

Assignment of Reference 5'-end 16S rDNA Sequences and Species-Specific Sequence Polymorphisms Improves Species Identification of *Nocardia*

Fanrong Kong¹, Sharon C.A. Chen^{1,*}, Xiaoyou Chen^{1,2}, Vitali Sintchenko¹, Catriona Halliday¹, Lin Cai³, Zhongsheng Tong^{1,4}, Ok Cha Lee¹, and Tania C. Sorrell¹

¹Centre for Infectious Diseases and Microbiology, The University of Sydney, Westmead Hospital, Westmead, New South Wales, Australia

²Department of Tuberculosis, Beijing Tuberculosis & Thoracic Tumour Research Institute, Beijing, P. R. China

³Department of Dermatology, Peking University People's Hospital, Beijing, P. R. China

⁴Research Laboratory for Infectious Skin Diseases, Department of Dermatology, Wuhan First Hospital, Wuhan, P. R. China

Abstract: 16S rDNA sequence analysis is the most accurate method for definitive species identification of nocardiae. However, conflicting results can be found due to sequence errors in gene databases. This study tested the feasibility of species identification of *Nocardia* by partial (5'-end 606-bp) 16S rDNA sequencing, based on sequence comparison with "reference" sequences of well-annotated strains. This new approach was evaluated using 96 American Type Culture Collection (n=6), and clinical (n=90) *Nocardia* isolates. Nucleotide sequence-based polymorphisms within species were indicative of "sequence types" for that species. Sequences were compared with those in the GenBank, Bioinformatics Bacteria Identification and Ribosomal Database Project databases. Compared with the reference sequence set, all 96 isolates were correctly identified using the criterion of $\geq 99\%$ sequence similarity. Seventy-eight (81.3%) were speciated by database comparison; alignment with reference sequences resolved the identity of 14 (15%) isolates whose sequences yielded 100% similarity to sequences in GenBank under >1 species designation. Of 90 clinical isolates, the commonest species was *Nocardia nova* (33.3%) followed by *Nocardia cyriacigeorgica* (26.7%). Recently-described or uncommon species included *Nocardia veterana* (4.4%), *Nocardia beijingensis* (2.2%) and *Nocardia abscessus* and *Nocardia arthritidis* (each n=1). *Nocardia asteroides sensu stricto* was rare (n=1). There were nine sequence types of *N. nova*, three of *Nocardia brasiliensis* with two each of *N. cyriacigeorgica* and *Nocardia farcinica*. Thirteen novel sequences were identified. Alignment of sequences with reference sequences facilitated species identification of *Nocardia* and allowed delineation of sequence types within species, suggesting that such a barcoding approach can be clinically useful for identification of bacteria.

Keywords: *Nocardia* spp., species identification, 16S rDNA, reference sequences, sequence polymorphisms.

INTRODUCTION

Nocardia species cause a range of infections including localised lung and skin infections, and disseminated disease. Speciation of clinical isolates is important to characterise associated disease manifestations, predict antimicrobial susceptibility and identify differences in epidemiology [1]. Since standard phenotypic identification methods are time-consuming and often imprecise [2, 3], nucleic acid-amplification tools targeting conserved gene regions have been developed to facilitate accurate species determination.

Of these, 16S rDNA sequence analysis is the most frequently-used method for definitive species identification of nocardiae [2, 4-6]. Polymorphisms within the 65-kDa heat shock protein gene (*hsp65*) target are also reported to enable speciation [7, 8]. These sequence-based identification

methods have led to substantial species re-assignment within the genus, especially among "*Nocardia asteroides*" isolates. Over 80 species have now been described of which at least 33 have been implicated in human disease (<http://www.ncbi.nlm.nih.gov/Taxonomy/>; <http://www.bacterio.cict.fr/n/nocardia.html>; [2]).

Numerous *Nocardia* 16S rDNA sequences have thus been deposited in public sequence databases; however, a substantial proportion of sequences, for example in GenBank, represents misidentified isolates or contains significant errors [9, 10]. Imprecise species identification may also result from the presence of multiple, but different, copies of 16S rDNA in certain *Nocardia* spp. such as *Nocardia nova* [11]. Further, sequence-based analyses are complicated by the lack of consensus regarding the degree of sequence similarity required for species definition of *Nocardia* [12].

To improve sequence-based species identification, there has been strong impetus to develop libraries of DNA sequences in order to designate, or link, standardised sequences, including nucleotide polymorphisms within these

*Address correspondence to this author at the Centre for Infectious Diseases and Microbiology, Westmead Hospital, Darcy Road, Westmead, New South Wales, 2145, Australia; Fax: +61-2- 9891 5317; E-mail: Sharon.chen@swahs.health.nsw.gov.au

sequences, with a particular species; this process requires the establishment of such “reference sequences” or “DNA barcodes” for species identification and for the recognition of intraspecies sequence polymorphisms or “sequence types” [13, 14]. Such an approach has not yet been applied to the identification of *Nocardia*. Since the few molecular analyses of *Nocardia* culture collections have reported significant species misidentification using phenotypic methods [5, 15], we re-examined the species identity of 96 *Nocardia* isolates in our collection using partial (5'-end 606-bp) 16S rDNA sequencing and the assignment of sequence types. The accuracy of three publicly-available gene databases for species identification was compared.

