Expression of Genetical Diversity of *Kappaphycus* Sp. Among Other Algae Using Bioinformatics Tools

P. Rajasulochana[1], SowmiyaVeronica[2]

[1]Associate Professor, Dept .Bioinformatics & Genetics, Bharath Institute of Higher Education and Research, Bharath University, Chennai, India

[2]Student, Dept .Bioinformatics., BharathUniversity, Chennai, India

Abstract: *Kappaphycus sp. belongs to the family of red algae and is one of the most important commercial sources of carrageenans. As the 18S gene of the whole DNA codes for the conserved segment in the eukaryotic organisms that speaks about the personal identity of the organisms, these sequences have been used to describe the genetic relation of Kappaphycus sp. with that of other algae. The Clustal X and Tree View analysis of these sequences has shown the diversity and convergence in the genetic identity matrix created for the selected reference sequences and the Kappaphycus sp. sequences used. This has enabled us to design the genetic identity matrix thus enabling the explanation of the relation that the algae Kappaphycus sp. demonstrated with other algae.*

Keywords: Red algae; *Kappaphycus sp*.; Genetic diversity; Clustal X; Tree View

1. Introduction

Kappaphycus sp. is a species of red alga and it is one of the most important commercial sources of carrageenans it is a family of gel-forming, viscosifying polysaccharides. They grow to two meters in length and is green or yellow in color. It is a very fast-growing algae and is known to double its biomass in every 15 days.The carrageenan secreted by these algae is of different types and differ in composition and conformation, resulting in a wide range of rheological and functional properties. Carrageenans are used in a variety of commercial applications as gelling, thickening, and stabilizing agents, especially in food products such as frozen desserts, chocolate milk, cottage cheese, whipped cream, instant products, yogurt, jellies, pet foods, and sauces [1-5]. They are also used in pharmaceutical formulations, cosmetics, and industrial applications such asmining. Carrageenan is extracted from this generally in two ways. In native extraction, the seaweed is made into an aqueous solution, and the residue is filtered, leaving nearly pure carrageenan. The alkaline-modified method is less expensive and easier. The seaweed is mixed in an alkali solution, leaving a mixture of carrageenan and cellulose that can be sold as semi-refined carrageenan. These seaweeds have been widely used for human consumption in many parts of the world. In general marine algae has (i) protein with all the essential amino acids-unlike most plant foods (ii) a high carbohydrate content (iii) an extensive fatty acid profile, including Omega 3 and Omega 6 and (iv) an abundance of vitamins, minerals and trace elements in a naturally-occurring synergistic design [6-7]. Macroalgae can be classified as red algae (*Rhodophyta*), brownalgae (*Phaeophyta*) or green algae (*Chlorophyta*) depending on their nutrient and chemical composition. Red and brown algae are mainly used as human food sources. Seaweed species are rich in beneficial nutrients, in countries such as China, Japan and Korea, they have been commonly utilized in human alimentation. Seaweeds have been consumed in Asia since ancient times. Further, marine algae have been utilized in Japan as raw materials in the manufacture of many seaweed food products, such as jam, cheese, wine, tea, soup and noodles and in the western countries, mainly as a source of polysaccharides for food and pharmaceutical uses. In Europe, there is an increasing interest in marine seaweeds as a food, nevertheless, at present there are no European union specific regulations concerning their utilization for human consumption. Various chemical constituents of the red alga *Grateloupiaturuturu* has been determined and has been submitted [8-10].

New technologies involving the removal of metals ions from waste waters have directed attention to biosorption based on metal binding capacities of various biological materials. Biosorption is an innovative technology that employs inactive and dead biomass for the recovery of heavy metals from aqueous solution. Research in the field of biosorption has mostly concerned itself with brown algae and to a less extent with red algae. Literature survey found that the marine red algae belonging to this family are rich sources of phenolic compounds, especially bromophenols. Phenolic compounds play an important role in one antioxidativeproperties of many plant derived antioxidants. Phenolic substances were also reported to possess a wide range of biological effects, including antioxidant, antimicrobial, anti-inflammatory and vasodilatory actions [11-15].

This algae was used in our study. The 18S sequence of this algae submitted in genebank (http://www.ncbi.nlm.nih.gov/) was collected. 18S rRNA is a component of the small eukaryotic ribosomal subunit (40S). 18S rRNA is the structural RNA for the small component of eukaryotic cytoplasmic ribosomes, and thus one of the basic components of all eukaryotic cells that stands for the genetic information of that particular species or family.

