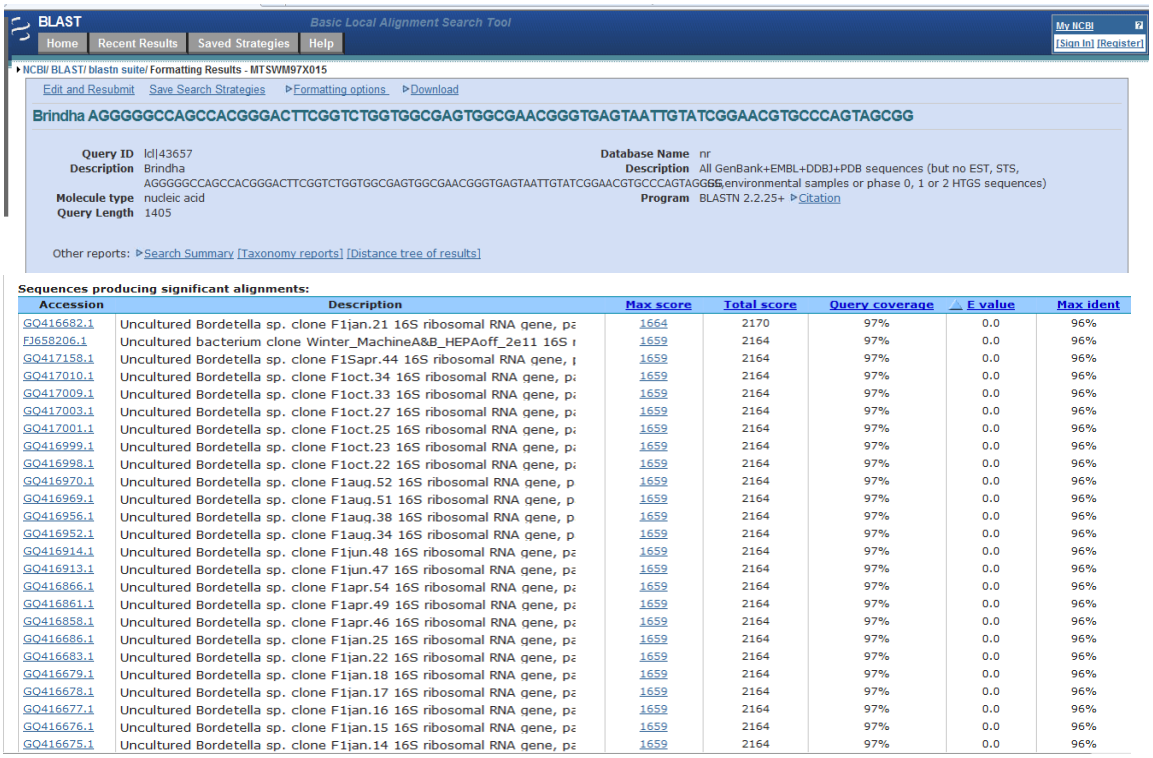

Figure 2

Similarity search of the sequence using BLAST

2.5. Genbank submission

The isolated species was submitted in Genbank for the identification of new species.

3. RESULTS AND DISCUSSION

3.1. Sequencing of amplified gene based on 16s rRNA

Sequence analysis of the 16s rRNA sequences is done with the help of several primers called universal primers. These primers target the conserved region of the 16s rRNA gene and amplify the target in parts. Finally the several amplified parts could be assembled together to have the entire sequence. The fasta format of the sequence was shown in Fig.1.

3.2. In – silico analysis of the sequence

The fasta format sequence was submitted in blast for similarity searches (Fig.2).