MATERIALS AND METHODS

Nocardia Organisms

Ninety-six *Nocardia* isolates were studied (supplementary Table S1; Table 1). These comprised six American Type Culture Collection (ATCC; Rockville, MD) strains (*N. asteroides* ATCC 19247^T, *Nocardia farcinica* ATCC 3308, *N. farcinica* ATCC 3318^T, *N. nova* ATCC 33726^T, *Nocardia otitidiscaviarum* ATCC 14629^T and *Nocardia paucivorans* ATCC BAA-278^T) and 90 clinical isolates (from the Clinical Mycology Laboratory, Centre for Infectious Diseases and Microbiology, Westmead Hospital, Sydney, Australia). Clinical isolates were cultured from separate patients from 1997-2005. All isolates were speciated using standard phenotypic methods and antibiotic susceptibility profiles [3] and, by 16S rDNA sequencing.

Isolates were cultured aerobically in brain heart infusion broth (Amyl Media, Dandenong, Australia) for 3-15 days at 37°C.

DNA Extraction

Cells from 2 ml-brain heart infusion broth cultures of *Nocardia* in late logarithmic phase were harvested by centrifugation at 14,000 X g for 10 min. The supernatant was removed and the pellet suspended in 150 µl of digestion buffer (10 mM Tris-HCl [pH 8.0], 0.45% Triton X-100 and 0.45% Tween 20). Bacterial suspensions were then heated for 10 min at 100°C to lyse the cells, followed by cooling at -20°C for 1 h. Cell lysates were centrifuged at 14,000 X g for 5 min to pellet the cell debris. Supernatants containing DNA were diluted in 350 µl TE buffer (5mM Tris HCl, 0.5 mM EDTA) and centrifuged for 2 min to remove cell debris. DNA was quantitated using a spectrophotometer and stored at -20°C until required.

PCR Amplification and Sequencing of the 16S rDNA

The 5'-end 606-bp fragment of the 16S rDNA gene was amplified using the universal bacterial primers 16S-27f0 (5' to 3': 1 TTA GAG TTT TGA TCM TGG CTC 21) and 16S-907r (5' to 3': 986 CCG TCA ATT CMT TRA GTT T 877) [16]. Each PCR reaction contained 5 µl template DNA, 0.25 µl (50 pmol/µl) each of forward primer and reverse primer, 1.25 µl dNTPs (2.5 mM of each dNTP: Roche Diagnostics, Mannheim, Germany), 2.5 µl 10x PCR buffer (Qiagen, Doncaster, Victoria), 0.1 µl HotStar *Taq* polymerase (5 U/µl) and water to a 25 µl final volume. Amplifica-

Table 1. Species Distribution of 96 *Nocardia* Isolates by Sequence-Based Alignment of 606-bp 16S rDNA Fragments with Reference Sequences and BLASTn Against Database Sequences

Species identification (Alignment with Reference Sequences)	No. Isolates/ No. Matched by Phenotypic Identification	No. Species with ≥99% Sequence Match (BLASTn)	GenBank Range % Similarity/ No. Matched Sequences ¹	BIBI Range % Similarity/ No. Matched Sequences ¹	RDP-II Range % Similarity/ No. Matched Sequences ¹
<i>N. asteroides</i> sensu stricto	2/2	1	99.0-100/3	100/3	100/2
<i>N. abscessus</i>	1/0	3 ²	100/15	100/20	100/13
<i>N. aobensis</i>	1/0	1	99.8/4	99.8/4	98.8/4
<i>N. arthritis</i>	1/0	1 ³	99.6/1 ³	99.6/1 ³	0
<i>N. beijingensis</i>	2/0	1	100/13	100/18	98.1/13
<i>N. brasiliensis</i>	3/3	1	100/2	100/4	100/2
<i>N. cyriacigeorgica</i>	24/0	1	99.8-100/35	99.8-100/21	99.1-100/15
<i>N. farcinica</i>	13/10	2 ⁴	99.8-100/14	100/12	100/8
<i>N. nova</i>	31/28	1	99.3-100/5	99.6-100/5	99.0-100/4
<i>N. otitidiscaviarum</i>	5/4	1	100/5	100/5	100/4
<i>N. paucivorans</i>	7/1	1	100/1	100/1	100/1
<i>N. transvalensis</i>	1/0	1	99.7/3	99.6/5	98.8/2
<i>N. veterana</i>	4/0	1	100/5	100/8	100/4
<i>N. vinacea</i>	1/0	1	100/3	100/3	100/2

¹Refers to the number strains in the database with a sequence similarity of ≥ 99% to that of the query sequence.

²Refers to 100% sequence similarity with *N. asteroides*, *N. asiatica* and *N. abscessus* sequences.

³99.6% sequence similarity to a single *N. beijingensis* sequence (GenBank accession no. AY756543) but <99% sequence similarity to other *N. beijingensis* strains. Since the sequence demonstrated 99.0% similarity to the reference *N. arthritis* sequence, it was assigned as such.

⁴Refers to 100% sequence similarity with *N. farcinica* sequences as well as with sequence of *N. otitidiscaviarum* strain DSM 43242^T (GenBank accession no. X80611).

tion was performed in a Mastercycler gradient thermocycler (Eppendorf; Netheler-Hinz GmbH, Germany). The cycling conditions were: 95°C for 15 min followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 90 s with a final extension step at 72°C for 10 min.

PCR products were purified (PCR Product Pre-sequencing Kit; USB Corporation, Cleveland, OH) and sequenced using the BigDye Terminator version 3.1 cycle sequencing kit (ABI PRISM 3100 genetic analyser; Applied Biosystems, Foster City, CA) and the primer 16S-27f (5' to 3': 3 AGA GTT TTG ATC MTG GCT CAA G 23) [16]. Each sequence was manually aligned and analysed to ensure high quality sequence data. Where the sequence of an isolate differed from the GenBank "reference" sequence for that species (see Results and Table 2), [7, 17, 18] or for novel sequences, additional primers (16S-27f0 and 16S-907r) were used to confirm the result of the sequence.

