Kasetsart J. (Nat. Sci.) 42 : 693 - 700 (2008)

In silico PCR-RFLP of *Bacillus* Species: Problem-Based Case of Teaching Bioinformatics

Ekachai Chukeatirote*, Komsit Wisitrasamewong and Jedsadapa Jongmahasavat

ABSTRACT

Bioinformatics is a multi-disciplinary subject that encompasses a wide range of fields including; genomics, biotechnology, information technology, algorithms and statistics. As a result, it is important to establish a skeletal set of courses capable of providing both a general and a detailed background on all aspects of bioinformatics. In this study, a problem-based project entitled "*in silico* PCR-RFLP (Polymerase Chain Reaction-Restriction Fragment Length Polymorphism) of *Bacillus* species" was assigned to students. The project aimed to cover the background of genetic information retrieval, multiple sequence alignment, phylogenetic analysis and *in silico* PCR-RFLP. Initially, the 16S rRNA genes of 34 *Bacillus* species were retrieved from the NCBI database. The sequence data obtained were then analyzed in terms of their similarities and variation using the ClustalW software. Subsequently, the phylogenetic tree was constructed to investigate their evolutionary relationship. The results indicated that these *Bacillus* species could be grouped into different clades. In addition, *in silico* PCR-RFLP was performed to introduce students to the principles of molecular taxonomy.

Key words: *in silico* PCR-RFLP, *Bacillus*, bioinformatics

INTRODUCTION

Bioinformatics, one of the fastest growing interdisciplinary sciences, has been established to manage and interpret biological data that has been generated in large quantities, particularly from a number of genomic sequencing projects. To date, the field of bioinformatics has been diverse and has covered a wide range of topics including; genomics, biotechnology, information technology, algorithms and statistics. Furthermore, both teaching and learning about bioinformatics can be even more complicated due to the range topics. These include *inter alia*: comparative and functional genomics; phylogenetic/phylogenomic analysis; biomolecular structure and its prediction; molecular modeling and thermodynamics; and structural genomics. Consequently, an effective strategy must be established to provide students with a clear understanding of one or more of the suitable options that could be pursued in this field. For example, students who are interested in the theme of evolution could then select coursework related to genomics and evolutionary biology. Students with a background in computing science, would initially be trained in basic (molecular) biology with an expectation that they could utilize their knowledge of and experience in information technology to analyze and/or create new algorithms to interpret the biological data.

School of Science, Mae Fah Luang University, Chiang Rai 57100, Thailand.
* Corresponding author, e-mail: ekachai@mfu.ac.th

Received date: 02/01/08 Accepted date: 12/06/08

The genus *Bacillus* is comprised of a large collection of aerobic or facultatively-anaerobic, rod-shaped, endospore-forming bacteria. They are widely distributed in nature and several species have a major impact on human beings. Some, such as *B. subtilis*, *B. stearothermophilus* and *B. amyloliquefaciens,* are useful in the food industry for enzyme production, whereas others, such as *B. anthracis* and *B. cereus* are harmful pathogens. Traditionally, the identification of *Bacillus* species has been mainly based on morphological and biochemical characteristics. Although these techniques remain useful, the processes involved are laborious and time-consuming. Over the past decade, the use of the 16S rRNA gene (rDNA) previously proposed by Woese *et al.* (1990) has been widely acknowledged as an alternative means of identifying bacteria. With the availability of the rRNA database and various kinds of useful computing software, a term assignment named "*in silico* PCR-RFLP (Polymerase Chain Reaction-Restriction Fragment Length Polymorphism) of *Bacillus* species" was used to introduce students to bioinformatics in the field of molecular taxonomy.

MATERIALS AND METHODS

Students were assigned to gather the names of *Bacillus* species from Bergey's Manual of Systematic Bacteriology (Sneath, 1986) or from the DSMZ catalogue (DSMZ, 2007). The 16S rRNA sequences of these representatives were then retrieved from the GenBank database (Benson *et al.*, 2005). The data of the 16S rRNA sequences were then analyzed for similarities using the ClustalW software (Higgins *et al.*, 1994). By this means, the phylogenetic tree used to classify the evolutionary relationships was also constructed.

