www.bioinformation.net

Issue 3

Hypothesis

Sequence analysis of 16S rRNA gene in Rhinosporidium seeberi shows similarity to plant chloroplast DNA

Karvita B Ahluwalia^{1*}, Dhara B Dhaulakhandi¹, Lalit C Garg²

¹Cell Biology and Electron Microscope Section, Department of Biophysics, All India Institute of Medical Sciences, New Delhi – 110029, India; ²Gene Regulation Laboratory, National Institute of Immunology, New Delhi – 110067, India; Karvita B Ahluwalia, E-mail: karvitaa@yahoo.com, Tel.: +91 11 26535914; *Corresponding author.

Received June 28, 2010; Accepted August 26, 2010; Published September 20, 2010

Abstract:

Identity of causative agent of rhinosporidiosis (*Rhinosporidium seeberi*) has been controversial since the disease was described in 1900. Extensive sequence alignments and phylogenetic analyses of 16S rRNA gene detected recently by us in *R. seeberi*, revealed 99% similarity with 16S rDNA in chloroplasts of flowering plants. Study demonstrates *R. seeberi* is a pigmented prokaryote displaying some characteristics of cyanobacteria, and contains 16S rDNA present in chloroplasts of all groups of land plants. This study and our recent publication of 2006 are the first molecular studies using purified organismal DNA extracted from *R. seeberi* free of infected tissue.

Key words: Rhinosporidium seeberi; 16S rRNA gene; sequence analysis; chloroplast DNA; molecular phylogeny.

Abbreviation: RB - Round body

Rackground:

Rhinosporidiosis produces polypoid masses in nose and other sites in individuals exposed to fresh water ponds. An outbreak of rhinosporidiosis occurred among people on holiday near Silver Lake in Serbia [1]. The disease has worldwide distribution and is endemic in southern India, Sri Lanka and Argentina. Rhinosporidium seeberi is spherical upto 300-400 μm in diameter, visible in histological sections as a round body (RB) filled with endospores that are discharged and enlarge into new RBs. All in vivo stages in R. seeberi 's life cycle (enlargement, maturation of round bodies and release of endospores) have been demonstrated in vitro using monolayers of a human rectal tumor cell line [2]. Immunological defense mechanism is not known, only phagocytosis is reported [3]. First described by Seeber in 1900 [4], taxonomy of R. seeberi debated throughout the century. No treatment is effective; multiple recurrences are common after surgery. The disease is not transmitted to family members of affected individuals. RBs have not been seen in natural habitats, nor in pond water samples from 3 endemic regions in India [5].

Recently we detected a 16S rRNA gene (GenBank accession no. AJ440719) in purified RBs carefully isolated from infected tissue, and determined the sequence [6]. This study is a detailed analysis of sequence of 16S rRNA gene.

Methodologies:

Individual uncontaminated RBs (*R. seeberi*) were isolated from infected tissue (**Supplementary Figure 1**) and pure organismal DNA extracted from a single RB (details in [6]). Briefly, between 15-25 ng DNA extracted from a single RB was checked for purity spectrophotometrically and by running on 0.85% agarose. Experiments were repeated 3 times using 3 different clinical samples. PCR primers identical to those used in a previous study [7] were procured from Yale University, Connecticut. Gene Amp 2400 PCR system (Perkin Elmer) was used. Clustal algorithm was

used for multiple sequence alignment in LASERGENE software (DNA Star). Phylogenetic trees were constructed using Saitou & Nei's [8] neighbor-joining method.

Results and Discussion:

R. seeberi 's 1458 bp 16S rRNA gene sequence (AJ440719) aligned in BLASTN with 16S rDNA sequences in GenBank (no domain specified) showed high similarity (99%) with DNA in chloroplasts (cpDNA) of flowering plants Solanum tuberosum, Nicotiana tabacum, N. plumbaginifolia, Arabidopsis thaliana, Carica papaya and Vitis vinifera. cpDNA in 70 genera of higher plants, liverwort Marchantia, green alga Chara, Spirogyra and related genera displayed 95-97% identity with AJ440719. Oryza sativa showed 96% identity with cpDNA as also with genomic DNA in chromosome 10. cpDNA in nonphotosynthetic alga Prototheca and amoeba Paulinella chromatophora showed 82% and 84% identity, respectively. BLAST search specifying domain bacteria indicated high similarity with many cyanobacteria. Phylogenetic tree (Figure 1) shows R. seeberi's sequence positioned unambiguously among cpDNA of flowering plants. AJ440719 clusters with cpDNA with high bootstrap support. Chloroplast genes in human apicomplexan pathogens are not phylogenetically close to R. seeberi 's sequence (Supplementary Figure

Pairwise alignment of 1458 bp sequence (AJ440719) in BLAST2 with 16S rDNA in chloroplasts of several plants indicates most differences with *R. seeberi* 's sequence in nucleotide positions 1 to 82. Thus, cpDNA in Solanum nigrum displayed 99% identity with nucleotides 79-1458 with 4 gaps, and identity of 93% for nucleotides 4-63 with no gap. Two gene fragments, 1-82 and 83-1458 were aligned separately in BLAST2 in 2 steps. Nucleotides 1- 82 aligned mostly with bacteria including 14 proteobacteria and 5 pathogenic bacteria (identity 90-95%), nucleotides

Volume 5

www.bioinformation.net

Issue 3

Hypothesis

83-1458 showed 95-99% alignment with cpDNA of numerous plants. G+C content of AJ440719 is 56.8.