16S rDNA Sequence Analysis

For each isolate, the amplified 5'-606 bp 16S rDNA fragment was examined using the BioManager facility (ANGIS, Sydney; <http://biomanager.angis.org.au/>). Consensus sequences were constructed from alignments of sequence data using ClustalW [19] after careful examination of each electrophoregram trace representation of data. Sequence data were queried against archived sequences in the GenBank (BLASTn 2.2.10; <http://www.ncbi.nlm.nih.gov/>), Bioinformatics Bacteria Identification (BIBI) version 0.2 (<http://umr5558-sud-str1.univ-lyon1.fr/lebibi/lebibi.cgi>; [20]) and Ribosomal Database Project-II (RDP-II) version 9.54 (<http://rdp.cme.msu.edu/index.jsp>; [21]) databases.

A list of the closest sequence matches was generated from the database comparisons with pair-wise distance scores indicating the percent similarity between the unknown (query) sequence and database sequences. Only consensus sequences with a minimum length of 606-bp were analysed. A percent similarity (or identity) score of $\geq 99\%$ [4, 5] was used as the criterion to classify an isolate to species level whilst a 97 to 98.9% similarity score identified an isolate as belonging to the genus *Nocardia* but to a different species [9, 10].

Nucleotide Sequence Accession Number

Thirty-four 606 bp partial 16S rDNA sequences including 13 novel *Nocardia* sequences generated in the study were deposited in GenBank with the following accession numbers (also see Table 3): FJ172101 through to FJ172134.

RESULTS

Establishment of a Reference Set of 16S rDNA Sequences

For this study, the 606-bp 16S rDNA sequences of 43 well-characterised isolates representing 43 taxonomically-authenticated *Nocardia* species (<http://www.bacterio.cict.fr/n/nocardia.html>) were chosen as the reference sequences for that species; these isolates were previously-characterised by both 16S rRNA and *hsp65* gene analyses [7, 8]. The sequences of two additional isolates, *Nocardia elegans* DSM 44890^T and *Nocardia aobensis* IFM 0137 [17, 18], were included in the reference sequence set (Table 2). Examination of the *N. asteroides* ATCC 14759 sequence

(*N. asteroides* type VI, GenBank accession no. DQ223862) found it to be identical to that of *Nocardia cyriacigeorgica* DSM 44484^T (GenBank accession no. AY756550). The sequence of *Nocardia beijingensis* JCM 10666^T (GenBank accession no. AY756543) reported in one study [6, 7] was 98.5% similar to that of this same strain (GenBank accession no. DQ659901) described in a separate study [17]. We assigned the sequence corresponding to accession no. DQ659901 as the reference sequence for *N. beijingensis* since this yielded the higher similarity (100%) to multiple *N. beijingensis* sequences in the GenBank, BIBI and RDP-II databases.

This collection of partial 16S rDNA sequences representing 45 *Nocardia* species formed the basis for re-evaluating the species identity of the study isolates and was important in the validation of the sequence analyses. Strains fulfilling the criterion for a species ($\geq 99\%$ sequence similarity) but which demonstrated sequence polymorphisms compared to the reference sequence for that species were considered as separate sequence types for the species.

16S rDNA Sequence-Based Identification of *Nocardia* Isolates

The details of 96 *Nocardia* isolates identified by phenotypic methods and partial 16S rDNA sequencing are given in supplementary Table S1. The species distribution of isolates, as determined by sequence comparison with the reference sequence set, and with sequences in the GenBank, BIBI and RDP-II databases, is shown in Table 1.

(a) Identification Based on Comparison with Reference 606-bp 16S rDNA Sequences

Following alignment of sequence data with the reference sequence set, partial 16S rDNA sequencing provided species identification for all 96 isolates using a criterion of $\geq 99\%$ sequence similarity for species definition; 83 (86.5%) isolates were identified if a criterion of 100% sequence similarity was used.

Of 90 clinical isolates, phenotypic methods correctly identified 42 (46.7%) strains to species level (Table 1; supplementary Table S1). Thirty-four isolates were assigned a phenotypic identification of *N. asteroides*/*N. asteroides* complex based on their drug susceptibility pattern but 16S rDNA sequencing recognised them as a number of distinct species e.g. *N. cyriacigeorgica*, *N. farcinica* and *N. paucivorans* among others (Supplementary Table S1). Molecular and phenotypic species identification methods were concordant for 80% (8 of 10), 75% (three of four), 90% (27 of 30) and 100% (all of three) of *N. farcinica*, *N. otitidiscaevarium*, *N. nova* and *Nocardia brasiliensis* clinical isolates, respectively. However, none of the six *N. paucivorans* isolates were correctly identified by phenotypic methods. All six ATCC strains were assigned by 16S rDNA sequencing to their respective species (Supplementary Table S1).

Discrepant results for isolates are summarised in Supplementary Table S1 (see also Table 1). In particular, only one of 35 phenotypic "*N. asteroides*/*N. asteroides* complex" isolates (other than strain ATCC 19247^T) had a sequence identical to sequences of *N. asteroides* sensu

Table 2. Reference 5' end 606-bp 16S rDNA Sequences of 45 *Nocardia* Species¹ used in the Present Study