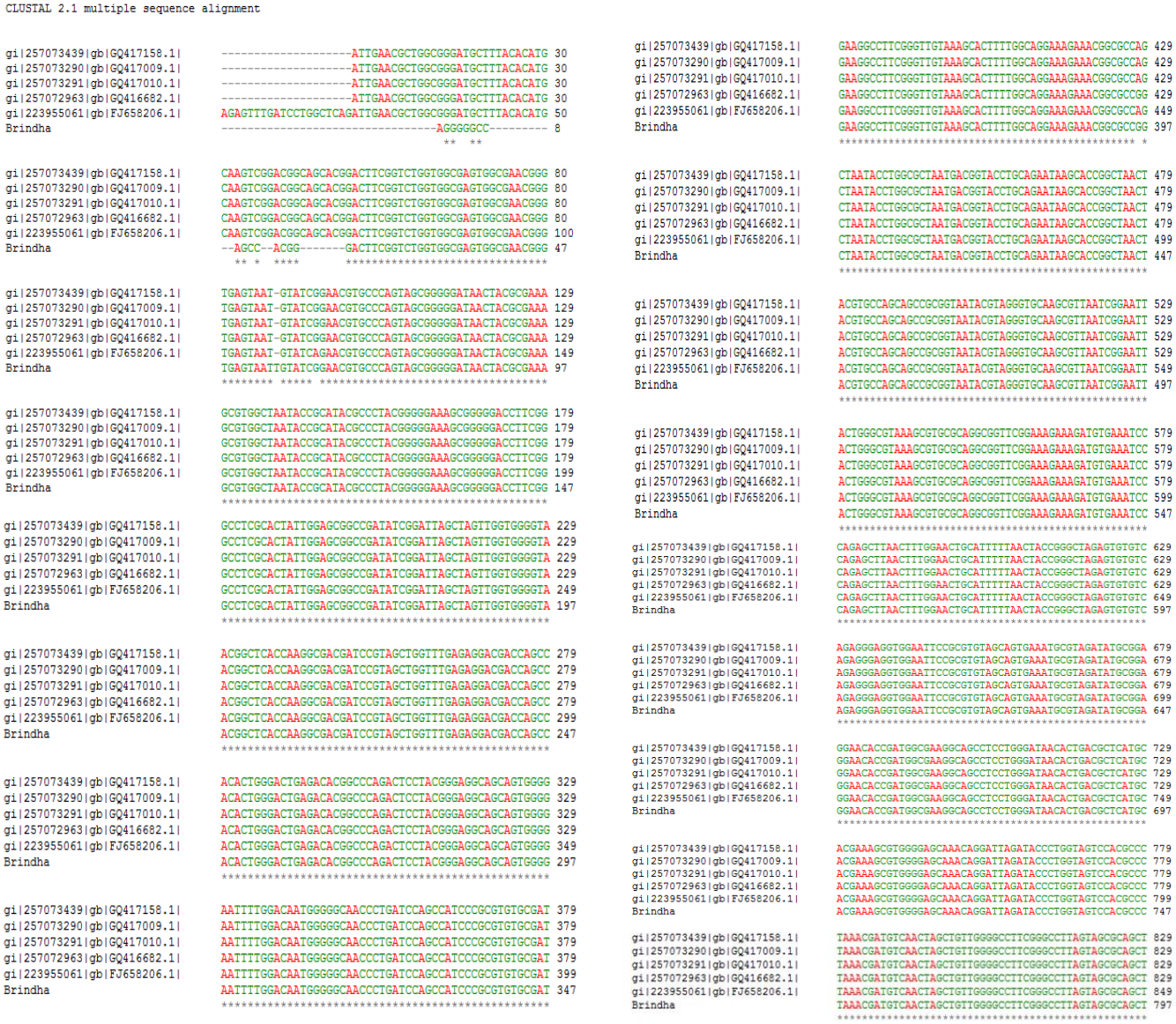

Figure 3

Evolutionary relationships between the species using clustalW

3.3. Multiple sequence analysis

Evolutionary relationship of the organism was identified using the tools clustalw and the Phylogenetic relationship between the organisms was identified using ATV software. The distance between the selected species was identified by dendrogram graphical representation and was shown in Fig.3. The phylogenetic relationship between the species was shown in Fig.4.

3.3.1. Genbank Submission

The identified sequence was directly submitted to Genbank (Fig.5). The Genbank provided accession number for the submitted nucleotide sequence. The accession number is BankIt1435630 Bordetella JF420884.

