BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(12)

Diversity of *Streptomyces* spp. in Eastern Himalayan region – computational RNomics approach to phylogeny

Kaushik Bhattacharjee, Subhro Banerjee & Santa Ram Joshi*

Microbiology Laboratory, Department of Biotechnology and Bioinformatics, North-Eastern Hill University, Shillong-793022; Santa Ram Joshi – E-mail: srjoshi2006@yahoo.co.in; Phone: 91-364-2722405, 91-9436102171; *Corresponding author

Received June 05, 2012; Accepted June 08, 2012; Published June 28, 2012

Abstract:
Isolation and characterization of actinomycetes from soil samples from altitudinal gradient of North-East India were investigated for computational RNomics based phylogeny. A total of 52 diverse isolates of *Streptomyces* from the soil samples were isolated on four different media and from these 6 isolates were selected on the basis of cultural characteristics, microscopic and biochemical studies. Sequencing of 16S rDNA of the selected isolates identified them to belong to six different species of *Streptomyces*. The molecular morphometric and physico-kinetic analysis of 16S rRNA sequences were performed to predict the diversity of the genus. The computational RNomics study revealed the significance of the structural RNA based phylogenetic analysis in a relatively diverse group of *Streptomyces*.

Keywords: Actinobacteria, Eastern Himalaya, *Streptomyces*, 16S rDNA, Molecular morphometrics

Background:
Actinomycetes comprise a large and diverse group of Gram-positive bacteria frequently filamentous and sporulating extensively obtained in environments with DNA rich in GC content from 57%-75%. These bacteria are primarily saprophytic and are known to contribute in nutrient turnover, using many available nutrient sources for their development **[1].** Many actinomycetes are recognized for their metabolic versatility that frequently is accompanied by the production of secondary metabolites of economic importance **[2].** During the last few decades, isolation of actinomycetes has been mainly performed as part of drug discovery programs and has focused on terrestrial sources. Considering the importance of *Streptomyces* and its relatives in terms of both biological behavior and metabolic products, it becomes essential to understand its evolutionary relationships to other species in the diverse actinomycetales order **[3].** An evolutionary study may help to explain how *Streptomyces* emerged and adapted to the soil environment. On the other hand, information obtained from a well-resolved phylogeny can be used for the comparison of genome sequences, comparative genome reannotation, and genome visualization. A robust phylogeny is central for ongoing efforts in many groups to reconstruct system-wide metabolic models of *Streptomyces* and related species **[4],** which are used for systematic strain-engineering in biotechnology. Molecular systematics, which includes both classification and identification, has its origin in the early nucleic acid hybridization studies, but has achieved a new status following the introduction of nucleic acid sequencing techniques **[5].** Significance of phylogenetic studies based on 16S rDNA sequences is increasing in the systematics of actinomycetes **[6].** However, RNA secondary structures have not been particularly used for such taxonomic studies, although they can provide meaningful information on systematics because of their inherent prospects of providing characteristics on "molecular morphometrics" usually not found in the primary sequences. The basic idea behind molecular morphometrics is to use the molecular structures as a direct source of measurable

information. This method is based on the assumption that, secondary structure can be as significant as primary sequence in deriving phylogenetic relationship. In other words, one can consider the secondary-structure elements of RNA molecules, i.e., the helices, loops, bulges, and separating single-stranded portions, as phylogenetic characters **[7]** instead of the conventional linear sequences. Sequence variations in RNA sequences maintain base-pairing patterns that give rise to double-stranded regions in the secondary structure of the molecule. Thus, alignments of two sequences that specify the same RNA molecules will show covariation at interacting base-pair positions. In addition to these covariable positions, sequences of RNA-specifying genes may also have rows of similar sequence characters that reflect the common ancestry of the genes. It is an established fact that rRNA structure is highly conserved throughout evolution as most of the folding is functionally important despite primary sequence divergence.