Strain no. ^{1,2}	Species Identification	GenBank Accession no. ³
DSM 44432 ^T	<i>N. abscessus</i>	AY544980
DSM 44491 ^T	<i>N. africana</i>	AY756540
IFM 0137	<i>N. aobensis</i>	AB126875 (bp positions 40-645)
DSM 44729 ^T	<i>N. araoensis</i>	AY903623
DSM 44731 ^T	<i>N. arthritidis</i>	AY903619
DSM 44668 ^T	<i>N. asiatica</i>	AY903617
ATCC 19247 ^T	<i>N. asteroides</i>	AY756541
ATCC 49872	<i>N. asteroides</i> type IV	AY756542
JCM 10666 ^T	<i>N. beijingensis</i>	DQ659901 ⁴ (bp positions 14-619)
JCM 10666 ^T	<i>N. beijingensis</i>	AY756543
ATCC 19296 ^T	<i>N. brasiliensis</i>	AY756544
DSM 43024 ^T	<i>N. brevicatena</i>	AY756545
DSM 43397 ^T	<i>N. carnea</i>	AY756546
DSM 44546 ^T	<i>N. cerradoensis</i>	AY756547
ATCC 700418 ^T	<i>N. crassostreae</i>	AY756548
DSM 44490 ^T	<i>N. cummidelens</i>	AY756549
DSM 44484 ^T	<i>N. cyriaciageorgica</i>	AY756550 ⁵
ATCC 14759	<i>N. asteroides</i> type VI	DQ223862 ⁵
DSM 44890 ^T	<i>N. elegans</i>	DQ659905 (bp positions 1-603) ⁶
DSM 43665 ^T	<i>N. farcinica</i>	AY756551
JCM 3332 ^T	<i>N. flavorosea</i>	AY756552
DSM 44489 ^T	<i>N. fluminea</i>	AY756553
DSM 44732 ^T	<i>N. higoensis</i>	AY903620
DSM 44496 ^T	<i>N. ignorata</i>	AY756554
DSM 44667 ^T	<i>N. inohanensis</i>	AY903611
CIP 108295 ^T	<i>N. mexicana</i>	AY903610
DSM 44717 ^T	<i>N. neocaledoniensis</i>	AY903614
DSM 44670 ^T	<i>N. niigatensis</i>	AY903615
CIP 104777 ^T	<i>N. nova</i>	AY756555
ATCC 14629 ^T	<i>N. otitidiscaviarum</i>	AY756556
DSM 44386 ^T	<i>N. paucivorans</i>	AY756557
DSM 44730 ^T	<i>N. pneumonia</i>	AY903622
DSM 44290 ^T	<i>N. pseudobrasiliensis</i>	AY756558
DSM 43406 ^T	<i>N. pseudovaccinii</i>	AY756559
DSM 44599 ^T	<i>N. puris</i>	AY903618
JCM 4826 ^T	<i>N. salmonicida</i>	AY756560
DSM 44129 ^T	<i>N. seriolae</i>	AY756561
DSM 44733 ^T	<i>N. shimofusensis</i>	AY903621
DSM 44488 ^T	<i>N. soli</i>	AY756562
DSM 44704 ^T	<i>N. tenerifensis</i>	AY903613
DSM 44765 ^T	<i>N. testacea</i>	AY903612
DSM 43405 ^T	<i>N. transvalensis</i>	AY756563

Table 2. contd....

Strain no. ^{1,2}	Species Identification	GenBank Accession no. ³
JCM 3224 ^T	<i>N. uniformis</i>	AY756564
ATCC 11092 ^T	<i>N. vaccinii</i>	AY756565
DSM 44445 ^T	<i>N. veteran</i>	AY756566
JCM 10988 ^T	<i>N. vinacea</i>	AY756567
DSM 44669 ^T	<i>N. yamanashiensis</i>	AY903616

¹All species designations were cross-checked against two websites - <http://www.ncbi.nlm.nih.gov/Taxonomy/> and <http://www.bacterio.cict.fr/n/nocardia.html>. Abbreviations: ATCC, American Type Culture Collection; CIP, Collection Institut Pasteur, France; DSM, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany; JCM, Japan Collection of Microorganisms, Wako-Shi, Japan. Adapted from [7, 17, 18].

²T refers to the previously-designated current type strains for the particular species.

³Unless otherwise specified refers to first 1-606 bp of the 16S rDNA sequence from the 5' end.

⁴Different sequence results obtained for *N. beijingensis* JCM 10666^T in two separate studies. The sequence with GenBank accession no. DQ659901 is chosen as the reference sequence.

⁵The sequence of *N. cyriacigeorgica* strain DSM 44490^T (GenBank accession no. AY756550) is identical to the sequence of *N. asteroides* ATCC 14759 (accession no. DQ223862).

⁶The sequence is based on a 603-bp 16S rDNA fragment.

stricto (represented by *N. asteroides* ATCC 19247^T); 22 clinical isolates had sequences with 100% similarity to the reference *N. cyriacigeorgica* sequence. The remaining strains were *N. paucivorans* (n=3), *N. nova* (n=2), *N. beijingensis* (n=2), *Nocardia abscessus* (n=1), *Nocardia arthritidis* (n=1), *N. farcinica* (n=1) and *Nocardia veterana* (n=1). Other discrepant results included two *N. farcinica* (phenotypic identification) isolates yielding sequences with 100% similarity to the reference *N. cyriacigeorgica* sequence and a *N. brasiliensis* strain with 100% sequence similarity to *N. otitidiscavarium*.

