Molecular Identification of Textile Azo Dye Degrading Bacterial Strains by 16S rRNA Sequencing and Phylogenetic Analysis

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE IN BIOTECHNOLOGY

B.S. THESIS

Submitted by

Taizina Momtareen

Student ID: 12336001

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University

Bangladesh

September 2016

Dedicated to

My family and to every individual who has been there for me and wished good things for me

Declaration

I hereby solemnly declare that the research work embodying the results reported in this thesis entitled "**Molecular Identification of Dye Degrading Bacterial Strains by 16S rRNA Sequencing and Phylogenetic Analysis**" submitted by the undersigned has been carried out under the supervision of Ms. Romana Siddique, Senior Lecturer and Associate Coordinator, Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University, Dhaka. It is further declared that the research work presented here is original and any part of this thesis has not been submitted to any other institution for any degree or diploma.

Candidate:

_____

Taizina Momtareen

Certified

_____

Ms. Romana Siddique

Supervisor

Associate Coordinator
Biotechnology Program
Department of Mathematics and Natural Sciences
BRAC University, Dhaka

Acknowledgement

First and foremost, I would like to express my deepest gratitude to the Almighty, for providing me with the strength, patience, and perseverance required to successfully complete this project.

I express my sincere gratitude to **Professor A. A. Ziauddin Ahmad**, Chairperson, Department of Mathematics and Natural Sciences, BRAC University and to **Professor Naiyyum Choudhury**, former Coordinator of the Biotechnology and Microbiology program, Department of Mathematics and Natural Sciences, BRAC University for their excellent guidance, support, and inspiration during my tenure as a student at BRAC University.

This work was completed under the meticulous guidance of my Research Supervisor **Romana Siddique,** Associate Coordinator, Department of Mathematics and Natural Sciences, BRAC University. She believed in me, encouraged me, and offered her expert guidance throughout the course of this research. I would like to express my sincere appreciation to her for being an exemplary personnel during my tenure as a researcher and a student.

I am much obliged to Ms. Eusra Mohammad, Lecturer, Department of Mathematics and Natural Sciences, BRAC University, for providing me with valuable insights regarding the technical aspects of this project. I am indebted to her for all the help and support she provided to me.

I also extend my sincere gratitude to all the other teachers of BRAC University, who shared their knowledge and experiences during the coursework of the program. I would specially like to mention the name of Dr. Mahboob Hossain Associate Professor of the Department of Mathematics and Natural Sciences, BRAC University.

I would also like to extend my gratitude to Salman Khan Promon, Teaching Assistant, BRAC University for extending support throughout the research work I am indebted to all the other Teaching Assistants the Lab Officers and the Lab Assistants who had always been there to guide me through the intricacies of laboratory work.

I thank my friends Cinderella Akbar Mayaboti, SihintaShembil, Marium Kibtia, Dipanjana Datta for always offering me help and mental support. Last but not the least; I would like thank my parents for their unwavering support and dedication without which I would be helpless. They

have persevered through all my expeditions in life, held me up when I was down, and helped me overcome all the struggles. I am eternally grateful for having them in my life.

Taizina Momtareen

ABSTRACT

Colours in the form of dyes are an absolute necessity in the manufacturing of commodities. Synthetic dyes and mordants are used by the food, textiles, tanning, and paper industries to name a few. Used to enhance the aesthetic appeal of products, these dyes cause severe damage to the environment. The discharge of synthetic dyes into the environment without proper effluent treatment results in inevitable harm to all the components of the ecosystem. Moreover, the high expense of actively maintaining an Effluent Treatment Plant (ETP) by industries is considered a huge factor in heavily reducing profits, especially in third world countries like Bangladesh. Due to this, factories often evade the necessary effluent treatment and thereby cause high degrees of environmental pollution. Bioremediation offers a great and affordable solution to this seemingly unavoidable dilemma. When a microorganism is found in a particular environment, it indicates that that microorganism is able to feed on and metabolize the ingredients constituting the environment. Similarly, microorganisms present in industrial effluents containing high quantities of synthetic dyes should be able to breakdown and metabolize those dyes. Keeping this in mind, two bacterial strains have been isolated from the effluent of a textile industry near Dhaka. Molecular tests such as PCR, agarose gel electrophoresis, 16S rRNA sequencing, and phylogenetic tree construction and analysis were performed to identify one of the strains as a species under the genus *Providencia*. This bacterium has performed well in degrading the azo dye 'Reactive Red 3BX'. If used in combination with other dye degrading bacteria, this bacterium is expected to provide success in the biodegradation of harmful azo dyes. Thus it is a prime candidate to contribute to the cleaning of environmental pollution caused by azo dyes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Descriptions
COD	Chemical Oxygen Demand
BOD	Biochemical Oxygen Demand
µL	Microliter
Ml	milliliter
Nm	Nanometer
Psi	pounds per square inch
ETP	Effluent Treatment Plant
g	gram(s)
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
Rpm	Rotations per minute
Min	Minute(s)
mM	Milimolar
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediamenetetraaceticacid

Kb	Kilobase
Bp	Basepair
E-value	Expect value
NA	Nutrient Agar
LB	Luria-Bertani broth
MSA	Multiple Sequence Alignment

Chapter One

Introduction

1. Introduction

Colours are the wavelengths in the visible light spectrum that are reflected by objects and intercepted by the retina, but whose uplifting nature is beyond the grasp of the mind. These colours give a discernible characteristic to the elements of our ecosystem: the land, the water, and everything contained in them. Colours are prime contributors to the beauty of this world.

Apart from the nature, colour is imperative for clothing, fabrics, and pretty much every other commodity. Dyeing of fabrics dates back to 2600 B.C when natural dyes were extracted from plants and animals (Donatelli, 2016).

Fig 1.1 (left): Colouring cloth with natural red cabbage dye;

(right): pieces of fabric dyed with natural dyes (Pereira, 2012)

However, the colouration from natural dyes faded after washing, had poor light fastness, and the dyes themselves were difficult and time-consuming to produce. Progress in science led to the advent of synthetic dyes that were brighter, cheaper, and more colourfast. These massive upsides over natural dyes incited their widespread use and with that, brought about a revolution in the dyeing and textile industry.

1.1 Synthetic Dyes

Synthetic dyes can be chemically classified based on their chromophores that are groups of atoms responsible for the dye colour. Acridine, anthraquinone, azo, nitroso, quinine-imine, safranin, diazonium, cyanene, xanthenes, indophenol, and arylmethane dyes are some of the many synthetic dyes and are named after their chromophores (Chequer et al., 2013; Text-Team, 2011).

Fig 1.2 (left): Dyeing with synthetic dyes in a factory (Business Nonstop Desk, 2014); (right): Vivid colours of synthetic dyes (Blanken, 2016)

1.2 Azo Dyes

Azo dyes are a notorious class of aromatic dyes with the general formula R-N=N-R, where the –N=N- is the azo group and the Rs are aromatic functional groups to which they are bonded. They result in vibrant colours like red and yellow and are used widely for colouring commercial goods such as foods, leather, cosmetics, and predominantly textiles. Composed of aromatic hydrocarbons derived from benzene, toluene, aniline, phenol, and naphthalene, these dyes are recalcitrant compounds difficult to be broken down by natural or chemical means (Puvaneswari, N., Muthukrishnan, and J., Gunasekaran, P., 2006).

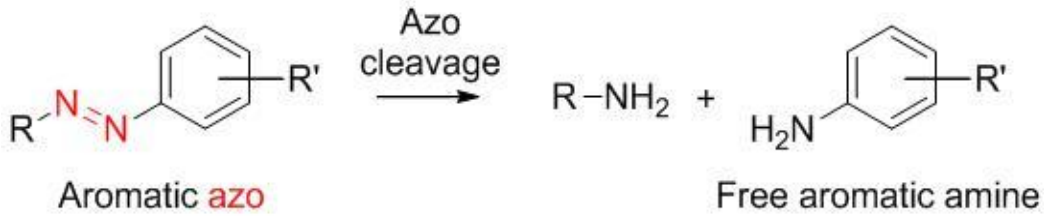


Fig 1.3a) Conversion of aromatic azo dyes to aromatic amines

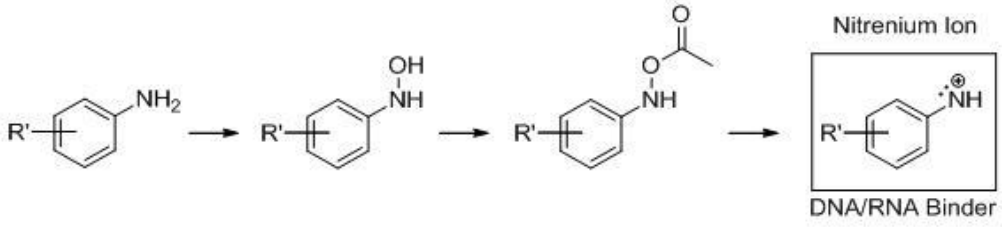


Fig 1.3b) **Metabolic activation of the aromatic amine and formation of electrophilic reactants (Environment and Climate Change Canada, 2012)**

1.3. Health and Environmental Hazards associated with azo dyes

Due to their colour and wet fastness, coupled with lower price and bright vivid colouration, azo dyes constitute 60-70% of all organic dyes in the world (Chequer et al., 2013). According to Carmen and Daniela (2012), every year, up to 25% of these dyes are lost to effluents during the dyeing and finishing operations. 2-20% is directly released into the water as aqueous effluent. Textile mills and tanneries use large volumes of water for dyeing, washing, and finishing purposes as a result of which large volume of wastewater is released. However, a lot of the toxic chemical dyes used escapes wastewater treatment processes and are discharged into the environment. This makes the textile sector one of the most polluting industrial sectors.

The damage to environment by toxic textile effluents is a problem that plagues many countries. It is much worse in Bangladesh as it is the second largest ready-made garment

(RMG) producer in the world but where almost no wastewater treatment is done (Mirdha, 2016). Moreover, the recalcitrant tendency of the textile chemicals requires complex and expensive treatment procedures. To avoid these high-costs, industries discharge large volume of untreated wastewater directly into the aquatic environment, ignorant and unconcerned about the detrimental aftermath.

1.3.1 Soil Pollution

There is no doubt that the release of untreated textile effluents directly into soil or using the effluents for land filling leads to soil pollution. Heavy metals like Zn, Cu, Cr, Cd, Pb, As, Hg, and Fe contained in these effluents are deposited in the soil. This soil nourishes plants and vegetation, however, most of the heavy metals are either toxic to the plants or to the animals/humans that feed on them. As a result, these poisonous metals degrade soil quality and reduce crop productivity. Studies on effect of different qualities of water on the growth and productivity of rice plant has shown that plants treated with textile wastewater are the shortest in height and resulted in the lowest grain yield compared to those grown in fresh and mixed water (Table 1) (Begum, Zaman, Mondol, Islam, & Hossain, 2011).

Field	Treatment	Plant height (cm)		Grain yield (t/ha)	
		1999	2000	1999	2000
Uncontaminated	Fresh water	73.2	72.2	5.23	5.40
	Mixed water	68.7	69.1	4.19	4.24
	Industrial Effluent	63.8	64.7	2.89	2.91
Contaminated	Fresh water	72	72.7	2.49	2.23
	Mixed water	66.9	66.0	1.86	1.82
	Industrial Effluent	62.1	61.3	1.24	1.16

Table 1.1: Effect of textile wastewater on plant height and grain yield in rice (Begum et al., 2011)

1.3.2 Water Pollution

Wastewater and effluents from dyeing units are rich in colour, containing residue of dyes and chemicals, high COD and BOD concentration, as well as non-biodegradable materials. It is usually highly alkaline with strong unpleasant odour. The alkalinity of the chemicals can raise the pH of water to as high as 11 (Lakherwal,2014).

The layer of discharged dyes on water surface causes turbidity that interferes with the penetration of sunlight to the plants below. This impairs the process of photosynthesis. The

contaminants in the layer also decrease the amount of dissolved oxygen, thereby creating a high oxygen demand. Chemical Oxygen Demand (COD) and Biochemical Oxygen Demand (BOD) are indirect measures of organic pollutants in water. High quantity of organic compounds in water mean high amount of oxygen needed by the aquatic organisms to degrade them, therefore high oxygen demand. Inability to fulfil this demand leads to depletion of oxygen resulting in harm and eventual death of aquatic flora and fauna. On the other hand, nitrates and phosphates in the effluent can act as necessary nutrients to result in algal blooms. However, algal blooms can also give rise to oxygen-depleted dead zones (Lakherwal,2014; Pereira and Alves, 2012).

Fig 1.4: Effluent from a tannery in Gazipur released directly into water of Turag river. The effluent also constitutes numerous pieces of cloth (Rashid, 2011)

1.3.3 Health Issues

The discharge of toxic azo dyes in natural ecosystem results in the conversion of the azo group to aromatic amines. Aromatic amines are established carcinogens that have been linked with several cancers such as urinary bladder cancer, breast cancer, splenic sarcomas, and hepatocarcinomas. Some xenobiotic azo dyes, such as benzidine- and naphthalene-containing dyes get reduced to toxic aromatic amines inside living organisms. Around twenty four aromatic amines have been confirmed as, or expected to be, carcinogens in humans (The Parliamentary Office of Science and Technology, 2014). Furthermore, some aromatic amines are highly mutagenic and can cause chromosomal aberrations in mammalian cells. Allergenicity, lung and urinary bladder cancers have been reported in workers with high exposure to dye materials (Puvaneswari et al., 2006).

Apart from high solid content and low COD: BOD ratio, dye components are very high in conductivity. This indicates that the effluent is rich in ions such as nitrates, phosphates, and other ions at a level higher than that approved by WHO (Sivakumar, Balamurugan, Ramakrishnan, and Bhai, 2011). Nitrate in drinking water gets converted to nitrite by endogenous bacteria, which then transforms haemoglobin and causes methemiglobinemia (Fewtrell, 2004).

1.4 Bioremediation

There are many treatment processes for the cleanup of textile wastewater and sludge. Three tiers, primary, secondary, and tertiary, are employed to remove contaminants. These tiers incorporate many procedures to achieve decolourization which include physiochemical methods like filtration, activated carbon, electrodialysis, chemical flocculation, adsorption, and coagulation. Most of them are partially effective but are also quite expensive and result in huge quantity of sludge (The World Bank Group, 2016; Lakherwal, 2014).

As a result of all these drawbacks, a biological alternative leading to lower environmental pollution is highly sought-after. Due to their ubiquitous nature and the ability to metabolize and decolourize dyes, the use of microorganisms in wastewater treatment can solve a lot of the problems. This use of microorganisms to degrade environmental pollutants in order to clean a contaminated site is known as bioremediation. There are plenty of microorganisms in nature that can decolourize, transform, or even mineralize the recalcitrant compounds in dyestuff. With the help of enzymes and biochemical reactions, particular microbes breakdown the azo dye molecules into their constituent aromatic amines. Then, in combination of aerobic and anaerobic conditions, the bacteria metabolize and mineralize those aromatic amines (Puvaneswari et al., 2006).

According to Chengalroyen and Dabbs (2013), *Pseudomonas*, *Bacillus*, and white rot fungi are some of the highly competent biological agents of dye degradation. Bacteria use several azoreductases for the reductive azo bound cleavage. First, bacteria reduce azo dyes into colourless aromatic amines under anaerobic conditions. Due their carcinogenic and mutagenic character, these amines are then further degraded in aerobic conditions.

1.5 About this project

Bacteria from textile sludge were previously isolated and screened for their ability to degrade Reactive Red 3BX and Yellow 4GL azo dyes. With the help of many biochemical tests it was found that two of the bacterial species were the most effective in degrading the dyes. These bacteria were hypothesized to be *Brevibacillus laterosporus* and *Staphylococcus nepalensis* by previous researcher.