Further tasks were then performed by relating the data obtained to the principles of molecular taxonomy used in identifying bacteria and in analyzing their phylogenetic relationships. Initially, the probes specific to different taxa (i.e., group, genus or species) could be designed based on multiple sequence alignment results. The Probe Match software available from the Ribosomal Database Project (RDP-II) was used to confirm the probe specificity (Cole *et al.*, 2005). To perform *in silico* PCR, the sequences at the 5'- and 3'-ends of the 16S rRNA sequences were screened for identical areas in which they could be used as universal primers of the bacterial groups. The selected sequences that referred to the 'amplified' fragments from the PCR were subjected to a series of restriction endonucleases using the Webcutter 2.0 software (Heiman, 1997). The RFLP pattern was then constructed using the Microsoft Excel© software package.

RESULTS AND DISCUSSION

As bioinformatics represents an array of themes such as genomics, phylogenetic analysis, and molecular modeling, teaching such a course is an interesting and challenging task. This study considered a problem-based case that aimed to cover certain specialized areas including: biological database access and retrieval; sequence comparison and phylogenetic analysis. In addition, probe design and *in silico* PCR-RFLP were also introduced as examples of applications in molecular genetics. While there were many topics to deal with, because they were related, it was possible for them to be taught in the one course.

The Genus *Bacillus* was selected for the study, with students assigned to gather the names of *Bacillus* species as well as their 16S rRNA sequences, as the use of rRNA genes in determining the phylogenetic relationships among living organisms is now widely accepted, since: they are present in all living organisms; have conserved structures and functions; and they have been collated into a comprehensive database. Based on the nucleotide NCBI database, thirty four

Bacillus species were selected and their 16S rRNA gene sequences were retrieved by their accession number (Table 1). The 16S rRNA sequences of *Escherichia coli*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Staphylococcus aureus* were also compiled to use for outgroup analysis. These 16S rRNA sequences were then subjected to multiple sequence analysis in which the dendogram was

Table 1 List of *Bacillus* species used in the study.

	Strain	Accession number
B. acidocaldarius	DMS 446	AJ496806
B. alkalophilus	Domuvar	AJ277904
B. alvei	ATCC 6344	X57034
B. anthracis	Steme	AF176321
B. azotoformans	ATCC 29788	X60609
B. badius	ATCC 14574	X77790
B. brevis	-	M10111
B. cereus	-	AJ853737
B. circulans	WSBC 20061	Y13065
B. coagulans	T 154	AB116143
B. fastidiosus	DSM 91	X60615
B. firmus	-	AJ509007
B. globisporus	NCIM B 11434	X60644
B. insolitus	DSM 5	X60642
B. larvae	ATCC 9548	X60619
B. laterosporus	ATCC 6344	X57307
B. lentimorbus	ATCC 14707	AF071861
B. lentus	JCM 2511	D78315
B. licheniformis	MC 20	AB05506
B. macerans	ATCC 8244	X57306
B. macquariensis	ATCC 23464	X57305
B. marinus	ATCC 29841	AB021190
B. megaterium	NRRL B 21660	AY739901
B. mycoides	DSM 2048	X55061
B. pantothenticus	NCDO 1765	X60627
B. pasteurii	NCIMB 8841	X60631
B. polymyxa	DSM 36	X57307
B. popilliae	NRRL B 4081	AF071860
B. pumilus	8N-4	AY548949
B. schlegelii	DSN 2000	Z26924
B. sphaericus	S 33	AB116123
B. stearothermophilus	NCDO 1768	X60640
B. subtilis	-	AB177641
B. thuringiensis	NCIM B 9134	X55062
Staphylococcus aureus	ATCC 12600	L37597
Lactobacillus delbrueckii subsp. *bulgaricus*	LG1	AY735407
Escherichia coli	0157:H7	AY513502

also constructed (Figure 1). According to Figure 1, the selected *Bacillus* species could be divided into three different clades: B1, B2 and B3. The results obtained were then further analyzed to see if they were related to morphology and the biochemical data (Table 2). Although the interpretation of both data sets seemed to be in agreement, it should be noted that some information was not available for all species investigated.