Chloroplast sequence in R. seeberi displayed cyanobacterial sequence characteristics. Two palindromic repeats and sequence signature (AATTTTCCG) of cyanobacteria were found in RB's 16S rRNA gene [6]. In this study, multiple sequence alignments showed that sequence signature is shared with 16S rDNA in 11 cyanobacteria and cpDNA in plant Carica papaya, and flanking nucleotides are also similar (Supplementary Figure 3). Among 826 conserved nucleotide positions in AJ440719, 773 (92.4%) are shared with 11 cyanobacteria and cpDNA in Carica, while conserved nucleotides in segment 116-225 are dissimilar. The single conserved nucleotide T is in position 799, and A is in position 1233 as reported in cyanobacteria [9]. Pairwise alignment of 1458 bp sequence (AJ440719) in BLAST2 with several cyanobacteria showed identity range of 79-86%, with most differences in first segment of 82 nucleotides. Low identity of 79-86% may be result of comparison with chloroplast-like 16S rRNA sequence in RB. Another different copy of 16S rRNA gene, that is not chloroplast-like, could possibly be present in R. seeberi 's genome. Different 16S rRNA sequences could be present in a single cell genome [10, 11].

Taxonomic Placement of R. seeberi

Ashworth [12] considered organism in rhinosporidiosis to be a fungal sporangium filled with spores, and named it *Rhinosporidium seeberi*. Mycology textbooks describe *R. seeberi* as a fungus of uncertain taxonomy. We used fungal-specific primers for 18S and 28S rRNA genes, no amplification was seen with DNA from RBs as template, negating fungus in *R. seeberi* [6]. Presence of chlorophyll led Vanbreuseghem [13] to suggest RB as an alga that was refuted by some authors [3] as chloroplasts have not been seen in RB. Photosynthetic pigments, also found by us [14], rule out fungus.

The in vivo life cycle of *R. seeberi* is initiated with the entry of endospores into host epithelium [15, 16]. Study of pond water samples revealed striking similarity between mature endospores in RB and cyanobacterium Microcystis [5]. PCR using genomic DNA of *R. seeberi* showed a sequence that is strongly amplified by 16S rRNA primers specific for cyanobacterium Microcystis [6]. For a pathogen as little understood as *R. seeberi*, accurate assessment of ultrastructure could complement molecular data. Our findings, and of others indicate RBs phenotypic relationship with cyanobacteria. These include dispersed chromatin [15-17], polyhedral bodies, wall peptidoglycan [5], photosynthetic pigments [13, 14], and red auto fluorescence excited with

green (510-540 nm) wavelengths (this study) that is typical of photosynthetic pigments [18].

Another study in two laboratories, based on DNA from infected tissue, proposed R. seeberi 's relationship with the protist Dermocystidium [7, 19]. Since rhinosporidiosis develops in individuals exposed to fresh water ponds, the presence of aquatic protists like Dermocystidium in infected tissue is not ruled out. Cysts and spores of Dermocystidium sloughed from gills of fish "could be a significant source of infection in fresh water systems" [20]. Even a low number of Dermocystidium spores in infected tissue could be detected by the sensitive PCR. Authors in both laboratories selected 18S SSU rRNA primers F1-fw F2-rev and NS1 NS8 (degenerated) for fungi that also anneal to some protists [19, 7]. Therefore, DNA of Dermocystidium, if present in infected tissue, would be amplifiable with these primers. Although primer F2-rev differs from the sequence in 2 nucleotide positions and primer F3-rev differs in 3 nucleotide positions, these primers gave amplification, "data not shown" [19]. Primer Rhino-fw differs from the sequence at 4 nucleotide positions and Rhino-rev differs in 2 nucleotide positions, they are reported to give amplification [19]. In one study [19], since control samples of nasal polyps from patients without rhinosporidiosis came from humans not exposed to fresh water ponds, the samples are not expected to harbor aquatic protists, and this explains why the authors [19] got negative results with PCR. Controls are lacking in the other study [7]. We procured the same 18S SSU rRNA primers mentioned in their study, for using with organismal DNA extracted from purified RBs, PCR did not show amplification with those primers. As 16S rRNA sequence can be compared with 18S rRNA [10], a comparison made between R. seeberi 's sequence AJ440719 and 18S rDNA in [19] did not show similarity.

Conclusion:

R. seeberi is a pigmented prokaryote with cyanobacterial features, and contains 16S rDNA identical to that in chloroplasts of land plants. This is the first report of a chloroplast sequence in R. seeberi, and since humans do not harbor chloroplast, can be used as target for development of chemotherapy. Detection of a sequence similar to AJ440719 in a cyanobacterial genus, or species/strain of Microcystis, might be useful for taxonomic identification of RB. DNA from RB generated a strong band with primers for Cyanobacterium microcystis [6]. Taking a cue from this observation, the presence of cyanobacterium-specific repeat sequences and sequence signature found in the sequence of R. seeberi 's chloroplast-like 16S rRNA gene [6] should now be searched elsewhere in the genome. Sequencing the entire genome of R. seeberi would be rewarding.

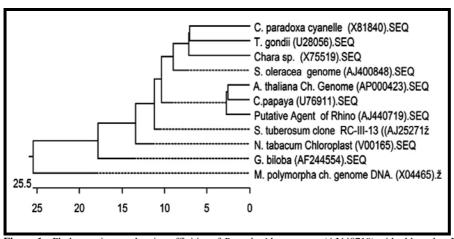


Figure 1: Phylogenetic tree showing affinities of *R. seeberi* 's sequence (AJ440719) with chloroplast DNA in genera from different taxonomic groups. Flowering plants: Arabidopsis thaliana, Carica papaya, Solanum tuberosum, S. oleracea, Nicotiana tabacum. Gymnosperm: Ginkgo biloba. Bryophyte: Marchantia polymorpha. Streptophyte: Chara. Cyanelle of Cyanophora paradoxa. Human pathogen: Toxoplasma gondii.

