Isolation, identification and comparative analysis of 16S rRNA of *Bacillus subtilis* grown around *Rhazya stricta* roots

Jamal S.M. Sabir,[1] Salah E.M. Abo-Aba,[1,2,*], Ayman Sabry,[3,4] Refaei M. Hussein,[1,5] Ahmed Bahieldin[1,6], and Nabeeh A. Baeshen[1]

[1]Biological Science Dept., Faculty of Sciences, King Abdul-Aziz Univ., KSA
[2] Microbial Genetics Dept., Genetic Engineering & Biotechnology Div., National Research Center, Dokki, Giza, Egypt
[3]Biotechnology and Genetic Engineering Unit, Scientific Research Deanship, Taif Univ., Taif, Zip Code 21974, P.B 888, KSA
[4]Cell Biology Dept., National Research Center, Dokki, Giza Egypt
[5]Genetics and Cytology Dept., Genetic Engineering and Biotechnology Div., National Research Center, Dokki, Giza, Egypt
[6]Genetics Dept., Fac. Agriculture, Ain Shams Univ., Cairo, Egypt
*Corresponding Author: Salah_aboaba@yahoo.com

Abstract: A total of 26 *Bacillus* strains grown around *Rhazya stricta* roots were isolated from Saudi Arabia soil, and identified according to their morphological, physiological, biochemical characters as well as phylogenetic analysis using 16S gene sequences. The phylogenetic tree of the twenty six strains based on 16S rDNA sequences was generated in comparison with 48 Genbank 16S *Bacillus subtilis* partial sequences. *In silico* analysis the generated showed that the 16S sequence size for the 26 isolates ranged between 995 to 1233 nt, while their counterparts in the Genbank ranged between 1153-1559 nt. The percentage for GC content for all 68 16S rRNA gene sequences ranged between 55-56%. A total of five genomic conserved sequences were detected with the smallest being 106 nt and the largest is 437 nt. The range size (±standard deviation) of these five sequences is 251.8±142 nt.
[Jamal S.M. Sabir, Salah E.M. Abo-Aba, Ayman Sabry, Refaei M. Hussein, Ahmed Bahieldin, and Nabeeh A. Baeshen. **Isolation, identification and comparative analysis of 16S rRNA of *Bacillus subtilis* grown around *Rhazya stricta* roots.** *Life Sci J* 2013;10(12s):980-986] (ISSN: 1097-8135).http://www.lifesciencesite.com. 158

Keywords: Genbank, Phylogeneitc tree, rDNA, *Bacillus*

1. Introduction

Soil is an excellent source for unknown microorganisms; the most frequently isolated genus has been *Bacillus*. The use of 16S rRNA gene sequences to study bacterial evolution has been the most common house-keeping genetic marker used for a number of reasons. They include (i) existence in almost all bacteria, often as a multigene family, or operon; (ii) the function has not changed over time, suggesting that random sequence changes are a more accurate measure of time (evolution); and (iii) the 16S rRNA gene (1,500 bp) is large enough for bioinformatics purposes (Patel, 2001).

The cumulative results from a limited number of studies to date suggest that 16S rRNA gene sequencing provides identification on the genus level (90%) but less so with regard to species level (65 to 83%), with from 1 to 14% of the isolates remaining unidentified after testing (Drancourt *et al*., 2000; Mignard and Flandrois, 2006; Woo *et al*., 2003). 16S rRNA and its gene have proven to be useful and powerful markers for the presence of bacteria in biological samples (Rappe and Giovannoni, 2003). Several *Bacillus* species were reclassified based on 16S rDNA and separated into different phylogenetically distinct clusters (Dong and Jean-charles 2003). Partial 16S rDNA sequence (Goto *et al*., 2000) and rRNA gene restriction pattern (Joung and Coˆte 2002) have been used for the rapid identification or classification of *Bacillus* species and related genera (Dong and Jean-charles 2003). *Bacillus* species is ubiquitous and broadly adapted to grow in diverse settings within the biosphere. *B. subtilis* can be isolated in greater numbers than most other spore-forming bacteria from the rhizosphere of a variety of plants like *Rhazya stricta*. There is an evidence that through these associations, *B. subtilis* can promote plant growth (Fall *et al*., 2004; Cazorla *et al*., 2007; Nagorska *et al*., 2007). The traditional methods for isolating *B. subtilis* require that the organism be in its spore form. Besides, there is no guarantee that the isolated strain from a particular environment actually grow at that location (Earl *et al*., 2008). It is important to realize that libraries of PCR-amplified 16S rRNA and 16S rRNA genes may not represent a complete or accurate picture of the bacterial community in a given biological sample. Firstly, the species diversity is so great, the available

