

Identification of *Yersinia pestis* as the causative organism of plague in India as determined by 16S rDNA sequencing and RAPD-based genomic fingerprinting

S. Shivaji *, N. Vijaya Bhanu, R.K. Aggarwal

Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India

Received 17 April 2000; received in revised form 22 June 2000; accepted 25 June 2000

Abstract

Eighteen isolates of bacteria obtained from the sputum of pneumonic plague patients and from the liver and spleen of rodents from the plague-affected areas of India during 1994–1995 when analyzed by 16S rDNA analysis clearly demonstrated that all 18 isolates exhibit an average similarity of 98.5% with the genus *Yersinia* and 99.1% with *Yersinia pestis*, thus identifying the isolates as *Y. pestis*. The isolates from the human plague patients were found to be genetically more homogeneous compared to the isolates from the rodents which were more heterogeneous. An epidemiological linkage among the rodents and human patients is also indicated by 16S rDNA analysis, which suggests that only a sub-population of the rodents was probably the source of the infectious pathogen to the humans initiating the outbreak of the epidemic. The results of the randomly amplified DNA polymorphisms (RAPD)-based DNA fingerprinting are in agreement with the above conclusions. © 2000 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: *Yersinia pestis*; Plague; 16S rDNA; Randomly amplified DNA polymorphism; Genomic fingerprinting; Phylogeny

1. Introduction

Plague is an acute zoonotic disease caused by infection with *Yersinia pestis*. In a recent outbreak of plague in India in the states of Maharashtra and Gujarat between August and October, 1994, *Y. pestis* was suggested to be the causative organism [1]. The isolated cultures were bacteriologically characterized as *Y. pestis* [2,3] and 18 of them were studied in detail. All the isolates were found to be positive for the ‘*pla*’ and ‘*ft*’ genes and had plasmids of three sizes viz., 9.5, 70 and 110 kb [1], similar to those reported earlier for *Y. pestis* [4–6]. In contrast to the above observations which suggested the plague outbreak in India was due to *Y. pestis*, the ribotype patterns of the isolates [1] did not match with any of the 16 ribotypes reported earlier for *Y. pestis* [7]. The latter observation necessitated further studies for validation of the Indian isolates as *Y. pestis*. In order to address this issue, in the present study we amplified and sequenced the 16S rDNA from all 18 Indian isolates to further confirm their species status and

ascertain their phylogenetic relationship with other related microorganisms. In addition, the isolates were also fingerprinted using the polymerase chain reaction (PCR)-based approach of RAPD (randomly amplified DNA polymorphisms) to evaluate the extent of genetic heterogeneity.

2. Materials and methods

2.1. Bacterial cultures and preparation of DNA

Pure cultures of bacteria were isolated and established from human patients and rodent samples collected from the affected areas of the plague outbreak in India, during 1994–1995 [2,3]. Eighteen such pure cultures, 11 from the sputum of patients (isolates 04, 08, 09 and 101–108), six from the liver and spleen of *Rattus rattus* (isolate 111 and 113–117) and one from the liver and spleen of *Tatera indica* (isolate 112) suspended in 70% ethanol, henceforth mentioned as Indian isolates, were obtained from the Defence Research and Development Establishment (DRDE, Gwalior, India) as part of a National effort in India to identify and characterize the causative bacterial isolates responsible for the 1994 plague epidemic [8]. High molec-

* Corresponding author. Tel.: +91 (40) 7172241;
Fax: +91 (40) 7171195; E-mail: shivas@cemb.ap.nic.in

ular mass DNA was prepared from each of the 18 isolates as described earlier [1].

2.2. PCR amplification and sequencing of 16S rRNA gene

The 16S rDNA of the Indian isolates, was amplified individually using a set of primers complementary to the conserved regions of both 5' and 3' ends of 16S rRNA gene [9]. The forward and reverse primers used were: 5'-GAGTTTGATCCTGGCTCAG-3' and 5'-ACGGCTACCTTGTTACGACTT-3', respectively. These primers have earlier been successfully used to amplify the 16S rRNA gene of strains representing all members of the genus *Yersinia* [10]. The PCR reaction was set up using 100 ng of genomic DNA, 10 pmol of both forward and reverse primers and 1 µl of standard 10× PCR buffer containing 20 mM MgCl₂, 200 µM dNTP mix and 0.5 U Taq DNA polymerase (Bangalore Genei Pvt. Ltd., Bangalore, India) in 10 µl reaction volume. The PCR amplification was carried out in glass capillary tubes using the hot air Rapid Thermocycler (Model 1002, Idaho Technologies, ID, USA), for 35 cycles after an initial denaturation for 15 s at 94°C. Each PCR cycle comprised three steps: 5 s at 94°C (denaturation), 10 s at 48°C (annealing) and 20 s at 72°C (extension). A final extension of 5 min was given at 72°C. After PCR amplification, the 1.5 kb amplified fragment was purified from 1% agarose gel and used directly for DNA sequencing using an ABI Prism 377 automated sequencer.