North-East India being one of the mega biodiversity hotspots of the world has natural ecosystems still prevalent in pristine conditions which offer scope for bioprospection of novel organisms hitherto unknown to scientific world. Exploration of actinomycetes diversity from diverse ecological niches of the Indo-Myanmar hotspot holds promise for isolation of biotechnologically significant strains of actinomycetes and even novel species **[8].** In this study, cultivation-based and molecular approach is employed to compare the culturable actinobacterial diversity along the altitudes of eastern Himalayan range. The phylogenetic affiliation of the actinobacterial isolates were assessed by 16S rDNA and molecular morphometrics approach.

Methodology:

Site description and Sample collection

A total of 18 soil samples were collected aseptically from different sites (located in between latitude 26 55′N and 28 40′ and longitude 92 40′ and 94 21′) at a depth of 5-15 cm in sterile containers from May 2010 to April 2011 along the Eastern Himalayan range of India. The soil samples were brought to the laboratory and stored in dark at 4°C for not more than one week.

Morphological and Biochemical studies

One gm of each sample was taken for actinomycete isolation. Samples were suspended in sterile saline solution (0.85% NaCl), treated by thermal shock at 60°C for 6 min and serial diluted to 10^{-6}. Aliquots (0.1 ml) of each dilution were spread on Starch casein agar, Actinomycetes isolation agar, ISP-2 and ISP-3 medium (Himedia Laboratories Pvt. Ltd., India) **[9]** amended with nalidixic acid (10 µg/ml) and nystatin (25 µg/ml) to inhibit fungal growth and the fast growing Gram-negative bacteria. Plates were incubated at 28°C for one to three weeks and checked regularly for the emergence of presumptive actinobacteria. Actinomycetes were subcultured onto new Starch casein agar, Actinomycetes isolation agar, ISP-2 and ISP-3 medium plates supplemented with the above antibiotics. Isolates were compared based on colony and morphological characteristics and two representative colonies for each sample were selected. Coverslip culture of all the isolates were prepared on CSPY-ME and ISP-3 medium and incubated at 28°C for 4-8 days **[10].** Periodical observations on Gram staining and mycelloid structure were recorded using Leica microscope model DM 5500. Biochemical characterizations were performed following standard protocols **[9, 10].**

DNA isolation, 16S rRNA gene amplification and Sequencing

The representative 6 isolates were selected and cultured on ISP-2 liquid medium at 28° C for 3 to 5 days in a orbital shaking incubator (New Brunswick Scientific, USA). Genomic DNA was isolated according to a method described by Selvin *et al.* **[10]** and for 16S rDNA gene amplification, genomic DNA was amplified with actinomycetes specific primers T7-pA, 5′-TAA TAC GAC TCA CTA TAG AGA GTT TGA TCC TGG CTCAG -3′, and 16Sact1114r, 5′-GAG TTG ACC CCG GCR GT -3′ **[11].** The reactions were performed on a GeneAmp PCR system 9700 (Applied Biosystems, USA). The PCR mixture consisted of 5 µl 10X Buffer (with $Mg2^{+}$), 8 µl dNTP mixture (1.25 mM each), 0.5 µl of each primer, 1 µl of template DNA, and 1.0 µl of *Taq* polymerase (Fermentas, USA) in a final volume of 50 µl [10]. PCR amplification parameters were as follows: 94°C for 3 min of initial melting; 30 cycles of 94°C, 45 s; 55°C, 30 s; and 72°C, 90 s; and a final extension at 72°C for 7 min **[11].** The PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen, Germany) according to manufacturer's protocols. The DNA content of the PCR products was estimated using a NanoVue Plus Spectrophotometer (GE Healthcare's Life Sciences, Sweden). Sequencing reactions of the 16S rDNA fragments were performed with the Big Dye Terminator v 3.1 Cycle Sequencing Kit (Applied Biosystems, USA).