To further investigate the principles and applications of molecular taxonomy in bacterial identification, the study then introduced an *in situ* hybridisation technique. By observing the similarities and the variation amongst the multiple-sequence-alignment results, specific probes at the species level were then designed and confirmed using Probe Match software. Some representatives of the specific probes are shown in Table 3. Additionally, the PCR-RFLP technique was also introduced to the class. This technique provided a good example to stress the importance of the 16S rRNA gene sequence in bacterial identification. Students were asked to locate the identical areas of the nucleotide sequences at the 5′- and the 3′-ends to be used as primers. This activity identified the sequence of the 5′-primer as GAGTTT GATCTGGCTC and that of the 3′-primer as ACGGGCGGTGTGTC. Consequently, the amplified fragments could be obtained and were subjected to several restriction endonucleases using the Webcutter software. A few examples of the restriction profiles performed are shown in Figure 2. The *in silico* results are in agreement with those previously reported (Ash *et al*., 1993; Goto *et al*., 2000).

The results obtained were then interpreted and students were prompted to carry out a literature search on *Bacillus* species identification. Like other microbes, the identification of *Bacillus* species had previously been performed mainly by morphological and

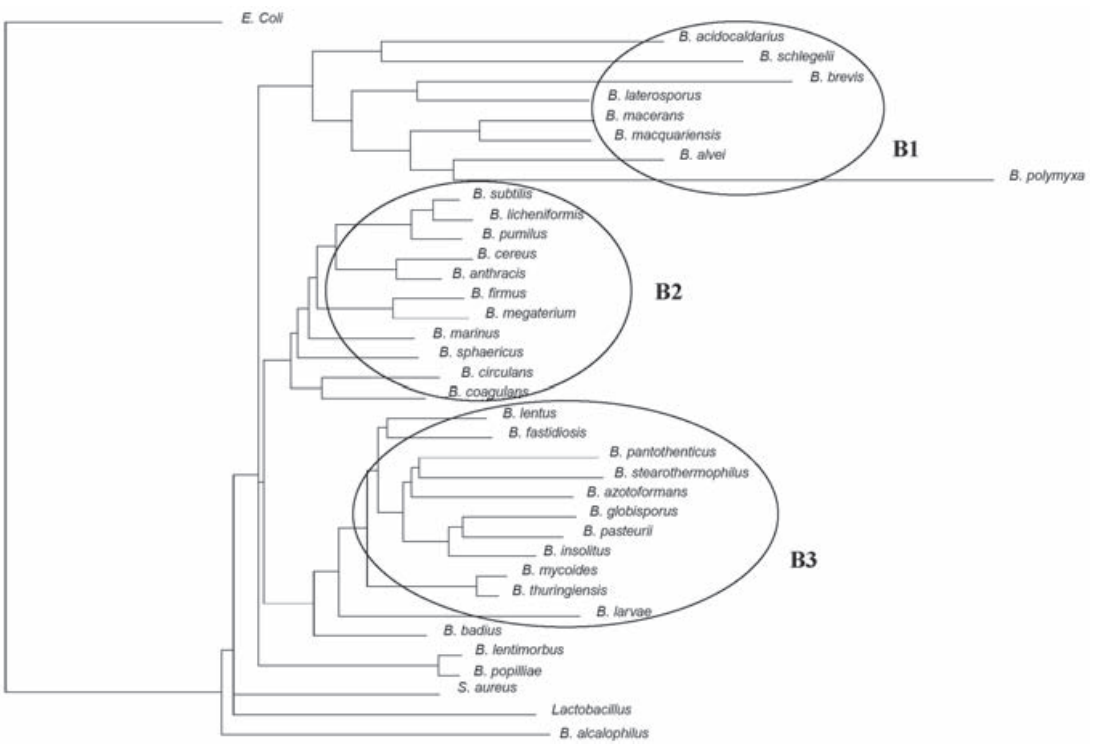


Figure 1 Phylogenetic trees of *Bacillus* species based on their 16S rRNA similarity.

biochemical analyses. However, it has been shown that these analyses are inconsistent and many microbes (especially *Bacillus* species) often yield different biochemical results depending on strains and variations (Sneath, 1986, Chukeatirote, personnel communication). With the advent of molecular taxonomy, several techniques such as RAPD (Random Amplification of Polymorphic DNA), RFLP (Restriction Fragment Length Polymorphism) and ARDRA (Amplified rDNA Restriction Analysis) have currently been established and used in microbial identification

Table 2 Some physiological and biochemical characteristics of the *Bacillus* species investigated in the study.