The PCR product was sequenced using the two primers given above and also the following set of five forward and two reverse primers [11]. The forward primers used were: TAACACATGCAAGTCGAACG (50–70); CTACGGGAGGCAGCAGTGGG (341–361); CAGCAGCCCGCGGTAATAC (518–536); AAACCTCAAAGGAATTGACGG (908–928) and CATGGCTGTCGTCAGCTCGT (1053–1073) and the reverse primers were GTATTACCGCGGCTGCTG (536–518) and ACGAGCTGACGACAGCCATG (1073–1053). The numbers given in parentheses refer to the nucleotide positions of the primers corresponding to those in the 16S rDNA of *Escherichia coli* [12]. Sequencing of the purified PCR product (~200 ng/reaction) was carried out using the Big Dye Terminator sequencing kit (Perkin Elmer, USA) or Thermo Sequenase Dye terminator cycle sequencing kit (Amersham, USA) in a total volume of 20 µl according to the manufacturer's instructions using thermocycler PE 9600.

2.3. Phylogenetic analysis based on 16S rDNA sequence analysis

The 16S rDNA sequences obtained for the 18 Indian isolates (EMBL Accession Numbers AJ232222 to AJ232239) and one unknown bacterial isolate (AJ232240) were aligned with 12 reference sequences from the EMBL database using the multiple sequence

alignment program CLUSTAL V [13]. The reference sequences viz., *Y. pestis*, Accession Number Z-75317; *Yersinia pseudotuberculosis*; Z-21939; *Yersinia enterocolitica*, X-68672; *Yersinia ruckeri* X-75275; *Yersinia rohdei*, X-75276; *Yersinia kristensenii*, X-75278; *Yersinia frederiksenii*, X-75273; *Yersinia mollaretti*, X-75280; *Rahnella aquatilis*, X-79937; *Hafnia alvei*, M-59155; *Yersinia intermedia*, X-75299 and *E. coli*, X-80725^T, were initially identified through GenBank searches by BLAST [14] as having similarity to the Indian isolates.

The aligned sequences were then manually checked for gaps. Finally, 1303 non-ambiguously aligned positions (corresponding to positions 125–1423 of the rDNA of *E. coli*, EMBL Accession Number X-80725) were analyzed by computation of pairwise genetic distances calculated with DNADIST using the Jukes–Cantor equation [15]. To obtain the confidence values for the rDNA sequence-based genetic affiliations, the original sequence data set was subjected to bootstrap analysis. The sequence data was re-sampled 100 times using SEQBOOT, and each of the re-sampled data set was then used to calculate distances using DNADIST. The multiple distance matrices were then used to reconstruct the distance trees/topologies showing the genetic relationship between the isolates and reference microorganisms, using Fitch and Margoliash's (FITCH) and UPGMA methods. In all cases, the input order of species added to the topology being constructed was randomized via the 'jumble' option. Majority-rule (50%) consensus trees were constructed for the topologies found by each method using CONSENSE. All of these analyses were done using the PHYLIP package version 3.5C [16].

2.4. RAPD analysis

RAPD-based genomic fingerprinting of the Indian isolates of *Y. pestis* was carried out using the following primers (Operon, USA): OPA-10, 5'-GTGATCGCAG-3'; OPB-1, 5'-GTTTCGCTCC-3'; OPB-10, 5'-CTGCTGGGAC-3'; OPR-1, 5'-TGCGGGTCT-3'; (inhouse synthesized) AP-3, 5'-AACGCGCAAC-3'; AP-4, 5'-CCG-CAGCCAA-3'; AP-5, 5'-CCCCTCAGCA-3' and AP-6, 5'-AAGAGCCCGT-3'; AP-7, 5'-GCGATCCCCA-3'; AP-8, 5'-GTGGATGCGA-3' and M-13, 5'-GAGGGTGGTGGCTCT-3'. The RAPD amplifications were carried out in glass capillary tubes as described above but using 50 ng of genomic DNA. After initial denaturation of DNA for 15 s at 94°C the amplification was carried out for 35 cycles. Each cycle consisted of denaturation for 5 s at 94°C, annealing for 10 s at 34°C and extension for 30 s at 72°C. A final extension cycle was performed for 5 min at 72°C after the completion of 35 cycles. Afterwards, the amplified products were resolved on a 1.5% agarose gel in 1× TBE buffer and visualized under UV light after staining with ethidium bromide and photographed. The RAPD profiles of different isolates were compared manually and