3.3.2. Gene Prediction

Functional region of the selected organism was identified by submitting the sequence to orf finder, Genemark and Promoter finder. The functional regions were shown in Fig. 6, 7 and 8. The Gene Mark graphical output depicts the coding potential in the six possible reading frames. An unbroken horizontal line at the 0.5 level indicates an open reading frame (ORF). Large vertical lines above the 0.5

level indicate “ATG” codons. Vertical lines under the 0.5 level represent one of the three stop codons, “TAA,” “TGA,” or “TAG.” The thick gray horizontal line indicates a region of interest.

Figure 4
Phylogenetic relationships between the species using ATV software

4. CONCLUSION

There are many practical applications for identifying bacteria from waste water. One of the main responsibilities is to determine the identity of pathogenic bacteria. It is very important to determine the specific type of bacterium causing disease so that the physician able to correctly treat the patients. *Bordetella sps.* colonize the mammalian respiratory tract and cause whooping cough in infants and young children. It also causes hypoglycemia, increased histamine and endotoxin sensitivity. The toxin inhibits many leukocyte functions, including chemotaxis, phagocytosis and respiratory burst and impairs NK cell killing. Treatment is by routine acellular vaccine immunization. Present study investigates only to identify the pathogenic organism from waste water based on its molecular characters.

SUMMARY RESEARCH

1.Water plays a very important role in supporting all forms of life. However if it is contaminated, it has great a potential for transmitting a variety of diseases and illnesses to human beings via pathogens like bacteria.
2.In the proposed work, DNA of *Bordetella parapertussis* was isolated and its molecular characterization was analyzed by bioinformatics tools.

FUTURE ISSUES

1.Predict the 3D structure of the protein/ toxin which causes the disease whooping cough.
2.Design a drug for the disease whooping cough using medicinal plants.

DISCLOSURE STATEMENT

There is no financial support for the proposed research work.

ACKNOWLEDGMENTS

We wish to express our deep felt gratitude and heartiest thanks to our Director, Principal and Lecturers those who all have given enormous support and their inspiring suggestions for our research work. Moreover, I thank the Almighty, for their blessings.