2. Materials & Methods

The gene sequences used in the study was collected from the NCBI website. This comprises the database of all submitted sequences to the Gene Bank collections. The related sequences to the given string used for the study are given by the search engine from where the sequences can be picked to establish a selected set of sequences that can be used as references sequences for the construction of the genetic distance matrix and the tree. The sequences of the *Kappaphycus* sp. used for the study was also collected from the NCBI database [16-20].

2.1 Clustal X software

Clustal X is the software that is used to determine the genetic diversity between the sequences used for the study. The sequences were loaded to the Clustal X software. This tool follows the heuristic algorithm with progressive alignment method. Progressive alignment is a multiple sequence alignment strategy with a stepwise approach to assemble an alignment. It first performs all possible pair wise alignment using the dynamic programming approach and determines the relative distance between each pair of sequence to construct a distance matrix, which is subsequently used to build a guide tree. It then re-aligns the two most closely related sequence using the dynamic programming approach. Other sequences are progressively added to the alignment according to the degree of similarity suggested by the guide tree. The process proceeds until all sequence are used in building alignment.

2.2 Pair wise alignment of the sequences

An aligned sequence file is created after the pair wise sequence alignment. The method involves aligning sequences using the Neighbor Joining (N-J) method. First, the sequences were aligned and distances were calculated (percent divergence) between all pairs of sequences from a multiple alignment. Finally, the Neighbor Joining (N-J) method was applied to the distance matrix.

2.3 Bootstrap

Bootstrap is performed to designate the node point of the distance matrix with 100 iterations to demonstrate the diversity and identity matrix between the sequences used. The tree was created using the Bootstrap N-J Tree method using 1000 iterations.

2.4 Tree View 2.0 software

The Bootstrap file opened using the Tree View 2.0 software will be converted to a graphical tree connecting the nodes that have been assigned using 1000 iterations matrix to know the divergence of the sequences from each other.

2.5 Result of gene selection

The gene sequences of *Kappaphycus* sp. and the reference strains were collected from the NCBI Gene Bank using the portal (http://www.ncbi.nlm.nih.gov/) are as follows. The sequences that were used for the construction of the genetic diversity and the organism names are given in the Table 1.

Typical sequence is given below.
>**JN984925**

acgaaactgcgaatggctcggtaaaacagctatagtttcttcggtggtaaatactactt
ggatacccgtagtaattctagagctaatacatgccataaaagcggcgcttgcgccgt
ggtataaattggaggtacaagccaatgtttggtgattcacaatttatttctgattgcacat
gagtgcgacacaccgttcaaatttctgacctatcaactttggatggtaaggtattgtctt
accatggttgtgacgggtaacggaccgtgggtgcgggactccggagagggagcct
gagaaacggctaccacatccaaggaaggcagcaggcgcgcaaattacccaatccg
gacaccgggaggtagtgacaagaaatagcaatagagggcccaacgggttttctaat
tggaatgagaacaaggtaaacagcttatcgaggagccagcagagggcaagtctgg
tgccagcagccgcggtaattccagctctgtaagcgtataccaaagttgttgcagttaa
aacgctcgtagtcggactttggcgggcgggctggggcgtctttacggacgggtccg
gggccggccgcctttgtggaggggggcccggccgtcgcttcactgcgacggcgtgg
tcgctgccaccgtttactgtgagaaaattagagtgttcaaagcgggcgtttgccatga
atacattagcatggaataatagaataggacctgtctcctattttgttggtttgtgaggaac
gggtaatgattaagagggacggttgggggcatttgtattccggcgtcagaggtgaaa
ttcttggattgtcggaagacaaacggctgcgaaagcgtctgccaaggacgttttcatt
gatcaagaacgaaagtaaggggatcgaagacgatcagataccgtcgtagtctttact
ataaacgatgaggactggggatcgggcaggacttgaattttgggctgcccggcacc
cttcgggaaaccaaagtgtttgctttctgggggggagtatggtcgcaaggctgaaactt
aaaggaattgacggaagggcatcaccgggtgtggagcctgcggcttaatttgactca
acacggggaaacttaccaggtccggacatagtgaggattgacagattgagagctctt
tcttgattctatggttggtggtgcatggccgttcttagttggtggagtgatctgtctggtta
attccgttaacgagcgagacctgggcgtgctagctaggcgcgactaccgctttttggt
agtgaggctgccttcctagacggactgcgggcgtctagtccgcggaagctccaggc
aataacaggtctgagatgcccttagatgttctgggccgcacgcgtgctacactgagc
gggtcaacgggtgaggatgtgcgagagcgcttcccaatctctaaatccgctcgtgct
ggggatcgaggcttgcaattttccctcttgaacgaggaataccttgtaagcatgggtc
atcatcccgtgctgaatacgtccctgccctttgtacacaccgcccgtcgctcctaccga
ttgagtggtccggt