Molecular morphometric and phylogenetic analysis:

Sequence alignment based on primary sequences and phylogenetic analysis

The 16S rRNA gene sequence of the isolates and their closely related species were retrieved from EzTaxon-e server (http://eztaxon-e.ezbiocloud.net/) **[12]** and aligned using the ClustalX2 programme **[13].** The tree topologies were evaluated by bootstrap analyses based on 1,000 replications with ProfDistS 0.9.9 **[14]** and consensus phylogenetic tree was inferred using the distance based method neighbor-joining i.e., ProfDist neighbor-joining (PNJ) method considering *Escherichia coli* and *Bacillus subtilis* as outgroups with Kimura's 2 parameter model (K2P) as distance correction **[15].** The threshold for profile generation was adjusted to 75% in case of bootstrap value lower and 90% in each case of uncorrected p-distances. The generated alignments and the consensus trees were deposited in TreeBASE under submission ID number 12801 **[16].**The Weblogo program **[17]** was used to generate sequence logos for assessing sequence conservation within 16S rRNA gene and the relative frequency of the nucleotides at each position. To generate the sequence logos the sequences for the phylogenetic outgroups were removed from the alignment.

Prediction of secondary structure of RNA and analysis

Open source sequence conversion tools were used to convert the DNA sequences to RNA sequences in upper case and lower case format and a single .fas file was created containing all the 16S rDNA sequences of *Streptomyces* spp. Different physical and kinetic properties for these RNA sequences were calculated using *in silico* open source tool OligoCalc **[18].**

Restrictions and constraints of the RNA sequences of the selected *Streptomyces* spp. were checked and secondary folding were predicted using MFOLD 3.0 **[19]** at a default temperature

of 37°C, [Na^+ = 1 M] and [Mg^{++} = 0 M]. The structure chosen from different output files were with the highest negative free energy value but essentially identical for all the species (e.g., ring model structure). The .ct files and Vienna files were downloaded and used for secondary structure display, Loop Free-Energy Decomposition (LFD) and minimum free energy (MFE) determination purpose. The energy dot-plots **[20]** of the sequences were also drawn using the MFOLD 3.0 **[19].** Structure annotation followed in the study was according to Zuker and Jacobson method **[21].** The Maximum Weight Matching (MWM) analysis and circle plot of individual RNA sequences were drawn using Circle 0.1.1 package **[22, 23]** which gives the assumptions regarding the pseudoknots. The structure annotated alignment, RNAalifold consensus structure and consensus minimum free energy (MFE) of the six *Streptomyces* isolates was also interpreted by LocARNA 1.5.2 **[24].**