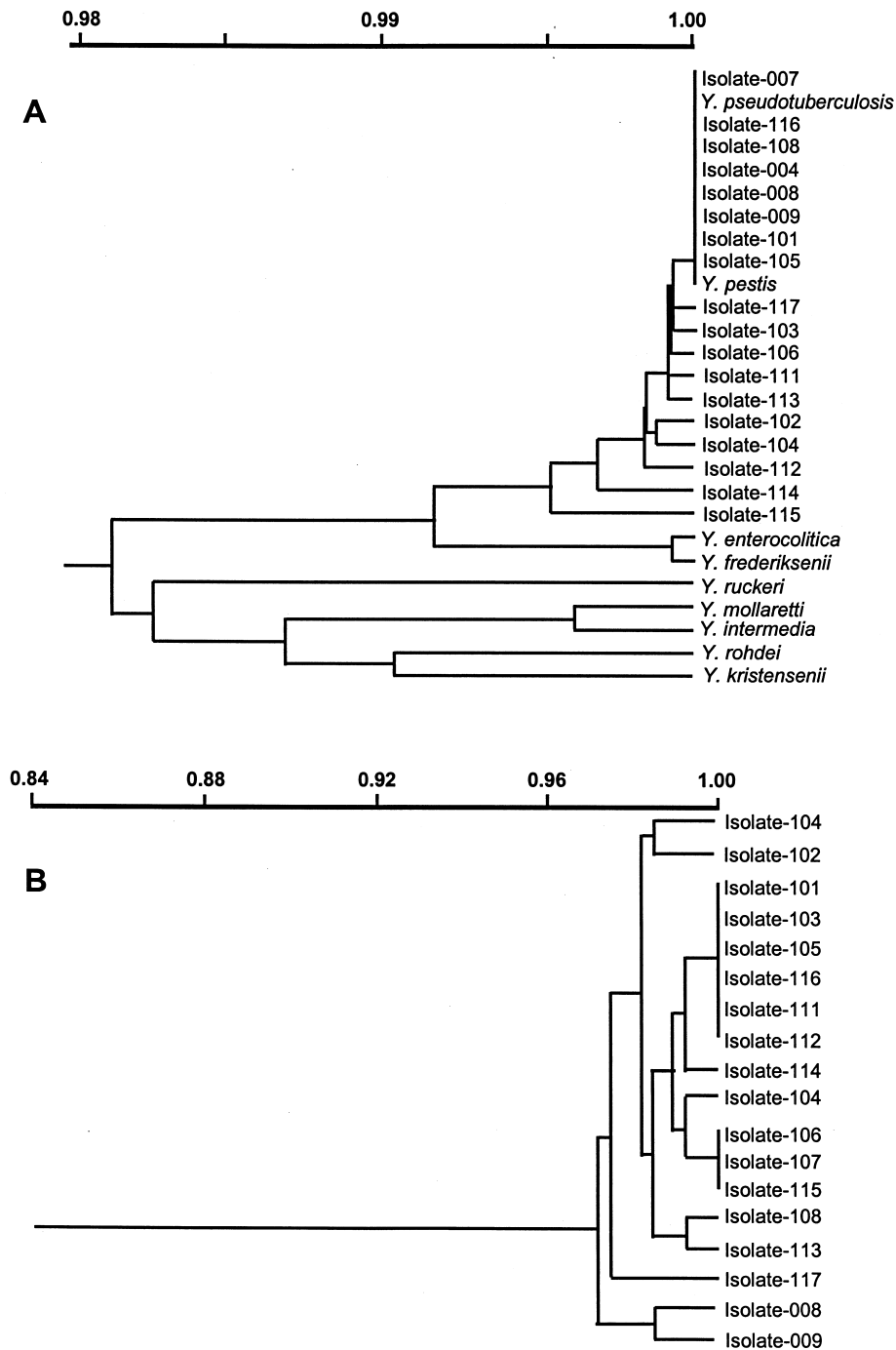


Fig. 1. A: UPGMA phenogram showing the phylogenetic relationship between Indian bacterial isolates (obtained from plague-affected human subjects and rodents) and reference microorganisms based on 16S rDNA sequence analysis. The % bootstrap values are given at the nodes to which they apply. Values below 50 and for the closely related strains are omitted. B: UPGMA phenogram showing the genetic relationship between the Indian isolates, based on the distance estimates from RAPD data.