gaggccttgggagggaaagatggactgtgtttcacggaccgtctggctcaaacttgg
tcaaaccttatcacttagaggaaggagaagtcgtaacaaggtttccgtaggtgaacct
gcggaaggatc

Table 1: The sequences used for the study

Gene Bank Accession Numbers	Name of the algae
KC218498	*Chlorella* sp.
KC166137	*Chlamydomonas reinhardtii*
KC218488	*Scenedesmus maegelii*
KC218482	*Chlorococcum* sp.
KC218500	*Dunaliella* sp.
JN934686	*Chlamydomonas* sp.
AJ000708	*Cyanothece* sp.
AJ000709	*Euhalothece* sp.
AJ000710	*Euhalothece* sp.
AJ000711	*Dactylococcopsis* sp.
AJ000712	*Euhalothece* sp.
AJ000713	*Euhalothece* sp.
AJ000724	*Halothece* sp.
JN984925	*Kappaphycus* sp.
GQ869847	*Kappaphycus* sp.
GQ869848	*Kappaphycus* sp.
GQ869846	*Kappaphycus* sp.
GQ869844	*Kappaphycus* sp.

Paper ID: SUB141153

952

3. Result of the Tree View 2.0 software

Using the sequences fed in the software. The dendogram, phylogram, cladogram and the unrooted radial tree.A dendrogram (Figure 1)in Greek is a tree drawings representing a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering. Dendrograms are often used in computational biology to illustrate the clustering ofgenes or samples. The dendogram of the sequences of this study has also depicted the same picture explaining the hierarchical Using the Clustal X2 and the Tree Vies 2.0 soft ware the phylogenetic trees were constructed using the Bootstrap Tree Method from Clustal X software. Four types of trees can be created for a single input set namely the Dendogram, Phylogram, Cladogram and the unrooted radial tree. These trees were developed using the Tree View 2.0 software, which reads the aligned sequence distance matrix with 1000 iterations to show the distance diversity and similarity between the *Kappaphycus*sp. and the other algae used in the study. Initially the sequences are formatted in a note pad with sequence accession numbers, which is then opened using the Clustal X2 software and were completely aligned. This step creates a file with all the aligned sequences. Then it is allowed for bootstraping where the node-joint iterations are decided and assigned. This also generates a file of Bootstraped sequence file. Then these files are opened using the Tree View 2.0 software where these alignments are graphically represented as phylogenetic tree.In biologyphylogenetics is the study of evolutionary relationships among groups of organisms (e.g. species, populations),which are discovered through molecular sequencing data and morphological data matrices. The term is derived from the Greek terms *phylé* and *phylon*, denoting "tribe", "clan", "race" and the adjectival form, *genetikós* , of the word *genesis* "origin", "source", "birth". The result of phylogenetic studies is a hypothesis about the evolutionary history of taxonomic groups: their phylogeny.

Evolution is regarded as a branching process, whereby populations are altered over time and may split into separate branches, hybridize together, or terminate by extinction. This may be visualized in a phylogenetic tree, a hypothesis of the order in which evolutionary events are assumed to have occurred.Phylogenetic analyses have become essential in researching the evolutionary tree of life. The overall goal of National Science Foundation's Assembling the Tree of Life activity (AToL) is to resolve evolutionary relationships for large groups of organisms throughout the history of life, with the research often involving large teams working across institutions and disciplines. Investigators are typically supported for projects in data acquisition, analysis, algorithm development and dissemination in computational phylogenetics and phyloinformatics. Forexample, RedToL aims at reconstructing the Red Algal Tree of Life.Taxonomy, the classification, identification, and naming of organisms, is usually richly informed by phylogenetics, but remains methodologically and logically distinct. The degree to which taxonomy depends on phylogenies differs between schools of taxonomy: numerical taxonomy ignored phylogeny altogether, trying to represent the similarity between organisms instead; phylogenetic systematics tries to reproduce phylogeny in its classification without loss of information; evolutionary taxonomy tries to find a compromise between them in order to represent stages of evolution.