libraries of 16S rRNA and 16S rRNA genes permit an initial survey of the global soil bacterial community structure. The majority (79 to 89%) of 16S rRNA gene sequences are from bacteria that are not affiliated with known genera (Janssen, 2006). The gene encoding the small subunit rRNA serves as a prominent tool for the phylogenetic analysis and classification of bacteria and archaea owing to its high degree of conservation and its fundamental function in living organisms (Salman *et al*., 2012).

In the present study, we performed an investigation for *Bacillus subtilis* strains by isolating and identifying twenty-six strains grown in Saudi Arabia soil around *R. stricta* roots. Theses isolates were sequenced for its 16S rRNA. The purpose of this study is bi-fold, (i) to carry out morphological physiological and biochemical characters of Saudi soil Bacteria (*Bacillus subtilis)* isolates around *R. stricta* roots (ii) *in silico* comparative analysis of the nucleotide sequence, and differentiation to its counterparts *Bacillus subtilus* collected form NCBI Genbank.

2. Materials and Methods

Data collection

Twinty six soil samples were collected from Bahra, Jeddah region, Saudi Arabia around roots of *R. stricta*. Soil samples were collected using sterile forceps in sterile falcon tubes with little manipulation and stored directly in dry ice. Before using the soil samples, large debris (stones or plant materials) was removed carefully, using sterilized fine forceps. The sample taken from each site for *Bacillus* isolation, physiological and biochemical analysis was placed separately in clean sterile Falcon tubes, stored overnight at 4 °C.

Isolation of *Bacillus* isolates from field

One gm of soil sample from each collected soil sample transferred into 10 ml tube containing saline solution (NaCl 0.1 & w/v) and mixed vigorously with vortexes for 5 min. until soil particles settled and left for 30 min at room temperature, sample solutions were diluted ten- fold with 0.1% NaCl, spread onto Luria-Bertani (LB) (Luria and Burrous 1995) agar plates, and incubated overnight at 37^{o}C. Bacilli-like colonies were isolated according to their morphological characters. Individual colonies from each site were picked up and purified by re-streaking, numbered on LB media as shown in Fig (1).

Culture media and growth conditions

Bacillus isolates were cultivated at 37^{o}C in Luria-Bertani (LB) medium composed of 1% NaCl, 1% tryptone, and 0.5% yeast extract. A single colony of an isolate was three times streaked over LB agar and incubated at 37^{o}C for 24 h. Purified colonies cultured under various pH (between pH 5 and pH 9). Salt tolerance experiments were performed on LB broth with NaCl, in various concentrations (between 0.5 and 10%, w/v) and temperature (between 25^{o}C and 50^{o}C) at 37^{o}C. Purified *Bacillus* growth was numbered as indicated in Table (2).

Identification of selected bacterial isolates

Identification of the isolated *Bacillus* isolates was performed based on their morphological, physiological, and biochemical characteristics, as described in Bergey's Manual of Systematic Bacteriology (Claus and Berkeley, 1986). The morphological characteristics of the isolates were observed with a microscope (Olympus BX51) using cells grown on LB media. The experiments were performed in duplicate with inoculums subcultures at least once under the same test conditions.

Extraction of DNA and PCR amplification and sequencing of 16S rDNA

Extraction of DNAs was carried out at Macrogen Company Korea, with macrogen universal primer (Table 1). PCR amplification and sequencing of the 16S rRNA gene were performed. The 16S rRNA genes were PCR-amplified from the genomic DNA using the bacterial Macrogen universal primer set of 518 F (5'- CCAGCAGCCGCGGTAATACG -3') and 800 R (5'- TACCAGGGTATCTAATCC -3').

In silico sequence analysis

Nucleotides sequences were analyzed using Biostrings package (Pages *et al*., 2013) under the R Project for Statistical Computing (R Core Team, 2013). DnaSP (version 5.10.01) software was used to analyze the haplotype diversity (Hd), the average number of nucleotide differences (Tajima, 1983), and the nucleotide diversity($\boldsymbol{\pi}$), (Lynch & Crease 1990). The phylogenetic tree among species was constructed using maximum likelihood metohd implemented in Mega 5 software (Tamura *et al*., 2007). Bootstrap analysis was used to evaluate the tree topology of the maximum likelihood data by means of 1000 resamplings.