The objective of this project was to use multiple molecular tests to verify this hypothesis and to accurately determine the genotypes of the bacteria. Identification of bacteria has been done with the help of 16S rDNA sequencing and phylogenetic tree analysis.

.

Chapter Two

Materials and Methods

2. Materials and methods

2.1 Place of study

The study was carried out in the Biotechnology laboratory of the Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh.

2.2. Handling of laboratory apparatus and glassware

All the glassware used in this project such as conical flasks, petri dishes, and beakers were washed once with tap water and then with distilled water. Pipette tips and centrifuge tubes were autoclaved at 121 °C at 15 psi for 15 minutes prior to use. Clean lab coat was worn while carrying out all the experiments and hand gloves were used to avoid contamination.

2.3 Collection of soil sample and isolation of dye decolourizing bacteria

Collection of soil and the isolation of bacteria had already been done and stored in the laboratory before the start of this project. Soil was collected from the ETP treatment plant of a textile factory located in Bhaluka industrial area just outside of Dhaka.

Different concentrations of dye solutions were made and inoculated with 5g soil sample. Next, the absorbance for each dilution was measured because absorbance decreases with increased dye degradation. The wavelengths for Reactive Red and Yellow 4GL were set as 534 and 485nm respectively. From the solutions with highest dye degradation (lowest absorbance), bacterial colonies were isolated. Then, their dye decolourization ability was tested by growing them on different dye concentrations. Those colonies that tested positive were then isolated for further biochemical tests. Among the six bacterial colonies that were excellent in decolourizing both the dyes, two best ones were chosen. Their biochemical test results were input into the Advanced Bacterial Identification Software (ABIS) which suggested the possible names of the two bacterial strains as *Brevibacillus laterosporus* and *Staphyloccus nepalensis*.The bacterial samples were stored in -20°C in the laboratory before the commencement of this project.

2.4 Genotypic Identification of the bacterial strains

In order to fully confirm the identity of bacterial species, molecular tests such as PCR and 16S rRNA sequencing are required. First and foremost, DNA was extracted from the two bacterial strains for molecular identification.

2.4.1 DNA Extraction

The procedure performed in order to extract high quality DNA from the two bacterial samples has been given below. It was made using the protocol by He, F. (2011).

1. 1.5 ml of the overnight culture (grown in LB medium) was transferred to a 1.5 ml Eppendorf tube and centrifuged at 13,500rpm for 3min to pellet the cells.
2. The supernatant was discarded without disturbing the pellet.
3. The cell pellet was then resuspended in 600µl lysis buffer and vortexed to resuspend completely.
4. The resuspended cell pellet was then incubated for 1 h at 37 °C.
5. After incubation was complete, 600µl of phenol/chloroform (1:1) was added and mixed by inverting the tubes until the phases were completely mixed.
6. Following second round of centrifuge for 5min, three distinct layers were visible: bottom layer of phenol/chloroform, intermediate layer of proteins, and the top aqueous layer of nucleic acids.
7. The top layer was carefully transferred to a new tube.
8. To remove phenol, an equal volume of chloroform to the aqueous layer. This was again inverted to mix well.
9. The tubes were spinned at 13,500 rpm for 5 min.
10. Approximately 200µl of the upper aqueous layer containing DNA was transferred to a new tube.
11. To precipitate the DNA, 600µl of cold ethanol was added and mixed gently
12. The tubes were incubated at -20°C for 30 min.
13. Next, the tubes were centrifuged at 13,500 rpm for 15 min.
14. The supernatant containing ethanol was discarded and the DNA pellet rinsed with 1 ml 70% ethanol.
15. Another round of centrifuge was done at 13,500 rpm for 2 min.
16. The supernatant was discarded and the DNA pellet was air-dried.
17. The DNA was resuspended in 50µl TE buffer.

2.4.2. Polymerase Chain Reaction to amplify extracted DNA

From the extracted DNA, the 16S rRNA gene was amplified using Polymerase Chain Reaction (PCR). PCR is a powerful technique in molecular biology that uses Taq DNA Polymerase enzyme to amplify the quantity of a DNA sample. First, high temperatures in the denaturation phase denature the double stranded template into single strands. Then in the annealing phase, short oligonucleotides called primers bind to the template strands. Then the Taq Polymerase binds to the primers and starts forming new strands by adding complementary nucleotides and thus forming double stranded DNA again. This occurs in the renaturation phase. The aim of this PCR reaction was to amplify only the 16S rRNA segment of the bacterial DNA. Hence universal primers for 16S rRNA were used.

In this project, PCR was done using the following procedure:

1. Autoclaved PCR tubes were taken
2. 50µl master mix was made with the following components:

Component	Amount
10X *Taq* Reaction Buffer	5 µl
10 mM dNTP	1 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
Template DNA	5.0 µl
Taq DNA Polymerase	0.25 µl
Nuclease Free Water	36.75 µl

Table 2.1: Reaction set up for PCR carried out for 50 µl reaction volume

Reaction Setup:

3. After preparation, the tubes were placed in the PCR machine. The reaction followed the thermal cycle given in Table 2.2. The PCR was carried out for 35 cycles.

PCR condition	Temperature	Time
Initial denaturation	94°C	5 minutes
Denaturation	94°C	45 seconds
Annealing	55°C	1 minutes
Extension	72°C	1 minute
Final extension	72°C	10 minutes
Final Hold	4°C	

Table 2.2: The thermal cycle followed for PCR reaction

The details of the universal 16S rRNA primers used in this PCR reaction are given below in Table 2.3:

Sample	Specimen 1: *Brevibacillus laterosporous* (probable)
	Specimen 2: *Staphylococcus nepalensis* (probable)
Specimen 1	
Forward Primer	F:fD1=(5'-AGAGTTTGATCCTGGCTCAG-3')
Reverse Primer	R: rP2= (5'-ACGGCTACCTTGTTACGACTT-3')
Specimen 2	
Forward Primer	F: 27F=(5' AGAGTTTGATCMTGGCTCAG 3')
Reverse Primer	R: 1492R=(5' TACGGYTACCTTGTTACGACTT 3')

Table 2.3: Primers used in the PCR reaction

All the steps of master mix synthesis were performed on ice. Once the reaction was complete, the PCR product was stored at -20°C for further work.

2.4.3. Agarose gel electrophoresis for the detection of amplified DNA

After PCR reaction, 1% agarose gel electrophoresis was performed to check for PCR product amplification. Agarose gel electrophoresis is a standard laboratory procedure used to separate amplified PCR product into bands based on size. Amplified DNA is applied in wells in the gel close to the negative electrode. In the presence of an electrical field, negatively charged DNA moves toward the positive pole through the small holes that make up the gel matrix. These holes allow the shorter fragments of DNA to migrate faster than their longer counterparts. Once the reaction is complete, the length of the amplified DNA can be accurately determined by comparing with a DNA ladder.

The process followed during electrophoresis experiment is given below:

1. 0.30g agarose was dissolved in 30ml 1X Tris-EDTA (TE) buffer of pH 8.0 and heated to dissolve in a microwave oven for about 30 seconds.
2. The mixture was allowed to cool down to about 50^0C.
3. To the cooled agarose gel, 1.5µl Ethidium Bromide (EtBr) stain was added and mixed in order to stain the DNA bands.
4. The gel was then poured on the gel casting tray previously set with the comb and allowed to solidify.
5. 5µl of the PCR product was mixed with 2µl of loading dye and was loaded into the individual wells of the gel.
6. A ladder of size 1kb plus (Invitrogen, USA) was used to ensure amplification of the desired 16S rRNA gene and measure the exact product size which was estimated to be within 1,500bp.
7. After loading of the samples, the lid was placed on the gel box.
8. The electrodes were connected to their respective wires and the wires were connected to power supply. Black is the negative electrode and red is the positive electrode. DNA runs from negative pole (black) toward the positive pole (red). Power supply was turned to 90 volts.
9. After 30min, the power supply was turned off and then gel was transferred to be visualized under UV light.

2.4.4. Purification of DNA

After observing the presence of the desired 16S rDNA in the gel, DNA was purified from the PCR amplicon using the Wizard® SV Gel and PCR Clean-Up System (Promega, USA). The manufacturer's protocol followed for the purification is given below:

Processing PCR Amplicon

1. 20 µl of the PCR-amplified DNA was taken into a microcentrifuge tube.
2. To this, 20µl Membrane Binding solution was added and vortexed.

Binding of DNA

3. An SV Minicolumn was inserted into a Collection Tube.
4. Then the PCR product was transferred to the Minicolumn assembly and incubated at room temperature for 1 min.
5. This was then centrifuged at 11,000rpm for 1min.
6. The flowthrough was discarded and the Minicolumn reinserted into Collection Tube.

Washing

7. 700µl Membrane Wash Solution was added to the Minicolumn assembly. This was centrifuged at 11,000 rpm for 1 min.
8. Again the flowthrough was discarded and the Minicolumn reinserted into Collection Tube.
9. Step 7 was repeated with 500µl Membrane Wash Solution and centrifuged at 11,000rpm for 5 min.
10. The Collection Tube was emptied and the column assembly recentrifuged for 1 minute with the microcentrifuge lid open to allow evaporation of any residual ethanol.

Elution

11. The Minicolumn was carefully transferred to a clean 1.5ml microcentrifuge tube.
12. 50µl of Nuclease-Free Water was added to the Minicolumn.
13. This was incubated at room temperature for 1 minute and then centrifuged at 16,000 × g for 1 minute.

14. The Minicolumn was discarded and the purified DNA stored at –20°C.

2.5.6. DNA sequencing

The DNA samples were sequenced by the Sanger method using Applied Biosystems'3500 Dx Genetic Analyzer in the **DNA Solution Laboratory**, Panthapath, Dhaka.

Through the Sanger method, the nucleotide sequence of the PCR-amplified 16S rRNA gene segment was determined.

16S ribosomal RNA sequencing is an important tool in molecular biology. 16S rRNA is one of the constituents of the bacterial organelle ribosome. The gene of the 16S rRNA (or 16S rDNA) is an important molecular marker because of many reasons. First, the 1,550bp long rDNA is universal, meaning that it exists in all bacteria. Second, this DNA segment has both highly conserved and variable regions. The 16S rDNA does not evolve at the same rate in all organisms. This indicates that bacterial species that evolve faster will have more differences in sequence that those that evolve slower. These differences can be analyzed to predict evolutionary relationships between different bacteria using bioinformatics (Patwardhan, Ray, Roy, 2014).

Fig. 2.1 Applied Biosystems'3500 Dx Genetic Analyzer

After completion of the sequencing reactions, a forward and reverse sequence and a consensus sequence were provided for each strain by DNA Solution Laboratory. A consensus sequence is one that is constructed by calculating the most frequently expressed nucleotide at each position of a multiple alignment (consensus sequence, n.d.).

2.5.7 Bioinformatics: *in silico* analysis

In order to identify the genotypes of the bacterial strains, phylogenetic trees were constructed with the given consensus sequences. Below is a short description of each tool used for this purpose.

- **BLAST**

The first step in constructing a phylogenetic tree is to use the consensus sequence and find similar sequences from the NCBI database using the Basic Local Alignment Search Tool (BLAST). BLAST is an algorithm that can compare and align a query nucleotide or protein sequence with a number of sequences contained in its database. It finds regions of local similarity between the sequences by calculating the statistical significance of matches. It is both rapid and sensitive and hence is used by millions of biologists. It is available online at the National Center for Biotechnology Information (NCBI) website (Lobo, 2008). As this project involved DNA sequences, BLASTn (nucleotide BLAST) program has been used.

BLAST URL: http://blast.ncbi.nlm.nih.gov/Blast.cgi

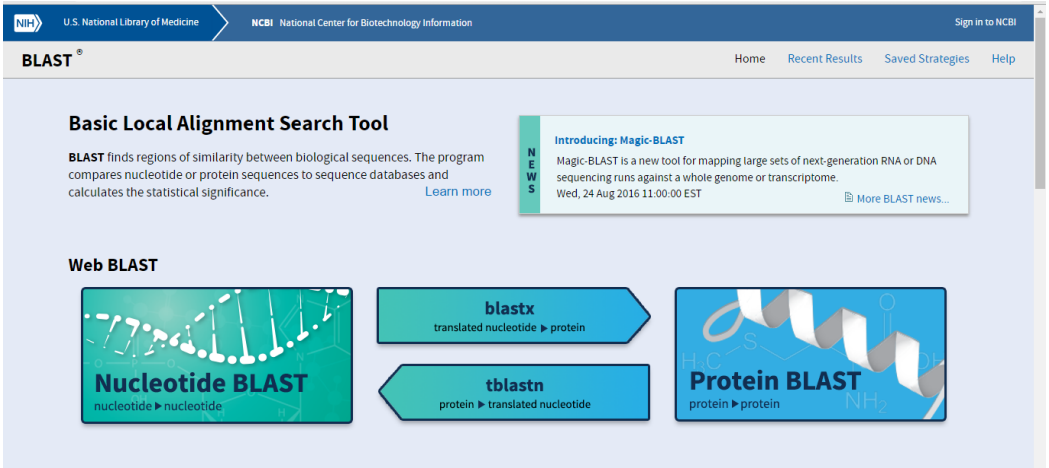


Fig 2.2: BLAST Homepage

- **Clustal Omega**

Clustal Omega is the latest multiple sequence alignment tool of the Clustal series of programs. It aligns multiple sequences to highlight areas of similarity that may be associated with specific features that have been more highly conserved than other regions. The Clustal Omega web form is available at http://www.ebi.ac.uk/Tools/msa/clustalo/

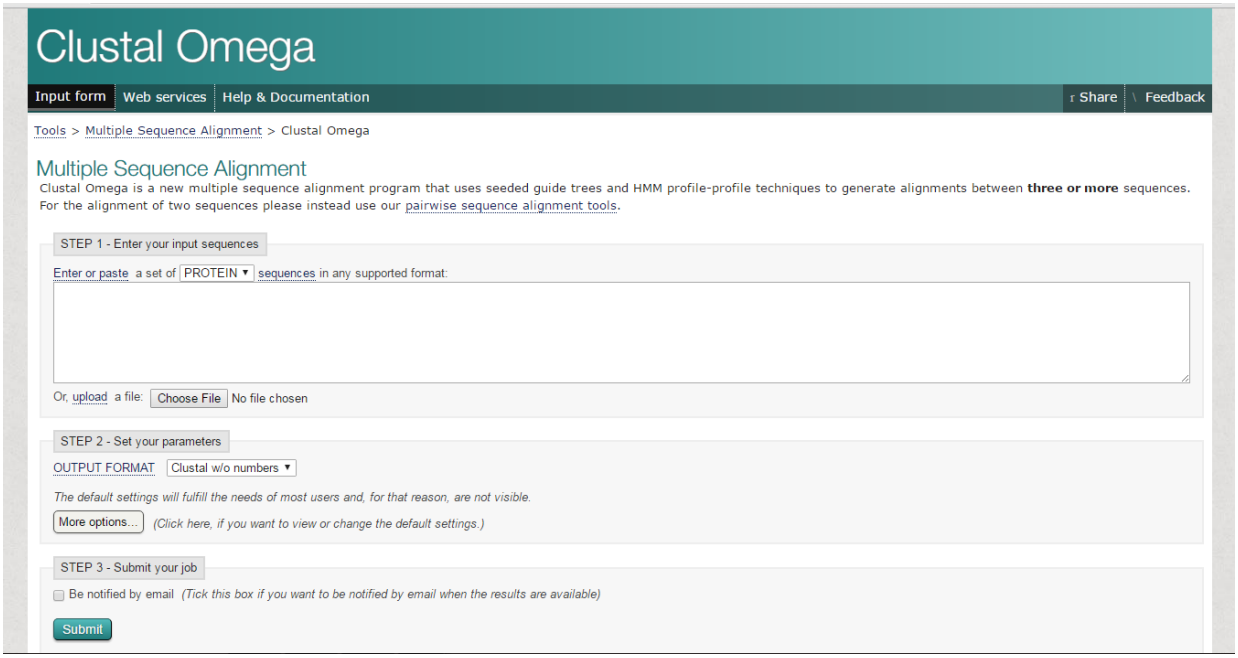


Fig 2.3: Clustal Omega Homepage

- **Molecular Evolutionary Genetics Analysis (MEGA)**

Molecular Evolutionary Genetics Analysis (MEGA) is an integrated software for conducting sequence alignments, estimating divergence times, inferring phylogenetic trees, estimating molecular evolution rate, inferring ancestral sequences and testing evolutionary hypotheses. It is used by biologists for reconstruction of evolutionary histories of species and hypothesizing the extent and nature of the selective forces that shape the evolution of genes as well as species. Many versions of the software are available online and can be downloaded. For this project, MEGA version 6 was used.