Volume 5

www.bioinformation.net

Issue 3

Hypothesis

Acknowledgements:

K.B.A. thanks Prem Chopra, Professor of Pathology, AIIMS for useful discussions. Department of Science and Technology, Government of India and All India Institute of Medical Sciences provided research grants.

References:

- Z Vukovic et al. J. Trop. Med. Hyg. 98:333 (1995) [PMID: 756 3 26 [1]
- MG Levy et al. Science. 234: 474 (1986) [PMID: 3764422]
- [3] M Thianprasit et al. Curr. Topics Med. Mycol. 5:64-85 (1989) [PMID: 2688923]
- GR Seeber (1900) Thesis, Universite Nacionale de Buenos Aires
- KB Ahluwalia et al. Amer. J. Rhinol. 11: 479 (1997) [PMID: 94380621
- DB Dhaulakhandi et al. Infec. Genetics Evol. 6: 331 (2006) [PMID: 16418014]
- [7] RA Herr et al. J. Clin. Microbiol. 37: 2750 (1999) [PMID: 104494461

- [8] N Saitou et al. Mol. Biol. Evol. 4:406 (1987) [PMID: 3447015]
- CR Woese. Microbiol. Rev. 51: 221 (1987) [PMID: 2439888]
- JE III Clarridge. Clin. Microbiol. Rev. 17: 840 (2004) [PMID: [10] 154893511
- J Viezens et al. Microbiology 154: 2881 (2008) [PMID: 18759822]
- [12] JH Ashworth, Trans. Roy. Soc. Edinburgh (1923) 2: 301.
- R Vanbreuseghem. Int. J. Dermatol. 12: 20 (1973) [PMID: [13] 45716041
- KB Ahluwalia. J. Laryngol. Otol. 113: 513 (1999) [PMID: 10605581]
- M Kannan-Kutty *et al. Pathology* **6**: 63 (1974) [PMID: 4423221] DF Savino *et al. Ophthalmology* **90**: 1482(1983) [PMID: 6677849]
- FA Kennedy et al. J. Med. Vet. Mycol. 33: 157 (1995) [PMID:
- [18] JW Hofstraat et al. J. Plankton Res. 16: 1197. (1994)
- DN Fredricks et al., Emerg. Infec. Dis. 6: 273 (2000) [PMID: 10827117]
- [20] RE Olson et al. Dis. Aquatic Organisms 12: 41 (1991)

Edited by P. Kangueane

Citation: Ahluwalia et al. Bioinformation 5(3): 89-96 (2010)

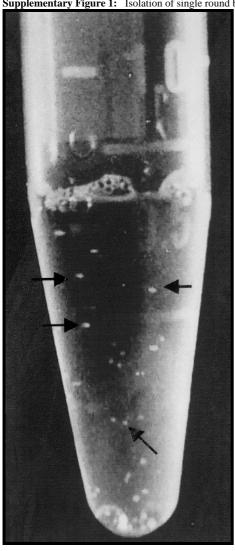
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Volume 5

www.bioinformation.net Issue 3 **Hypothesis**

Supplementary material:

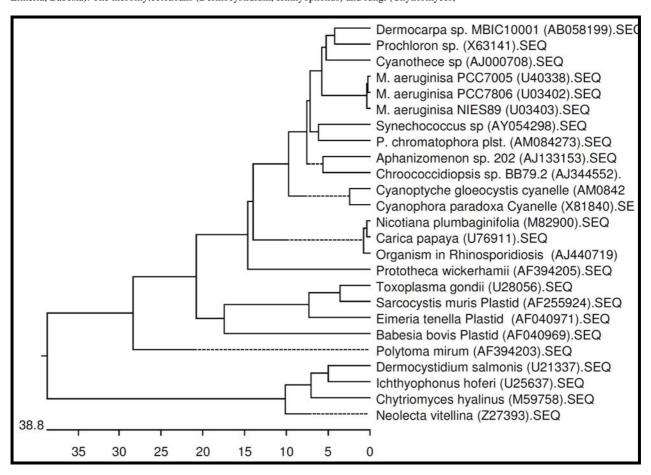
Supplementary Figure 1: Isolation of single round bodies. Purified single round bodies after isolation from infected tissue in Eppendorf tube.



Volume 5

www.bioinformation.net Issue 3 **Hypothesis**

Supplementary Figure 2: Extended phylogenetic tree showing affinities of *R. seeberi* 's chloroplast sequence (AJ440719) with relevant taxa. AJ440719 clusters with chloroplast DNA in flowering plants. Tree also shows 16S rDNA in cyanobacteria, and cpDNA in amoeba Paulinella chromatophora, cyanelle of Cyanophora, nonphotosynthetic Prototheca and Polytoma, as well as cpDNA in apicomplexan pathogens (Toxoplasma gondii, Sarcocystis, Eimeria, Babesia). The mesomyteetozoans (Dermocystidium, Ichthyophonus) and fungi (Chytriomyces,



Volume 5

www.bioinformation.net

Issue 3

Hypothesis

Supplementary Figure 3: Multiple sequence alignments of *R. seeberi* 's sequence (AJ440719) with 16S rDNA in 11 cyanobacteria and cpDNA in flowering plant Carica papaya. Base numbering according to sequences available in NCBI nucleotide sequence database. Box shows sequence signature AATTTTCCG; stars indicate conserved bases; nucleotides flanking AATTTTCCG are identical in all the aligned sequences. Accession nos. are of following cyanobacteria: X59559.1 – Anabaena; AJ133161.1 – Nostoc; X84811.1 – O. agardhii; AF013028.1 - Trichodesmium; AF091109.1 – Spirulina; X63`141.1 – Prochloron; AJ000708.1 – Cyanothece; U03403.1 - Microcystis; AF091110.1 – Plectonema; AB0150162.1 – Synechococcus; AB058219.1 – Phormidium; U76911.1 – chloroplast of Carica papaya; AJ440719 – *Rhinosporidium seeberi* .