Four types of output trees were obtained similarity. A phylogram is a tree that indicates the relationships between the taxa and also conveys a sense of time or rate of evolution. A cladogram is a diagram used in cladistics which shows relations among organisms. A cladogram is not, however, an evolutionary tree because it does not show how ancestors are related to descendants or how much they have changed; many evolutionary trees can be inferred from a single cladogram. A cladogram uses lines that branch off in different directions ending at groups of organisms. There are many shapes of cladograms but they all have lines that branch off from other lines. The lines can be traced back to where they branch off. These branching off points represent a hypothetical ancestor (not an actual entity) which would have the combined traits of the lines above it. This hypothetical ancestor might then provide clues about what to look for in an actual evolutionary ancestor. Although traditionally such cladograms were generated largely on the basis of morphological characters, DNA and RNA sequencing data and computational phylogenetics are now very commonly used in the generation of cladograms. Unrooted trees illustrate the relatedness of the leaf nodes without making assumptions about ancestry at all. While unrooted trees can always be generated from rooted ones by simply omitting the root, a root cannot be inferred from an unrooted tree without some means of identifying ancestry; this is normally done by including an outgroup in the input data or introducing additional assumptions about the relative rates of evolution on each branch, such as an application of the molecular clock hypothesis.The scientific methods of phylogenetics are often grouped under the term cladistics. The most common ones are parsimony, maximum likelihood, and MCMC-based Bayesian inference. All methods depend upon an implicit or explicit mathematical model describing the evolution of characters observed in the species included; all can be, and are, used for molecular data, wherein the characters are aligned nucleotide or amino acid sequences, and all but maximum likelihood (seebelow) can be, and are, used for morphological data.Phenetics, popular in the mid-20th century but now largely obsolete, uses distance matrix-based methods to construct trees based on overall similarity, which is often assumed to approximate phylogenetic relationships.

3.1Result of a dendogram

Schematic representation of the genetic similarity of the sequences using a dendogram is shown below (Fig. 1).
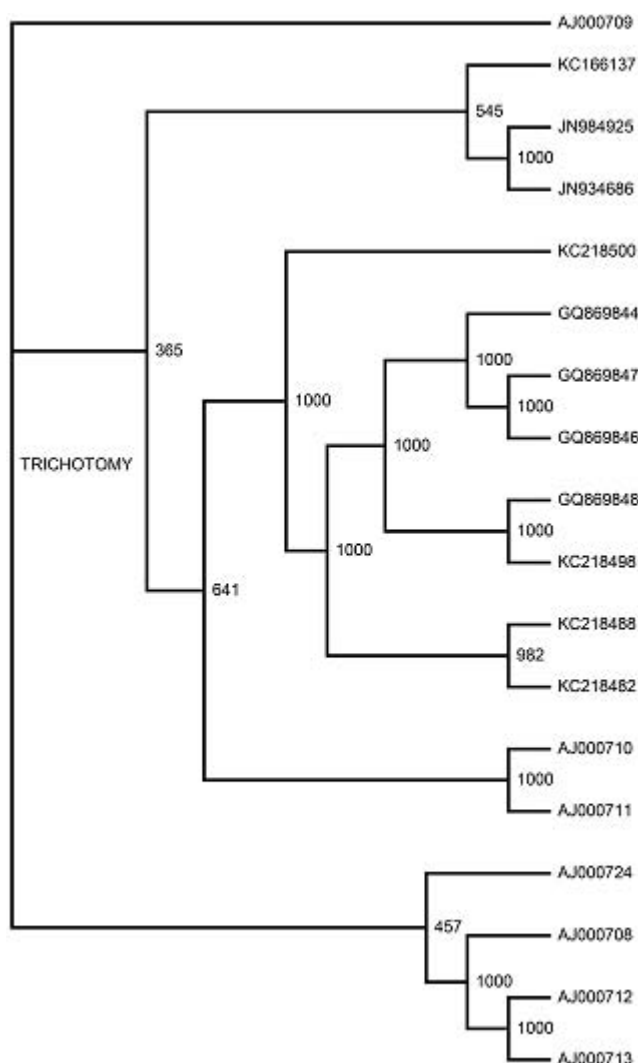
Paper ID: SUB141153