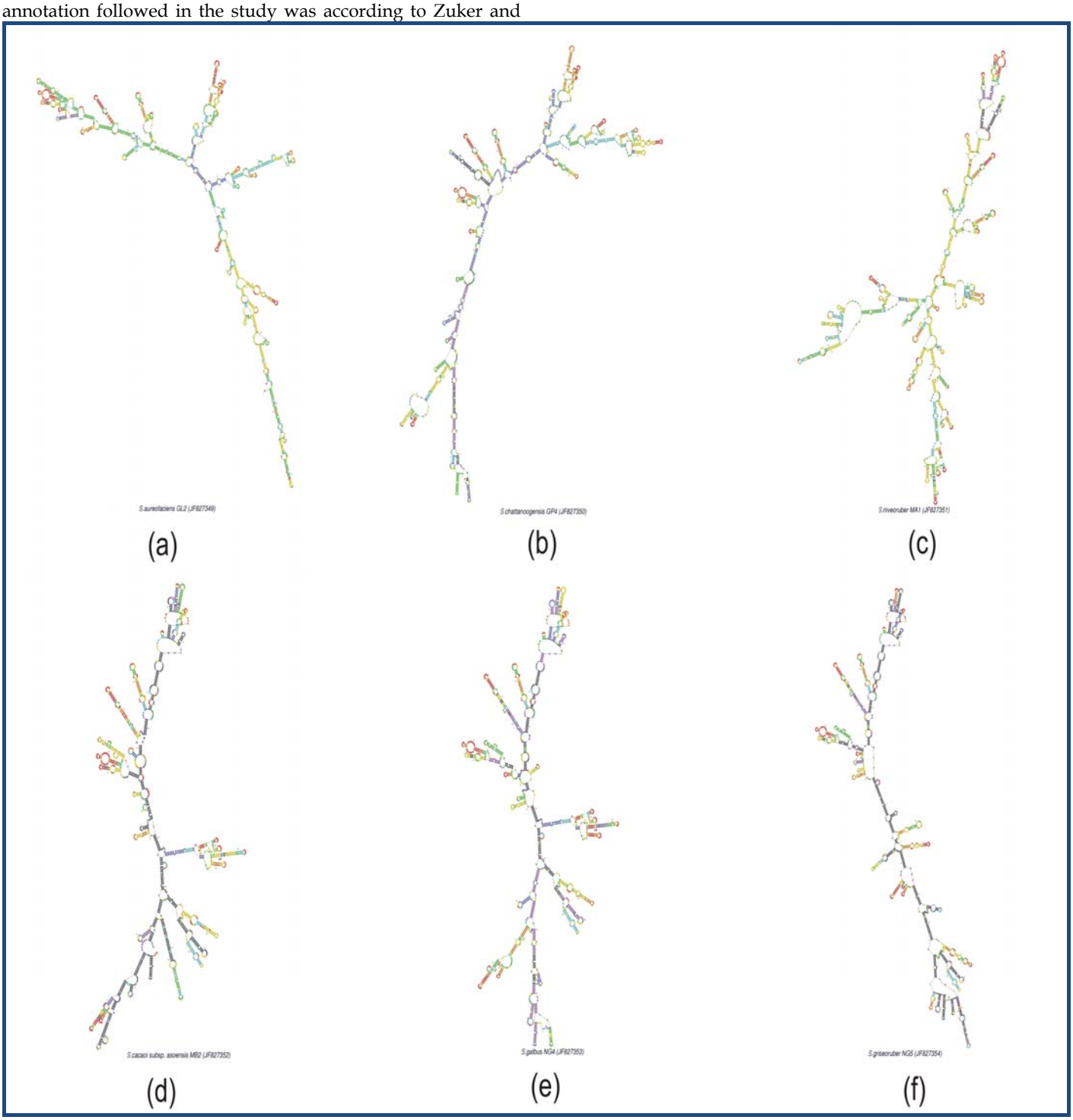

Figure 1: Result of MFOLD 3.0's prediction on 16S rRNA of the *Streptomyces* isolates [**(a)** *S.aureofaciens* GL2, **(b)** *S.chattanoogensis* GP4, **(c)** *S.niveoruber* MA1, **(d)** *S.cacaoi subsp. asoensis* MB2, **(e)** *S.galbus* NG4; **(f)** *S.griseoruber* NG5]. Different colours indicates different probability of occurrence of base paring in the secondary structure with red colour as highest (0.999) and black as lowest (0.001) [please refer 19 for colour scheme].

Sequence alignment based on secondary structure and phylogenetic analysis

Using the Q-INS-i of MAFFT programme a multiple structural alignment of RNA sequences was constructed taking both multiple sequence alignment and secondary structure prediction in account, which was achieved with McCaskill routine and McCaskill-MEA algorithm implemented in Vienna RNA package and MXSCARNA **[25].** The output file was converted in order for it to be recognized by ProfDistS 0.9.9 and phylogenetic analysis was performed as mentioned above. Here also the generated alignments and the consensus trees were deposited in TreeBASE under submission ID number 12802 **[16].** Sequence logos were also generated for 16S rRNA structural alignment to assess sequence conservation and the relative frequency of the nucleotides at each position using the Weblogo program. Analytical runs were performed on two different HP Workstations, one with Windows XP and another with CentOS 6.2 which has one quad-core 2.93 GHz Intel Xeon processors, Intel 5520 chipset, 16 Gb of RAM at the NUMA and NVIDIA Quadro FX 3800 1.0 GB PCIe Graphics Card.

Nucleotide sequence accession numbers

The 16S rRNA gene sequences determined for the 6 representative isolates in this study were deposited in the GenBank database with the accession numbers JF827349 to JF827354.