X59559.1	119 TTCAGGTCGGGGACAACCACTGGAAACGGTGGCTAATACCGGATGTGCCGAAAGGTGAAAGAT-TTATT-GCCTGAAGATGAGCTCGCG	nama amma aama	ammaamamaamaa aa aaaa aa aa	
AJ133161.1	25 TTCAGGTTCGGGACAACCACTGGAAACGGTGGCTAATACCGAATGTGCCGAGAGGTGAAAGGC-TTGCT-GCCTGAAGATGAGCTCGCTCTGATTAGCTAGTAGGTGTGAAAGGCGC 142			
X84811.1	24 TCTAGGACGGGGACAACAGTTGGAAACGACTGCTAAACCCGGATGAGCCGAAAGGTAAAAGAT-TAATC-GCCTAGAGAGGGGCTTGCGTCTGATTAGCTAGTTGGTGAGGAGGCCC 141			
AF013028.1				
AF091109.1	92 TCTAGGTCGGGGACAACAGTTGGAAACGACTGCTAATCCCCGATGAGCCTAACGGTAAAAGAT-TTATT-GCCTAGAGAAGAGCTCGCG			
X63141.1	92 TCAAGGACGGGGACAACAGCAGGAAACTGGTGCTAATACCCGATAAGCTGAAAAGTTGAAAGAT-TTATT-GCCGAGAGAGGAGCTCGCG			
AJ000708.1	96 TTCAGGATGGGGACAACCGTTGGAAACGACGGCTAATCCCCAATAGGCCGAGAGGTGAAACAGGTTTCT-GCCTGAAGAGGAGCTCGCG	ICCGATTAGTTA	GTTGGTGGGGTAAAAGCGT 214	
U03403.1	94 TTCAGGACGGGACAACAGTTGGAAACGACTGCTAATACCCGATGTGCCGCGAGGTGAAACCTAAT-TG-GCCTGAAGAAGAGCTTGCG	TCTGATTAGCTA	GTTGGTGGGGTAAAGGCCT 211	
AF091110.1	94 CAAAGGACGGGACAACAGTGGGAAACCGCTGCTAAGACCCGATGTGCCGAGAGGTGAAACAT-TTATG-GCCTTTGGATGAACTCGCG	TCTGATTAGCTA	GTTGGTGTGGTAAGAGCAT 211	
AB015062.1	91 TATAGGATGGGGACAACAGGGAGAAATTCTTGTTAACACCCGATGTGCCGAGAGGTGAAATAT-TAATA-GCCTGTAGATGAGCTCGCG	ICTGATTAGCTA	GTTGGTGGGGTAAAGGCTT 208	
AB058219.1	08 CTTAGGTGGGGACAACAGTTGGAAACGACTGCTAATACCGCATATGTCGAGAGATGAAAGAT-TTATT-GCCTGAGGATGAACTCGCG	ICTGATTAGCTA	GTAGGTGAGGTAAGAGCTT 125	
U76911.1	103 CTTGGGAGGGAACAACAGCTGGAAACGGCTGCTAATACCCCGTAGGCTGAGGAGCAAAAGGAGGAATCCGCCCGAGGAGGGGCTCGCG	TCTGATTAGCTA	GTTGGTGAGGCAATAGCTT 222	
AJ440719.1	.1 116 CTTGGGAGGGGAACAACAGCTGGAAACGGCTGCTAATACCCCGTAGGCTGAGGAGCAAAAGGAGGAATCCGCCCGAGGAGGAGGGCTCGCTTCGATTAGCTAGTTGGTGAGGCAATAGCTT 235			
	** ** ****	**** **** *	* ** ** **	
X59559.1	${\tt ACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGA-CTCCTACGGGAGGCAGCAGTGGGGAGCAGTGGGGAGGAGCAGTAGGAGAGCAGTAGGAGAGAGA$	AATTTTCCGC	AATGGGCGAAAGCCTG 355	
	ACCTAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGA-CTCCTACGGGAGGCAGCAGTGGG		AATGGGCGAAAGCCTG 261	
X84811.1	ACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACCGCCCAGA-CTCCTACGGGAGGCAGCAGTGGGG		AATGGGCGAAAGC-TG 259	
	ACCAAGGCAACGATCAGTAGCTGGTCTGAGAGGATGAGCAGCCACACTGGGACTGAGACACGGCCCAGA - CTCCTACGGGAGGCAGCAGTGGGG		AATGGGCGAAAGCCTG 328	
	ACCAAGACAGCGATCGGTAGCTGGTCTGAGAGGATGACCAGTCACACTGGGACTGAGACACCGCCCAGA-CTCCTACGGGAGGCAGCAGTGGGG		AATGGGCGAAAGCCTG 328	
	ACCAAGGCGACGATCGGTACCTGGTCTGACAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGA -CTCCTACGGGAGGCACCACTGGGG		AATGGGCGAAAGCCTG 328	
