

Optimized genotyping method for identification of bacterial contaminants in pharmaceutical industry

BORCHE STAMATOSKI^{1*}
MIROSLAVA ILIEVSKA¹
HRISTINA BABUNOVSKA¹
NIKOLA SEKULOVSKI²
SASHO PANOV²

¹ *Quality Assurance
Profit Center Pharmaceuticals
Alkaloid AD, Skopje
Republic of Macedonia*

² *Laboratory for Molecular Biology
Institute of Biology
Faculty of Natural Sciences and
Mathematics
Ss. Cyril & Methodius University
Skopje, Republic of Macedonia*

Accepted October 13, 2015
Published online March 24, 2016

Microbiological control is of crucial importance in the pharmaceutical industry regarding the possible bacterial contamination of the environment, water, raw materials and finished products. Molecular identification of bacterial contaminants based on DNA sequencing of the hypervariable 16SrRNA gene has been introduced recently. The aim of this study is to investigate the suitability of gene sequencing using our selection of PCR primers and conditions for rapid and accurate bacterial identification in pharmaceutical industry quality control.

DNA was extracted from overnight incubated colonies from 10 bacterial ATCC strains, which are common contaminants in the pharmaceutical industry. A region of bacterial 16SrRNA gene was analyzed by bidirectional DNA sequencing. Bacterial identification based on partial sequencing of the 16SrRNA gene is the appropriate method that could be used in the pharmaceutical industry after adequate validations. We have successfully identified all tested bacteria with more than 99 % similarity to the already published sequences.

Keywords: 16SrRNA, genotyping, bacterial identification, pharmaceutical industry

Bacterial contamination of pharmaceutical products is one of the main reasons for their recall from the market (1). In the case of suspicious contamination, and in order to find its origin, identification of the bacterial species level is mandatory for sterile production failure spotting and is important for the production environment, raw materials, pharmaceutical water or the release of the finished non-sterile products onto the market.

Morphological and biochemical determination of an unknown microorganism is traditionally given through pharmacopoeial tests. Recently, different state-of-the-art methods for bacterial identification have emerged; they include instrumental complex biochemical tests, characterization of fatty acid composition, restriction endonuclease banding

* Correspondence; e-mail: bstamatoski@alkaloid.com.mk

patterns, matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and genotyping of the 16SrRNA sequence (2, 3).

Although all these methods have their advantages and disadvantages, DNA sequencing of the bacterial genome region, including the 16SrRNA gene, are evidently becoming the gold standard overcoming the majority of other identification techniques (4). These genotyping-based methods have a very favorable cost-benefit ratio, particularly considering the ongoing downward cost trend of DNA sequencing. Namely, the 16SrRNA gene is evolutionary conserved in bacterial genomes and very slight variation in the nucleotide sequence between different species is used for the identification named ribotyping. In addition, ribotyping is used for establishing the phylogenetic connections between the species (5).

Genotyping of the 16SrRNA gene involves polymerase chain reaction (PCR)-based amplification and automated DNA sequencing. Considering that the pair of oligonucleotide primers determines the exact amplification region, primer selection is crucial for the overall identification success (6). Furthermore, the primers could be selected to be species-specific, group-specific or universal amplifying almost any bacterial species.

The aim of this study was to investigate the suitability of partial 16SrRNA gene sequencing with our optimization of amplification with selected oligonucleotide primers for the purpose of rapid and accurate bacterial identification in the pharmaceutical industry.

EXPERIMENTAL

Strains preparation

A total of 10 bacteria (4 Gram-positive, 5 Gram-negative and one Gram-positive to Gram-variable) from the American Type Culture Collection (ATCC) were used in the study. The following strains were used: *Escherichia coli* ATCC 8739, *Pseudomonas aeruginosa* ATCC 9027, *Staphylococcus aureus* ATCC 6538, *Staphylococcus epidermidis* ATCC 12228, *Salmonella typhimurium* ATCC 14028, *Enterococcus faecalis* ATCC 29212, *Burkholderia cepacia* ATCC 25416, *Bacillus subtilis* subsp. *spizizenii* ATCC 6633, *Proteus mirabilis* ATCC 12453, *Kocuria rhizophila* ATCC 9341. These bacterial strains (KWIK-STIK™, Microbiologics, Microbiologics, Inc., Saint Cloud, USA), were inoculated on Trypticase soy agar (TSA, bioMérieux, France) and incubated overnight. A loop-full of colonies from the second passage, also incubated overnight, was used for DNA isolation.