3. Results and Discussion

Morphological, physiological and biochemical characters of *Bacillus subtilis* isolates

The genus *Bacillus* is generally defined according to classical phenotypic characterization based primarily on morphology, nutrition, growth characteristics; and various substrate utilization and physiological assessments. Physiological reactions are generally used to determine the species of the genus (Claus and Berkeley, 1986; Slepecky and Hemphill, 1992).).

All cells are Gram-stain-positive, straight rods occurring singly, as pairs or as short chains, that

produce subterminal endospores. Motile Colonies are pale yellow-pigmented, flat and translucent, aerobic, growth occurs in 25°C–50°C temperature (optimum 30°C–37°C), at pH 5.0–9.0 (optimum pH 7.0), and 0.5-10% (w/v) NaCl (optimum 0.5–4.0 %). Oxidase and catalase are positive, nitrate reduced, gelatin, casien and starch are hydrolyzed, acids are produced from D-glucose, D-mannitol, glycerol, sucrose, starch.

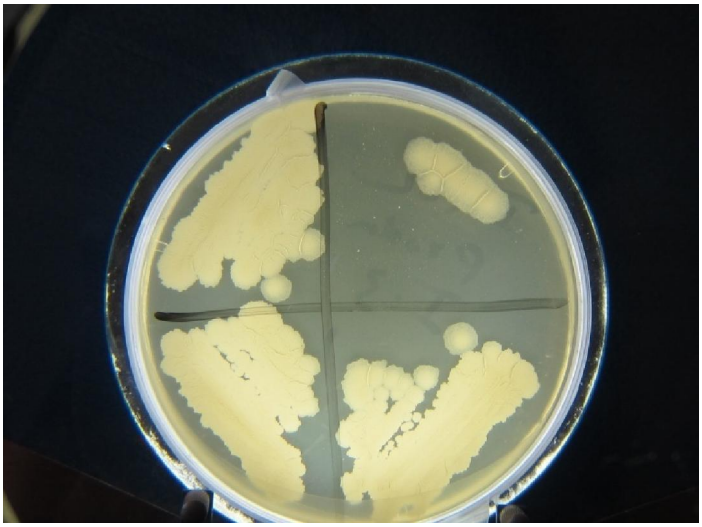

Fig. (1): Some purified isolates of *Bacillus subtilis*.

In the present studyو we determined the 16S rRNA gene sequence of 26 isolates form Makka road near Jeddah site around roots of *R. stricta*. (Table 2). Using BLAST search, we found that all strains belonged to species *Bacillus subtilis*. The identities of the 26 soil Bacillus isolates were determined by comparing them to the available 16S rRNA sequences found in Genbank and with high-scored rRNA sequences in BLAST searches. BLAST similarity scores ranged between 97% to 100%, where 42 matches of 16S rRNA gene sequence were included in the analysis. That is, the *in silico* analysis comprised a total of 68 sequences of 16S rRNA gene (26 Saudi soil isolates and 42 Genbank sequences). These 68 sequences along with their base pair length and GC content are shown in Table (2). The gene size for the 26 isolates ranged between 995 to 1233 nt, where that for the Genbank counterparts ranged between 1153-1559 nt. The percentage for GC content for all 68 16S rRNA gene sequences ranged between 55-56%. The alignment of 68 sequences with region of 1828 nt and containing gaps was carried out using Clastal W. The results of the DnaSp analysis indicated that the selected region (1828) of the 68 sequences from different species have 922 sites excluding sites with gaps (905). There are 702 invariable sites and 220 variable sites include 184 singletone variable site and 36 parsimony informative sites. The nucleotide diversity (π=0.01) and the average number of nucleotide differences (K=9.9) are shown in Table (2). Polymorphic information and haplotype diversity, and stochastic variance of k is 20.6 of *Bacillus* species for informative species are presented in Table (3).

Table (1): Primers of 16S rDNA gene used in the study.

	Primer Name	Type	Type2	Sequence (5 to 3)
1	518F	Universal	Forward	CCAgCAgCCgCggTAATACg
2	800R	Universal	Forward	TACCAgggTATCTAATCC

Table (2): Genetic diversity for 16S of *Bacillus subtilis* gene.