	ACCANGGCACGATCGGTAGCTGGTCTGAGAGGATGAGCAGCACACTGGGACTGAGACACGGCCCAGA-TTCCTACGGGAGGCAGCAGTGGGG		AATGGGCGAAAGCCTG 328	
	ACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGAGCACCACACTGGGACTGAGACACGCCCAGA-TCCCTACGGGAGGCAGCAGTGGG		AATGGGCGAAAGCCTG 333	
	ACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGACGATCAGCACCACTGGGACTGAGACACGCCCAGA-CTCCTACGGGAGGCAGCAGTGGG	AATTTTCCGC	AATGGGCGAAGC-TG 329	
	ACCAAGGCGACCATCAGTAGCTGGTCTGAGAGGACGATCAGCCACACTGGGACTGAGACACGGCCCAGA-CTCCTACGGGAGGCAGCAGTGGGGACCAGCGGGCCCAGA-CTCCTACGGGAGGCAGCAGTGGGG		AATGGGCGCAAGCCTG 329	
			AATGGGGCAACCCTG 327	
	ACCTAGGCGACCATCAGTAGGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACCACACGGCCCAGA-CTCCTACGGGAGGAGCAGCAGTGGGG			
	ACCAAGGCGATGATCAGTAGCTGGTCCGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGATCTCCTACGGGAGGCAGCAGTAGGG		AATGGGCGAAAGC-TG 341	
AJ440/19.1	ACCAAGGCGATGATCAGTAGCTGGTCCGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGA-CTCCTACGGGAGGCAGCAGTGGGG	AATTTTCCGC	AATGGGCGAAAGCCTG 354	
	*** ** * **** ******* ****** ** *** ****	******	****** * ** * **	
AJ133161.1 X84811.1 AF013028.1 AF091109.1 X63141.1 AJ000708.1 U03403.1 AF091110.1 AB015062.1 AB015062.1 U76911.1	ACGGAGCANTACCGCCTGAGGAGGAAGGCTCTTGGGTTGTAAACCTCTTTTCTCAGGGAATAAAAAATGAAGGTACCTGAGGAATAAGCATC ACGGAGCAATACCGCCTGAGGGAGGAGGAGGCTCTTGGGTCGTAAACCTCTTTTCTCAGGGAATAAAAAATGAAGGTACCTGAGGAATAAGCATC ACGGAGCAAGACCGCGTGGGGGAGGAAGGTTCTTGGATTGTCAACCCCTTTTCTCAGGGAAGAACAATGACGGTACCTAGGGAAAAAACCAC ACGGAGCAAGACCGCCTGGGGGAGGAAGGCCTTGAGGTTGTAAACCCCTTTTCTTTTGGGAAGAATA-ACTGACGGTACCAAAGGATCACCAC ACGGAGCAATACCGCGTGCGGGAGGAAGGCCTTTGGGTTGTAAACCCCTTTTTCTCAGGGAACAAG-CTCTGACGGTACCTGAAGAAAAAACCCCC ACGGAGCAACCCCGCTGTGGGAAGAAGAGGCTCTTGGGTTGAAACCACTTTTTCTCAGGGAAGAG-TCTCTGACGGTACCTGAAGAAAAAACCCCC ACGGAGCAACGCCGCCTGTGGGAAGAAGAGGCTCTTGGGTTGAAACCACTTTTTCTCAGGGAAGAAG-TCCTGACGGTACCTGAAGAAAAAACCCCC ACGGAGCAACGCCGCCTGAGGGAGGAAGAGCTCTTTGGTTGTAAACCTCTTTTCTCAGGGAAGAAG-TTCTGACGGTTCTGAGGATCAGCCTCC ACGGAGCAACGCCGCGTGAGGGAGGAAGGCCTGTGGGTTGTAAACCTCTTTTCTCAGGGAAGAAG-TTCTGACGGTACCTCAGGGAATCAGCCTCC ACGGAGCAACGCCGCGTGAGGGAAGAGCCTGTGGGTTGTAAACCTCTTTTCTCAGGGAAGAAG-TTCTGACGGTACCTCAGGGAATCAGCATCC ACGGAGCAACGCCGCGTGAGGGAAGGAAGGCCTGTGGGTTGTAAACCTCTTTTCTCAGGGAAGAAGATCTTGACGGTACCTCAAGGAATCAGCATCC ACGGAGCAACGCCGCGTGAGGGAAGAAGGCCTGTGGGTTGTAAACCTCTTTTCTCAGGGAAGAAG-TTCTGACGGTACCTGAGGAATCAGCATCC ACGGAGCAACGCCGCGTGAGGGAAGAAGACCCCGGTGGGTTGTAAACCTCTTTTCTCAGGGAAGAAGA-TTCTGACGGTACCTGAGGAATCAGCATCC ACGGAGCAACGCCGCGTGAGGGAAGAAGGCCCTGTGGTTTTAAACCTCTTTTCTCAGGGAAGAAG-TTCTGACGGTACCTGAGGAATCAGCATCC