Eleven isolates (supplementary Table S1) identified as "*Nocardia* spp." by phenotypic methods were identified as *N. paucivorans* (n=3), *N. farcinica*, *N. veterana* (each n=2) and *N. aobensis*, *N. nova*, *Nocardia transvalensis* and *Nocardia vinacea* (each n=1).

b) Species Identification Based on BLASTn Alignments

Isolates were also identified to species level by comparison with database sequences with the following exceptions: firstly, *N. farcinica* could not be definitively speciated; all 13 strains (100% sequence similarity to the reference *N. farcinica* sequence) were identified as either *N. farcinica* or *N. otitidiscavarium* (Table 1). The three databases contained a sequence corresponding to *N. otitidiscavarium* strain DSM 43242^T (GenBank accession no. X80611). This sequence was indistinguishable from *N. farcinica* sequences but only had 94.7% similarity to the reference sequence of *N. otitidiscavarium* (GenBank accession no. AY756556). Secondly, a phenotypic *N. asteroides* isolate was identified as *N. asteroides*/*N. abscessus*/*Nocardia asiatica* (Table 1). Since its sequence was identical to the reference sequence of *N. abscessus*, it was assigned as such. The reference 16S rDNA sequences of *N. abscessus* and *N. asiatica* (Table 2) differ only by a single nucleotide polymorphism (SNP) at position 527 ("G" for *N. abscessus* but "C" for *N. asiatica*). Finally, an isolate (strain 01-320-2714; supplementary Table S1) was identified as *N. beijingensis* by database comparisons (99.6% sequence similarity to a single *N. beijingensis* sequence; GenBank accession no. AY756543). However, as it yielded 99% sequence similarity (5-bp difference) to the reference sequence of *N. arthritidis*, it was assigned as *N. arthritidis* (Table 1).

Comparison of Species Identification Using GenBank, BIBI and RDP-II Databases

For all isolates, the same species identification result was obtained by comparison of their sequences with those in the GenBank, BIBI and RDP-II databases (Table 1). The distribution of percent similarity scores according to database is shown in Fig (1). Using the criterion of ≥99% sequence similarity for species designation, all 96 isolates were identified by the GenBank and BIBI databases and 91 (96.7%), by the RDP-II system. The length of sequences employed for sequence alignment in the BIBI, RDP-II and GenBank databases ranged from 512-548 bp, 572-590 bp and ≥606 bp, respectively. Perfect matches (100% similarity) were observed for 77%, 79% and 93% of sequence alignments against the GenBank, BIBI and RDP-II databases, respectively (Fig. 1).

Species Distribution of Clinical Isolates

Fourteen *Nocardia* species were identified amongst 90 isolates, the most common being *N. nova* (30 isolates; 33%) followed by *N. cyriacigeorgica* (n=24; 27%), *N. farcinica* (n=11; 12%) and *N. paucivorans* (n=6; 7%). There were four isolates each of *N. otitidiscavarium* and *N. veterana*, three strains of *N. brasiliensis*, two of *N. beijingensis* and one each of *N. asteroides sensu stricto*, *N. transvalensis*, *N. vinacea*, *N. aobensis*, *N. arthritidis* and *N. abscessus*.

Intra-Species Variation and Sequence Types of Clinical Isolates

Partial 16S rRNA sequences of *N. veterana*, *N. abscessus*, *N. beijingensis* and *N. vinacea* isolates were identical to the reference sequence for that species, thus there was a single sequence type. Intraspecies sequence heterogeneity was evident in the remaining 10 species, and varied with species (Table 3). The largest number of sequence types, including that identical to the reference sequence, was noted for *N. nova* (nine sequence types) followed by *N. brasiliensis* (three sequence types). *N. cyriacigeorgica*, *N. farcinica*, *N. asteroides sensu stricto* and *N. otitidiscavarium* exhibited two sequence types. The sequences of three *N. cyriacigeorgica* isolates (indistinguishable from one another) differed from the reference *N. cyriacigeorgica* sequence by a SNP at position 576 (substitution of "A" for "G", see Table 3).

Table 3. Partial (5'-end 606-bp) 16S rDNA Sequence Polymorphisms in 10 *Nocardia* Species

Strain Identification no.	Identification Based on Comparison with Reference Sequence	% Similarity to Reference Sequence	Site of Nucleotide Polymorphisms (bp Position: Reference Sequence → Isolate Sequence)	100% Similarity to Another GenBank Sequences Rather Than Reference Sequences (13 Novel Sequence GenBank Accession no.)
04-303-0576	<i>N. aobensis</i>	99.67	137 G→A	novel sequence ¹ (FJ172101)
01-320-2714	<i>N. arthritidis</i>	99.0	132 C→T, 240 G→A, 251 C→T, 341 C→G, 566 A→G, 588 G→T	novel sequence ¹ (FJ172102)
02-071-3627	<i>N. asteroides</i>	99.0	133-135 TTC→ACA, 148-150 GAG→TGT	novel sequence ¹ (FJ172103)
00-194-3516	<i>N. brasiliensis</i>	99.84	203 T→C	Z36935
03-273-2825	<i>N. brasiliensis</i>	99.67	203 T→C, 328 G→A	AY245543
99-167-2395	<i>N. brasiliensis</i>	99.67	203 T→C, 328 G→A	AY245543
00-159-1584	<i>N. cyriaciageorgica</i>	99.84	576 G→A	novel sequence ¹ (FJ172112)
04-181-3939	<i>N. cyriaciageorgica</i>	99.84	576 G→A	
05-111-2308	<i>N. cyriaciageorgica</i>	99.84	576 G→A	
01-109-2248	<i>N. farcinica</i>	99.84	67 A→A/G ²	novel sequence ¹ (FJ172117)
05-053-4454	<i>N. nova</i>	99.84	86 C→C/T ²	novel sequence ¹ (FJ172126)
02-352-3316	<i>N. nova</i>	99.84	136 G→A/G ²	novel sequence ¹ (FJ172120)
00-130-2170	<i>N. nova</i>	99.84	260 A→A/G ²	novel sequence ¹ (FJ172121)
04-110-3287	<i>N. nova</i>	99.67	86 C→C/T ² , 136 G→A/G ²	novel sequence ¹ (FJ172125)
00-056-3529	<i>N. nova</i>	99.67	86 C→T, 136 G→A	AF430030
01-067-1349	<i>N. nova</i>	99.67	86 C→T, 136 G→A	
01-097-0996	<i>N. nova</i>	99.67	86 C→T, 136 G→A	
02-199-2723	<i>N. nova</i>	99.67	86 C→T, 136 G→A	
00-025-0538	<i>N. nova</i>	99.67	135 G→T, 148 T→G	AF430032
00-314-1789	<i>N. nova</i>	99.67	135 G→T, 148 T→G	
04-150-0614	<i>N. nova</i>	99.67	135 G→T, 148 T→G	
01-066-1903	<i>N. nova</i>	99.51	86 C→T, 136 G→A, 578 T→G	novel sequence ¹ (FJ172122)
02-021-0419	<i>N. nova</i>	99.51	135 G→T, 148 T→G, 328 A→G	novel sequence ¹ (FJ172119)
01-114-2816	<i>N. otitidiscaviarum</i>	99.85	133 A→G	novel sequence ¹ (FJ172127)
ATCC BAQ-278	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	AF179865
03-141-3073	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	
03-185-3304	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	
03-240-2758	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	
05-200-1797	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	
97-114-0609	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	
97-16298	<i>N. paucivorans</i>	99.67	33 C-ins ³ , 37 G-ins ³	
03-310-2776	<i>N. transvalensis</i>	99.67	402, 403 AG→CA	novel sequence ¹ (FJ172131)