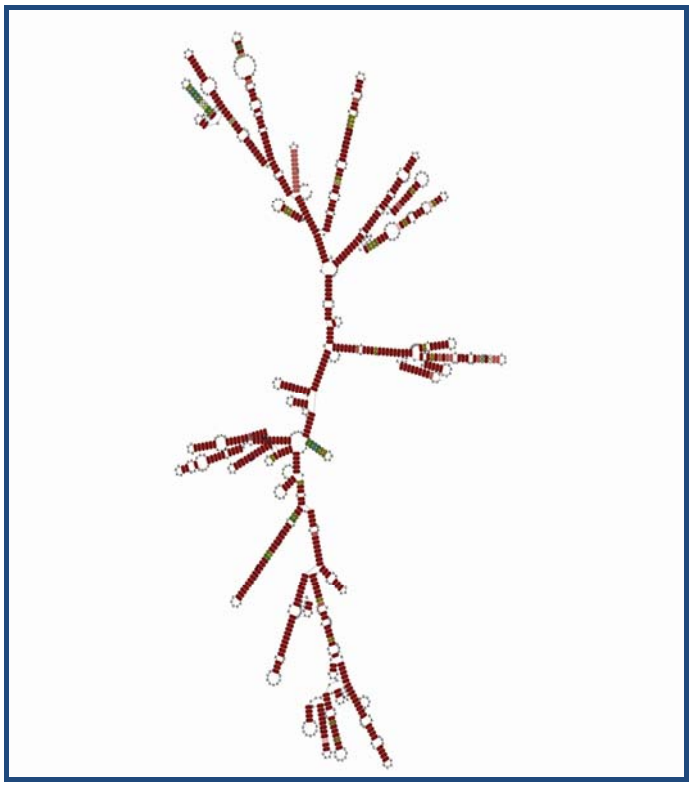

Figure 2: Predicted RNAalifold consensus secondary structure of the six *Streptomyces* isolates.

Discussion:

Discrete colonies were observed in various media used for isolation of actinomycetes **(Supplement Figure 1).** The use of standard culture media targeting *Actinobacteria* resulted in the isolation of 52 isolates not previously cultured from this region. In all cases, morphotypes obtained in the study were regarded as novel if they had not been observed in previous culture-based studies. The largest diversity of bacterial morphotypes was observed on plates of actinomycete isolation agar (AIA), while starch casein agar yielded the lowest diversity. Microscopic observation revealed the organisms to be Gram-positive and filamentous with distinct biochemical reactions. PCR of total DNA isolated from different sites using 16S rRNA primers specific for actinobacteria yielded a band of the expected size of ~1500 bp. An almost-complete 16S rDNA sequence containing less than 1% undetermined positions were obtained for all of the isolates. Out of 52 bacterial isolates, 6 isolates selected based on colony, morphology and biochemical characters were identified upto species level using 16S rRNA phylogenetic analysis. The isolates belonged to the *Streptomyces*, supported by two treeing algorithms and a high bootstrap value in the neighbour-joining analysis. The Tajima's D test was used to measure the allele frequency distribution of nucleotide sequence data. This statistics can be influenced by both population history and natural selection. The calculated Tajima's D value is -1.566278, so the null hypothesis of mutation-drift equilibrium and constant population size is rejected.

The *in silico* analysis of the physical and kinetic properties of the six *Streptomyces* isolates showed a similar values in most of the cases shown in **Table 1 (see supplementary material).** The high GC content of structural RNA also hypothesized that *Streptomyces* favors high adaptability to temperature change in the environment **[26]**. The calculated GC content of structural RNA and basic melting temperatures were positively correlated ($p<0.05$) which is also supported by the findings of Laurence and Merchant **[26]**. Except the *S.aureofaciens* GL2, all the other isolates showed a similar secondary structure of RNA **(Figure 1).** Comparison of *Streptomyces* 16S rRNA secondary structure with other available actinobacteria sequences showed that it was consistent with the structure of the other known actinobacterial 16S rRNA gene (data not given). The consensus structure also showed a consistent conserved secondary structure, but with a great difference in case of consensus MFE **(Table 1, Figure 2).** DNA and RNA secondary structure based sequence logos **(Supplement Figure 2)** were constructed for 16S rRNA gene in order to easily visualize conserved regions and local sequence biases, such as expansion or contraction segments and mono and di-nucleotide repeats. They clearly indicate variation in case of linear DNA sequence conservation and secondary structural conservation. Molecular phylogenetic and molecular morphometric analysis based upon 16S rRNA genes clearly and consistently showed the diversity of the genus *Streptomyces* **(Figure 3).**