URL for download: http://www.megasoftware.net/

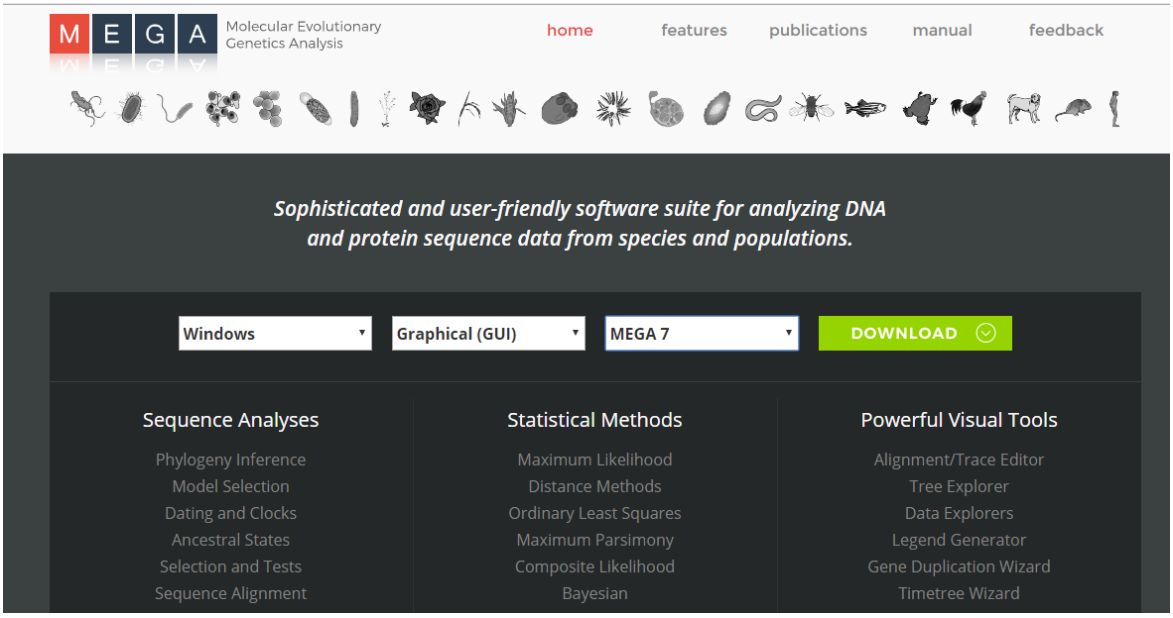


Fig 2.4.1: MEGA download page

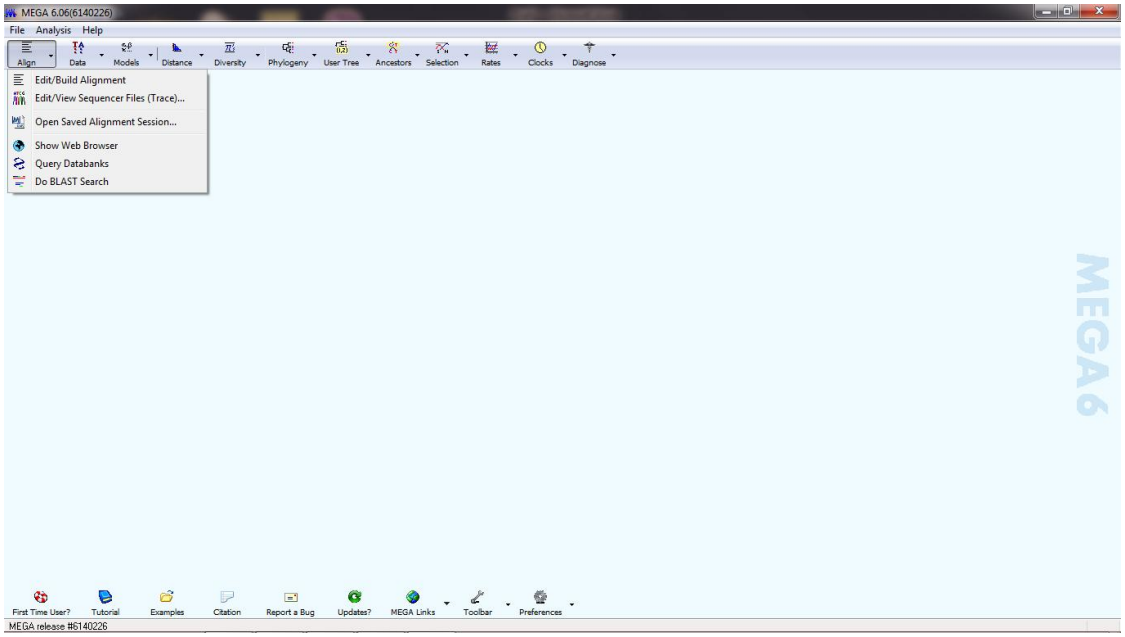


Fig 2.4.2: MEGA 6 software

- **Phylogenetic trees**

A phylogenetic tree, also known as an evolutionary tree, serves to visually represent the evolutionary relationships among a set of organisms or groups of organisms, called taxa. It infers relationships based on similarities and differences in genetic characteristics of organisms and portrays them in a branching diagram. A phylogenetic tree has several features through which

evolutionary relationships can be inferred. The leaf or tip of a tree designates groups of descendent taxa, such as a species. Leaves are often connected to nodes by branches. A node represents a common ancestor from which the leaves descended. The term sister group often refers to two species that have descended from the same node and are each other's closest relatives (Understanding Evolution team, n.d).

From BLAST to phylogenetic tree construction, the following protocol describes all the steps taken to identify the genotype of the bacteria:

1) First, the NCBI BLASTn website was opened.
2) Under the "Enter Query Sequence" box, the consensus sequence was uploaded. Most of the parameters were set as default [Database: Others; Optimize for: "Highly similar sequences (Megablast)"] but "Uncultured/environmental sample sequences" were excluded. BLAST button selected.

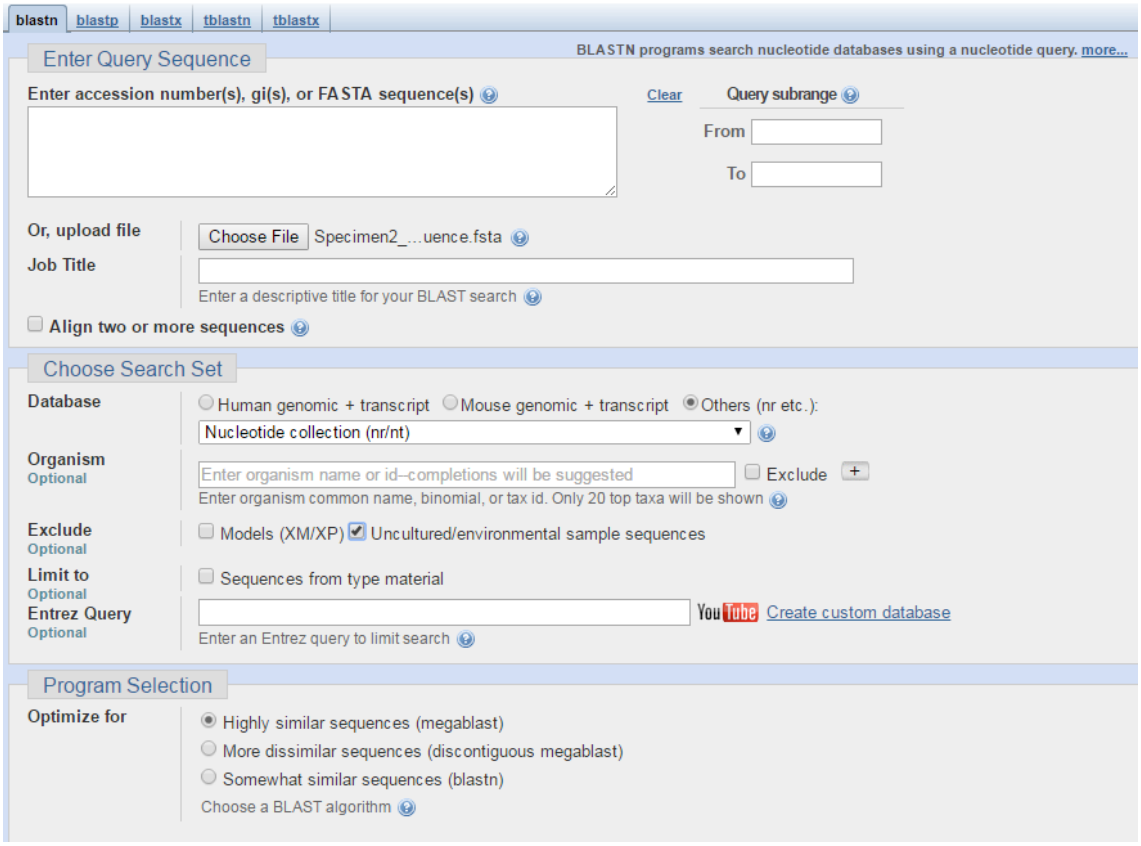
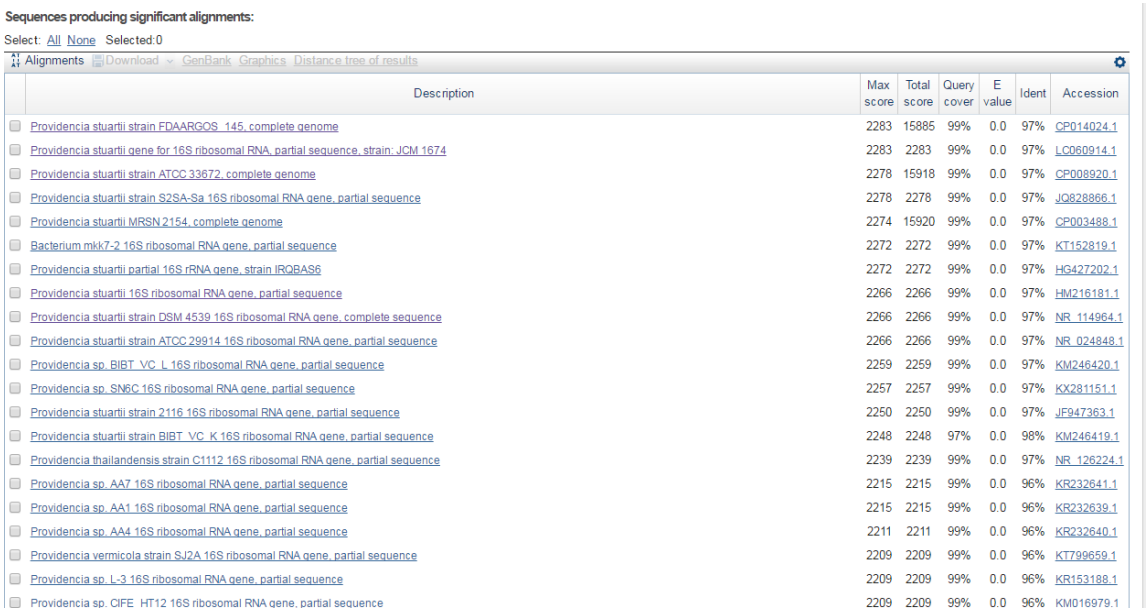


Fig 2.5: Uploading query sequence into BLASTn

3) The sequences for the top 8-10 results with low E-values were downloaded



Sequences producing significant alignments:

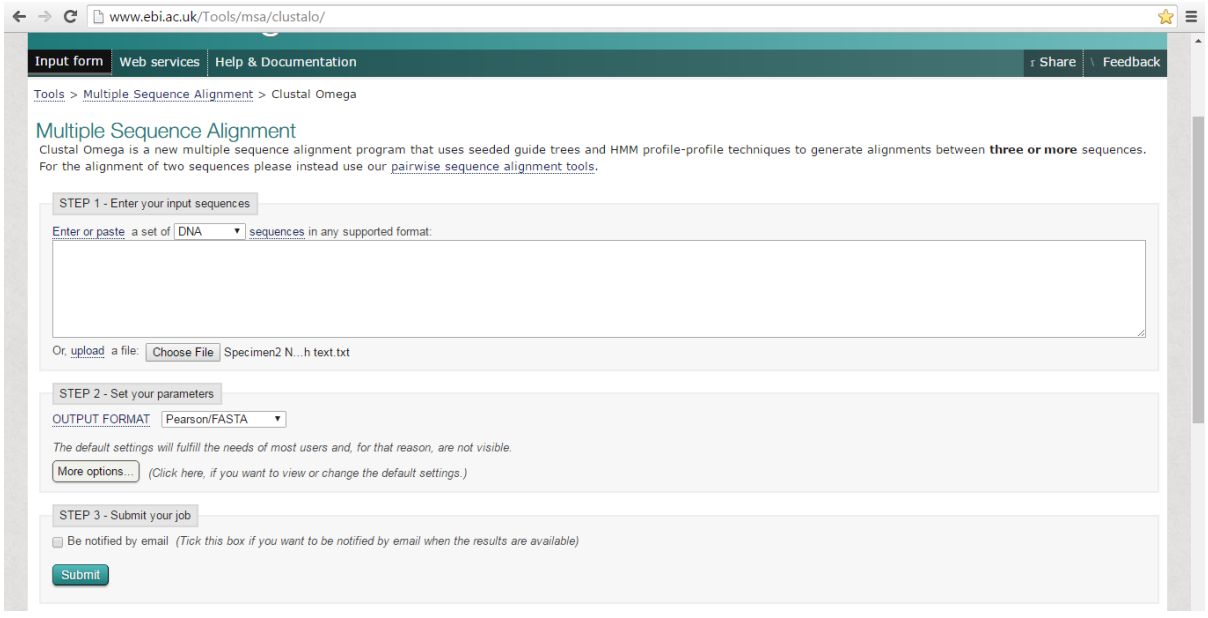
Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Providencia stuartii strain FDAARGOS_145, complete genome	2283	15885	99%	0.0	97%	CP014024.1
Providencia stuartii gene for 16S ribosomal RNA, partial sequence, strain: JCM 1674	2283	2283	99%	0.0	97%	LC060914.1
Providencia stuartii strain ATCC 33672, complete genome	2278	15918	99%	0.0	97%	CP008920.1
Providencia stuartii strain S2SA-Sa 16S ribosomal RNA gene, partial sequence	2278	2278	99%	0.0	97%	JQ828866.1
Providencia stuartii MRSN 2154, complete genome	2274	15920	99%	0.0	97%	CP003488.1
Bacterium mkk7-2 16S ribosomal RNA gene, partial sequence	2272	2272	99%	0.0	97%	KT152819.1
Providencia stuartii partial 16S rRNA gene, strain IRQBAS6	2272	2272	99%	0.0	97%	HG427202.1
Providencia stuartii 16S ribosomal RNA gene, partial sequence	2266	2266	99%	0.0	97%	HM216181.1
Providencia stuartii strain DSM 4539 16S ribosomal RNA gene, complete sequence	2266	2266	99%	0.0	97%	NR_114964.1
Providencia stuartii strain ATCC 29914 16S ribosomal RNA gene, partial sequence	2266	2266	99%	0.0	97%	NR_024848.1
Providencia sp. BIBT_VC_L 16S ribosomal RNA gene, partial sequence	2259	2259	99%	0.0	97%	KM246420.1
Providencia sp. SN6C 16S ribosomal RNA gene, partial sequence	2257	2257	99%	0.0	97%	KX281151.1
Providencia stuartii strain 2116 16S ribosomal RNA gene, partial sequence	2250	2250	99%	0.0	97%	JF947363.1
Providencia stuartii strain BIBT_VC_K 16S ribosomal RNA gene, partial sequence	2248	2248	97%	0.0	98%	KM246419.1
Providencia thailandensis strain C1112 16S ribosomal RNA gene, partial sequence	2239	2239	99%	0.0	97%	NR_126224.1
Providencia sp. AA7 16S ribosomal RNA gene, partial sequence	2215	2215	99%	0.0	96%	KR232641.1
Providencia sp. AA1 16S ribosomal RNA gene, partial sequence	2215	2215	99%	0.0	96%	KR232639.1
Providencia sp. AA4 16S ribosomal RNA gene, partial sequence	2211	2211	99%	0.0	96%	KR232640.1
Providencia vermicola strain SJ2A 16S ribosomal RNA gene, partial sequence	2209	2209	99%	0.0	96%	KT799659.1
Providencia sp. L-3 16S ribosomal RNA gene, partial sequence	2209	2209	99%	0.0	96%	KR153188.1
Providencia sp. CIFE_HT12 16S ribosomal RNA gene, partial sequence	2209	2209	99%	0.0	96%	KM016979.1

Fig 2.6: Matching sequence results

4) Next, the Clustal Omega website was opened and there this file containing the consensus and BLAST hits was uploaded. For the output format, Pearson/FASTA was selected.



www.ebi.ac.uk/Tools/msa/clustalo/

Input form | Web services | Help & Documentation | Share | Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

STEP 1 - Enter your input sequences

Enter or paste a set of DNA sequences in any supported format:

Or, upload a file: Choose File Specimen2 N...h text.txt

STEP 2 - Set your parameters

OUTPUT FORMAT Pearson/FASTA

The default settings will fulfill the needs of most users and, for that reason, are not visible.