ACGGAGCAATGCCGCCGCTGGAGGAAGAAGGCCCACGGGTCGTGAACTTCTTTTCCCGGGAAGAAG-TATCAACGGTACCTGAGGAATCAGCATCC ACGGAGCAATGCCGCCTGGAGGTAGAAGGCCCACGGGTCGTGAACTTCTTTTCCCGGGAAGAAG-CAATGACGGTACCTGAGGAATCAGCATCC ACGGAGCAATGCCGCCTGGAGGTAGAAGAGCCCACGGGCCGTGAACTTCTTTTCCCGGGAAGAAG-CAATGACGGTACCTGGGGAATAACCATCC ****************************	GCTAACTCCGT GGCTAACTCCGT	GCCAGCAGCCGCGG 381 GCCAGCAGCCGCGG 379 GCCAGCAGCCGCGG 447 GCCAGCAGCGCGGG 447 GCCAGCAGCGCGGG 447 GCCAGCAGCGCGGG 447 GCCAGCAGCGCGGG 449 GCCAGCAGCGCGGG 449 GCCAGCAGCGCGGG 449 GCCAGCAGCGCGGG 448 GCCAGCAGCGCGGG 447 GCCAGCAGCCGCGG 447 GCCAGCAGCCGCGG 447 GCCAGCAGCAGCGCGGG 447 GCCAGCAGCCGCGG 447	
X59559.1	TAATACGGAGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCG-CAGGTGGC ACTGTAAGTCTGCTGTTAAAGAGCAAGGCTCAA	CTTGTAAAGGC	ACTGGAAACTACAGA 594	
	TAATACGGAGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCG-CAGGTGGC TGTGTAAGTCTGCTGTTAAAGAGTCATGCTCAA			
X84811.1	TAATACGGGGGATGCAAGCGTTATCCGGAATGATTGGCCGTAAAGAGTCCG-TAGGTAGT CATCCAAGTCTGCTGTTAAAGAGCGAGGCTTAA			
	TAATACGGAGGAGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCG-CAGGTGGC CATGTAAGTCTGCTGTTAAAACCCAGGGCTTAA			
	${\tt TAATACGGAGGAGCAAGCGTTATCCGGAATTATTGGCGTAAAGCGTCG-CAGGTGGAATTATTAGTCTGCTGTCAAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGGAAGCTCAAGACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACGAAGCACACACGAAGCACGAAGCACACACACACACACACACACACACACACACACACACACA$			
	${\tt TAATACGGAGGATGCAAGCGTTATCCGGAATCATTGGGCGTAAAGGGTCCG-CAGGTGGC} \ \ {\tt GAATAAAGTCTGCTGTTAAAGACCGGGGCTAAACGGGGGCTAAACGGGGGGCTAAACGGGGGGGG$			
	${\tt TAATACGGAGGCAGCCTTATCCGGAATTATTGGCGTAAAGCGTCCG-CAGGGGGC \ GCATCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCTGCTGTTAAAGGTCGGAGCTCAAGTCAGTC$			
U03403.1	${\tt TAATACGGGGGAGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGTCCG-CAGGTGGT \ CAGCCAAGTCTGCTGTCAAATCAGGTTGCTTAACCGGGGGGAGGAGGCTAAATCAGGTTGCTTAACCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG$			
	TAAGACGGAGGATGCAAGCGTTATCCGGAATTATTGGCGTAAAGCGTCCG-TAGGTGGT TATACAAGTCTGTTGTTAAAGCCCATAGCTCAA			
	TAAGACGGAGGATGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGTCCG-TAGGCGGT TTAGAAAGTCAGTTGTTAAAGCCCACAGCTCAA			
	TAAGACGAGGATGCAAGCGTTATCCGGAATTATTGGGCGTAAACCGTCCG-TAGGCGGT TAATCAAGTCTGTTGTCAAAGACTGCAGCTCAA			
U76911.1	TAATACAGAGGATGCAAGCCTTATCCGGAATGATTGGGCCTAAAGCGTCTG-TAGGTGGC TTTTTAAGTCCGCCGTCAAATCCCAGGGCTCAA			
	TAATACAGAGGATGCAAGCCTTATCCGGAATGATTGGGCCTAAAGCGTCTGGTAGGTGGC TTTTTAAGTCCGCCGTCAAATCTCAGGGCTCAA			
1.0 110 / 17 . 1	1min.agaggaige.ma.giinicegganigaiiggaeginmaegicitgaiagiage			