the observed DNA polymorphisms were scored as dominant markers and converted to a binary matrix [17]. The data were used to derive similarity measures in terms of Dice's coefficient in all possible pairwise combinations. The similarity matrix thus obtained was used for cluster

analysis using the UPGMA method in NTSYS-PC 1.8 software package (F.J. Rohlf State University of New York, Stony Brook, NY, USA). The reliability of the phylogenetic tree was tested by bootstrapping using the software package Winboot [18].

3. Results

3.1. 16S rDNA sequence analysis of the Indian isolates

The PCR primers amplified a ~1.5 kb 16S rDNA fragment in all the 18 *Y. pestis* isolates and the control isolate, 1129. The amplified rDNA was directly sequenced, the sequences aligned and a total of 1303 nucleotide positions were analyzed by computation of genetic distances (data not shown), and then distance trees were generated to establish the genetic relationship between the isolates and the reference microorganisms.

The topologies of evolutionary trees generated by FITCH, UPGMA and CONSENSE methods using the 16S rDNA sequence data were almost similar. Fig. 1 is a representative UPGMA phenogram which reveals the genetic affiliations of all the Indian isolates (analyzed in the present study) between themselves and with the 12 reference microorganisms. The inferred phylogeny revealed that all the known *Yersinia* species form a phylogenetically coherent cluster. Notably, all the 18 Indian isolates appeared closely clustered with *Y. pestis* reference strain (Z-75317) with an average similarity of 99.91% (range being 99.38–100%) in all the phylogenetic trees (Fig. 1A). The confidence values determined through bootstrap analysis, for the cluster having Indian isolates and the standard *Y. pestis* was invariably very high (>97%). *Y. pseudotuberculosis* also appeared in the same cluster alongside the Indian isolates. The sequencing analysis also revealed low but significant nucleotide changes in 16S rDNA sequence of the Indian *Yersinia* isolates with respect to the reference *Yersinia* strain (Accession Number Z-75317).

The isolates purified from the liver and spleen tissues of six *R. rattus* and one *T. indica* rodents exhibited significantly more average sequence variability of 0.28% (ranging 0–0.62%) corresponding to 1–5 base differences, compared to the 11 isolates purified from the sputum of human patients that showed average variability of 0.07% corresponding to 0–2 base differences from the reference *Yersinia* strain (Table 1).

3.2. RAPD analysis

Out of the 11 primers tested, only five primers (OPA-10, OPB-1, OPB-10, AP-4 and M-13) generated RAPD profiles of all the 18 Indian isolates of *Y. pestis* and of isolate 1129 an unknown internal control. The RAPD profiles generated by the five random primers markedly differed between primer to primer, but for any given primer the inter-isolate variability was minimal except compared to the unknown control (data not shown). The total number of scorable bands generated by the five successful primers varied from a minimum of 11 bands (in case of OPB-10) to a maximum of 17 (OPB-1). UPGMA analysis based on the distance estimates from the RAPD data comprising of 68 well-resolved DNA fragments, revealed a robust tight cluster of all the Indian isolates, with a confidence/bootstrap value of 98% (Fig. 1B). Moreover, the RAPD-based genomic fingerprinting revealed that all the 18 Indian isolates were genetically very similar to each other (average 99% similarity) but uniquely distinct from the unknown control sample, which had a maximum of ~13% similarity with any given Indian isolate (data not shown).

Table 1

Nucleotide changes in the 16S rDNA sequence of Indian isolates of bacteria, with respect to the reference *Y. pestis* sequence obtained from EMBL database