More options... *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit

Fig 2.7: Uploading file containing consensus sequence and top BLAST matches

5) The result of the multiple alignment was gives as follows:

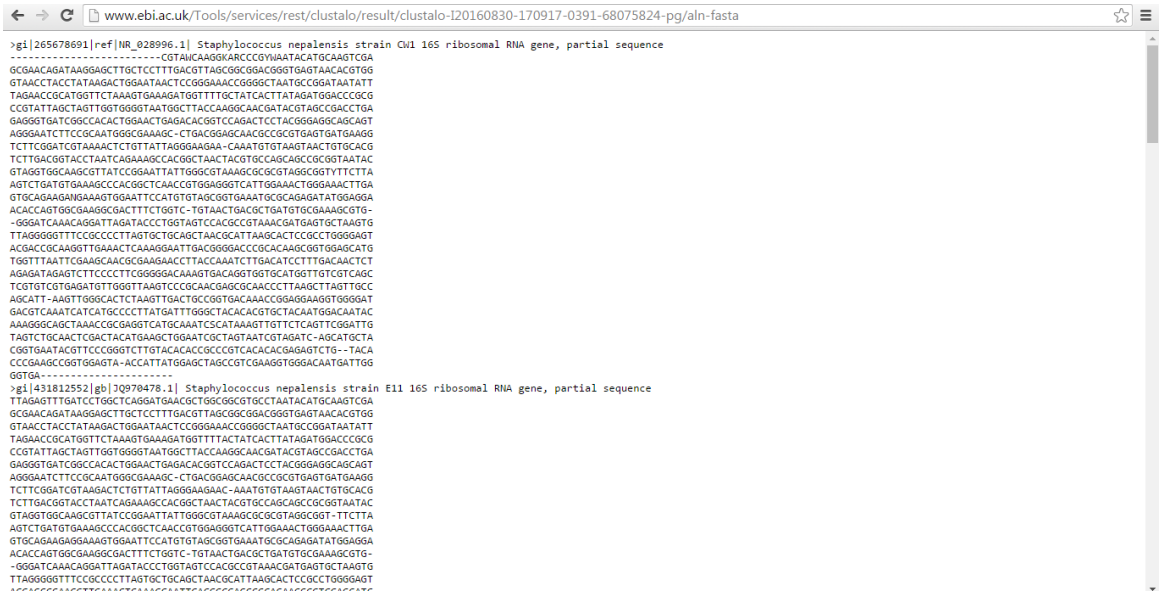


Fig 2.8: Clustal Omega multiple sequence alignment result

6) All these aligned sequences were downloaded and then opened in MEGA6 software

Fig 2.9: Multiple alignments opened in Mega6

7) Next, in the main Mega6 window, a maximum likelihood tree was constructed using default Analysis Preference parameters:

8) Similarly Maximum Parsimony and Neighbour Joining trees were also constructed for both Specimen1 and Specimen 2. For neighbour joining tree, ‘Bootstrapping method’ was selected for test of phylogeny with 500 bootstrap replications.
9) Once separate trees were constructed for the two strains, all the Clustal Omega aligned sequences for both bacterial strains were accumulated into one file from which cumulative Maximum likelihood, Maximum parsimony, and Neighbour joining trees were built.

Chapter Three

Results

3. Results

3.1 DNA Extraction

a) In order to extract DNA, the bacteria had to be revived by streaking in nutrient agar (NA). The streaking results are as follows:

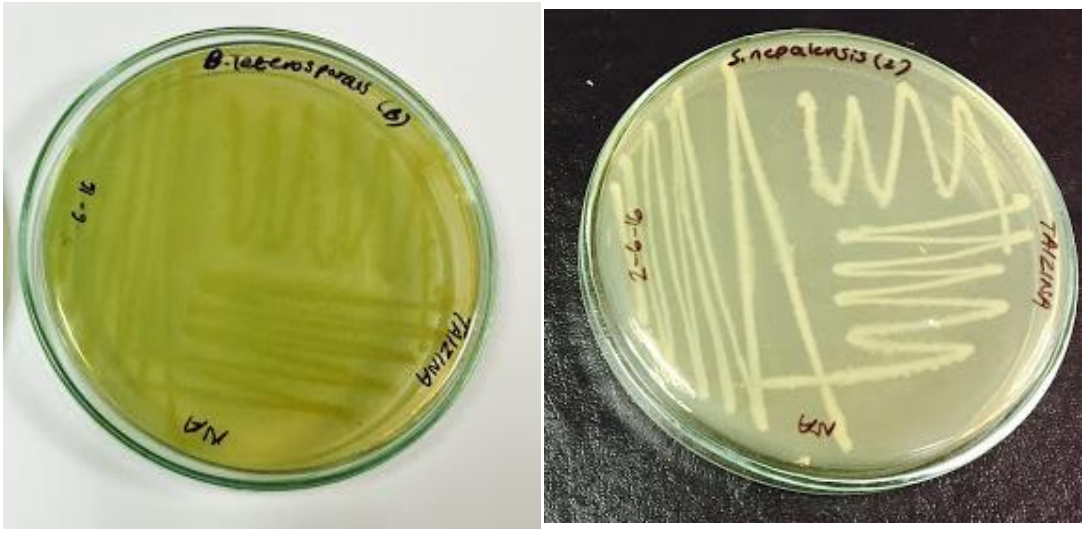


Fig 3.1.1: (left) Specimen 1 streak plate; (right) Specimen 2 streak plate

b) For the next step, bacteria from NA plates were cultured in LB broth by inoculating with a loop and incubated for 24 hours. Results after 24 hours of incubation is shown below:

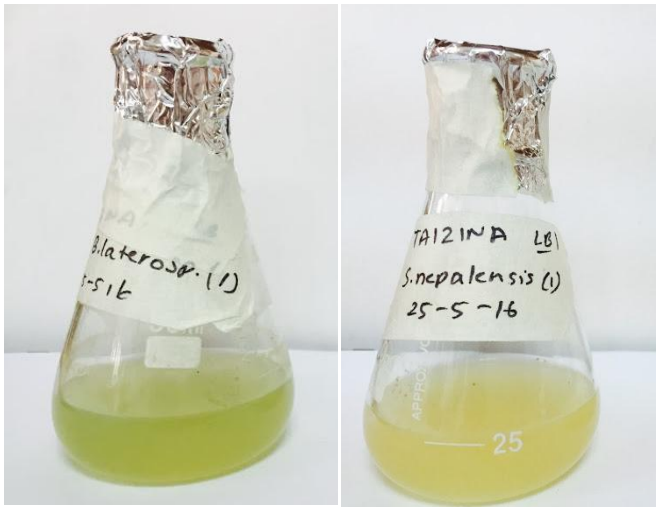


Fig 3.1.2: (left) Specimen 1 culture; (right) Specimen 2 culture

c) DNA was extracted from the liquid cultures in Fig 15. An image of precipitated DNA is given below:

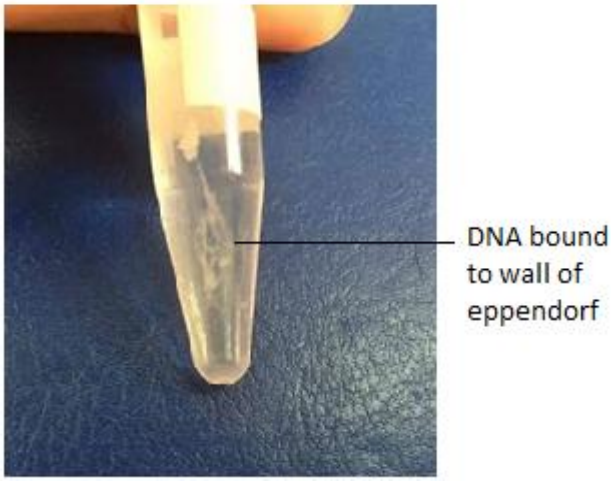


Fig 3.1.3: Extracted DNA on the wall of eppendorf tube

3.2 Agarose gel electrophoresis

In order to test whether PCR was successful, the amplified DNA was run through 1% agarose gel. To determine the size of resulting DNA bands, the "GeneRuler 100bp Plus DNA Ladder" by Invitrogen was used.

A **DNA ladder** consists of a set of known DNA fragments of different sizes that are separated and visualized as DNA bands on a gel. They are used in gel electrophoresis to determine the size and quantity of testing DNA fragments of PCR DNA (Expression Technologies, 2003).

a)

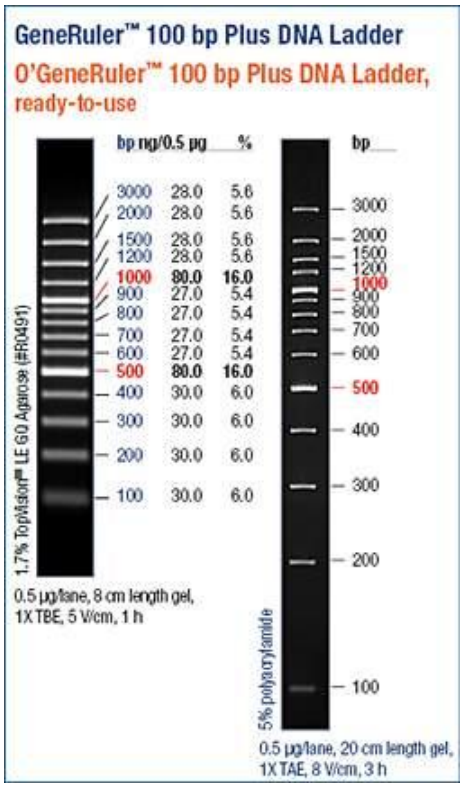


Fig 3.2.1 a) GeneRuler provided by manufacturer

The results of gel electrophoresis are portrayed in fig. 3.2.1 b and c for specimen 1 and 2 respectively. The figures show the stained DNA bands as observed under UV light. Each number on the top of the images indicates a well. The letter 'L' stands for ladder, indicating that DNA Ladder was applied on that well. In fig 3.2.1b, the PCR product for specimen 1 was applied on the first well. After comparing with the ladder it can be deduced from the figure that the DNA band on well 1 is approximately 1,500 bp in size.

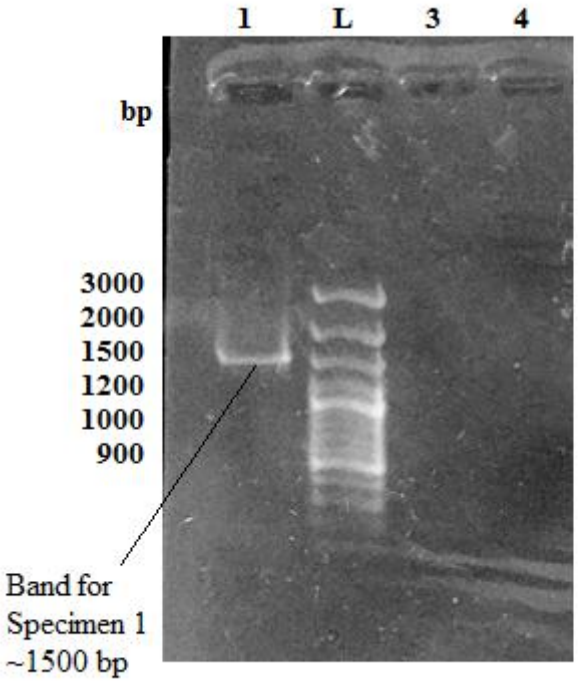


Fig 3.2.1 b) 16S rDNA band for Specimen 1

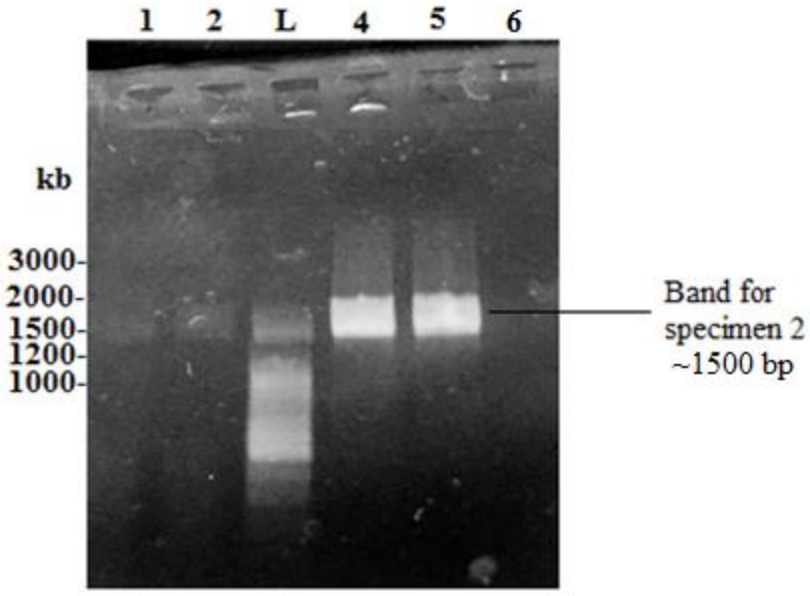


Fig 3.2.1 c) 16S rDNA band for Specimen 2

In fig 3.2.1c, the 16S rDNA PCR product for specimen 2 was applied in wells 1, 2, 4, and 5. The band can be clearly observed in wells 4 and 5 but not as well in 1 and 2. Similar to fig 3.2.1b, the target DNA band here is also deduced to be approximately 1,500bp in size. From

these results in can be inferred that almost the entire 1,550 bp long 16S rDNA was amplified by PCR. Hence, the PCR experiments were successful.

3.3 *In Silico* Analysis

Following sequencing of PCR-amplified DNA using 16S universal primers, the resulting consensus sequences were used for *in silico* analysis.