The phylogenetic data revealed clear differences between the isolates and divided these isolates into six different groups in both linear DNA sequence and 16S rRNA secondary structure based phylogeny **(Figure 3).** This variation among the isolates is also reflected by the relatively high amount of informative sites present. But with RNA secondary structure based phylogeny, cluster 6 containing *S.chattanoogensis* GP4 and cluster 7 containing *S.aureofaciens* GL2 are more closely related when compared to linear DNA based phylogeny. The structure of RNA is evolutionarily more conserved than its sequence and is thus key to its use in this phylogenetic analysis **[27].** The function of non-coding RNA (ncRNA) genes like 16S rRNA

largely depends on their secondary structure and the interaction with other molecules. Thus, the secondary structure of a 16S rRNA is more conserved than its counterpart DNA sequence **[28].** The tree obtained from the analysis of the concatenated 16S rRNA indicated that this gene determined the general topology of the tree and resolved the main lineages consistently (data not shown). The molecular phylogeny of the genus *Streptomyces* presented here is consistent with the earlier analysis by Kim *et al*. **[28].** However, the study is more comprehensive than earlier ones, in terms of the computational RNomics approach and depth of phylogeny of taxa resolved. More specifically, this is the first report on the secondary structure of 16S rRNA being used for sequence alignment in a molecular phylogenetic analysis of the genus *Streptomyces*.