Volume 5

www.bioinformation.net

X59559.1

Issue 3

Hypothesis

GCTAGAGTACGTTCGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATCA	
GGAAGAACACCGGTGGCGAAAGCGCTCTGCTAGGCCGTAACTGACACTGAGGGACGAAAG	714
AJ133161.1	
GCTAGAGTGCGTTCGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATCA	
GGAAGAACACCGGTGGCGAAAGCGCTCTGCTAGGCCGCAACTGACACTGAGGGACGAAAG	620
X84811.1	
ACTAGAGTGTAGTAGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATCA	C10
GGAAGAACACCGGTGGCGAAGGCGCTCTGCTGGGCTATAACTGACACTGAGGGACGAAAG	018
AF013028.1 GCTAGAGTTCGGTAGGGGCAAAGGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCG	
GCIAGAGIICGGIAGGGCAAAGCGGAAIICCCCGGIGIAGCGGIGAAAIGCGIAGAIAICG	606
AF091109.1	080
ACTAGAGTACAGTAGGGGTAGAGGGAATTCCCAGTGTAGCGGTGAAATGCGTAGAGATTG	
GGAAGAACACCGGTGGCGAAAGCGCTCTACTGGGCTGTAACTGACACTGAGGGACGAAAG	686
X63141.1	
GCTAGAGTACGGTAGGGGTAGAGGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTG	
GGAAGAACACCGGTGGCGAAAGCGCTCTACTAGGCCGAAACTGACACTGAGGGACGAAAG	686
AJ000708.1	
GCTAGAGGGCGACAGGGGTAGAGGGAATTCCCAGTGTAGCGGTGAAATGCGTAGAGATTG	
GGAAGAACACCGGTGGCGAAAGCGCTCTACTGGGTCGCACCTGACCCTCAGGGACGAAAG	691
U03403.1 ACTAGAGAGCAGTAGGGGTAGCAGGAATTCCCAGTGTAGCG-	
TGAAATGCGTAGAGATTG	
GGAAGAACATCGGTGGCGAAAGCGTGCTACTGGGCTGTATCTGACACTCAGGGACGAAAG	687
AF091110.1	
ACTTGAGTGTGGTAGGGGTAGAGGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTA	
GGAAGAACACCAGTGGCGAAGGCGCTCTACTGGGCCATAACTGACACTGATGGACGAAAG	687
AB015062.1	
ACTAGAGAGTGGTAGGGGTAGAGGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTA	
GGAAGAACACCAGTGGCGAAGGCGCTCTACTGGGCCATTACTGACGCTGATGGACGAAAG	686
AB058219.1	
ACTGGAGTGCGGTAGGGGTAGAGGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTG	
GGAAGAACACCAGTGGCGAAGGCGCTCTACTGGGCCGTAACTGACGCTGAGGGACGAAAG	602
U76911.1	
GCTGGAGTACGGTAGGGCAGAGGGAATTTCCGGTGGAGCGGTGAAATGCGTAGAGATCG	
GAAAGAACAGGAACGGCGAAAGCACTCTGCTGGGCCGACACTGACACTGAGAGACGAAAG	699
AJ440719.1	
GCTGGAGTACGGTAGGGGCAGAGGGAATTTCCGGTGGGCGGTGAAATGCGTAGAGATCG GAAAGAACACCAACGGCGAAAGCACTCTGCTGGGCCGACACTGACACTGAGAGACGAAAG	712
****** *** *** * ***** * **** * *** **	
* ******	
X59559.1	
CTAGGGGAGCGAATGGGATTAGATACCCCAGTAGTCCTAGCCGTAAACGATGGATACTAG	0.2.4
GCGTGGCTTGTATCGACCCGAGCCGTGCCGGAGCCAACGCGTTAAGTATCCCGCCTGGGG	834
CTAGGGGAGCGAATGGGATTAGATACCCCAGTAGTCCTAGCCGTAAACGATGGATACTAG	740
GCGTGGCTTGTATCGACCCGAGCCGTGCCGTAGCTAACGCGTTAAGTATCCCGCCTGGGG	740
CTAGGGGAGCGAATGGGATTAGATACCCCAGTAGTCCTAGCGGTAAACGATGGAAACTAG GTGTGGCCTGTATCGACCCGGGCCGTGCCGAAGCAAACGCGTTAAGTTTCCCGCCTGGGG	720
AF013028.1	730
AP 013020.1 CTAGGGGAGCGAATGGGATTAGATACCCCAGTAGTCCTAGCCGTAAACGATGGATACTAG	
GTGTTGCCTGTATCGACCCAGGCAGTGCCGTAGCTAACGCGTTAAGTATCCCGCCTGGGG	806
AF091109.1	000
APOJITOJ.T CTAGGGGAGCGAATGGGATTAGATACCCCAGTAGTCCTAGCCGTAAACGATGGAAACTAG	
GCGTGGCTTGTATCGACCCGAGCCGTGCCGAASSTAACGCGTTAAGTTTCCCGCCTGGGG	806
X63141.1	
CTAGGGTAGCGAAAGGGATTAGATACCCCTGTAGTCTTAGCGGTAAACGATGGATACTAG	
GTGTAGCTTGTATCGACCCGAGCTGTGCCGAAGCTAACGCGATAAGTATCCCGCCTGGGG	806