The length of the consensus for specimen 1 is just 582 bases, which is very short for 16S rDNA. This is due to the denaturation of the forward primer that may have occurred either during the process of transportation or the heating and cooling cycles of the sequencing reaction itself. A sequence so small cannot be accurately ascribed to any particular genus of bacteria.

On the other hand, the sequencing of specimen 2 was carried out perfectly. It presented excellent results in terms of phylogenetic and evolutionary relationship with the help of which the particular genus of the bacteria has been identified.

After BLASTn result analysis, top 16S rDNA sequences based on their high identity values and low E-values were selected for both bacteria. For sequence selection, the range set for identity values was 90% to 100%. This means that the selected sequences had a genomic configuration that was 90% to 100% identical to that of the query sequence. The E-value or the expected value is defined as the number of times the database match may have occurred by chance. Hence, a sequence with low E-value is considered a good match as it is unlikely to occur by chance.

3.3.1 Results for Specimen 1

The nucleotide consensus sequence derived after sequencing of the rDNA of Specimen 1 is given below:

a) Nucleotide sequence for 16S rDNA of Specimen 1 (582 bases):

>Consensus for segment for specimen Specimen1

TGTCGATTTGGAGGTTGTKCCCCWTGAGGGWCGTGGCTTCCGGAGCTAACGCRT
TAAATCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGAC

GGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAA
CCTTACCTRSYCTTGACATCCASAGAAYTTWSCAGAGATGSWTAWGGTGCCTTC
GGGAACTSTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTT
GGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGRTTCGGYC
GGGAACTCAAAGGAGACTGCCRGTGATAAACCGGAGGAAGGTGGGGATGACGT
CAAGTCATCATGGCCCTTACGASYAGGGCTACACACGTGCTACAATGGCGYATA
CAAAGAGAAGCGACCTCGCGAGAGCAAGCGGAACCTCAWAAAGTGCGTCGTAG
TCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCKT
WGAWMAAARTGCWCCGGTRATACGTTCCCGGSCCTTKTTCTC

b) Nucleotide Analysis:

BLAST results of the nucleotide sequence for Specimen 1

A graphic summary of the BLAST results for the target 16S rDNA sequence was obtained where the top red bar indicated the query sequence. The query sequence (in this case 16S rDNA) is represented by the numbered red bar at the top of the figure. Database hits are shown aligned to the query, below the red bar. Of the aligned sequences, the most similar are shown closest to the query. In this case, all the hits are high scoring database matches that align to most of the query sequence.

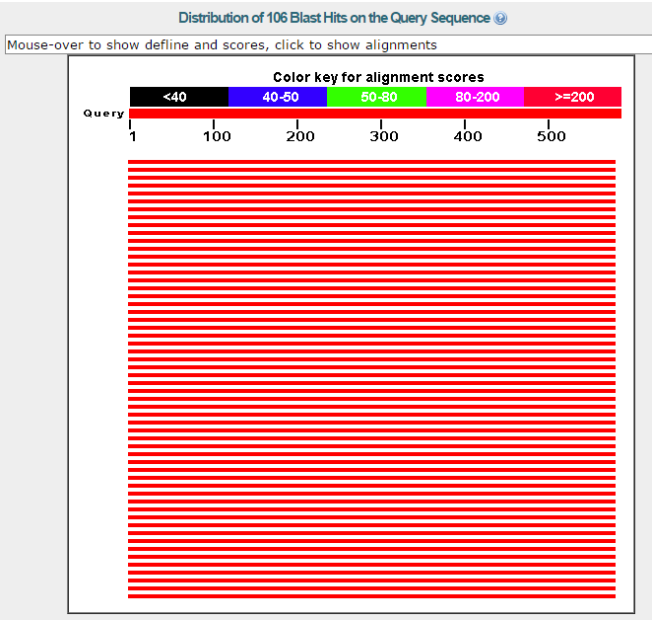
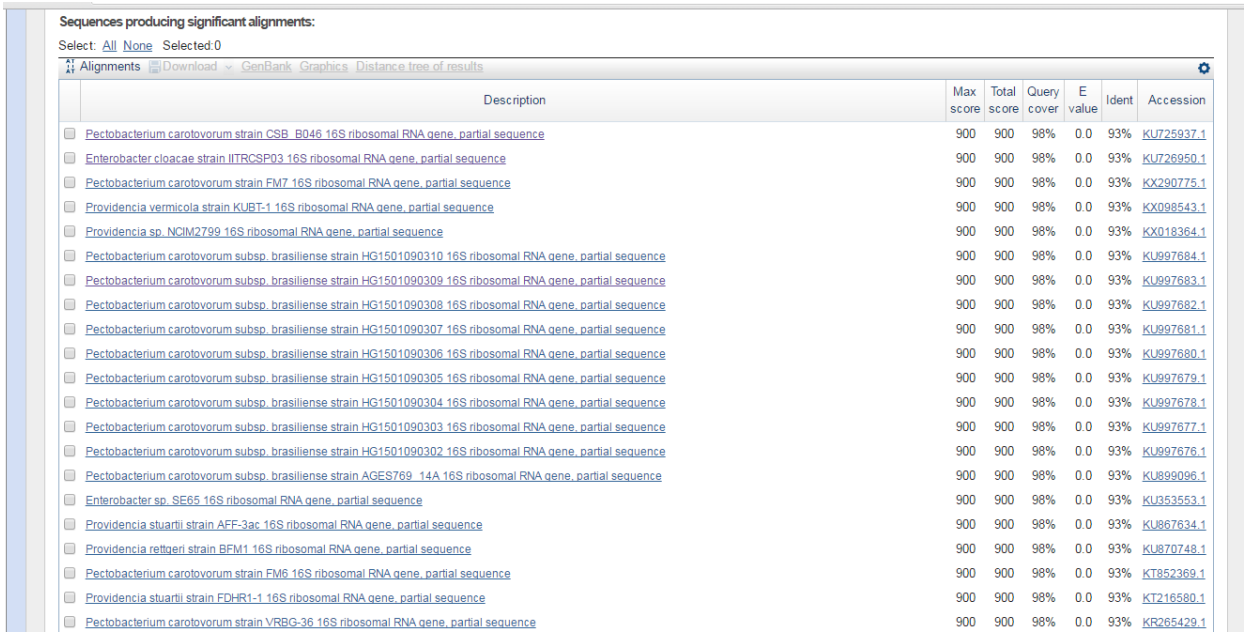


Fig 3.3.1.1 Graphic Summary of BLAST hits on query sequence of Specimen 1



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Pectobacterium carotovorum strain CSB_B046 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU725937.1
Enterobacter cloacae strain IITRCSP03 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU726950.1
Pectobacterium carotovorum strain FM7 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KX290775.1
Providencia vermicola strain KUBT-1 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KX098543.1
Providencia sp. NCIM2799 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KX018364.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090310 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997684.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090309 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997683.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090308 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997682.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090307 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997681.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090306 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997680.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090305 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997679.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090304 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997678.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090303 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997677.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090302 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU997676.1
Pectobacterium carotovorum subsp. brasiliense strain AGES769_14A 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU899096.1
Enterobacter sp. SE65 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU353553.1
Providencia stuartii strain AFF-3ac 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU867634.1
Providencia rettgeri strain BFM1 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KU870748.1
Pectobacterium carotovorum strain FM6 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KT852369.1
Providencia stuartii strain FDHR1-1 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KT216580.1
Pectobacterium carotovorum strain VRBG-36 16S ribosomal RNA gene, partial sequence	900	900	98%	0.0	93%	KR265429.1

Fig 3.3.1.2 Top BLAST hits that produced significant alignments with query sequence of Specimen 1

Organisms	**Accession ID**
Pectobacterium carotovorum strain CSB_B046	KU725937.1
Enterobacter cloacae strain IITRCSP03	KU726950.1
Pectobacterium carotovorum strain FM7	KX290775.1
Providencia vermicola strain KUBT-1	KX098543.1
Pectobacterium carotovorum subsp. brasiliense strain HG1501090310	KU997684.1
Brevibacillus sp. GI-9	FR686596.2

Table 3.1 List of organisms whose nucleotide sequences were selected

The identity value of all the selected sequences that matched with specimen 1 was 93% and the E-value was 0.0. For this bacterial strain, only the top 5 matches were selected. This is because below these 5, a series of sequences from *Pectobacterium* genus with the same nucleotide sequences were shown. As the sequence was very short in length and contained a

number of errors, the hits it matched with were not specific. In spite of these problems, further work has been carried out and is presented below.

Multiple Sequence Alignment (MSA) results for Specimen 1

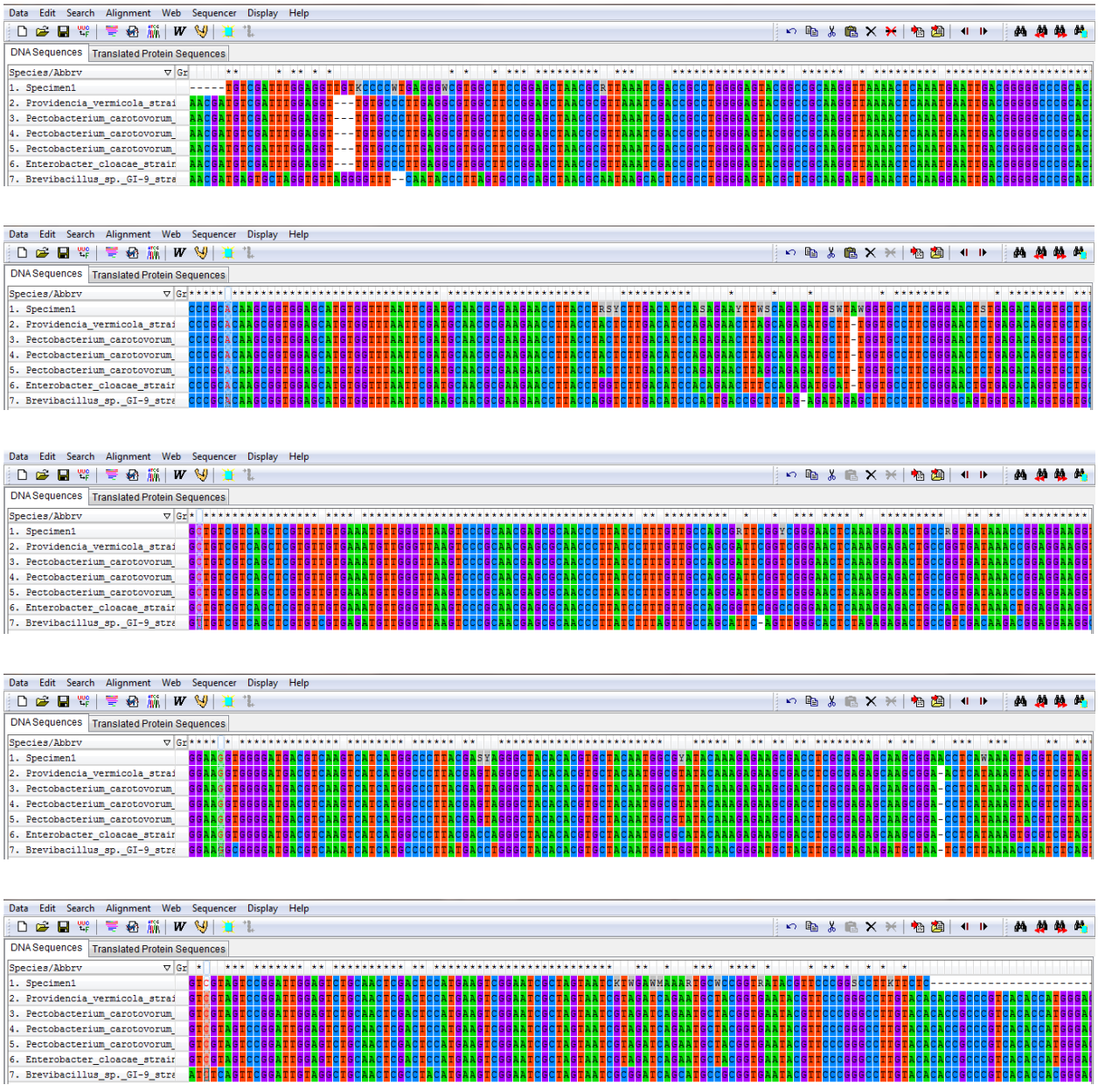


Fig 3.3.1.3 Entire sequence alignment of Specimen 1 with top 5 BLAST matches performed by Clustal Omega

In figure 3.3.4, the asterisk above each column indicates that homologous residues are aligned in every column. It is to be noted that the sequence for *Brevibacillus sp.* was not one of the hits in BLAST. It was added to this alignment because Specimen 1 was hypothesized to be of *Brevibacillus* genus from previous biochemical test results. However, this bacterial 16S rRNA gene was not anywhere in the BLAST hits list. Moreover, after studying the above alignment, it can be clearly observed that the query sequence of Specimen 1 has way more similarity with the BLAST sequences than with that of *Brevibacillus*. From this it can be assumed that Specimen 1 is not of *Brevibacillus* genus. For further verification, phylogenetic analysis has been done.

Phylogenetic Results

- **Decoding a Phylogenetic Tree**

The horizonal lines of a phylogenetic tree are branches that represent evolutionary lineages changing over time. The longer the branch in the horizonal dimension, the more amount of genetic changes occurred over time. It can be observed that the trees below have numbers on every branch. These numbers indicate the amount of genetic change in terms of nucleotide substitutions per site. These numbers, also known as units of branch length, result when the number of nucleotide substitutions in a DNA sequence is divided by the entire length of its sequence (Rambaut, 2013). For 16S rDNA, if two species or strains have 0.00 units of branch length, it usually means that they have evolved at the same rate which is why their 16S rDNA have no differences in nucleotide sequences. If they are under the same tree branch (which is usually the case), it confirms that these two strains/species have a recent common ancestor.

Phylogenetic trees can be of many types. As mentioned before, three types of trees were constructed for the analysis of each bacterial specimen: Maximum likelihood tree, Neighbour Joining tree, and Maximum Parsimony.

Maximum likelihood: a consistent and efficient method for deducing evolutionary relationships. Maximum likelihood method works by comparing species based on their most common traits and finding the maximum likelihood of the occurrence of evolutionary relationships shown in the resultant tree.

Neighbour joining: this method is a clustering algorithm that is capable of making trees very quickly. Neighbour joining can incorporate bootstrapping to increase reliability of the constructed tree.

Maximum parsimony: this method tries to derive a phylogenetic tree with the maximum parsimony or the least amount of divisions or splits in branching; to create a tree in a way that minimizes the number of mutations and evolutionary steps.

All of the above tree types were constructed for each specimen in this project. This is done primarily to increase reliability of the generated data so that the deduction of the species' genotype can be done more accurately and confidently.