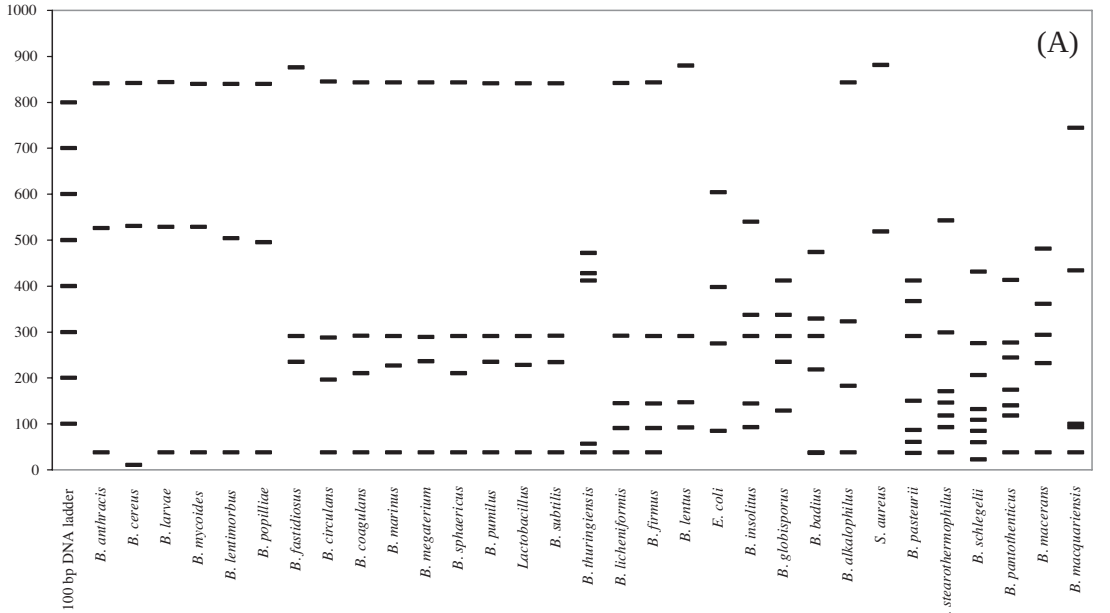
Characteristics	B1	B2	B3
% G-C	40 - 60	33 - 47	33 - 39
Anaerobic growth	+	na	na
Sporangium swollen	+	-	na
Utilization Citrate	-	+	na
Growth in NaCl			
2%	na	+	na
5%	-	+	na
7%	-	+	na
Growth at			
10 °C	-	+	na
50 °C	na	na	-
Acid from			
D-xylose	na	+	-
D- mannitol	na	na	-
Growth at pH 5.7	na	+	-
Degradation of tyrosine	na	na	-
Formation of dihydroxyacetone	na	na	-

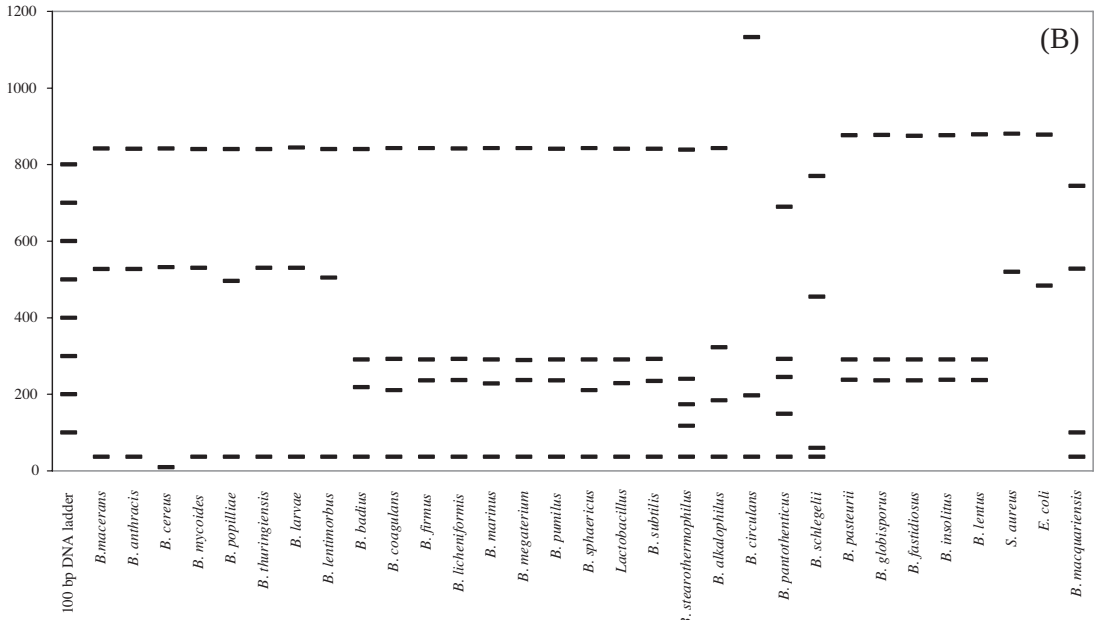
Note: +, positive; -, negative; na, the data are not available for all species.

