

16S BACTERIAL IDENTIFICATION FOR GMP ENVIRONMENTAL MONITORING PROGRAMMES

*Identification of bacterial isolates provides vital information for pharmaceutical companies undertaking environmental monitoring programmes within GMP production facilities. NCIMB's Identification Services Manager, **VIKKI MITCHELL**, takes a look at look at 16S rDNA sequencing, the options available and when it is most appropriate to use them.*



Although phenotypic methods continue to provide a valuable and widely used tool for initial identification of bacteria, the introduction of genotypic techniques has really revolutionised this aspect of microbiology, with some species being reclassified as a result of the additional information that sequencing data provides.

16S ribosomal DNA (16s rDNA) sequencing is now widely accepted as the “gold standard” for

identification of unknown bacterial isolates, and there are two options available – 500 bp and full gene sequencing.

Prokaryotic ribosomes contain a large (70S) and a small (30S) subunit, and 16S rRNA is a structural component of the 30S small subunit. The term 16S rDNA refers to the genes that encode it, but the method is also sometimes referred to as 16S rRNA sequencing.

16S rRNA is essential for cell function and consequently, is highly conserved between different bacterial species. It is also relatively quick to sequence and the variations that do exist within the sequences can be used to determine the relationships between different organisms and build phylogenetic trees.

In practice, when people refer to the 16S ribosomal DNA technique they are usually referring to the sequencing of the initial 500bp – approximately a third of the full gene. This is usually, though not always, sufficient to identify bacteria at the species level for the purposes of identification of environmental isolates or system contaminants.

Sequencing of the full 16s rDNA gene is more commonly used within research projects, or for patenting purposes, but there are some occasions where full 16s sequencing is required for identification of common environmental isolates.

NCIMB Limited

Ferguson Building, Craibstone Estate, Bucksburn, Aberdeen, AB21 9YA, United Kingdom
Tel: +44 (0)1224 711100 Email: enquiries@ncimb.com Website: www.ncimb.com

A good illustration of this, from my own experience, is with the genus *Bacillus*.

Bacillus is a ubiquitous and diverse genus of bacteria and an extremely common environmental contaminant. Consequently it is one of the genera that we are called on to identify most frequently at NCIMB, and 500bp 16s rDNA sequencing is generally our first port of call. While it is a reliable method in the vast majority of cases, we have occasionally had some difficulties with the use of this technique when identifying *Bacillus* species.

Specifically, during sequencing we have encountered issues relating to the alignment of the forward and reverse strands of DNA, due to the presence of indel mutations. This creates difficulties for the sequencing software, which will only use the sections of DNA where it can get a good match between the forward and reverse strands.

The net result of this is that the system will be trying to identify the bacterial species on the basis of less than 500 base pairs, and this is not a long enough sequence to obtain an accurate bacterial identification.

Under these circumstances it is therefore necessary to revert to full 16s sequencing in order to identify the bacterial species. We have also noticed this phenomenon with a number of our culture collection strains, but, from our experience it does seem to be an issue that mainly arises with the genus *Bacillus*.

Full 16s sequencing can also be used to help clarify whether environmental isolates identified as being same species of bacteria are in fact the same strain, and therefore likely to have arisen from the same source. Although full 16s gene sequencing results are not always conclusive for this purpose, sequencing the full gene does provide additional useful data that can highlight differences between bacterial isolates. It can provide a useful approach

when data is not available for other, potentially more definitive, strain typing techniques such as multilocus sequence typing (MLST). Although the amount of sequence data available for MLST analysis is continuously increasing, it is inevitable that we will come across species for which no data is available from time to time.



Full 16s sequencing can also be useful for distinguishing between closely related species of bacteria.

When undertaking 500bp sequencing of bacterial isolates, we occasionally find that the sequences obtained give a 100% match to more than one species. In this situation, sequencing of the full gene can sometimes give enough additional information to obtain a more precise result.

An interesting example of the requirement for full 16S gene sequencing that I have encountered is the bacterial species *Enterococcus faecium*.

There has been interest in the use of strains of this species of bacteria as a probiotic and in this context being able to accurately identify its presence is essential. However, we have found that 500bp sequencing can mistakenly identify strains of *Enterococcus faecium* as *Enterococcus durans*.

Sequencing of the full gene identified the isolate as *Enterococcus faecium* and consequently when confirming the presence of *E.faecium*, we would always advise this approach.

NCIMB Limited

Ferguson Building, Craibstone Estate, Bucksburn, Aberdeen, AB21 9YA, United Kingdom
Tel: +44 (0)1224 711100 Email: enquiries@ncimb.com Website: www.ncimb.com

This kind of occurrence, where two closely related species of bacteria cannot be identified accurately based on the 500 bp sequence, really serves to highlight the importance of taking time to review the sequencing data rather than just reporting the top match given by the software. It also emphasises the importance of using a validated database that is updated regularly and incorporates user feedback with respect to these kinds of issues.

There are a number of commercial and public sources of data that can be used for the identification of bacterial isolates from their 16S rDNA sequences. Validated commercial databases have been built using reference strains from recognised culture collections, and these generally include a good selection of commonly isolated species.

The alternative is the much more comprehensive, but unvalidated, EMBL-EBI (the European Bioinformatics Institute – part of the European Molecular Biology Laboratory) nucleotide database. This database is continuously expanding as it allows researchers to share sequence data with the wider scientific community. EMBL is part of the International Nucleotide Sequence Database Collaboration (INSD) which also includes GenBank at the National Centre for Biotechnology Information in the United States and the DNA DataBank of Japan.

These three organisations regularly exchange data and the INSD has a policy of free and unrestricted access to all of the data records that their database contains in order to provide the scientific community with the most up to date and comprehensive DNA sequence information available.

In practice, we regularly encounter bacterial isolates from environmental monitoring programmes that cannot be identified to the species level using a validated database. The

options in these circumstances are either to report on the level of identification obtained using the validated database – often genus level – or use the more comprehensive public databases to obtain a species match.

When using public databases to identify unknown bacterial isolates, I would always recommend referring to any relevant published papers for additional supporting information, in order to ensure that the most accurate and reliable available identification is obtained.



ABOUT THE AUTHOR: Vikki Mitchell joined NCIMB in 2005. She leads a team of scientists responsible delivering NCIMB's Identification Services and sequencing new deposits to the UK's National Collection of Industrial Food and Marine Bacteria. Vikki holds a BSc (hons) degree in Applied Biosciences and Management, and an MSc in Instrumental Analytical Techniques; DNA Analysis, Proteomics and Metabolomics from the Robert Gordon University in Aberdeen.

NCIMB Limited

Ferguson Building, Craibstone Estate, Bucksburn, Aberdeen, AB21 9YA, United Kingdom
Tel: +44 (0)1224 711100 Email: enquiries@ncimb.com Website: www.ncimb.com