Volume 5

www.bioinformation.net

Issue 3

Hypothesis

```
CCAGC-ATTAAGTTGGGCACTCTAGAGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAGTCAGCATGCCCCTTACGTCTTGGGCTACACACGTACTACAATGCTACGG 1192
AJ133161.1 CCATC-ATTAAGTTGGGCACTCTAGAGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAGTCAGCATGCCCCTTACGTCTTGGGCTACACACGTACTACAATGCTCCGG 1098
X84811.1 CCATC-ATTAAGTTGGGAACTCTAGCGAGACTGCCGGTGACAAACCGGAGGAAGGTGAGG
ATGACGTCAAGTCAGCATGGCCCTTACGTCCTGGGCGACACACGTACTACAATGCGAGGG 1096
AF013028.1 CCATC-ATTAAGTTGGGCACTCTGGAGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAGTCAGCATGCCCCTTACGCTCTGGGCTACACACGTACTACAATGGTTGGG 1164
AF091109.1 CTAGCGATTTAATCGCGCACTCTAGGGAGACTGCCGGGGAAAACTCGGAGGAAGGTGGGG
ATGACGTCAAGTCAGCATGCCCCTTACGTCCTGGGCTACACACGTACTACAATGGTCGAG 1165
         CCAAT-ATTAAGTTAGGCACTCTAGGGAGACTGCCGGGGACAACTCGGAGGAAGGTGGGG
ATGACGTCAAGTCAGCATGCCCCTTACGTCCTGGGCGACACGCGTACTACAATGGTTGGG 1164
AJ000708.1 CCAGC-ATTAGGTTGGGCACTCTAGGGAGACTGCCGGTGACAAACCGGAGGAAGGTGAGG
ATGACGTCAAGTCAGCATGCCCCTTATGCTCTGGGCAACACACGTCCTACAATGGTCGGG 1169
          CCAGC-ATTAAGTTGGGGACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAGTCAGCATGCCCCTTACGTCTTGGGCGACACACGTACTACAATGGTCGGG 1164
AF091110.1 CCATC-ATTAAGTTGGGCACTTTGAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAGTCATCATGCCCCTTACATCCTGGGCTACACACGTACTACAATGCAACGG 1165
AB015062.1 CCATC-ATTAAGTTGGGCACTTTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAGTCATCATGCCCCTTACATCCTGGGCTACACACGTACTACAATGCAACGG 1164
AB058219.1 CCATC-ATTAAGTTGGGCACTCTGGGGAGACTGCCGGGGACAACTCGGAGGAAGGTGTGG
ATGACGTCAAGTCATCATGCCCCTTACGTTCTGGGCTACACACGTACTACAATGCTTCGG 1080
U76911.1
          CCACC-GTTGAGTTTGGAACCCTGAGCAGACTGCCGGTGATAAGCCGGAGGAAGGTGAGG
ATGACGTCAAGTCATCATGCCCCTTATGCCCTGGGCGACACACGTGCTACAATGGCCGGG 1177
AJ440719.1 CCACC-GTTGAGTTTGGAACCCTGAACAGACTGCCGGTGATAAGCCGGAGGAAGGTGAGG
ATGACGTCAAGTCATCATGCCCCTTATGCCCTGGGCGACACACGTGCTACAATGGCCGGG 1192
                                     ******* ** **
          ACAGAGGGCAGCAAGCTAGCGATAGCAAGCAAATCCCG-TAAACCGTAGCTCAGTTCAGA
TCGCAGGCTGCAACTCGCCTGCGTGAAGGAGGAATCGCTAGTAATTGCAGGTCAGC-ATA 1310
AJ133161.1 ACAAAGGGCAGCAAGCTAGTGATAGCAAGCAAATCCCA-TAAACCGGAGCTCAGTTCAGA
TCGAAGGCTGCAACTCGCCTTCGTGAAGTCGGAATCGCTAGTAATTGCAGGTCAGC-ATA 1216
          ACAGAGAGCAGCCAACCCGCGAGGGAGAGCGAATCTCA-TAAACCCTGGCACAGTTCAGA
TTGCAGGCTGCAACTCGCCTGCATGAAGGAGGAATCGCTAGTAATCGGAGGTCAGC-ATA 1214
AF013028.1 ACAGAGGGCAGCGAACCCGAGAGGGGAAGCCAATCCCC-AAAACCCAGCCTCAGTTCAGA
TTGCAGGCTGCAACTCGCCTGCATGAAGGAGGAATCGCTAGTAATCGCAGGTCAGC-ATA 1282
AF091109.1 ACAAAGGGCAGCAAACTGGCAACAGCAAGCGAATCCCAGCAAACTCGGTCACAGTTCGGA
TTGCAGGCTGAAACTCGCCTGCATGAAGGAGGAATCGCTAGTAATCGCAGGTCAGC-ATA 1284
          ACAAAGGGCAGCAAACTCGCGAGAGTAAGCGAATCTCATCAAACCCAGCCACAGTTCAGA
TTGCAGGCTGCAACTCGCCTGCATGAAGTAGGAATCGCTAGTAATCGCAGGTCAGC-ATA 1283
AJ000708.1 ACAAAGGGGAGCAAGCTCGCGAGAGTAAGCGAATCCCAGCAAACTCGGCCTCAGTTCAGA
TAGCAGGCTGCAACTCGCCTGCTTGAAGGAGGAATCGCTAGTAATCGCAGGTCAGA-ATA 1288
1103403.1
          A CA A AGGGCAGCGA A CTCGCGAGAGCCAGCGA A TCCCAGCA A A CCCGGCCTCAGTTCAGA
TTGCAGGCTGCAACTCGCCTGCATGAAGGAGGAATCGCTAGTAATCGCCGGTCAGC-ATA 1283
AF091110.1 ACAAAGGGCAGCCAACTAGCGATAGTGCGCTAATCCCA-TAAACCGTTGCTCAGTTCAGA
TTGCAGGCTGCAACTCGCCTGCATGAAGGCGGAATCGCTAGTAATCGCAGGTCAGC-ATA 1283
AB015062.1 ACAAAGGGCAGCCAACTAGCGATAGTGAGCTAATCCCA-TAAACCGTTGCTCAGTTCAAA
TTGCAGGCTGCAACTCGCCTGCATGAAGGCGGAATCGCTAGTAATCGCAGGTCAGC-ATA 1282
AB058219.1 ACAAAGGGCAGCAAGCGTGCGAGCGCAAGCTAATCCCA-TAAACCGAGGCACAGTTCAGA
TTGCAGGCTGCAACTCGCCTGCATGAAGGAGGAATCGCTAGTAATCGCAGGTCAGC-ATA 1198
         ACAAAGGGTCGCGATCCCGCGAGGGTGAGCTAACTCCA-AAAACCCGTCCTCAGTTCGGA
U76911.1
TTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCCGGTCAGCCATA 1296
AJ440719.1 ACAAAGGGTCGCGATCCCGCGAGGGTGAGCTAACTCCA-AAAACCCGTCCTCAGTTCGGA
FTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCCGGTCAGCCATA 1311
```

CTGCAGTGAATTCGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGAAGCTGGCA 1370