Table 3 Representatives of specific probes of *Bacillus* species.

Probe name	Probe sequence (5′-3′)	Position No.[a]	Specificity Species	Remark
BC006	AAGCACTCATCAAGTCCG	23-40	+	*Bacillus alvei*
BC015	AACCTTTTAAAAGCAA	40-55	+	*Bacillus coagulans*
BC022	GCTCTCATCATTC	25-37	+	*Bacillus globisporus*
BC023	CTCTCATCATTCCGCTCG	19-36	+	*Bacillus globisporus*
BC039	ACCTTTTATGATTG	151-164	+	*Bacillus licheniformis*
BC049	ACGGTTACTCCA	1396-1407	+	*Bacillus megaterium*
BC056	CATCCGGGAGCAAGC TCCCTTCTGTCC	24-50	+	*Bacillus pumilus*
BC060	TGACCGGTCAGA	956-967	+	*Bacillus polymyxa*
BC063	TTGAGTCCGCTCGACTTGCA	12-31	+	*Bacillus popilliae*
BC070	GGCTTAACCTCGCGGTTTC	1216-1234	+	*Bacillus subtilis*

[a] Positions of nucleotides in accordance with those of the *Escherichia coli* 16S rRNA gene.


(A)
1000
900
800
700
600
500
400
300
200
100
0
100 bp DNA ladder
B. anthracis
B. cereus
B. larvae
B. mycoides
B. lentimorbus
B. popilliae
B. fastidiosus
B. circulans
B. coagulans
B. marinus
B. megaterium
B. sphaericus
B. pumilus
Lactobacillus
B. subtilis
B. thuringiensis
B. licheniformis
B. firmus
B. lentus
E. coli
B. insolitus
B. globisporus
B. badius
B. alkalophilus
S. aureus
B. pasteurii
B. stearothermophilus
B. schlegelii
B. pantotheniticus
B. macerans
B. macquariensis


(B)
1200
1000
800
600
400
200
0
100 bp DNA ladder
B.macerans
B. anthracis
B. cereus
B. mycoides
B. popilliae
B. thuringiensis
B. larvae
B. lentimorbus
B. badius
B. coagulans
B. firmus
B. licheniformis
B. marinus
B. megaterium
B. pumilus
B. sphaericus
Lactobacillus
B. subtilis
B. stearothermophilus
B. alkalophilus
B. circulans
B. pantotheniticus
B. schlegelii
B. pasteurii
B. globisporus
B. fastidiosus
B. insolitus
B. lentus
S. aureus
E. coli
B. macquariensis