953

Figure 1: Schematic representation of the genetic similarity of the sequences using a dendogram

4. Discussion

The results of all the four trees infer the ratio of similarity of the *Kappaphycus* sp. with that of the other selected algae as reference sequences (Table 1). The *Kappaphycus* sp. GQ869847, GQ869848, GQ869846 and GQ869844 has expressed a maximum similarity of 100% to the neighborhood sequence of Chlorella sp. (KC218498) a blue-green algae. This again has a close similarity of 100% with that of *Scenedesmusmaegelii*(KC218488) and *Chlorococcum* sp. (KC218482) that expressed a similarity of 98.2%. *Scenedesmusmaegelii* and *Chlorococcum* sp. further demonstrated a 100% similarity with the *Dunaliella* sp. (KC218500) that belonged to the Chlophyta phylum and is a green algae. *Euhalothece* sp. (AJ000710) and *Dactylococcopsis*Chlophyta phylum and is a green algae. *Euhalothece* sp. (AJ000710) and *Dactylococcopsis* sp. (AJ000711) that had an internal similarity of 100% exhibited a similarity of 64.1% with the *Dunaliella* sp. and it also exhibited a similarity of 36.5% with that of the *Kappaphycus* sp. (JN984925) and *Chlamydomonas* sp. (JN934686) and *Chlamydomonasreinhardtii* (KC166137).

These algae are tough, fleshy, firm up to 2 m tall. Thalli coarse, with axes and branches 1-2 cm diameter, heavy with major axes relatively straight, lacking secondary branches near apices. They are generally frequently and irregularly branched, most branches are primary and secondary branches intercalated between primary or mostly lacking. They are generally shiny green or yellow orange. Morphologically they have variable thalli forms, form gnarled forms with few small branches in shallow areas to large, intricately tangled, fleshy mats in deeper waters. They are generally found in reef flat and reef edges at 1 -17 m depth. Loosely attached to broken corals or unattached fragments floating in shallow and deep waters. It also forms large moving mats with unattached thalli.

Arround five 18S sequences of the Kappaphycus sp. was collected for the study and the reference sequence list consisted of about thirteen different algal 18S sequences.They were analyzed for their genetic similarity by constructing a model tree based on the genetic distance and identity matrix using the Clustal X and Tree View 2.0 software that equipped the pair wise alignment algorithem and node selection based on the 1000 iteration bootstrap method. Then the tree was graphically created using the Tree View 2.0 software engaging the Neighbor joining method. This clearly demonstrated the genetic similarity matrix with the help of the percentage similarity based on the 1000 iteration nodal point of all the reference 18S sequences with that of the Kappaphycus sp. 18S sequences collected from the NCBI web portal.

5. Summary

Arround five 18S sequences of the*Kappaphycus* sp. was collected for the study and the reference sequence list consisted of about thirteen different algal 18S sequences. They were analyzed for their genetic similarity by constructing a model tree based on the genetic distance and identity matrix using the Clustal X and Tree View 2.0 software that equipped the pair wise alignment algorithem and node selection based on the 1000 iteration bootstrap method. Then the tree was graphically created using the Tree View 2.0 software engaging the Neighbor joining method. This clearly demonstrated the genetic similarity matrix with the help of the percentage similarity based on the 1000 iteration nodal point of all the reference 18S sequences with that of the *Kappaphycus* sp. 18S sequences collected from the NCBI web portal.

References

[1] Ahmad A., Mukherjee P., Mandal D., Senapati S., Islam Khan M., Rajivkumar Sastry M. J. (2002), 'Enzyme Mediated Extracellular Synthesis of CdS Nanoparticles by the Fungus', *Fusarium oxysporum*, Am. Chem. Soc, Vol. 124, pp.12108–12109.
[2] Ahmad A., S., Senapati M. I., Khan R., Kumar M. and Sastry. (2003), 'Extracellular Biosynthesis of Monodisperse Gold Nanoparticles by a Novel Extremophilic Actinomycete', *Thermomonospora* sp., Langmuir, Vol. 19, pp. 3550 -3553.
[3] Ahmad A., Senapati S., Islam K., Rajiv K., Ramani R., Srinivas V. and Sastry M. (2003), 'Intracellular

synthesis of gold nanoparticles by a novel alkalotoleran actinomycete *Rhodococcus* species', Nanotechnology, Vol. 14, pp.824-828.