Source	Nucleotide position ^a :	39	117	166	221	261	454	516	531	542	550	559	692	706	712	762	844	891	1143	1249	
Reference strain ^b		A	A	G	G	G	C	A	A	G	G	G	G	G	G	T	G	–	C	G	
Isolate 04 (Human)	
Isolate 08 (Human)	
Isolate 09 (Human)	
Isolate 101 (Human)	
Isolate 102 (Human)		.	.	T	C
Isolate 103 (Human)		A
Isolate 104 (Human)		G	.	.	C
Isolate 105 (Human)	
Isolate 106 (Human)		A
Isolate 107 (Human)	
Isolate 108 (Human)	
Isolate 111 (<i>R. rattus</i>)		A
Isolate 112 (<i>T. indica</i>)		G	.	.	.	T
Isolate 113 (<i>R. rattus</i>)		T	.
Isolate 114 (<i>R. rattus</i>)		T	.	T	A	T
Isolate 115 (<i>R. rattus</i>)		.	G	.	.	.	G	.	G	.	.	T	.	.	A	.	.	G	.	.	T
Isolate 116 (<i>R. rattus</i>)	
Isolate 117 (<i>R. rattus</i>)		C

Dots represent nucleotides identical to that present in the reference strain.

^aNucleotide position corresponds to the 16S rDNA sequence of *E. coli* (Accession Number X-80725T).

^b*Y. pestis* (Accession Number Z-75317).

4. Discussion

In recent years, 16S rDNA sequence analysis has been used extensively for the reconstruction of bacterial phylogeny. Earlier studies based on 16S rDNA analysis demonstrated that all the *Yersinia* species form a phylogenetically coherent cluster exhibiting sequence similarities of 96.9 to 99.8% within the genus [10,19]. In this background, it was thought prudent to use the 16S rDNA analysis to establish the species status of the Indian isolates from the plague epidemic as *Y. pestis*. Accordingly, in the present study, the 16S rDNA sequence analysis was carried out that clearly establishes the genetic affiliation of the Indian isolates to the genus *Yersinia* and species *Y. pestis* in specific. Notably, all the 18 Indian isolates appeared closely clustered with the reference microorganism *Y. pestis* (Accession Number Z-75317) with an average similarity of 99.91% and all other *Yersinia* species with an average similarity of 98.5%, in all the phylogenetic trees (see Fig. 1, which shows a UPGMA phenogram).

The phylogenetic analysis also revealed *Y. pseudotuberculosis*, closely clustered along with the 18 Indian isolates and the reference *Y. pestis*. This was as per expectation as the rRNA gene in *Y. pseudotuberculosis* is known to be almost identical (99.5%) to that of *Y. pestis* [10]. Recently, Achtman et al. [20] have also shown that *Y. pseudotuberculosis* shares many identical genes with bacteriologically distinct *Y. pestis*. Further, none of the Indian isolates carried the DNA signatures characterized by 23 nucleotides [19] corresponding to the stem loop of 'V3' region between positions 455–477 of 16S rDNA of *E. coli* [21] specific to atypical *Yersinia* isolated from the raw milk [19].

All these results of the present study thus validate our earlier observations (based on various bacteriological and genetic characteristics) [2,3] that the 18 Indian isolates in question are essentially *Y. pestis*. The results, in turn, would suggest that the Indian isolates define a new ribotype [1] that is not represented in the ones described earlier [7]. The later is very much possible as the previously described 16 ribotypes are based on the analysis of only 85 of the ~6000 isolates of *Y. pestis* known to exist in collections worldwide.

Further, though all the 18 isolates closely resembled one another, it was observed that the 11 isolates which originated from the human plague patients, were genetically more homogeneous with seven being identical (to each other) and the remaining four showing only one to two base differences to the *Y. pestis* reference strain. In contrast, the Indian isolates from the rodents *R. rattus* and *T. indica* (isolates 111–117) appeared to be more genetically heterogeneous as they showed about one to five base differences (Table 1). These results suggest that the Indian isolates may not be clonal in origin specially in the rodent populations, though the human isolates probably are clonal representing the sub-population of a few rodents which might have passed the pathogen to the humans in

the initial stages of plague outbreak. This is plausible as seven of the 11 human isolates are identical to only one rodent isolate no. 116 and the variation of 1–2 bases in remaining four isolates might be the result of relatively high genomic instability of *Y. pestis* under in vitro conditions as reported earlier by Guiyoule et al. [7]. The high level of genetic similarity among the 18 Indian isolates became also evident in the RAPD-based genomic fingerprinting analysis (Fig. 1B).

In conclusion, the present study establishes the identity of causative organism of plague in India as *Y. pestis* based on the rRNA gene sequence analysis. It also demonstrates that the 18 isolates are genetically very similar within themselves, both based on 16S rDNA analysis and RAPD DNA fingerprinting, and further suggests that these may define a new ribotype but may not be clonal in origin. The results also indicate an epidemiological connection between rodents and man in the epidemic zone. Also with the present study, the EMBL database entries on the rDNA sequences specific to *Y. pestis* isolates have gone up from 1 to 19 (EMBL Accession Number Z-75317, AJ232222 to AJ232239).