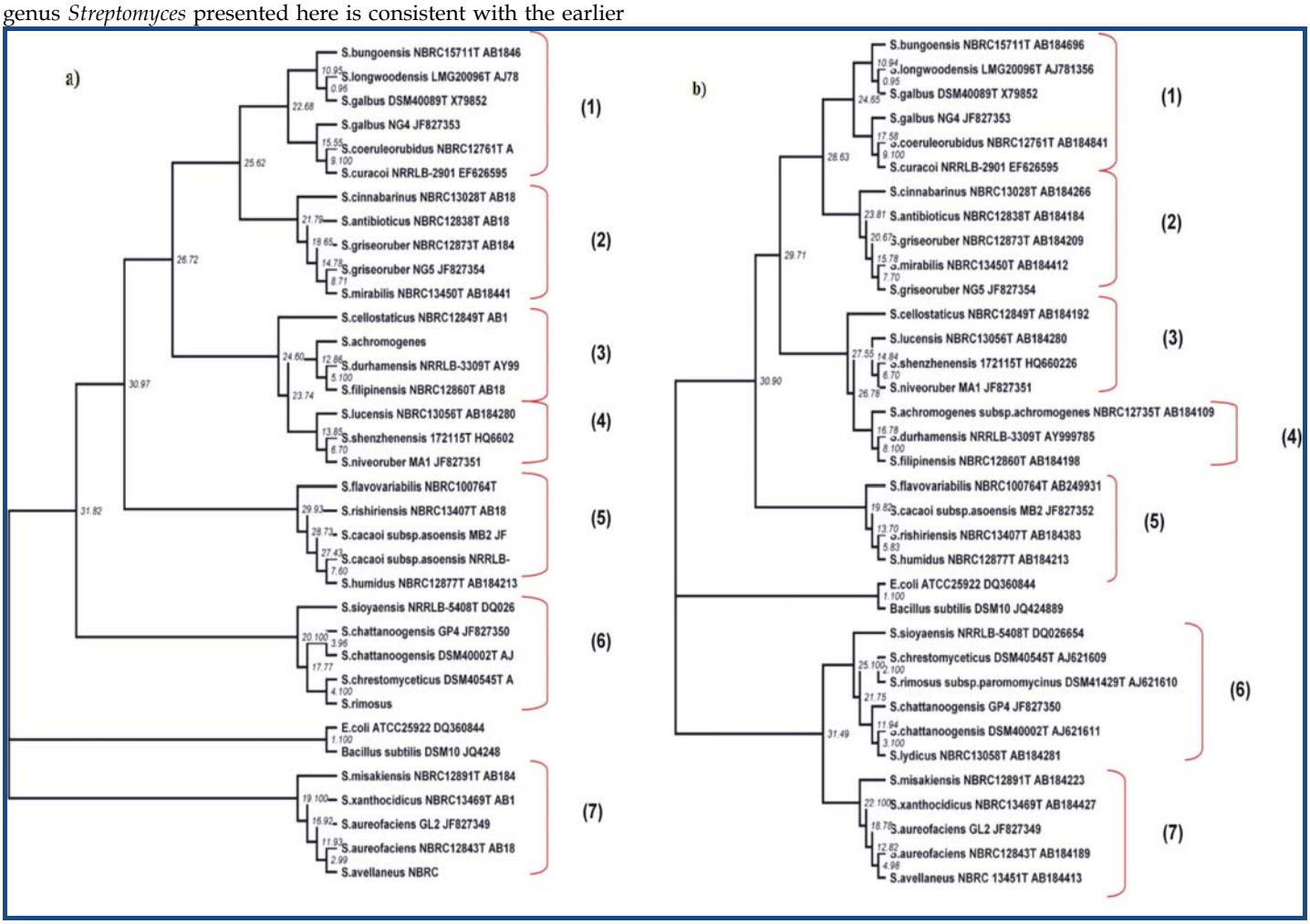

Figure 3: Phylogenetic tree showing the relationships among the 16S rDNA sequences of 6 isolates from Eastern Himalayan region and related organisms from the EzTaxon-e database. The tree was constructed by using the neighbour-joining method for calculation. Bootstrap values were calculated from 1000 replications of Kimura 2-parameter, and bootstrap values higher than 75% were shown on branches. Values on the braches are branch length. **(a)** Based on linear sequence alignment; **(b)** based on secondary structure of RNA alignment**.**

Acknowledgement:
Authors are grateful to Department of Information Technology, Ministry of Communications & Information Technology, Government of India for the research support. Authors acknowledge the support received from Mr. Shakti Kumar, Bioinformatics Center, North-Eastern Hill University, Shillong.

References:

[1] Mincer TJ *et al*. *Appl Environ Microbiol*. 2002 **68**: 5005 [PMID: 12324350]
[2] Bull AT *et al*. *Microbiol Mol Biol Rev*. 2000 **64**: 573 [PMID: 10974127]
[3] Alam MT *et al*. *Mol Phylogenet Evol*. 2010 **54**: 763 [PMID: 19948233]
[4] Borodina I *et al*. *Genome Res*. 2005 **15**: 820 [PMID: 15930493]
[5] Höfling JF *et al*. *Rev Inst Med Trop Sao Paulo*. 1997 **39**: 345 [PMID: 9674286]
[6] Anderson AS & Wellington EM, *Int J Syst Evol Microbiol*. 2001 **51**: 797 [PMID: 11411701]
[7] Billoud B *et al*. *Mol Biol Evol*. 2000 **17**: 1435 [PMID: 11018151]
[8] Myers N *et al*. *Nature*. 2000 **403**: 853 [PMID: 10706275]
[9] Shirling EB & Gottlieb D, *Int J Syst Bacteriol*. 1966 **16**: 313.
[10] Selvin J *et al*. *Appl Microbiol Biotechnol*. 2009 **83:** 435 [PMID: 19190903]
[11] Kyselkova M *et al*. *Antonie van Leeuwenhoek*. 2008 **94**: 439 [PMID: 18600470]
[12] Kim OS *et al*. *Int J Syst Evol Microbiol*. 2012 **62**: 716 [PMID: 22140171]