[4] Ahmad A., Mukherjee P., Mandal D., Senapati S., Islam Khan M., Rajivkumar Sastry M. J. (2002), 'Enzyme Mediated Extracellular Synthesis of CdS Nanoparticles by the Fungus', *Fusarium oxysporum*, Am. Chem. Soc, Vol. 124, pp.12108–12109.

[5] Alejandro H. Buschmann, Daniel Varela, Marcos Cifuentes, Maria del Carmen Hernandez-Gonzalez, Luis Henriquez, Renato Westermeier, Juan A. Correa, (2004), 'Experimental indoor cultivation of the carrageenophytic red alga Gigartina skottsbergii', Aquaculture, Vol. 241, pp. 357-370.

[6] Ana P., Ramos Rogeria R., Goncalves Osvaldo A., Serra Maria Elisabete D. and Zaniquelli Kenneth Wong, (2007), 'Europium ion as a probe for binding sited to carrageenans.', J. Luminescence Vol. 127, pp. 461-468.

[7] Anne-Valerie Galland-Irmouli, Joel Fleurence, Radia Lamghari, Michel Lucon, Catherine Rouxel, Oliver Barbaroux, Jean-Pierre Bronowicki, Christian Villaume and Jean-Louis Gueant, (1999), 'Nutritional value of proteins from edible seaweed Palmaria palmate (Dulse)', J. Nutr. Biochem., Vol. 10, pp. 353-359.

[8] Anthony Chiovitti, Antony Bacic, David J. Craik, Gerald T. Kraft and Ming-Long Liao., (2004), 'A nearly idealized 6'-o-methylated t-carrageenan from the Australian red alga Clavicloniumovatum (Acrotylaceae, Gigartinales)', Carbohydate research Vol. 339, pp. 1459-1466.

[9] Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

[10] Benson, D. A., M. S. Boguski, D. J. Lipman, and J. Ostell. 1997. GenBank. Nucleic Acids Res. 25:1–6.

[11] Bin, Y., Jiti, Z., Jing, W., Cuihong, D., Hongman, H., Zhiyong, S. &Yongming, B. (2004). Expression and characteristics of the gene encoding azoreductase from Rhodobactersphaeroides AS1.1737. FEMS Microbiol. Letters, 236, 129–136.

[12] Bin-Gui Wang, Wei-wei Zhang, Xiao-Juan Duan, Xiao-Ming Li, (2009), 'In vitro Antioxidative activities of extract and semi-purified fractions of the marine red alga', Rhodomelaconfervoides (Rhodomelaceae), Food Chemistry Vol. 113, pp. 1101-1105.

[13] Bu¨del, B. 1992. Taxonomy of lichenizedprocaryotic blue-green algae, p. 301–324. *In* W. Reisser (ed.), Algae and symbioses: plants, animals, fungi, viruses, interactions explored. Biopress Ltd., Bristol, England.

[14] Cavas L. and Yurdakoc K. (2005), 'An investigation on the antioxidant status of the invasive alga *Cauleraracemosa var. cylindracea (Sonder) Velaque, Huisman, et Boudoresque (Caulerpales, Chlorophyta)*', J. Exp. Mar. Biol. Ecol., Vol. 325, pp. 189-200.

[15] Chatterjee I. B. (1973), 'Vitamin C synthesis in animals', evolutionary trend. Sci. Cult. Vol. 39, pp. 210-212.

[16] Suzuki, Y., Yoda, T., Ruhul, A. & Sugiura, W. (2001). Molecular cloning and characterization of the gene coding for azoreductase from Bacillus sp. OY1-2 isolated from soil. J. Biol. Chem. 276, 9059–9065.

[17] Suzuki, M. T., and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62:625–630.

[18] Tasende M. G. (2000), 'Fatty acid and sterol composition of gametophytes and sporophytes of Chondruscrispus (Gigartinaceae, Rhodophyta)', Scientia Marina Vol. 64(4), pp. 421-426 Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

[19] position-specific gap penalties and weight matrix choice. Nucl.Acids Res. 22, 4673–4680.

[20] Viana M. T. Lopez L. M. and Salas A. (1993), 'Diet development for juvenile abalone Haliotisfulgens', Evaluation of two artificial diets and macro algae. Aqaculture Vol. 117, pp. 149-156.

[21] http://www.ncbi.nlm.nih.gov