Phylogenetic tree results for Specimen 1

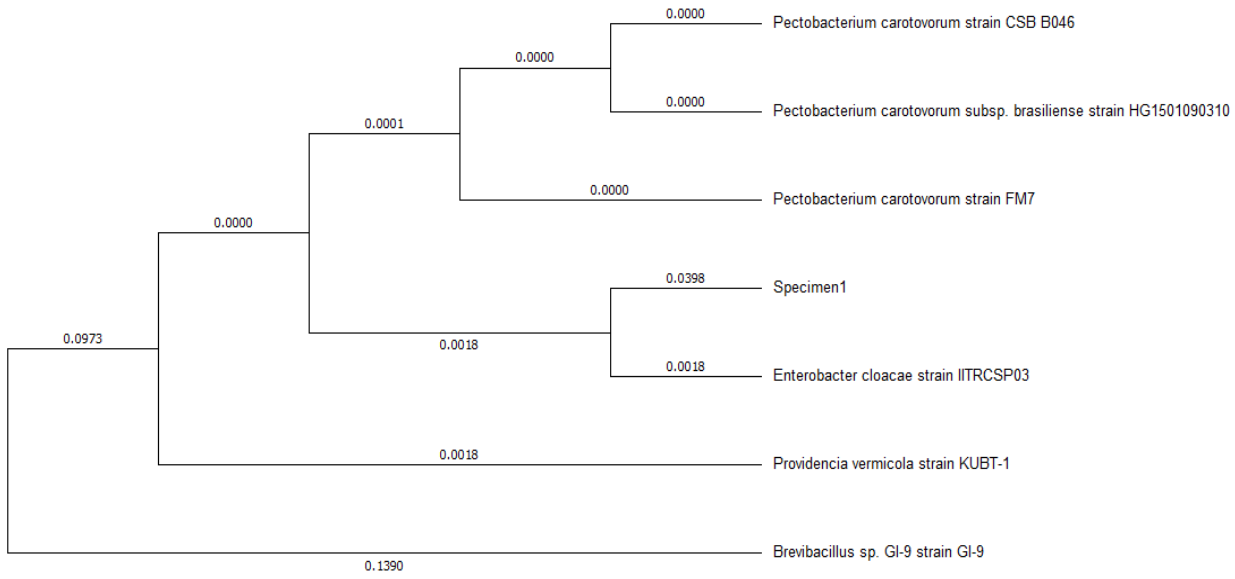
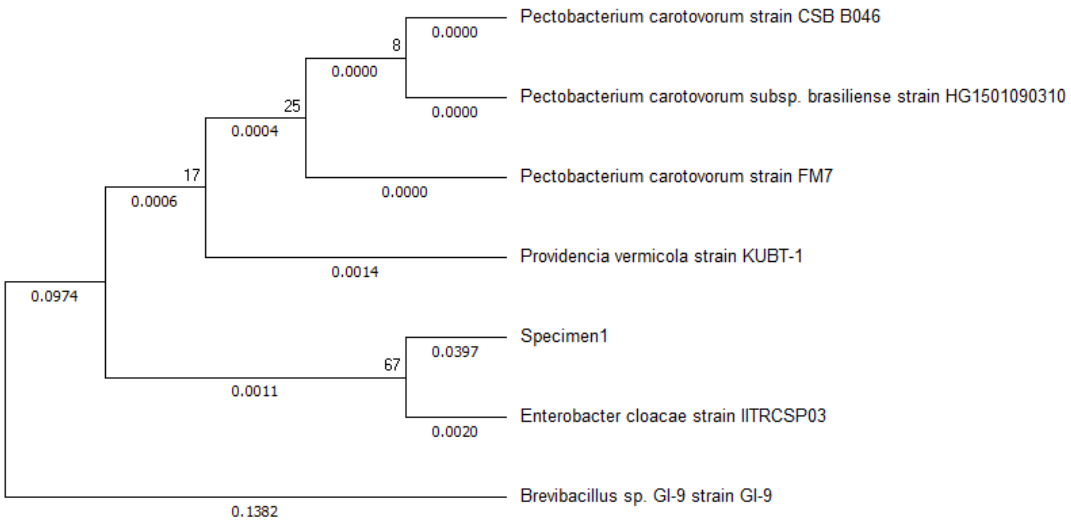
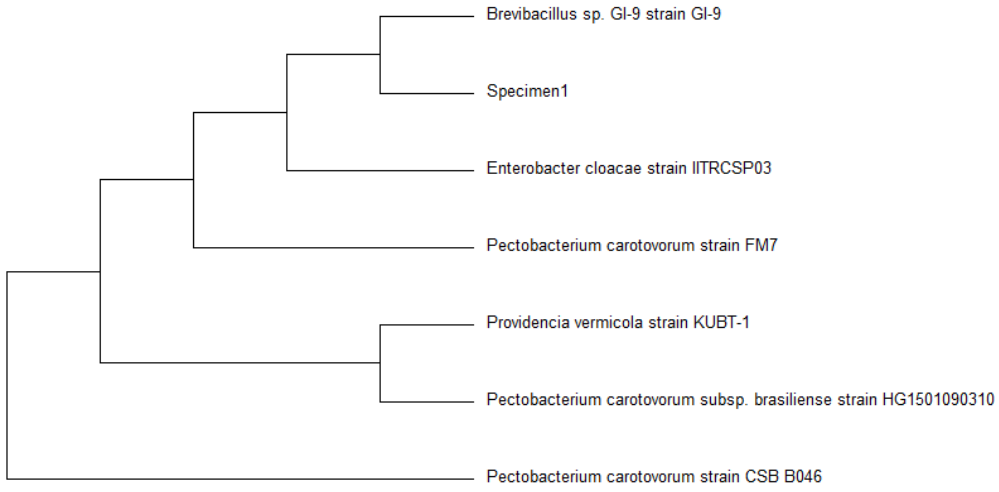


Fig 3.3.1.4 Maximum Likelihood tree for specimen 1



3.3.1.5 Neighbour Joining tree for specimen 1



.

Fig 3.3.1.6 Maximum Parsimony tree for specimen 1

Interpretation

Phylogenetic analysis of the 16S rRNA gene sequence for specimen 1 was performed. Fig 3.3.1(4-6) showed the phylogenetic relationship between the obtained aligned sequences by BLAST.

In the maximum likelihood and neighbour joining trees, it can be observed that the 16S consensus for specimen 1 seems to be evolutionary related to *Enterobacter clocae*. However, keeping in mind the diversity in the bacterial origins of the BLAST matches and the conciseness of the length of the target consensus, it will not be wise to form a hypothesis about the genotype of this bacterium.

3.3.2 Results for Specimen 2

The nucleotide consensus sequence derived after sequencing of the rDNA of Specimen 1 is given below:

a) Nucleotide sequence for 16S rDNA of Specimen 2 (1,337 bases**):**

>Consensus for segment for specimen Specimen2

YCYKCTGACGAGCGGCGGACGGGKAGWAGAATGGGGATCTGCCCGAWGAGGG
GGATAACTACTGKAAMCGGTGGCTAATACCGCATAATCTCTTAGGAGCAAAGC
AGGGGACCTTCGGGCCTTGCGCTGTCGGATGAACCCATATGGGATTAGCTAGTA
GGTAAGGTAATGGCTTACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGG
GGAATATTGCACAATGGGCGCAAGCCKTGATGCAGCCATGCCGCGTGTATGAA
GAAGGCCCTAGGGTTGTAAAGTACTTTCAGTCGGGAGGAAGGCGTTGATGTTAA
TACCATCAACGATTGACGTTACCGACAGAAGAAGCACCGGCTAACTCCGTGCCA
GCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAA
AGCGCACGCAGGCGGTTAATTAAGTTAGATGTGAAATCCCCGGGCTTAACCTGG
GAATGGCATCTAAGACTGGTTAGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCA
TGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGGTGGCGAAGCGGC
CCCCCTGGGACAAAGACTK????TCAGGTGCGAAAGCGTGGGGAAGCAAACAGG
GATTAGATACCCTGGTAGTCCACGCTGTAAWCGATGTCGATTTGGAGGTTGTTC
CCTAGAGGAGTGGCTTCCGSAGCTAWCGCGTTAAATCGWCCGCCTGGGGAGTA
CGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGG
AGCATGTGGTTTAATTCGATGCAACGCGAAGAMCCTTACCTACTCTTGACATCC
AGAGAATTTRGCAGAGATGCTTWAGTGCCTTCGGGAACTCTGAGACAGGTGCT
GCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAG
CGCAACCCTTATCCTTTGTTGCCAGCGATTCGGTCGGGAACTCAAAGGAGACTS
CCGGTGATAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTA
CGAGTAGGGCTACACACGTGCTACAATGGCGTATACAAAGAGAAGCGACCTCG
CGAGAGCAAGCGGAACTCATAAAGTACGTCGTAGTCCGGATTGGAGTCTGCAA
CTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTAGARTCAGAATGCTACGGT
GAATACGTTCCCGGGCCTTGTTCTCCCCCCGTCACACATGGAGTGGGTTGMAAA

b) BLAST results of the nucleotide sequence for Specimen 2

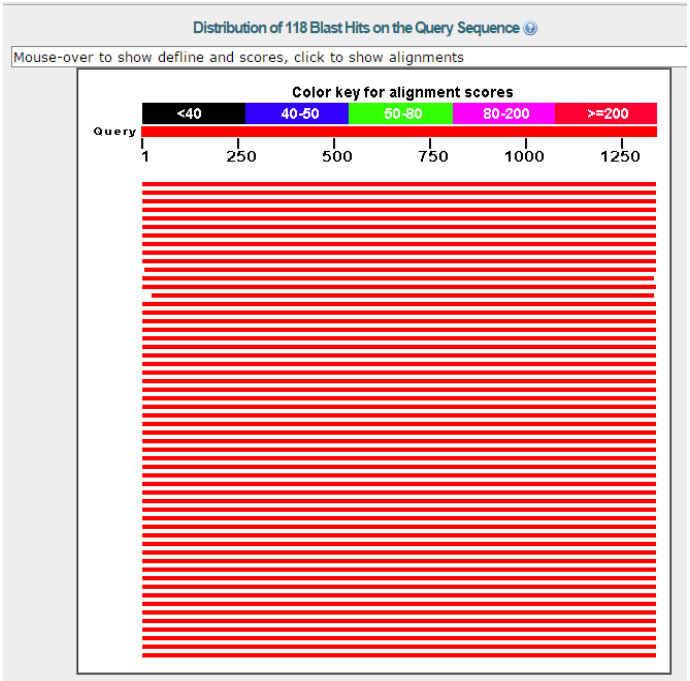
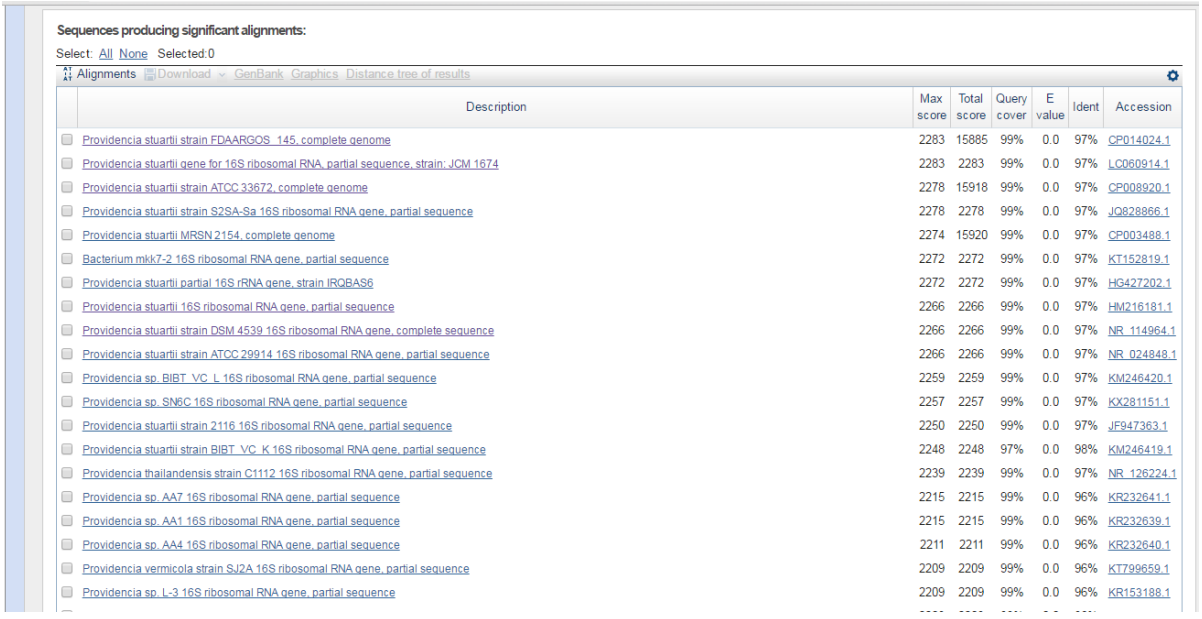


Fig 3.3.2.1 Graphic Summary of BLAST hits on query sequence of Specimen 2



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Providencia stuartii strain FDAARGOS_145, complete genome	2283	15885	99%	0.0	97%	CP014024.1
Providencia stuartii gene for 16S ribosomal RNA, partial sequence, strain: JCM 1674	2283	2283	99%	0.0	97%	LC060914.1
Providencia stuartii strain ATCC 33672, complete genome	2278	15918	99%	0.0	97%	CP008920.1
Providencia stuartii strain S2SA-Sa 16S ribosomal RNA gene, partial sequence	2278	2278	99%	0.0	97%	JQ828866.1
Providencia stuartii MRSN 2154, complete genome	2274	15920	99%	0.0	97%	CP003488.1
Bacterium mkk7-2 16S ribosomal RNA gene, partial sequence	2272	2272	99%	0.0	97%	KT152819.1
Providencia stuartii partial 16S rRNA gene, strain IRQBAS6	2272	2272	99%	0.0	97%	HG427202.1
Providencia stuartii 16S ribosomal RNA gene, partial sequence	2266	2266	99%	0.0	97%	HM216181.1
Providencia stuartii strain DSM 4539 16S ribosomal RNA gene, complete sequence	2266	2266	99%	0.0	97%	NR_114964.1
Providencia stuartii strain ATCC 29914 16S ribosomal RNA gene, partial sequence	2266	2266	99%	0.0	97%	NR_024848.1
Providencia sp. BIBT_VC_L 16S ribosomal RNA gene, partial sequence	2259	2259	99%	0.0	97%	KM246420.1
Providencia sp. SN6C 16S ribosomal RNA gene, partial sequence	2257	2257	99%	0.0	97%	KX281151.1
Providencia stuartii strain 2116 16S ribosomal RNA gene, partial sequence	2250	2250	99%	0.0	97%	JF947363.1
Providencia stuartii strain BIBT_VC_K 16S ribosomal RNA gene, partial sequence	2248	2248	97%	0.0	98%	KM246419.1
Providencia thailandensis strain C1112 16S ribosomal RNA gene, partial sequence	2239	2239	99%	0.0	97%	NR_126224.1
Providencia sp. AA7 16S ribosomal RNA gene, partial sequence	2215	2215	99%	0.0	96%	KR232641.1
Providencia sp. AA1 16S ribosomal RNA gene, partial sequence	2215	2215	99%	0.0	96%	KR232639.1
Providencia sp. AA4 16S ribosomal RNA gene, partial sequence	2211	2211	99%	0.0	96%	KR232640.1
Providencia vermicola strain SJ2A 16S ribosomal RNA gene, partial sequence	2209	2209	99%	0.0	96%	KT799659.1
Providencia sp. L-3 16S ribosomal RNA gene, partial sequence	2209	2209	99%	0.0	96%	KR153188.1

Fig 3.3.2.2 Top BLAST hits that produced significant alignments with query sequence of Specimen 2

Organism	**Accession ID**
Providencia stuartii strain S2SA-Sa	JQ828866.1
Providencia thailandensis C1112 (T)	-
Providencia thailandensis strain C1112	NR123224.1
Providencia stuartii strain JCM 1674	LC060914
Providencia stuartii	HM216181.1
Providencia stuartii strain ATCC 29914	NR024848.1
Providencia stuartii strain IRQBAS6	HG427202
Providencia stuartii strain DSM 4539	NR114964.1
Providencia vermicola strain SJ2A	KT799659.1
Providencia rettgeri strain ALK058	KC456564.1
Uncultured *Providencia* sp. clone F4feb.23	GQ417976.1

Table 3.2 List of organisms whose nucleotide sequences were selected

Other than *P. vermicola* and *P. rettgeri*, the identity value of all the selected sequences that matched with query sequence was 97% and the E-value was 0.0. For *P. vermicola* and *P. rettgeri*, the identity value was 96%. For this bacterial strain, ten matches were selected among which one (*Providencia thailandensis* C1112 (T)) was derived from ezbiocloud instead of NCBI database. Ezbiocloud is a similar program like NCBI where BLAST can be performed. *Providencia thailandensis* C1112 (T) was the top match in that program and hence was selected along with the NCBI BLAST hits for MSA.