Acknowledgements

We would like to thank Prof V. Ramalingaswami, Chairman of the Technical Advisory Committee on Plague for encouragement, and Ministry of Health, Government of India for financial assistance to carry out this work. Dr N. Vijaya Bhanu was supported by the Department of Biotechnology, Government of India. Our special thanks are to Mr G. Seshu Kumar and Mr K. Kannan for their involvement during the initial stages of this work.

References

- [1] Panda, S.K., Nanda, S.K., Ghosh, A., Sharma, C., Shivaji, S., Seshu Kumar, G., Kannan, K., Batra, H.V., Tuteja, U., Ganguly, N.K., Chakravarty, A. and Sharat Chandra, H. (1996) The 1994 plague epidemic of India: Molecular diagnosis and characterization of *Yersinia pestis* isolates from Surat and Beed. *Curr. Sci.* 71, 794–799.
- [2] Batra, H.V., Tuteja, U. and Agarwal, G.S. (1996) Isolation and identification of *Yersinia pestis* responsible for the recent plague outbreaks in India. *Curr. Sci.* 71, 787–791.
- [3] Aggarwal, G.S. and Batra, H.V. (1996) Passive haemagglutination tests for x 1996. Passive haemagglutination tests for *Y. pestis* infection in Surat pneumonic patients. *Curr. Sci.* 71, 792–793.
- [4] Bengurion, R. and Shafferman, A. (1981) Essential virulence determinants of different *Yersinia* species are carried on a common plasmid. *Plasmid* 5, 183–187.
- [5] Ferber, D.M. and Brubaker, R.R. (1981) Plasmids on *Yersinia pestis*. *Infect. Immun.* 31, 839–841.
- [6] Portnoy, D.A. and Falkow, S. (1981) Virulence associated plasmids from *Yersinia enterocolitica* and *Yersinia pestis*. *J. Bacteriol.* 148, 877–883.
- [7] Guiyoule, A., Grimont, F., Itean, I., Grimont, P.A.D., Lefevre, M. and Carniel, E. (1994) Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. *J. Clin. Microbiol.* 32, 634–641.

- [8] Ramalingaswami, V. (1996) The plague outbreaks of India, 1994, A prologue. *Curr. Sci.* 71, 781–782.
- [9] Lane, D.J. (1991) 16S/23S rRNA sequencing. In: *Nucleic Acid Techniques in Bacterial Systematics* (Stackebrandt, E. and Goodfellow, M., Eds.), pp. 115–147. Wiley, New York.
- [10] Ibrahim, A., Goebel, B.M., Liesack, W., Griffiths, M. and Stackebrandt, E. (1993) The phylogeny of the genus *Yersinia* based on 16S rDNA sequences. *FEMS Microbiol. Lett.* 114, 173–178.
- [11] Edwards, U., Rogall, T., Blocker, H., Emde, M. and Bottger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* 17, 7842–7843.
- [12] Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) Detailed analysis of the higher order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* 47, 621–669.
- [13] Higgins, D.G., Bleasby, A.T. and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* 8, 189–191.
- [14] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- [15] Jukes, T.H. and Cantor, C.R. (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (Munro, H.N., Ed.), Vol. III, pp. 21–132. Academic Press, New York.
- [16] Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) Version 3.5c. Distributed free of cost by the author. Department of Genetics, University of Washington, Seattle, WA, USA.
- [17] Denny, T.P., Gilmour, M.N. and Selander, R.K. (1988) Genetic diversity and relationships of two pathovars of *Pseudomonas syringae*. *J. Gen. Microbiol.* 134, 1949–1960.
- [18] Yap, I. and Nelson, R.J. (1996) Winboot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRRI Discussion Papers Series No. 14. International Rice Research Institute, Manila, The Philippines.
- [19] Ibrahim, A., Liesack, W., Steigerwalt, A.G., Brenner, D.J., Stackebrandt, E. and Robinsbrowns, R.M. (1997) A cluster of atypical *Yersinia* strains with a distinctive 16S rRNA signature. *FEMS Microbiol. Lett.* 146, 73–78.
- [20] Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. and Carniel, E. (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* 24, 14043–14048.
- [21] Neefs, J.M., Van de peer, Y., Hendriks, L. and De Wachter, R. (1990) Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* 18, 2237–2317.