[13] Larkin MA *et al*. *Bioinformatics*, 2007 **23**: 2947 [PMID: 17846036]
[14] Wolf M *et al*. *Bioinformatics*. 2008 **24**: 2401 [PMID: 18723521]
[15] Kimura M, *J Mol Evol*. 1980 **16**: 111 [PMID: 7463489]
[16] Anwar N & Hunt E, *BMC Evol Biol*. 2009 **9**: 93 [PMID: 19426482]
[17] Crooks GE *et al*. *Genome Res*. 2004 **14**: 1188 [PMID: 15173120]
[18] Kibbe WA, *Nucleic Acids Res*. 2007 **35**: W43 [PMID: 17452344]
[19] Zuker M, *Nucleic Acids Res*. 2003 **31**: 3406 [PMID: 12824337]
[20] Jacobson AB & Zuker M, *J Mol Biol*. 1993 **233**: 261 [PMID: 8377202]
[21] Zuker M & Jacobson AB, *Nucleic Acids Res*. 1995 **23**: 2791 [PMID: 7544463]
[22] Tabaska JE *et al*. *Bioinformatics*. 1998 **14**: 691 [PMID: 9789095]
[23] Page RD, *Bioinformatics*. 2000 **16**: 1042 [PMID: 11159317]
[24] Smith C *et al*. *Nucleic Acids Res*. 2010 **38**: W373 [PMID: 20444875]
[25] Katoh K & Toh H, *BMC Bioinformatics*. 2008 **9**: 212 [PMID: 18439255]
[26] Hurst LD & Merchant AR, *Proc Biol Sci*. 2001 **268**: 493 [PMID: 11296861]
[27] Athanasius F Bompfünewerer Consortium *et al*. *J Exp Zool B Mol Dev Evol*. 2007 **308**: 1 [PMID: 17171697]
[28] Kim *et al*. *Int J Syst Evol Microbiol*. 2004 **54**: 593 [PMID: 15023980]

Edited by P Kangueane
Citation: **Bhattacharjee *et al.*** Bioinformation 8(12): 548-554 (2012)

BIOINFORMATION

open access

Supplementary material:

Figure 1: Pure culture colonies on ISP-3 agar plates and microscopic observations performing Gram staining (100X magnification) of the *Streptomyces* isolates.

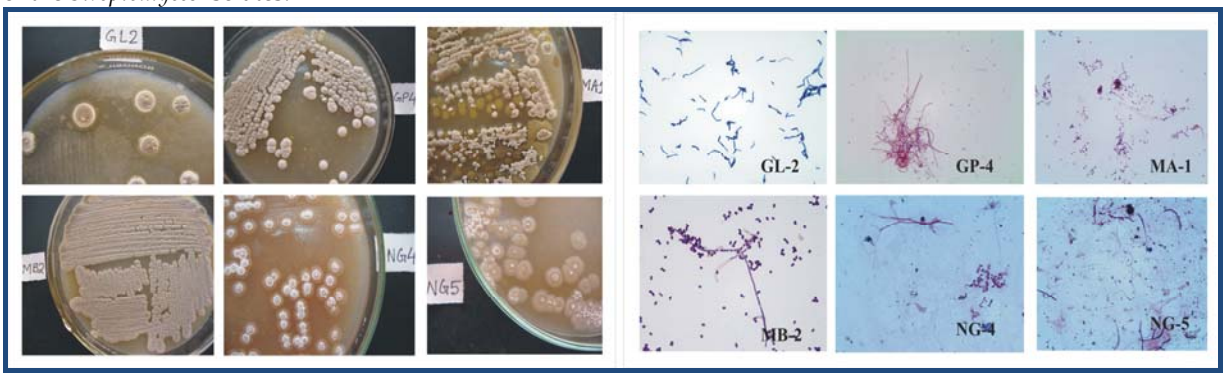

Figure 2: Sequence logos for assessing sequence conservation within 16S rRNA gene and the relative frequency of the nucleotides at each position generated using Weblogo program.

Table 1: Different physical and kinetic properties for these RNA sequences MFE and consensus MFE of structural RNA of the *Streptomyces* isolates (MFE- minimum free energy).

S.no	NCBI Acc. No.	Strain	Length	Melting Temperature T_m		GC %	MW (g/M)	MFE (Kcal/mol)	Consensus MFE (Kcal/mol)
				Basic	Salt adjusted				
1	JF827349	*S.aureofaciens* GL2	1464	88.6	93.7	59	475278.7	-640.3	
2	JF827350	*S.chattanoogensis* GP4	1487	88.5	93.8	59	482433.8	-648.7	-1015.74
3	JF827351	*S.niveoruber* MA1	1484	88.4	93	58	481419.2	-638.6	
4	JF827352	*S.cacaoi subsp. asoensis* MB2	1486	88.3	93	58	482064.5	-640.3	
5	JF827353	*S.galbus* NG4	1483	88.4	93	58	481097	-642.1	
6	JF827354	*S.griseoruber* NG5	1484	88.4	93	58	481489.3	-641.1	