Multiple Sequence Alignment (MSA) results for Specimen 2

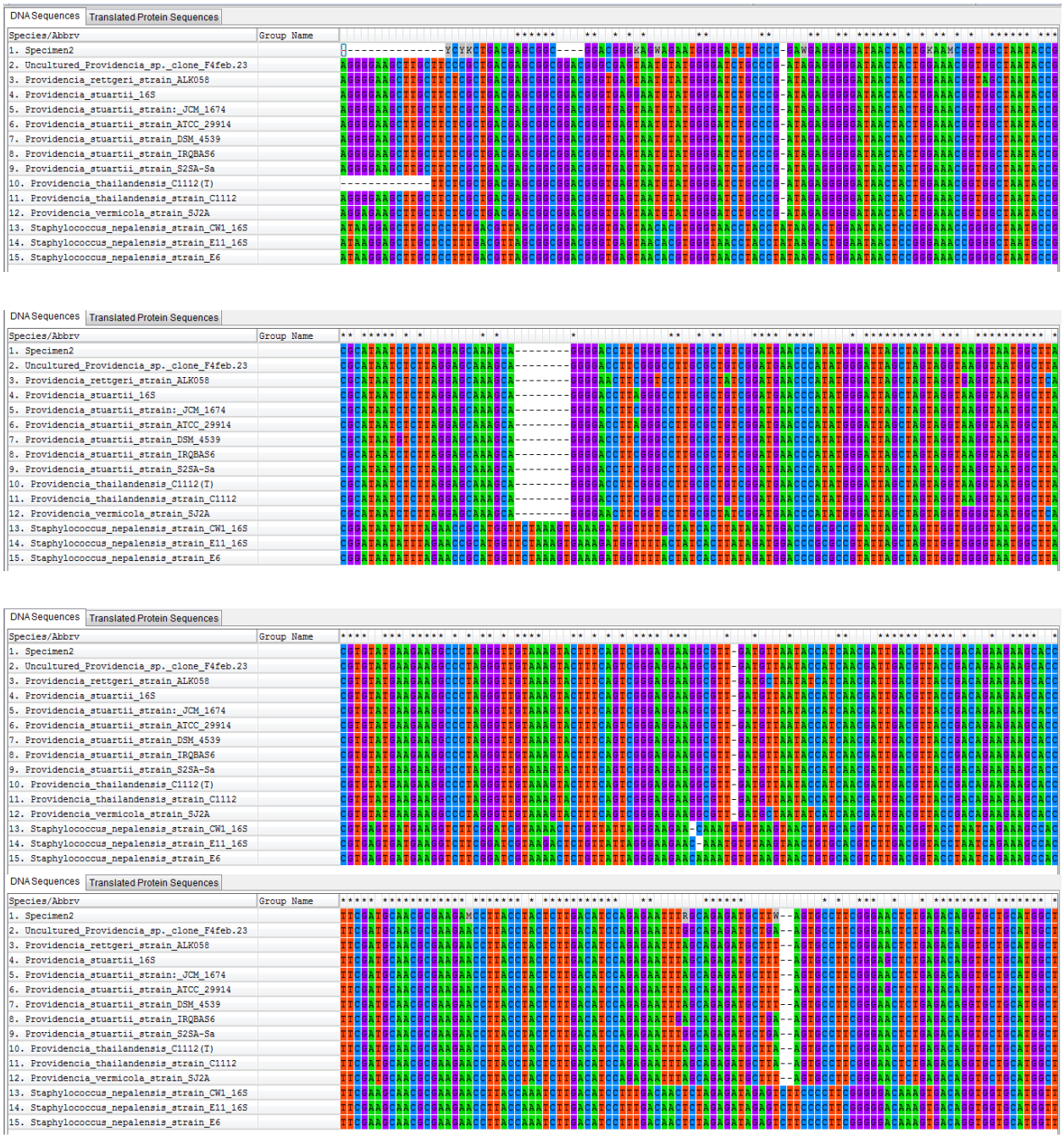


Fig 3.3.2.3 Sequence alignment of Specimen 2 with top BLAST matches performed by Clustal Omega

It is to be noted that the three sequences for *Staphylococcus sp.* in the bottom rows of the alignment in Fig 3.3.2.3 were not matches in BLAST. They were added to this alignment

because Specimen 2 was previously hypothesized to be of *Staphylococcus* genus from biochemical test results. However, like the *Brevibacillus* sequence for specimen 1, *Staphylococcus* 16S rRNA gene was also not anywhere in the BLAST hits list. Hence Specimen 2 obviously has way more similarity with the sequences that resulted from BLAST than with those of the *Staphylococcus*. From this it can be assumed that Specimen 2 is not of *Staphylococcus* genus. In fact there is very high possibility that it belongs to *Providencia* genus. For further verification, phylogenetic analysis has been done.

Phylogenetic tree results for Specimen 2

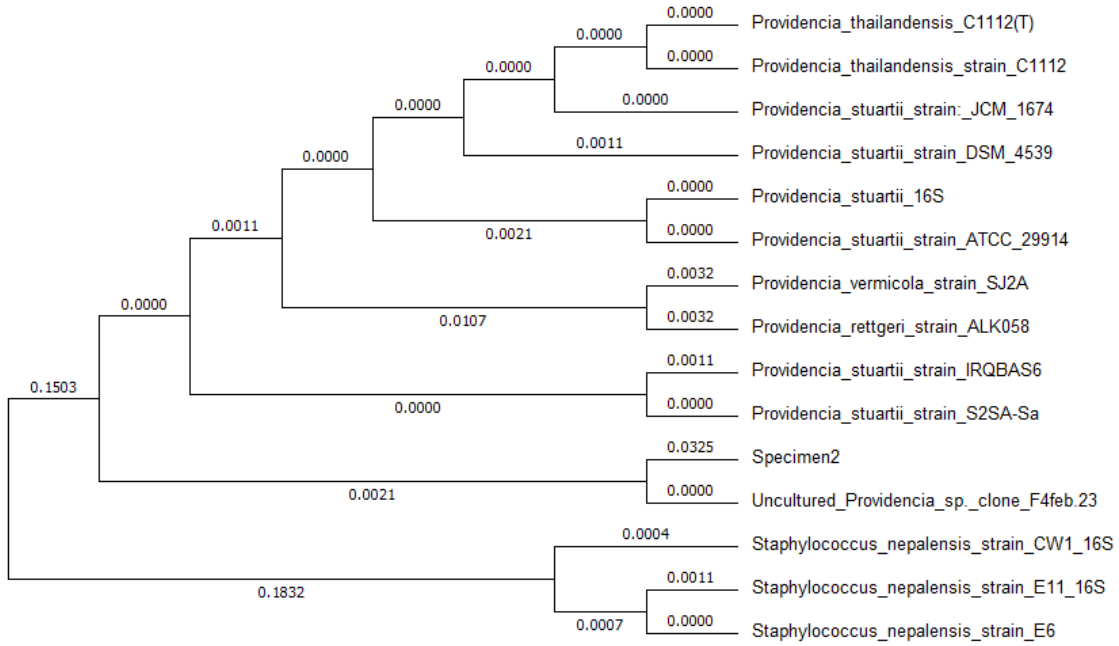


Fig 3.3.2.4 Maximum likelihood tree for specimen 2

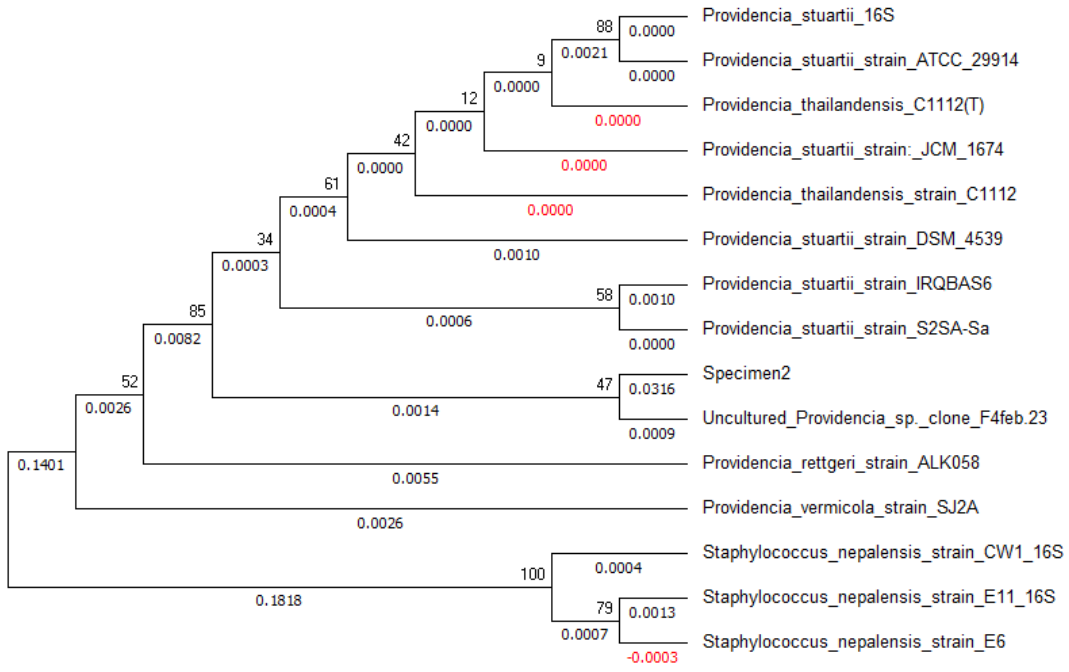


Fig 3.3.2.5 Neighbour Joining tree for specimen 2

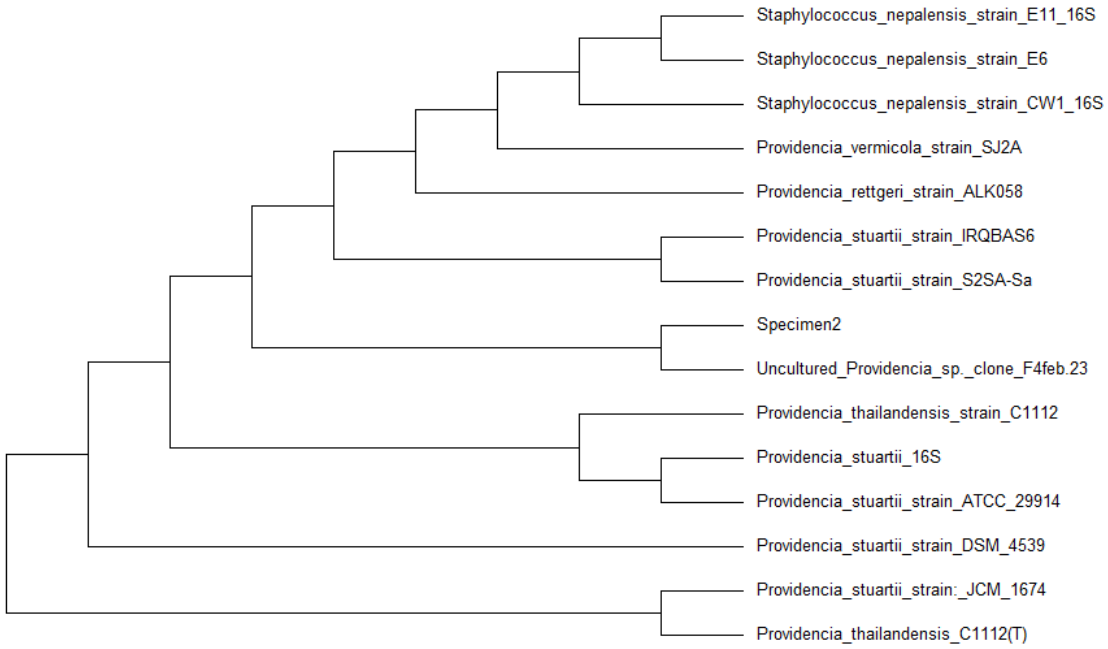


Fig 3.3.2.6 Maximum Parsimony tree for specimen 2

Interpretation

Phylogenetic analysis of the 16S rRNA gene sequence for specimen 2 was performed. Fig 3.3.2 (4-6) showed the phylogenetic relationship between the consensus and the obtained aligned sequences from BLAST. It is noteworthy that the sequence homology of bacterial 16S rDNA sequences as well as all three of the above trees place specimen 2 in the phylogenetic branch of "*Providencia*", particularly "Uncultured_*Providencia*_sp". The consistency in the BLAST hits and the tendency of grouping specimen 2 with *Providencia* sp. in the above trees enables to deduce the hypothesis that specimen 2 belongs to the genus *Providencia*. As three different methods of phylogenetic tree construction (Maximum Likelihood, Maximum Parsimony, and Neighbour Joining) were utilized in this project, the result representing the placement of specimen 2 under the genus "*Providencia*" is considered to be reliable.

Phylogenetic tree results for both specimens 1 and 2

All the MSA sequences were amassed from which trees were constructed to study the evolutionary relationships among the two specimens

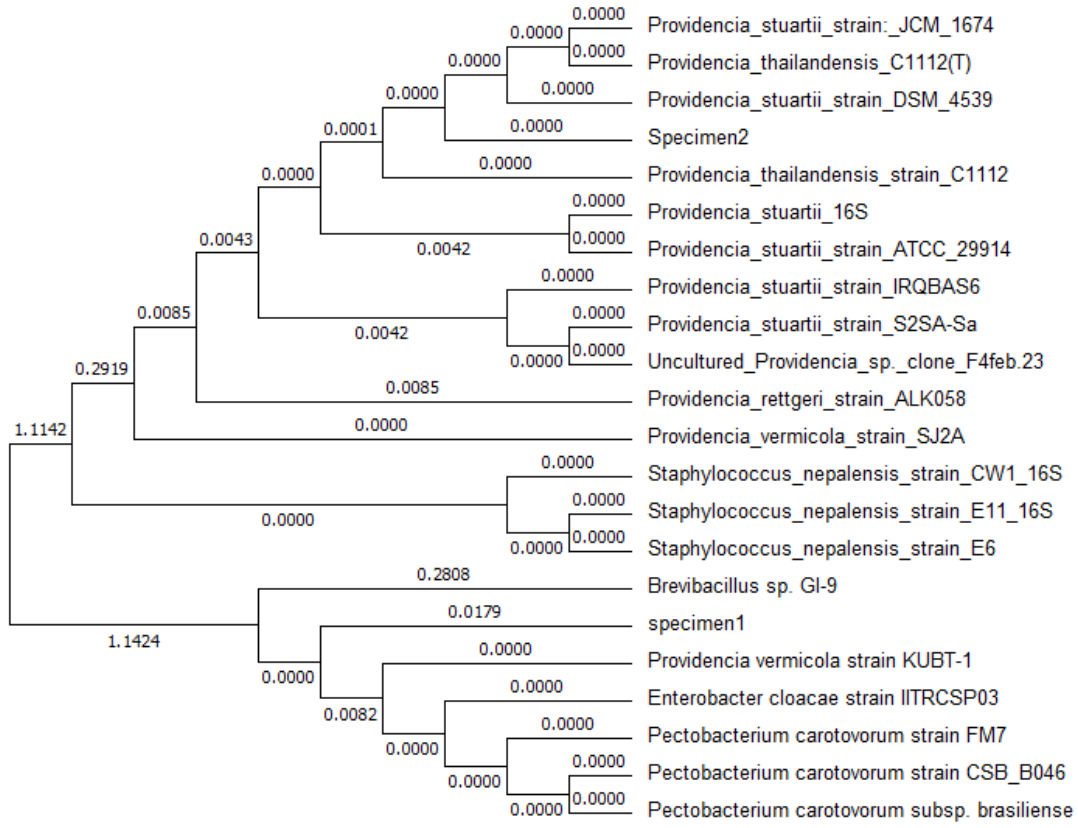


Fig 3.3.2.7 Maximum likelihood tree for both specimen and their matching 16S sequences