Diversity Parameters			
H	Hd	K	π
29	0.66	9.9	0.011

h, number of haplotypes; Hd, haplotype diversity; K, average number of nucleotide differences; π, nucleotide diversity

Table (3): Length of the 68 *B. subtilis* sequences and their GC contents.

Name/ Acc. number	Bp Length	GC Content	Name/ Acc. number	bp Length	GC Content
isolate-4	991	0.55	AY913755	497	0.55
isolate-14	983	0.55	JN987182	1501	0.55
isolate-17	985	0.55	JQ229696	1496	0.55
isolate-19	984	0.54	HE582781	1538	0.55
isolate-24	994	0.55	HE590857	1461	0.55
isolate-27	1233	0.54	X094283	J1500	0.55
isolate-33	995	0.56	HE681728	1490	0.55
isolate-49	1062	0.55	JQ361066	1496	0.55
isolate-53	967	0.55	JQ308588	1486	0.55
isolate-54	1007	0.55	KC519431	1455	0.55
isolate-55	990	0.55	KC506778	1511	0.55
isolate-57	969	0.55	KC443073	1522	0.54
isolate-61	997	0.55	KC443093	1502	0.54
isolate-62	985	0.55	KC179631	1466	0.54
isolate-64	970	0.55	JX997920	1473	0.55
isolate-69	994	0.55	JX845577	1496	0.55
isolate-70	1027	0.54	JX489167	1516	0.55
isolate-73	1043	0.54	JX177674	1458	0.55
isolate-76	967	0.55	JQ396173	1543	0.55
isolate-79	1056	0.56	JQ437542	1153	0.54
isolate-86	957	0.55	JQ308589	1498	0.56
isolate-88	969	0.55	JQ308567	1551	0.55
isolate-91	991	0.55	JQ308562	1491	0.55
isolate-95	963	0.55	JQ403532	1508	0.55
isolate-96	992	0.56	HE681738	1365	0.55
isolates-110	955	0.55	JN165753	1510	0.55
KC434971	1489	0.55	JF932296	1486	0.55
JX845578	1490	0.55	JQ229687	1517	0.55
JX402129	1559	0.55	JN366797	1456	0.55
JX852576	1487	0.55	JN366770	1447	0.55
KC428664	1466	0.55	JN587510	1513	0.55
KC310823	1451	0.55	JQ435698	1546	0.55
AB773829	1440	0.55	JQ361061	1496	0.55
KC443103	1280	0.55	JQ361055	1495	0.55

Using DnaSp, conserved regions were detected among the 68 sequences. A total of five genomic conserved sequences were detected along the 1828 nt (Table 4), with the smallest being 106 nucleotides size and the largest 437 nucleotides size. The range size (±standard deviation) of these five conserved sequences is 251.8±142 nucleotides. These conserved regions were distributed between first and 1025 nucleotides, with no conserved regions were detected in the last 800 nucleotides. It is also worth to mention that two regions were overlapped. However, this average did not take into account variations in length of individual sequences.

Table (4): Conserved regions of the 16S rDNA.

Region	Start-End	Nucleotides Length	Homozygosity	P value
1	1-437	437	0.99	< 0.001
2	449-555	107	0.99	0.003
3	496-860	465	0.99	0.02
4	661-860	200	0.99	0.001
5	868-1025	158	0.99	< 0.001

Molecular phylogenetic analysis

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993). The tree with the highest log likelihood (-2747) is shown. Bootstrapping (1000 replicates) was used to investigate support for groups in maximum likelihood analyses. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 68 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 921 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al*., 2011). Apart from isolates 49, 69, 70 and 79 the rest of the 22 isolates clustered together with Genbank sequences with varying degrees of bootstrap (Fig. 2). In similar context Earl *et al*. (2007) reported that analysis using 16S rRNA gene sequences from each strain failed to distinguish the *B. subtilis* subspecies or *B. vallismortis* as phylogenetically distinct taxa due to the limited number of informative sites at these loci. Moreover, other researchers (Freitas *et al*., 2008; Lima-Bittencourt *et al*, 2007; Pontes *et al*., 2007) reported that analysis of 16S rRNA gene sequences alone is not sufficient to identify *Bacillus* species. Identification of the bacteria using the API CHB50 system generally identified isolates to the genus level, and in some cases to the species level, but the identifications obtained using the API system and by 16S rRNA sequencing differed.