DNA Isolation

DNA was extracted using the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich, USA) according to the manufacturer's instructions.

PCR amplification of 16SrRNA gene

For oligonucleotide primer pair selection, *in silico* analysis was performed using the nucleotide sequence data from the specialized ribosomal RNA Database, Silva, release 119 (7). Modified touchdown amplification was performed using 1 U AmpliTaq Gold Poly-

merase (Applied Biosystems, USA) per reaction, 2.5 mmol L⁻¹ MgCl₂, 200 μmol L⁻¹ dNTPs, 0.4 μmol L⁻¹ of each forward primer 5'-TGG AGA GTT TGA TCC TGG CTC AG-3' and reverse primer 5'-TAC CGC GGC TGC TGG CAC-3' on a GeneAmp System 2400 thermocycler (Perkin Elmer, USA). Those primers were previously described by Hall *et al.* (2), Gomez *et al.* (8), and Nawaz and Ahmed (3). After initial denaturation at 94 °C for 3 min, a total of 37 cycles were used, each performed at 94 °C for 1 min, 62 °C for 30 s, 72 °C for 3 min. (2 cycles) and then the annealing temperature was reduced for each consecutive cycle by 1 °C up to 50 °C. After completion, a final extension was performed at 72 °C for 5 min. The PCR products were purified using the standard sodium acetate/ethanol protocol (9).

Negative controls were included in each experiment to verify the absence of contaminant bacterial DNA in PCR reagent or consumables.

Sequencing

Cycle sequencing was performed using the BigDye cycle sequencing kit v3.1 (Applied Biosystems, USA) on an ABI Prism Genetic Analyzer 310 (Perkin Elmer) according to the manufacturer's instructions. All sequencing experiments were performed twice on two different occasions.

Sequence data analysis

Sequence readings were manually verified using the Sequence Scanner software 2 ver. 2.0 (Applied Biosystems, USA). For each bacterial strain, the 16SrRNA gene sequence was created from both sequencing data (5' and 3') using the BioEdit Sequence Alignment Editor ver. 7.2.5 (10) and was trimmed to a length of 500 base pairs (bp). The percent of both similarity and divergence between 16SrRNA gene sequence for each of the ten ATCC bacterial strains used in the study was calculated using the iterative type of pairwise alignment by Similarity analyses and Distance Matrix, option of BioEdit software v.7.2.5 (Tom Hall, Ibis Biosciences, Carlsbad, California, USA), respectively.

Identification of bacterial species was performed by comparing the obtained sequences with all available similar sequences in the GeneBank database using the nucleotide BLAST on-line analysis in the NCBI GeneBank database (11).

RESULTS AND DISCUSSION

We have isolated genomic DNA from each of the ten bacterial strains and amplified it using modified touch-down PCR producing amplicons with the size range between 500 and 550 bp (Fig. 1).

Using DNA sequencing of the amplicons with both forward and reverse primers and subsequent nucleotide BLAST analysis, we have unambiguously identified all ATCC isolates to the species and/or subspecies level. The BLAST's identity score characterizes the percent of alignment of identical nucleotides between the submitted (query) and the database DNA sequences. In our experiments, this score was greater than 99 % for all ten species, signifying the similarity of our 16SrRNA gene sequences to the sequences already published in the NCBI nucleotide database (Table I).

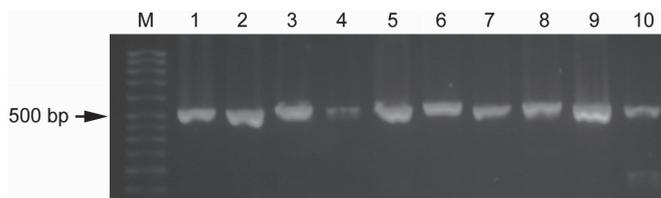


Fig. 1. Electrophoretic profile of PCR-amplified bacterial 16SrRNA gene from each isolate. The amplified DNA was resolved by agarose gel-electrophoresis.