REFERENCES

1. Ashok Gadgil. Drinking water in developing countries. *Annu.Rev.Energy Environ*. 1998, 23, 253-286
2. Blackall LL. Molecular identification of activated sludge foaming bacteria. *Water Sci Technol.* 1994, 29(7), 35-42
3. Boddinghaus B, Rogall T, Flohr T, Blocker H, Bottger EC. Detection and identification of mycobacteria by amplification of rRNA. *J.Clin.Microbiol*. 1990, 28, 1751-759
4. Lukashin AV, Borodovsky M. Gene Mark. hmm: new solutions for gene finding. *Nucleic acids Res*. 1998, 26, 1107-1115
5. Malik S, Beer M, Megharaj M, Naidu R. The use of molecular techniques to characterize the microbial communities in contaminated soil and water. *Environ.Int*. 2008, 34, 265-276
6. Wu L, Thompson DK, Li G, Hurt RA, Tiedje JM, Zhou J. Development and evaluation of functional gene arrays for the detection of selected genes in the environment. *Appl. Environ. Microbiol*. 2001, 67, 5780–5790

```
LOCUS       Bordetella              783 bp    DNA     linear   BCT 22-FEB-2011
DEFINITION  species Brindha-01 16S rRNA partial sequence.
ACCESSION   Bordetella
VERSION
KEYWORDS    .
SOURCE      Bordetella
  ORGANISM  Bordetella
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Alcaligenaceae.
REFERENCE   1  (bases 1 to 783)
  AUTHORS   Brindha,V.
  TITLE     Molecular charecterization and identification of an unknown
            organism from waste water
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 783)
  AUTHORS   Brindha,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2011) Department of Biotechnology, Hindustan
            College of arts and science, Padur, Chennai, Tamilnadu 603103,
            INDIA
COMMENT     Bankit Comment: TOTAL # OF SEQS:1.
            Sequences were screened for chimeras by the submitter using Pintail
            tool 2007.

FEATURES             Location/Qualifiers
     source          1..783
                     /organism="Bordetella"
                     /mol_type="genomic DNA"
                     /isolation_source="Waste water"
                     /db_xref="taxon:517"
                     /clone="Brindha-01"
                     /country="India"
                     /note="[uncultured (using universal primers)]"
     rRNA            <1..>783
                     /product="16S ribosomal RNA"
BASE COUNT      195 a    183 c    256 g    149 t
ORIGIN
        1 acttcggtct ggtggcgagt ggcgaacggg tgagtaattg tatcggaacg tgcccagtag
       61 cgggggataa ctacgcgaaa gcgtggctaa taccgcatac gccctacggg ggaaagcggg
      121 ggaccttcgg gcctcgcact attggagcgg ccgatatcgg attagctagt tggtggggta
      181 acggctcacc aaggcgacga tccgtagctg gtttgagagg acgaccagcc acactgggac
      241 tgagacacgg cccagactcc tacgggaggc agcagtgggg aattttggac aatgggggca
      301 accctgatcc agccatcccg cgtgtgcgat gaaggccttc gggttgtaaa gcacttttgg
      361 caggaaagaa acggcgccgg ctaatacctg gcgctaatga cggtacctgc agaataagca
      421 ccggctaact acgtgccagc agccgcggta atacgtaggg tgcaagcgtt aatcggaatt
      481 actgggcgta aagcgtgcgc aggcggttcg gaaagaaaga tgtgaaatcc cagagcttaa
      541 ctttggaact gcatttttaa ctaccgggct agagtgtgtc agagggaggt ggaattccgc
      601 gtgtagcagt gaaatgcgta gatatgcgga ggaacaccga tggcgaaggc agcctcctgg
      661 gataacactg acgctcatgc acgaaagcgt ggggagcaaa caggattaga taccctggta
      721 gtccacgccc taaacgatgt caactagctg ttggggcctt cgggccttag tagcgcagct
      781 aac
//
```

Figure 5
Sequence submitted to Genbank.