It is time for the study of bacterial ecology and evolution based on the recent rapid advances in genome sequencing platforms, such as 454 from Roche (www.454.com), Solexa from Illumina (www.illumina.com) and SOLiD from ABI (www3.appliedbiosystems.com). The taxonomic limitation can presently be overcome when there is a closely-related genome sequence available. The short sequence reads from a new strain can be readily assembled using the genome from a close relative as a scaffold. Given the distribution of genome sequences that will soon be available among *B. subtilis* and its close relatives, the *Bacillus* community is poised to take advantage of these faster and cheaper methods for whole genome sequencing (Earl *et al*., 2008).

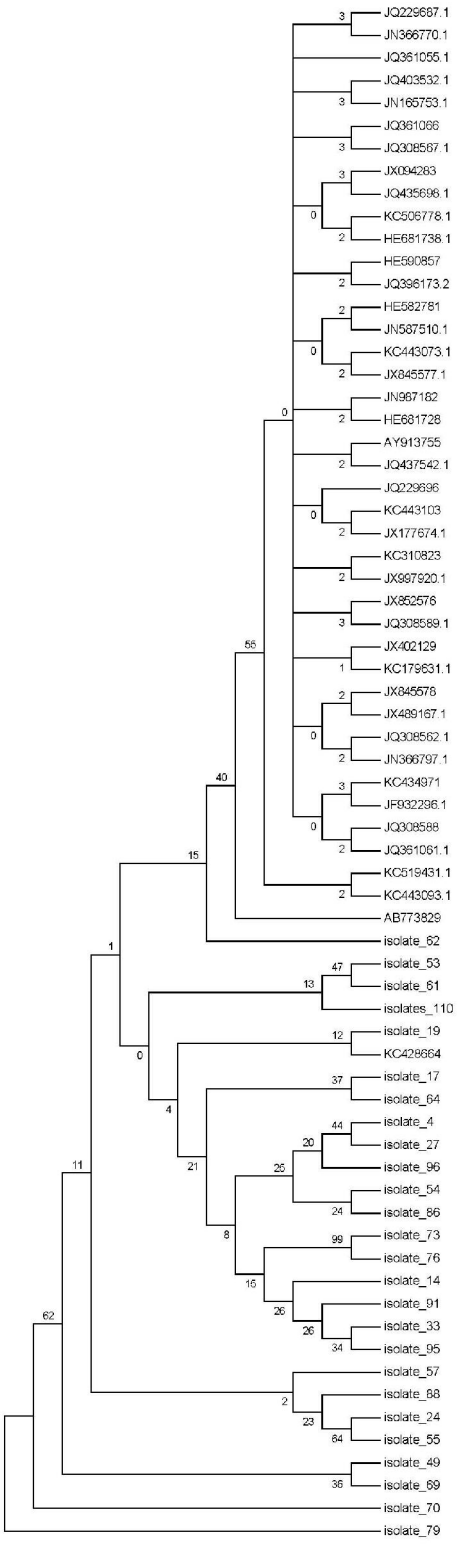

Fig. (2): Molecular Phylogenetic analysis by Maximum Likelihood method

Acknowledgements

The authors gratefully acknowledge the financial support from the KAU Vice President for Educational Affairs Prof. **PROF. DR. ABDULRAHMAN O. ALYOUBI** and the Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, represented by the Unit of Strategic Technologies Research through the Project number 431/008-D for the Project entitled: "Environmental Meta-Genomics and Biotechnology of *Rhazya stricta* and its Associated Microbiota".

References

Cazorla, F.M., Romero, D., Pe´ rez-Garcı´a. A., Lugtenberg, B.J.J., de-Vicente, A. and Bloemberg, G. (2007) Isolation and characterization of antagonistic *Bacillus subtilis* strains from the avocado rhizoplane displaying biocontrol activity. J. Appl. Microbiol. 103: 1950–1959

Claus D and Berkeley CW. (1986) The genus *Bacillus*. In: Bergey's Manual of Systematic Bacteriology. Vol 2. Sneath PHA (Ed). Williams, Wilkins, Baltimore. 34: 1105-1139.