M – molecular mass marker, 1 – *E. coli*, 2 – *P. aeruginosa*, 3 – *S. aureus*, 4 – *S. epidermidis*, 5 – *S. typhimurium*, 6 – *E. faecalis*, 7 – *B. cepacia*, 8 – *B. subtilis* sub. *spiz.*, 9 – *P. mirabilis*, 10 – *K. rhizophila*.

Table I. Results of the BLAST analysis of the nucleotide sequences of bacteria

| Bacterial species | Highest BLAST hit* | Identity score (%) | Max score ^a | Total score ^b | BLAST accession no. ^c |
|---|---|--------------------|------------------------|--------------------------|----------------------------------|
| <i>E. coli</i> ATCC 8739 | <i>Escherichia coli</i> strain ST540 | 99 | 917 | 9384 | CP007390.1 |
| <i>P. aeruginosa</i> | <i>Pseudomonas aeruginosa</i> strain CB3 | 99 | 898 | 898 | KJ854401.1 |
| <i>S. aureus</i> | <i>Staphylococcus aureus</i> subsp. <i>aureus</i> strain ATCC 25923 | 100 | 924 | 5546 | CP009361.1 |
| <i>S. epidermidis</i> | <i>Staphylococcus epidermidis</i> strain F2a | 100 | 924 | 924 | KM052202.1 |
| <i>Salmonella enterica</i> subsp. <i>enterica</i> | <i>Salmonella enterica</i> subsp. <i>enterica</i> | 99 | 904 | 6282 | CP008928.1 |
| <i>E. faecalis</i> | <i>Enterococcus faecalis</i> ATCC 29212 | 99 | 911 | 3646 | CP008816.1 |
| <i>B. cepacia</i> | <i>Burkholderia cepacia</i> strain ATCC 25416 | 100 | 924 | 924 | NR_114491.1 |
| <i>B. subtilis</i> subsp. <i>spizizenii</i> | <i>Bacillus subtilis</i> subsp. <i>spizizenii</i> strain NRS 231 | 100 | 902 | 7201 | CP010434.1 |
| <i>P. mirabilis</i> | <i>Proteus mirabilis</i> strain FBH-1 | 100 | 924 | 924 | KJ619638.1 |
| <i>K. rhizophila</i> | <i>Kocuria rhizophila</i> isolate LE 78 | 99 | 913 | 913 | FN908446.1 |

Only the most relevant hit is presented: the one with the highest identity score (%), max. score and total score.

^aScore of single best aligned sequence.

^bSum of scores of all aligned sequences. This score indicates that there is more than one part of the sequence that is aligned with the unknown sequence.

^cUnique number linked to the GeneBank record.

In order to evaluate further the informativeness of the obtained 16SrRNA gene sequences for ribotyping-based identification, similarity/divergence calculations were performed between the 10 bacterial species and presented in the diagonal matrix format (Table II).

It is obvious that the bacterial species belonging to the same genus have a higher percent of similarity: *S. aureus* and *S. epidermidis* from genus *Staphylococcus* (82.6 %), while the percent of divergence is very low (2.7 %). In addition, comparison of the values from different species (*E. coli* and *S. typhimurium*) of the same family (*Enterobacteriaceae*), reveals a very high percentage of similarity (91.2 %) and low divergence (4.1 %).