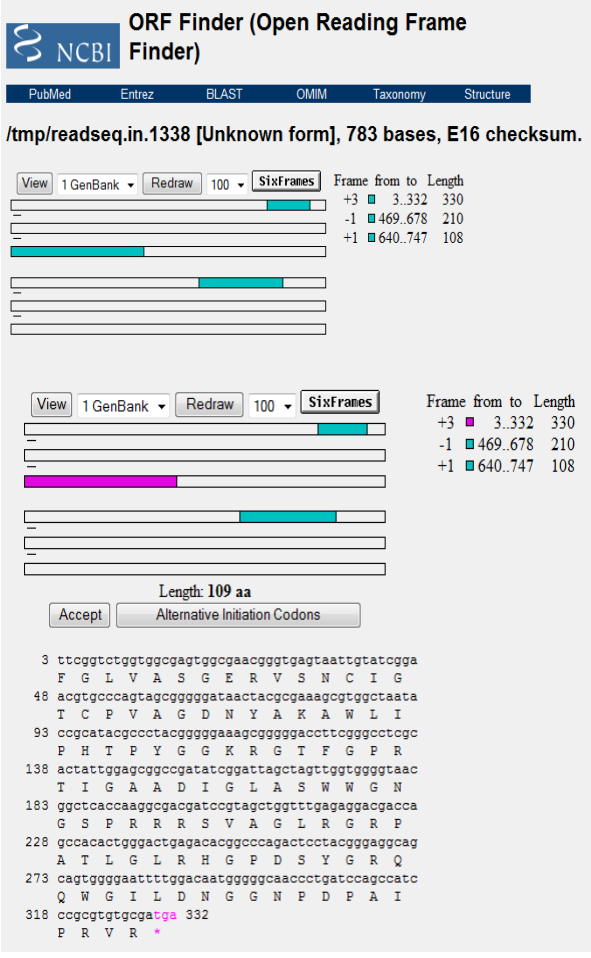

Figure 6
Functional region of the sequence using ORF Finder

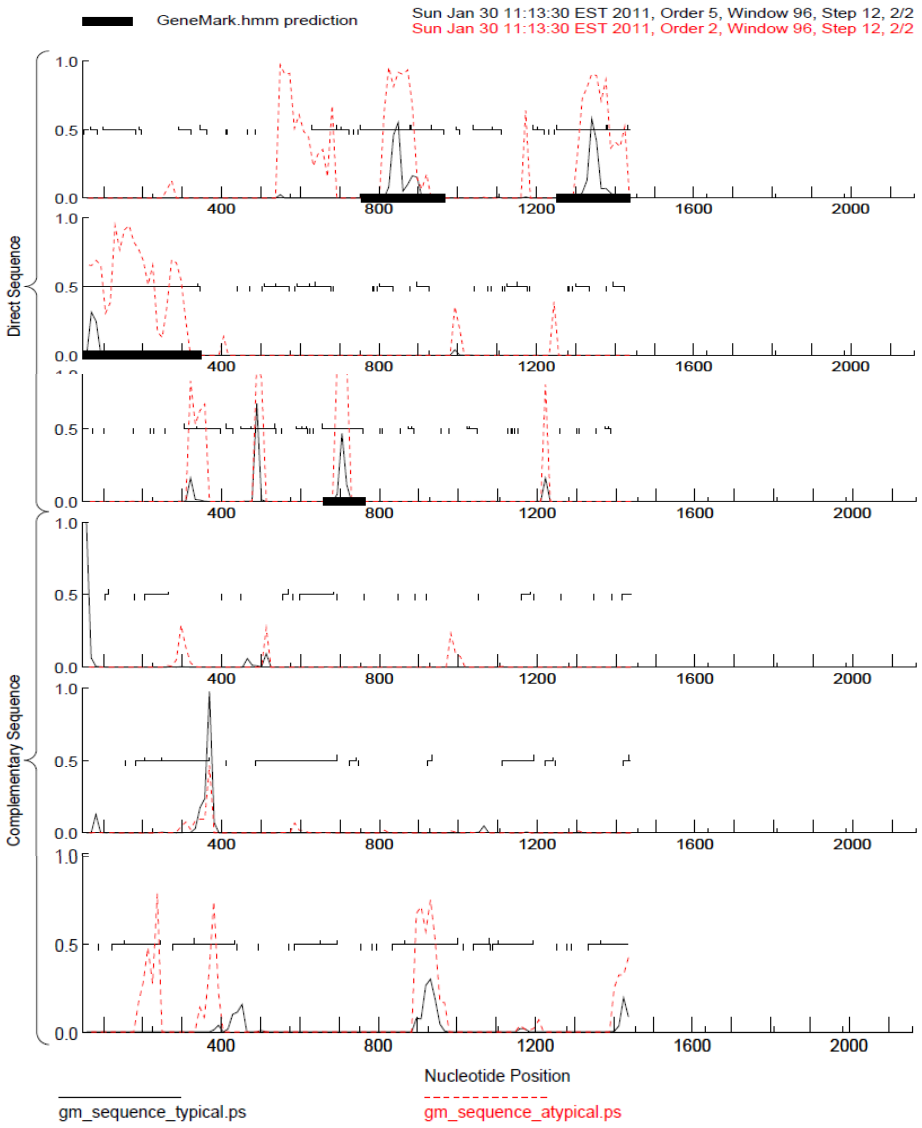

Figure 7

Coding region of the sequence using Gene Mark

Promoter predictions for 1 prokaryotic sequence with score cutoff 0.80 (transcription start shown in larger font):

Promoter predictions for /tmp/readseq.in.1338 :

Start	End	Score	Promoter Sequence
353	398	0.91	actttttggcaggaaagaaacggcgccggctaatacctggcGctaatgacg
541	586	0.94	ctttggaactgcatttttaactaccgggctagagtgtgtcAgagggaggt

Figure 8

Promoter site of the sequence using Promoter Finder