¹Sequences of isolates without a match with GenBank sequences are novel sequences.

²At the specified bp position, both nucleotides were present due to different multiple copies of the 16S rRNA gene.

³Refers to insertion of the specified base.

There was only one sequence type for *N. aobensis*, *N. transvalensis*, *N. arthritidis* and *N. paucivorans* but the sequences of these isolates all demonstrated SNPs when compared to the reference sequence for the species (Table 3). Thirteen novel *Nocardia* sequences were identified for 15 isolates (eight species).

DISCUSSION

Molecular-based identification of *Nocardia* spp. remains a challenge due to the increasing recognition of new species and changes in taxonomy [22]. Sequence analysis of the 16S rDNA is the current gold standard for identification of *Nocardia* spp. However few studies have explored the

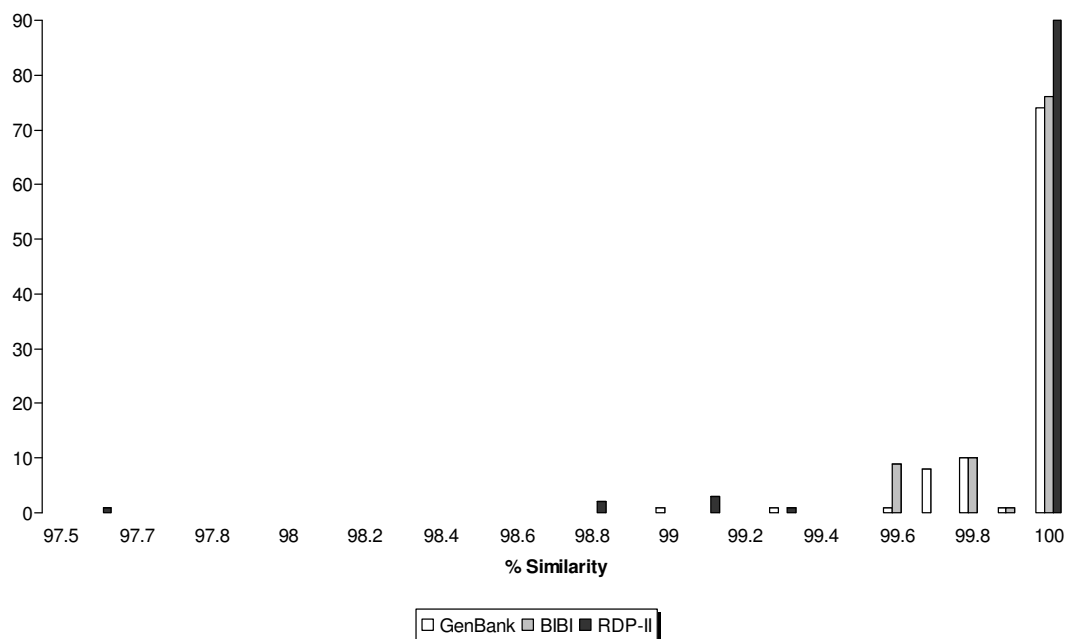


Fig. (1). Distribution of similarity scores for partial 16S rDNA sequence-based identification of *Nocardia* isolates by the GenBank, BIBI and RDP-II databases.

validity of sequence-based identification approaches using collections of phenotypically-characterised clinical isolates [9]. The present study proposes, and has tested the utility of a set of reliable reference *Nocardia* 16S rDNA sequences, derived from authoritatively-identified organisms [7, 17], as sequence standards for species identification. Further, by identifying and assigning sequence types to represent sequence polymorphisms within a species, the results have identified that such a “barcoding” approach can improve systematic and accurate species identification. DNA barcoding has been designed to provide rapid accurate species identification by using short, standardised gene regions (in this case, the 5'-end 16S rDNA region) as internal species tags [14]. Although it has been found to be effective in speciating eukaryote organisms and some parasites [13, 14, 16], there are few data on its application in the identification of bacterial pathogens.