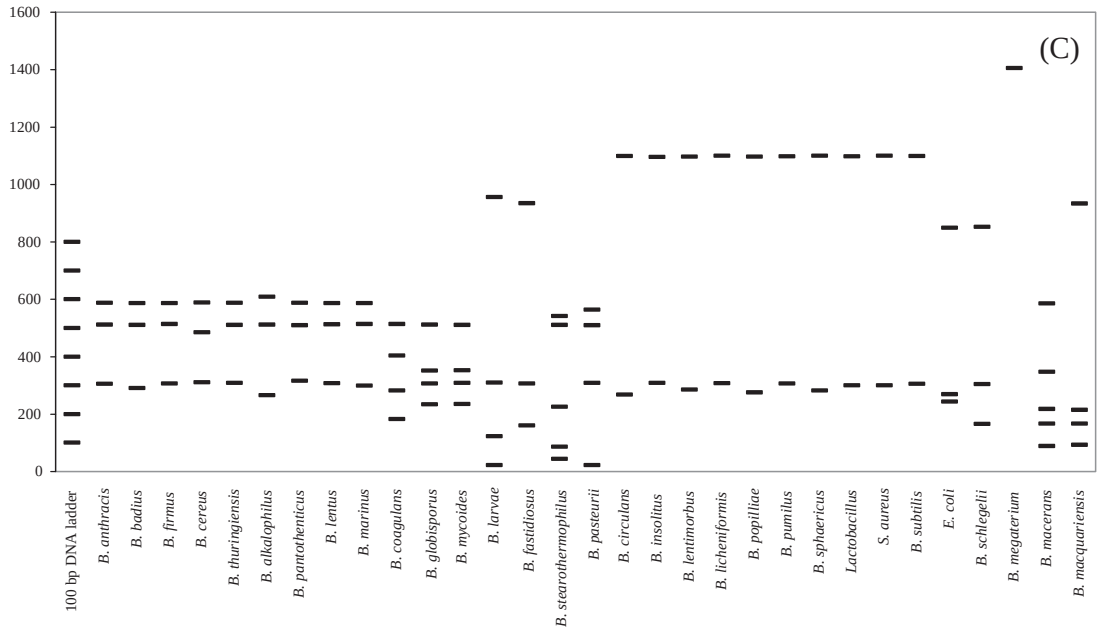


Figure 2 Representatives of the *in silico* analysis of *Bacillus* 16S rRNA genes using *Nsp*BII (A), *Sac*II (B) and *Cfr*I (C).

and characterization. The advantages and disadvantages between the two approaches were discussed in the class. For example, traditional biochemical tests could be carried out easily in any laboratory, but the technique is laborious (due to the many media used and the large number of biochemical reactions) and time-consuming. In addition, variations in the biochemical profiles within bacterial strains may lead to misidentification. In contrast, the molecular techniques offered a rapid means and the results were derived from the genetic information which was not affected by environmental factors such as culture conditions. This project was designed to familiarize students with the basic concepts of modern phylogenetic analysis. It should be noted that, apart from *Bacillus* species, other bacterial groups (i.e., lactic acid bacteria and Actinomycetes) could also be used in future work.

CONCLUSION

This present study described a problem-based case to introduce students to the field of bioinformatics. The *in silico* PCR-RFLP appeared to be a good example in which students could clearly appreciate the concepts behind the series of key molecular genetic techniques used. It is suggested that educators in this field would find students gained a better understanding from stimulating discussion as well as from the hands-on practical aspects.

LITERATURE CITED

Ash, C., F. G. Priest and M. D. Collins. 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. **Antonie van Leeuwenhoek** 64: 253-260.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and D. L. Wheeler. 2005. GenBank: Update. **Nucleic Acids Res.** 33: D34–D38.

Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity and J. M. Tiedje. 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. **Nucleic Acids Res.** 33: D294-D296.

DSMZ. 2007. Bacterial nomenclature up-to-date. Available at http://www.dsmz.de/microorganisms/bacterial_nomenclature_info.php?genus=BACILLUS (Last accessed 17 September 2007).

Goto, K., T. Omura, Y. Hara and Y. Sadaie. 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.

Heiman, M. 1997. Webcutter 2.0. Available at http://rna.lundberg.gu.se/cutter2 (Last accessed 17 September 2007).

Higgins, D., J. Thompson, T. Gibson, J. D. Thompson, D. G. Higgins and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22: 4673-4680.

Sneath, P. H. A. 1986. Section 13. Endospore-forming Gram-positive rods and cocci, pp. 1104-1139. *In* P. H. A. Sneath, N. S. Mair, M. E. Sharpe and J. G. Holt (eds.). **Bergey's Manual of Systematic Bacteriology Vol 2, 9th ed**. Williams & Wilkins, Baltimore, MD., USA.

Woese, C. R., O. Kandler and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. **Proc. Natl. Acad. Sci. USA** 87: 4576–4579.