Table II. Percent of DNA sequence similarity and divergence between the ten ATCC strains in the 16SrRNA gene sequence

| Strain name | Sequence similarity (%) | | | | | | | | | |
|-------------------------------|-------------------------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| <i>E. coli</i> | | 83.2 | 72.3 | 62.5 | 91.2 | 69.7 | 78.3 | 70.9 | 88.4 | 62.1 |
| <i>P. aeruginosa</i> | 14.3 | | 72.9 | 65.3 | 78.1 | 68.8 | 81.4 | 73.1 | 83.4 | 62.7 |
| <i>S. aureus</i> | 27.0 | 22.7 | | 82.6 | 72.1 | 80.6 | 71.0 | 86.8 | 70.4 | 65.7 |
| <i>S. epidermidis</i> | 26.4 | 24.3 | 2.7 | | 60.0 | 67.9 | 62.8 | 74.8 | 63.6 | 71.6 |
| <i>S. typhimurium</i> | 4.1 | 15.6 | 23.9 | 27.3 | | 70.1 | 75.3 | 69.8 | 83.7 | 63.2 |
| <i>E. faecalis</i> | 24.9 | 24.9 | 16.5 | 17.7 | 25.4 | | 69.7 | 80.5 | 68.4 | 64.8 |
| <i>B. cepacia</i> | 21.4 | 17.9 | 26.4 | 29.6 | 19.5 | 25.7 | | 72.3 | 78.9 | 65.4 |
| <i>B. subtilis</i> sub. spiz. | 26.9 | 23.4 | 11.6 | 12.2 | 26.4 | 16.0 | 26.8 | | 70.7 | 66.9 |
| <i>P. mirabilis</i> | 8.5 | 14.5 | 25.3 | 28.2 | 9.8 | 26.3 | 22.7 | 28.8 | | 63.4 |
| <i>K. rhizophila</i> | 27.5 | 27.4 | 27.6 | 23.9 | 28.8 | 26.7 | 21.9 | 26.0 | 26.3 | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | Sequence divergence (%) | | | | | | | | | |

These results suggest that even in the phylogenetically close bacteria, there are sufficient divergent nucleotides to be used for species identification in our modification of 16SrRNA genotyping. According to the presented data, this is applicable for evolutionary distant organisms such as *K. rhizophila* from phylum *Actinobacteria* displaying a relatively high divergence (28.8 %) from *S. typhimurium* from phylum *Proteobacteria*.

Genotyping of bacterial 16SrRNA gene is becoming a mainstream in microbiological quality control in the pharmaceutical industry, surpassing the biochemistry-based and traditional cultivation methods for bacterial identification. Many ribotyping methods have been described for different purposes including, but not limited to, clinical, ecological and phylogenetical identification of bacteria.

We have evaluated the suitability of our modified ribotyping method for its potential use in pharmaceutical quality control. One of the most critical components for PCR amplification is the primer pair selection. Each of the two oligonucleotide primers should anneal to the highly conserved sites of the bacterial 16SrRNA gene flanking the hypervariable gene sequence. We have performed extensive *in silico* analysis of the bacterial 16SrRNA gene database searching for primer pairs that could anneal most efficiently yet amplifying most of the variable sequence, hence enabling identification of as many as possible bacterial species. However, the pair of oligonucleotide primers that meet those criteria most efficiently was already published in the literature for other purposes (2, 3, 8). An almost identical primer sequence was originally described for detection of *Mycobacteria* in medical diagnostics (2, 12), as well as for molecular detection of different Gram-positive and Gram-negative, anaerobic or aerobic bacteria in clinical samples (8). In addition, the same pair of primers was used for amplification of DNA from unknown species of luminescent marine bacteria (3). This led us to the presumption that the same primers would be appropriate for identification of bacterial contaminants in pharmaceutical quality control samples. Ten

phylogenetically distant selected bacterial species from our study are used as test strains in many analytical applications (13) and are frequent contaminants in the manufacturing of pharmaceuticals (1).

While some authors emphasize that lack of sufficient sequence variations in 16SrRNA gene between some closely related species, *e.g.* in the genus *Bacillus* (14), could limit the usefulness of ribotyping, other studies have shown that only one nucleotide difference in the 16SrRNA gene is enough to distinguish between *B. cereus* and *B. anthracis* species (15). Our study also supports the conclusion that ribotyping of a partial sequence of this gene provides enough information for discrimination on even the subspecies level (subtypes and similar taxons). Considering the same genus, we have unambiguously allotted the *B. subtilis* strain to the subspecies level *spizizenii*.