As such, based on comparison with the reference sequence set, partial 16S rDNA sequencing provided clear species identification of all 96 isolates. In particular, the reference sequence set was useful in the speciation of 11 isolates identified only to genus level by phenotypic methods and assisted in resolving the identity of 14 (15%) clinically relevant isolates whose sequences aligned with 100% similarity to sequences assigned to more than one species in the GenBank, BIBI and RDP-II databases (Table 1). Without the resource of this sequence set, sequencing was unable to assign precise species identification to 13 *N. farcinica* isolates (Table 1). It is most likely that the *N. otitisdis-cavarium* DSM 43242^T sequence (GenBank accession no. X80611) with 100% identity to *N. farcinica* sequences represents a misidentification, underscoring the importance of adequate stewardship of database sequences. Of note, the reference sequence of *N. abscessus* and *N. asiatica* (GenBank accession nos. AY544980 and AY903617, respectively) differ only by a SNP - this low interspecies heterogeneity likely explains the inability to resolve species identifi-

cation for *N. abscessus* after alignment with databases sequences (Table 3). *N. abscessus* (previously *N. asteroides* antimicrobial susceptibility type I) represents ≈20% of isolates of the former *N. asteroides* complex; accurate species identification is important as MICs of imipenem, which is commonly used to treat nocardiosis, are high for this species [2]. There are few descriptions of *N. asiatica* as a pathogen.

Species identification of *Nocardia* by gene sequence analysis is heavily reliant on the entries in the gene repositories being queried [10, 23]. As noted in the present study, inappropriate and/or obsolete sequence entries are important potential limitations. Further, species identification based on data derived from a single or small numbers of strains representing a species must be interpreted with caution (see also Table 1). As such, the submission of carefully-annotated new sequences is critical to maintaining the accuracy of current gene repositories. Since the BIBI and RDP-II systems contain a larger proportion of shorter 16S rDNA sequences (512-548 bp, 572-590 bp, respectively vs. ≥606 bp in GenBank), this may have resulted in an artificially inflated number of perfect sequence matches (Fig. 1). The validity of comparisons with sequences of different lengths for species identification requires further study.

The approach undertaken in this study has further allowed us to distinguish between closely-related *Nocardia* species, to identify newly described or uncommon species and to determine the species distribution of clinical *Nocardia* isolates received in our laboratory. Elsewhere and in Australia, most human infections historically have been attributed to *N. asteroides* sensu stricto antimicrobial susceptibility class types I and VI, *N. nova*, *N. brasiliensis* and *N. farcinica* [1, 24]. Our results reconfirm that this generally remains the case. Given the prevalence of *N. cyriacigeorgica* (previously *N. asteroides* type VI as a pathogen (this study; [25]), adop-

tion of protocols by microbiology laboratories for its identification is important. Apart from distinguishing *N. abscessus* from *N. asiatica* (see above), partial 16S rDNA sequencing was able to differentiate between other closely-related species including *N. veterana* and *N. nova* sensu stricto (sequence similarity of 98.1%; [6, 26, 27]). Classified within the *N. nova* complex, *N. veterana* is an emerging pathogen capable of causing serious infection [27]. Of note, the re-assignment of all but one clinical "*N. asteroides*" strains to other taxonomic groups questions the validity of *N. asteroides* as a separate species.

Importantly, the results identified significant intraspecies sequence polymorphisms within the 16S rDNA for many (10 of 14) *Nocardia* species or for seven of nine species represented by more than one strain (Table 3), and that such nucleotide heterogeneity differed according to species, being most evident for *N. nova* (nine sequence types). Although there was no genetic heterogeneity amongst isolates of, for example, *N. farcinica*, the sequences of the isolates differed from the reference sequence for this species. Thus, if constructing a DNA template or "identification barcode" for identifying Australian *N. paucivorans* isolates, 16S rDNA position 33 should incorporate an extra C and at position 37, an extra G in relation to the reference sequence (Table 3). In the USA, three "genetic types" of *N. cyriacigeorgica* with SNPs at positions 448, 1427 and 1480 [25] have been reported; we identified a SNP at position 576 but not at position 448. Thus, comparison of sequence types of isolates from different countries to identify potential clinical and epidemiological associations may be warranted. As noted for other bacteria, substitutions of as little as 1-2 bp may correlate with unique *Nocardia* phenotypes and clinical significance [5, 28]. Therefore, documentation of sequence polymorphisms within species is relevant to delimiting species or highlighting genetically-distinct groups with levels of sequence divergence that are either suggestive or exclusive of species status.

Finally, 13 novel sequences from eight *Nocardia* species were identified and some of them may merit description as new species (Table 3). Species designation, however, is confounded by the lack of a consensus criterion for species definition based on percent similarity scores; isolates of distinct species of *Nocardia* have been reported to exhibit as much as 99.8% sequence similarity [12].

CONCLUSION

Partial 16S rDNA sequencing is a viable alternative to full-length sequencing for species identification of *Nocardia* in a diagnostic laboratory. The present approach encompassed comparison of the sequence of interest with a library of reference sequences and the assigning of sequence types to represent sequence polymorphisms within species, based on sequence similarity to the reference sequence for that species; unambiguous species identification was obtained for all study isolates. As affirmed in the present study, errors in sequence entries remain important potential limitations of public gene repositories [6, 10, 23]. The results of the study suggest that a barcoding approach [29, 30] can assist with species identification of clinically relevant *Nocardia*. Further studies are warranted to explore its wider application in

improving species differentiation and unravelling sequence data for phylogenetically-unresolved groups of bacteria.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

ACKNOWLEDGMENTS

We thank Ms. Maryann Pincevic for her assistance in performing the 16S rDNA sequencing and Ms. Ping Zhu for help with the Figure preparation.