Dong, Xu and Jean-Charles, Coˆte (2003) Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 39 end 16S rDNA and 59 end 16S–23S ITS nucleotide sequences. International Journal of Systematic and Evolutionary Microbiology. 53: 695–704.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., & Gayral, J-P.and Raoult, D. (2000)) 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentiable bacterial isolates. J. Clin. Microbiol. 38: 3623-3630..

Earl, A. M., Losick R and Kolter R (2008) Ecology and genomics of *Bacillus subtilis* Trends in Microbiology. 16 (6): 269-275.

Earl, AM., Losick, R., and Kolter, R. (2007) *Bacillus subtilis* Genome Diversity. Journal of Bacteriology. 189: 1163-1170.

Fall, R., Kinsinger, R.F. and Wheeler K.A. (2004) A simple method to isolate biofilm-forming *Bacillus subtilis* and related species from plant roots. Syst. Appl. Microbiol. 27: 372–379

Freitas, D.B., Mariana P Reis, M.P., Lima-Bittencourt, C.I.L., Costa, P.S., Assis, P.S., Chartone Souza, E., and Nascimento, AMA. (2008) Genotypic and phenotypic diversity of *Bacillus spp*. isolated from steel plant waste. BMC Research Notes. 1:92-103.

Goto, K., Omura, T., Hara, Y. and Sadaie, Y (2000) Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. J. Gen. Appl. Microbiol. 46: 1–8.

Joung, K.B. and Coˆte, J.C (2002) Evaluation of ribosomal RNA gene restriction patterns for the classification of *Bacillus* species and related genera. J. Appl. Microbiol. 92: 97–108.

Janssen P.H (2006) Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. Applied and Environmental microbiology. 1719–1728

Lima-Bittencourt, C.I., Astolfi-Filho, S., Chartone-Souza, E., Santos, F.R., and Nascimento, A.M.A. (2007) Analysis of *Chromobacterium sp*. natural isolates from different Brazilian ecosystems. BMC Microbiol. 7: 1-9.

Luria, S.E., and J.W. Burrous. (1995) Hybridization between *Escherichia coli* and *Shigella*. J. Bacteriol. **74**: 461-476.

Lynch, M. and Crease. T.J. (1990) The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. 7: 377-394.

Mignard, S., and Flandrois, J. P. (2006) 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. J. Microbiol. Methods. 67: 574-581.

Nagórska, K., Bikowski, M., Obuchowski, M. (2007) Multicellular behaviour and production of a wide variety of toxic substances support usage of *Bacillus subtilis* as a powerful biocontrol agent. Acta Biochim. Pol. 54: 495–508

Pages, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2013) Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.28.0.

Patel, J. B. (2001) 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. Mol. Diagn, 6: 313-321.

Pontes, DS., Lima-Bittencourt, CI., Chartone-Souza, E., and Nascimento, AMA. (2007) Molecular approaches: advantages and artifacts in assessing bacterial diversity. J. Ind. Microbiol. Biotechnol. 34:463-473.

R Core Team. (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rappe´, M. S., and S. J. Giovannoni. (2003) The uncultured microbial majority. Annu. Rev. Microbiol. 57: 369–394.

Salman V.; Amann R.; Shub D.A. and Schulz-Vogt H.N.(2012) Multiple self-splicing introns in

the 16S rRNA genes of giant sulfur bacteria. PNAS 109 (11) 4203–4208

Slepecky, R.A and Hemphill, H.E (1992) The Genus *Bacillus* Nonmedical in "The Prokaryotes". Vol II. Balows A, Trüper HG, Drowkin M, Tno WH, Schleifer KH (Ed). Springer-Verlag, New-York, Berlin, Heidelberg, London, Paris, Tokyo, Hong-Kong, Barcelona, Budapest. 48: 1697-1745.

Tajima, F. (1983) Evolutionary relationship of DNA sequences infinite populations. Genetics, 105: 437-460.

Tamura, K., and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 10: 512-536.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology Evolution, 24(8), 1596-1599.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28(10): 2731-2739.

Woo, P. C. Y., Ng, K. H. I., Lau, S. K. P., Yip, K.-T, Fung, A. M. Y., Leung, K.-W., Tam, D. M. W., Que, T.-L., & K.-Y., Yuen (2003) Usefulness of the MicroSeq 500 16S ribosomal DNA-based identification system for identification of clinically significant bacterial isolates with ambiguous biochemical profiles. J. Clin. Microbiol. 41: 1996-2001.

12/11/2013