CONCLUSIONS

The region of the 16SrRNA gene that we have amplified and sequenced has provided sufficient information for unambiguous identification of the ATCC strains at the species and subspecies level, with more than 99 % of sequence similarity to the already published sequences in the NCBI nucleotide database using BLAST analysis.

Sequence analyses of ten bacteria revealed that the nucleotide variations were adequate for identification of both phylogenetically related and distant bacteria with our modification of 16SrRNA genotyping. Bacterial identification based on partial sequencing of the 16SrRNA gene is a reliable, fast and very accurate method, which could be used in quality control of pharmaceuticals after adequate method validation. In contrast to classical microbiology, this molecular testing does not require viable bacteria, time-consuming cultivation or problematic identification of slow-growing and difficult-to-cultivate organisms.

REFERENCES

1. S. Sutton and L. Jimenez, A review of reported recalls involving microbiological control 2004–2011 with emphasis on FDA considerations of “objectionable organisms”, *Am. Pharm. Rev.* **15** (2012) 42–57.
2. L. Hall, K. A. Doerr, S. L. Wohlfiel and G. D. Roberts, Evaluation of the MicroSeq system for identification of mycobacteria by 16S ribosomal DNA sequencing and its integration into a routine clinical mycobacteriology laboratory, *J. Clin. Microbiol.* **41** (2003) 1447–1453. DOI: 10.1128/JCM.41.4.1447-1453.2003.
3. A. Nawaz and N. Ahmed, Isolation and characterization of indigenous luminescent marine bacteria from Karachi coast, *Acad. Res. Int.* **1** (2011) 74–83.
4. A. Zbinden, E. C. Böttger, P. P. Bosshard and R. Zbinden, Evaluation of the colorimetric VITEK 2 card for identification of Gram-negative nonfermentative rods: comparison to 16S rRNA gene sequencing, *J. Clin. Microbiol.* **45** (2007) 2270–2273; DOI: 10.1128/JCM.02604-06.
5. P. Hugenholtz, B. M. Goebel and N. R. Pace, Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity, *J. Bacteriol.* **180** (1998) 4765–4774.
6. A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn and F. O. Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, *Nucleic Acids Res.* **41** (2013) e1; DOI: 10.1093/nar/gks808.

7. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies and F. O. Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, *Nucleic Acids Res.* **41** (2013) 590–596; DOI: 10.1093/nar/gks1219; <http://www.arb-silva.de/>, last access: October 13, 2015.
8. E. Gomez, D. R. Gustafson, J. E. Rosenblatt and R. Patel, *Actinobaculum* bacteremia: a report of 12 cases, *J. Clin. Microbiol.* **49** (2011) 4311–4313; DOI: 10.1128/JCM.00798-11.
9. J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning*, 2nd ed., Cold Spring Harbor Press, New York 1989, p. 10.
10. T. A. Hall, Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT, *Nucl. Acids Symp.* **41** (1999) 95–98; DOI: 10.1021/bk-1999-0734.ch008; <http://www.mbio.ncsu.edu/bioedit/bioedit.html>, last access: October 13, 2015.
11. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* **215** (1990) 403–410, available at: <http://www.blast.ncbi.nlm.nih.gov/Blast.cgi> (last access date: March 2015).
12. R. Baracaldo, M. Foltzer, R. Patel and P. Bourbeau, Empyema caused by *Mycoplasma salivarium*, *J. Clin. Microbiol.* **50** (2012) 1805–1806; DOI: 10.1128/JCM.06839-11.
13. *European Pharmacopeia*, Vol. I, 8th ed., EDQM, Strasbourg 2013, pp. 557–558.
14. C. A. Petti, Detection and Identification of microorganisms by gene amplification and sequencing, *Clin. Infect. Dis.* **44** (2007) 1108–1114; DOI: 10.1086/512818.
15. C. T. Sacchi, A. M. Whitney, L. W. Mayer, R. Morey, A. Steigerwalt, A. Boras, R. S. Weyant and T. Popovic, Sequencing of 16S rRNA Gene: A Rapid Tool for Identification of *Bacillus anthracis*, *Emerg. Infect. Dis.* **8** (2002) 1117–1123.