REFERENCES

- [1] Saubolle MA, Sussland D. Nocardiosis: review of clinical and laboratory experience. *J Clin Microbiol* 2003; 41: 4497-501.
- [2] Brown-Elliott BA, Brown JM, Conville PS, Wallace RJ, Jr. Clinical and laboratory features of the *Nocardia* spp. based on current molecular taxonomy. *Clin Microbiol Rev* 2006; 19: 259-82.
- [3] McNeil MM, Brown JM. The medically important aerobic *Actinomycetes*: epidemiology and microbiology. *Clin Microbiol Rev* 1994; 7: 357-417.
- [4] Cloud JL, Conville PS, Croft A, Harmsen D, Witebsky FG, Carroll KC. Evaluation of partial 16S ribosomal DNA sequencing for identification of *Nocardia* species by using the MicroSeq 500 system with an expanded database. *J Clin Microbiol* 2004; 42: 578-84.
- [5] Roth A, Andrees S, Kroppenstedt RM, Harmsen D, Mauch H. Phylogeny of the genus *Nocardia* based on reassessed 16S rRNA gene sequences reveals underspeciation and division of strains classified as *Nocardia asteroides* into three established species and two unnamed taxa. *J Clin Microbiol* 2003; 41: 851-6.
- [6] Mellmann A, Cloud JL, Andrees S, et al. Evaluation of RIDOM, MicroSeq, and Genbank services in the molecular identification of *Nocardia* species. *Int J Med Microbiol* 2003; 293: 359-70.
- [7] Rodriguez-Nava V, Couble A, Devulder G, Flandrois JP, Boiron P, Laurent F. Use of PCR-restriction enzyme pattern analysis and sequencing database for *hsp65* gene-based identification of *Nocardia* species. *J Clin Microbiol* 2006; 44: 536-46.
- [8] Steingrube VA, Brown BA, Gibson JL, et al. DNA amplification and restriction endonuclease analysis for differentiation of 12 species and taxa of *Nocardia*, including recognition of four new taxa within the *Nocardia asteroides* complex. *J Clin Microbiol* 1995; 33: 3096-101.
- [9] Clarridge JE, III. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 2004; 17: 840-62.
- [10] Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* 2007; 45: 2761-4.
- [11] Conville PS, Witebsky FG. Analysis of multiple differing copies of the 16S rRNA gene in five clinical isolates and three type strains of *Nocardia* species and implications for species assignment. *J Clin Microbiol* 2007; 45: 1146-51.
- [12] Conville PS, Witebsky FG. Multiple copies of the 16S rRNA gene in *Nocardia nova* isolates and implications for sequence-based identification procedures. *J Clin Microbiol* 2005; 43: 2881-5.
- [13] Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philos Trans R Soc Lond B Biol Sci* 2005; 360: 1805-11.
- [14] Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. *Syst Biol* 2005; 54: 852-9.
- [15] Wauters G, Avesani V, Charlier J, Janssens M, Vaneechoutte M, Delmee M. Distribution of *Nocardia* species in clinical samples and their routine rapid identification in the laboratory. *J Clin Microbiol* 2005; 43: 2624-8.
- [16] Becker K, Harmsen D, Mellmann A, et al. Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of *Staphylococcus* species. *J Clin Microbiol* 2004; 42: 4988-95.

- [17] Conville PS, Zelazny AM, Witebsky FG. Analysis of *secA1* gene sequences for identification of *Nocardia* species. *J Clin Microbiol* 2006; 44: 2760-6.
- [18] Kageyama A, Suzuki S, Yazawa K, Nishimura K, Kroppenstedt RM, Mikami Y. *Nocardia aobensis* sp. nov., isolated from patients in Japan. *Microbiol Immunol* 2004; 48: 817-22.
- [19] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; 22: 4673-80.
- [20] Devulder G, Perriere G, Baty F, Flandrois JP. BIBI, a bioinformatics bacterial identification tool. *J Clin Microbiol* 2003; 41: 1785-7.
- [21] Cole JR, Chai B, Farris RJ, *et al.* The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* 2007; 35: D169-D72.
- [22] Patel JB, Wallace RJ, Jr., Brown-Elliott BA, *et al.* Sequence-based identification of aerobic *Actinomycetes*. *J Clin Microbiol* 2004; 42: 2530-40.
- [23] Petti CA. Detection and identification of microorganisms by gene amplification and sequencing. *Clin Infect Dis* 2007; 44: 1108-14.
- [24] Georghiou PR, Blacklock ZM. Infection with *Nocardia* species in Queensland. A review of 102 clinical isolates. *Med J Aust* 1992; 156: 692-7.
- [25] Schlaberg R, Huard RC, Della-Latta P. *Nocardia cyriacigeorgica* is an emerging pathogen in the United States. *J Clin Microbiol* 2007; 46: 265-73.
- [26] Gurtler V, Smith R, Mayall BC, Potter-Reinemann G, Stackebrandt E, Kroppenstedt RM. *Nocardia veterana* sp. nov., isolated from human bronchial lavage. *Int J Syst Evol Microbiol* 2001; 51: 933-6.
- [27] Pottumarthy S, Limaye AP, Prentice JL, Houze YB, Swanzy SR, Cookson BT. *Nocardia veterana*, a new emerging pathogen. *J Clin Microbiol* 2003; 41: 1705-9.
- [28] Tortoli E. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin Microbiol Rev* 2003; 16: 319-54.
- [29] Blaxter M, Mann J, Chapman T, *et al.* Defining operational taxonomic units using DNA barcode data. *Philos Trans R Soc Lond B Biol Sci* 2005; 360: 1935-43.
- [30] Ferri G, Alu M, Corradini B, Licata M, Beduschi G. Species identification through DNA "barcodes". *Genet Test Mol Biomarkers* 2009; [Epub ahead of print] PMID: 19405876.

Received: May 14, 2009

Revised: May 18, 2009

Accepted: May 20, 2009

© Kong *et al.*; Licensee Bentham Open.This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.