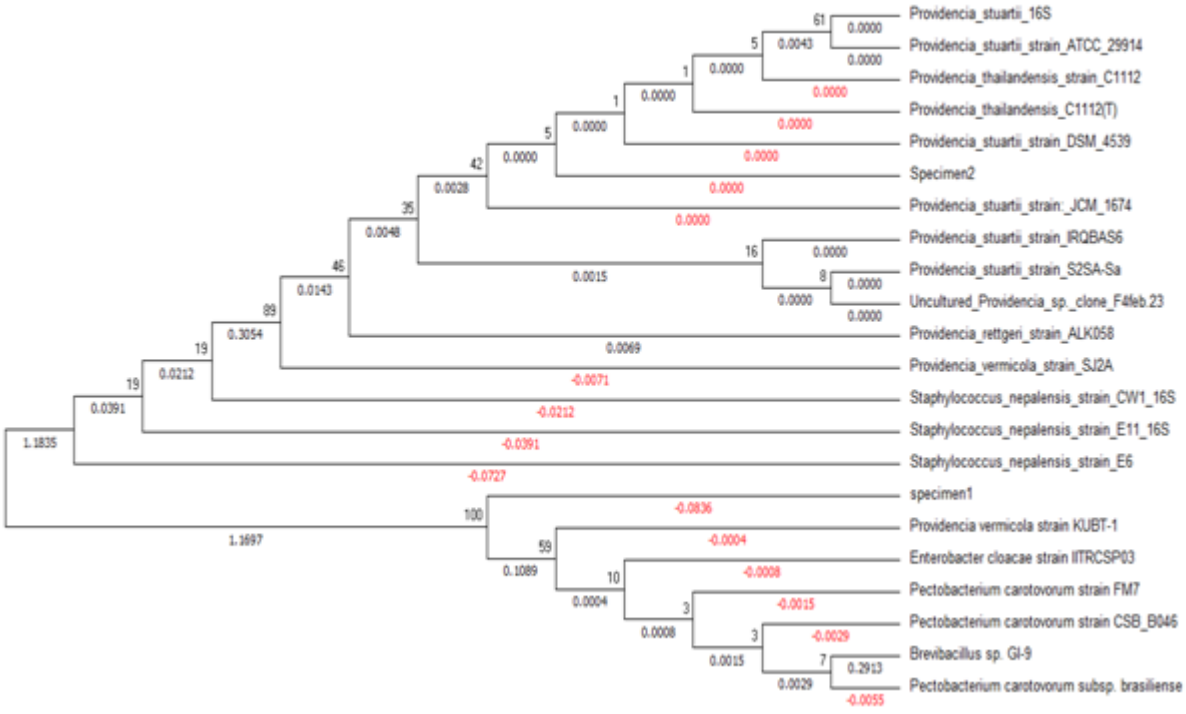


Fig 3.3.2.8 Neighbour Joining tree for both specimen and their matching 16S sequences

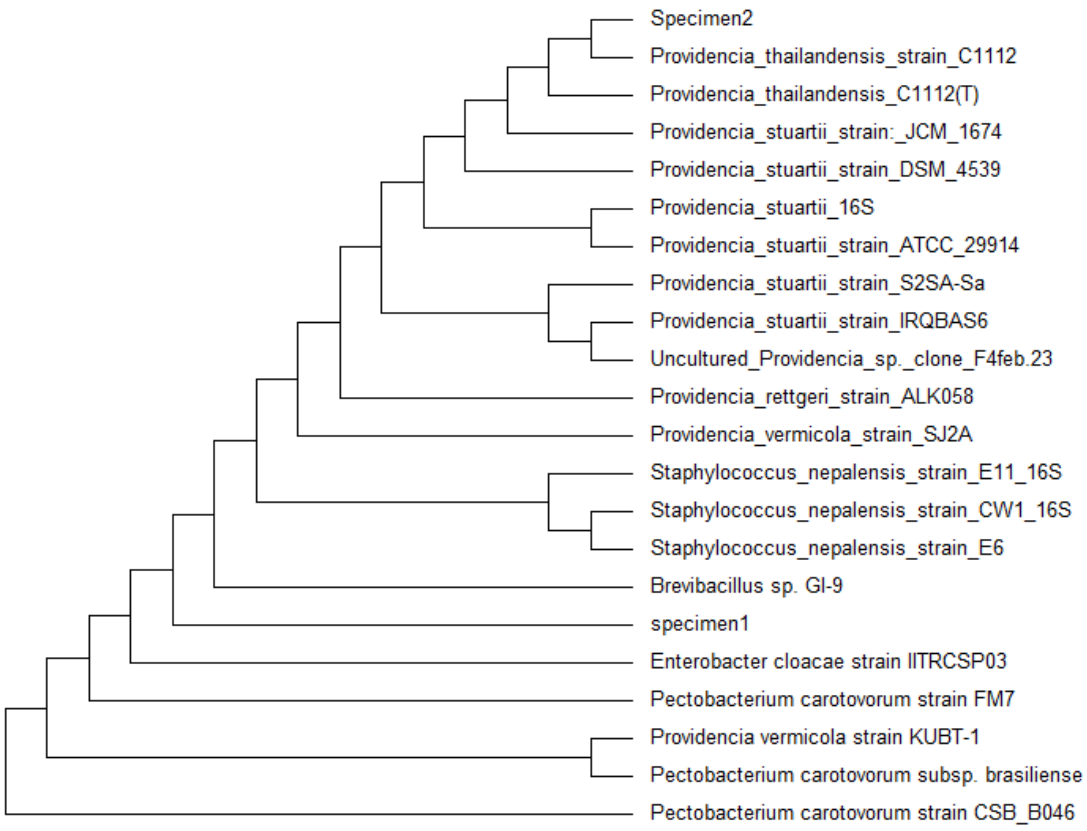


Fig 3.3.2.8 Maximum Parsimony tree for both specimen and their matching 16S sequences

Chapter Four

Discussion

4. Discussion

Reactive red is a common dye and very widespread in the textile industry. The two bacteria used in this project are both good at degrading this dye. However, they are not as successful at breaking down Yellow 4GL as it is a difficult chemical to biodegrade. Specimen 1 is able to degrade 5% concentration of reactive dye while specimen 2 can discolour 1% of it in 24 hours. In the first part of this project, the identities of the two bacteria were assumed to be *Brevibacillus laterosporus* and *Staphylococcus nepalensis* with 33% and 32% accuracy respectively. The percent accuracy here designates the amount of biochemical tests covered to assure the identity of the bacteria. Because the percent accuracy is low, it is not surprising that in the second part of the project, the identities of the bacteria turned out to be very different from that predicted.

Among the two samples, specimen 2 resulted in excellent outcome. After much analysis of the multiple sequence alignment and the maximum likelihood, maximum parsimony, and neighbour joining trees, specimen 2 can be hypothesized to belong to the genus "*Providencia*". The tendency of repeatedly clustering specimen 2 with "uncultured *Providencia* sp." in all the phylogenetic trees is a strong indication that Specimen 2 belongs to this genus.

From the trees cumulative of the two consensus and their respective matching sequences (Fig.3.3.2.7, 8, and 9), it can be assumed that the 16S rDNA of specimen 2 is more homologous with the 16S rDNA of *P. stuartii* and *P. thailandensis* than with those of *P. vermiocola* and *P. rettgeri*. In order to specifically identify the species of this *Providencia* strain, further molecular analysis needs to be done. It is possible that a whole genome analysis may help to determine the exact species of the *Providencia* strain isolated in this project.

The ability of members of genus *Providencia* to biodegrade textile dyes is not widely known. So far in current literature, only *Providencia rettgeri* has been documented to be capable of textile dye degradation (Lade, Kadam, Paul, Govindwar, 2015). Hence, if further analysis results in the identification of this species as any *Providencia* species other than *P. rettgeri*, the dye-degrading capability of a new *Providencia* species will be successfully discovered.

Due to the technical error in sequencing of Specimen 1, it cannot be attributed to any specific genus. The consensus sequence provided is so small that it matches with a great variety of bacterial species. Nevertheless, all the BLAST hits for this specimen were members of *Enterobacteriaceae*. Hence, it can be assumed that this specimen is also an enteric bacterium. Further molecular analysis needs to be done to help determine the identity of this specimen.

Conclusion

Hence after phylogenetic analysis of the 16S rDNA, this project forms the hypothesis that the bacteria of Specimen 2 belongs to the genus ***Providencia*** and that Specimen 1 is an enteric bacteria. Further studies will reveal more information about the exact identification of both strains of bacteria.

The aim of biodegradation is to use indigenous bacteria to decontaminate a polluted site. The land and water bodies that are being polluted every day with textile effluents can be cleaned up by biodegradation. This easy, cheap, and environmentally-friendly solution to textile pollution will definitely help RMG dependent countries like Bangladesh. It is believed that if both the bacteria isolated for this project are used in a consortium of other dye-degrading bacteria, high biodegradation of azo dyes will be achieved. In this project, each bacterium took 24-48 hours for the breakdown of Reactive red. However, this breakdown is expected to occur much faster in combination with a wide variety of other dye degraders. By degrading hazardous azo dyes like Reactive Red into their non-toxic constituents, these bacteria can detoxify textile effluents. This will greatly reduce the cost of effluent treatment and offer a safer environment for everyone.

Providencia species is well known for causing urinary tract infections. The finding in this project that strains of *Providencia*, hypothetically of *P. stuartii*, can degrade azo dyes may lead to the discovery of more unknown characteristics of the bacterium.

Chapter Five

References

References

1. Begum, R.A., Zaman, M. W., Mondol, A. T. M. A. I., Islam, M. S., Hossain, K. M. F. (2011). Effects Of Textile Industrial Waste Water And Uptake. *Bangladesh Journal of Agricultural Research* , 319-331.

2. Business Nonstop Desk. (2014). Tamil Nadu Government Announce RS.700 Crore Scheme for Small and Medium Dyeing Units. *Business Nonstop*.

3. Blanken, R. (2016). *What is Fiber Reactive Dye?* Retrieved from About Style: http://diyfashion.about.com/od/dyingandscreenprinting/f/What-Is-Fiber-Reactive-Dye.htm

4. Carmen, Z., Daniela, S. (2012). Textile Organic Dyes – Characteristics, Polluting Effects and Separation/Elimination Procedures from Industrial Effluents – A Critical Overview, Organic Pollutants Ten Years After the Stockholm Convention - Environmental and Analytical Update, Dr. Tomasz Puzyn (Ed.), ISBN: 978-953-307-917-2, *InTech* , DOI: 10.5772/32373

5. Chengalroyen, M. D. &Dabbs, E. R. (2013). The microbial degradation of azo dyes: minireview. *World Journal of Microbiology and Biotechnology* ,29:389. Doi: 10.1007/s11274-012-1198-8

6. Chequer et al (2013). Textile Dyes: Dyeing Process and Environmental Impact, Eco-Friendly Textile Dyeing and Finishing, Dr. Melih Gunay (Ed.), *InTech* , DOI:10.5772/53659.

7. consensus sequence. (n.d.) *Collins Dictionary of Biology, 3rd ed.*. (2005). Retrieved September 1 2016 from http://medical-dictionary.thefreedictionary.com/consensus+sequence

8. Donatelli, J. (2016). *The History of Fabric Dye*. Retrieved from Zady: https://zady.com/features/the-history-of-fabric-dye

9. Environment and Climate Change Canada. (2012). *The Chemicals Management Plan Substance Groupings Initiative Aromatic Azo- and Benzidine-Based Substances*. Retrieved from Environment and Climate Change Canada: https://www.ec.gc.ca/ese-ees

10. He, F. (2011). *E.coli* Genomic DNA Extraction. *Bio-protocol* Bio101: http://www.bio-protocol.org/e97

11. Lade, H., Kadam, A., Paul, D., Govindwar, S (2015). Biodegradation and detoxification of textile azo dyes by bacterial consortium under sequential microaerophilic/aerobic processes. *EXCLI* , 158-174.

12. Lakherwal, D. (2014). Adsorption of Heavy Metals: A Review. *International Journal of Environmental Research and Development* , Volume 4, pp. 41-48.

13. Lobo, I. (2008) Basic Local Alignment Search Tool (BLAST), *Nature Education* 1(1):215

14. Mirdha, R. U. (July 17, 2016). Bangladesh remains second largest garments exporter, against all odds. *The Daily Star*.

15. Patwardhan A, Ray S, Roy A (2014) Molecular Markers in Phylogenetic Studies – A Review. J Phylogen Evolution Biol 2: 131. doi:10.4172/2329- 9002.1000131

16. Pereira. (2012, July Tuesday). Natural Fabric Dyes.

    Available at: http://twentyfirstcenturylady.blogspot.com/2012/07/natural-fabric-dyes.html?m=1

17. Pereira, L., Alves, M. (2012). Dyes-Environmental Impact and Remediation . In E. G. Abdul Malik, *Environmental Protection Strategies for Sustainable Development* (pp. 112-154). Springer .

18. Puvaneswari, N., Muthukrishnan, J., Gunasekaran, P. (2006). Toxicity assessment and microbial degradation of azo dyes. *Indian Journal of Experimental* Biology, 44(8):618-26.

19. Rambaut, A. (2013, 08 12). *epidemic*. Retrieved September 2016, from How to read a phylogenetic tree: http://epidemic.bio.ed.ac.uk/how_to_read_a_phylogeny

20. Rashid, P. (2011, February 25). *Industrial Pollution in Bangladesh*. Retrieved from Getty Images: http://www.gettyimages.fr

21. Sivakumar, K.K., Balamurugan, C., Ramakrishnan, D., Bhai, L.H., (2011). Assessment Studies On Wastewater Pollution By Textile Dyeing And Bleaching Industries At Karur, Tamil Nadu. *Rasayan Journal of Chemistry* , Vol.4, 264-269.

22. The Parliamentary Office of Science and Technology. (2014). *The Environmental, Health and Economic Impacts of Textile Azo Dyes.* London.

23. The World Bank Group. (2016). *Introduction to Wastewater Treatment Processes*. Retrieved from The World Bank: http://water.worldbank.org/shw-resource-guide/infrastructure/menu-technical-options/wastewater-treatment

24. Understanding Evolution team. (n.d.). *Phylogenetic systematics, a.k.a. evolutionary trees.* Retrieved September 2016, from Understanding Evolution: http://evolution.berkeley.edu/evolibrary/article/0_0_0/phylogenetics_01

Appendix-I

Reagents

The reagents used in the above procedure were made using the following composition (He,F., 2011):

Composition of reagents

LB medium

1% tryptone

0.5% yeast extract

200 mMNaCl

TE buffer

10 mMTris-Cl (pH 8.0)

1 mM EDTA (pH 8.0)

Lysis buffer (10 ml)

9.34 ml TE buffer

600 µl of 10% SDS

60 µl of proteinase K (20 mg ml^{-1})

Appendix-II

Instruments

Equipment	Source Company
Autoclave	SAARC
Freeze(-20°C)	Siemens
Incubator	SAARC
Micropipette (10-100µl)	Eppendorf, Germany
Micropipette (20-200µl)	Eppendorf, Germany
Oven, Model:MH6548SR	LG, China
pH meter, Model: E-201-C	Shanghai Ruosuaa Technology Company China
Refrigerator (4ºC) Model: 0636	Samsung
Safety cabinet	SAARC
Class II Microbiological Shaking Incubator, Model: WIS-20R	Daihan Scientific, Korea
Vortex Mixture	VWR International
Water bath	Korea
Weighing balance	ADAM EQUIPMENT™, United Kingdom