JCM
Journals.ASM.org

Identification, Typing, and Phylogenetic Relationships of the Main Clinical *Nocardia* Species in Spain According to Their *gyrB* and *rpoB* Genes

Gema Carrasco, Sylvia Valdezate, Noelia Garrido, Pilar Villalón, Maria J. Medina-Pascual, Juan A. Sáez-Nieto

Laboratorio de Taxonomía, Servicio de Bacteriología, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain

This study compares the identification, typing, and phylogenetic relationships of the most prevalent clinical *Nocardia* species in Spain, as determined via sequence analysis of their housekeeping genes *gyrB* and *rpoB*, with the results returned by the gold standard 16S rRNA method. *gyrB* and *rpoB* analyses identified *Nocardia abscessus*, *N. cyriacigeorgica*, *N. farcinica*, and the *N. nova* complex, species that together account for more than half of the human nocardiosis cases recorded in Spain. The individual discriminatory power of *gyrB* and *rpoB* with respect to intraspecies typing, calculated using the Hunter-Gaston discriminatory index (HGDI), was generally high (HGDI, 0.85 to 1), except for *rpoB* with respect to *N. farcinica* (HGDI, 0.71). Phylogenetically, different degrees of intra- and interspecies microheterogeneity were observed for *gyrB* and *rpoB* in a group of 119 clinical strains. A single 16S haplotype was obtained for each species, except for the *N. nova* complex (8 types), while *gyrB* and *rpoB* were more polymorphic: *N. abscessus* had 14 and 18 haplotypes, *N. cyriacigeorgica* had 17 and 12, *N. farcinica* had 11 and 5, and the *N. nova* complex had 26 and 29 haplotypes, respectively. A diversity gradient was therefore seen, with *N. farcinica* at the bottom followed by *N. abscessus* and *N. cyriacigeorgica* in the middle and *N. nova* complex at the top. The complexity of the *N. nova* complex is highlighted by its six variations in the GyrB ^{126}AAAPEH motif. *gyrB* sequencing (with or without *rpoB* sequencing) offers a simple means for identifying the most prevalent *Nocardia* species in Spanish medical laboratories and for determining the intraspecific diversity among their strains.

Members of the genus *Nocardia* are ramified Gram-positive bacilli that normally live in dust, sand, soil, decaying vegetation, and stagnant water (1). To date, nearly 99 *Nocardia* species have been identified (see NCBI taxonomy for *Nocardia*, http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1817 and http://www.bacterio.cict.fr/n/nocardia.html), but this number undergoes constant modification. Some one-third of *Nocardia* species are known human pathogens, and new pathogenic species causing different clinical problems are constantly being discovered (2, 3). Infection occurs mainly via the respiratory tract, later disseminating to other locations, such as the central nervous system. Localized infection is caused by traumatic injury and gives rise to abscesses (4).

Partial 16S rRNA gene sequencing is the gold standard for identifying *Nocardia* spp. (1). However, the conservation of 16S can be an obstacle to distinguishing between closely related species, such as those of the *Nocardia nova* complex (5, 6). To overcome this disadvantage and to establish phylogenetic relationships at the intra- and interspecies levels, other protein-encoding genes have been studied, such as the 65-kDa heat shock protein *hsp65* (7), the essential secretory protein *secA1* (8), *gyrB*, which is the β-subunit of DNA gyrase and a type II DNA topoisomerase (9), and *rpoB*, which is the β-subunit of DNA-dependent RNA polymerase (RNAP) (6).

gyrB promotes negative supercoiling in the bacterial chromosome during DNA replication, while *rpoB* is involved in transcription. These protein-encoding genes offer advantages over RNA-encoding genes as molecular markers; as housekeeping genes, they are less susceptible to horizontal gene transfer and can be analyzed at the nucleotide and amino acid levels to determine phylogenetic relationships (10).

The majority of phylogenetic studies on *Nocardia* spp. have involved just one strain that is representative of each species, taking into account the combinations of many genes (6, 11). No studies, however, have been performed that have examined large numbers of clinical strains per species.

The aims of the present work were to (i) compare a sequence analysis of *gyrB* and *rpoB* against partial 16 rRNA gene sequencing (the gold standard) for identifying, typing, and determining the phylogenetic relationships between clinical strains of *N. abscessus*, *N. cyriacigeorgica*, *N. farcinica*, and the *N. nova* complex (collectively the most prevalent clinical *Nocardia* species in Spain), and (ii) to gain insight into the intraspecific diversity of these strains via the analysis of these housekeeping genes.

MATERIALS AND METHODS

***Nocardia* strains.** A total of 119 strains of *Nocardia* spp., belonging to the four species most commonly collected in Spain, *N. abscessus*, *N. cyriacigeorgica*, *N. farcinica*, and the *N. nova* complex, were isolated from clinical samples submitted for identification to the Spanish National Center of Microbiology (Majadahonda, Madrid, Spain) between 2006 and 2010. Isolates were grown in heart infusion Columbia agar supplemented with 5% (vol/vol) sheep blood and buffered charcoal-yeast extract (BCYE) for 48 to 72 h at 37°C under aerobic conditions.

Received 25 February 2013 Returned for modification 3 April 2013
Accepted 15 August 2013
Published ahead of print 21 August 2013
Address correspondence to Sylvia Valdezate, svaldezate@isciii.es.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.00515-13.

doi:10.1128/JCM.00515-13

TABLE 1 Characteristics of the four most prevalent species of *Nocardia* isolated from clinical samples in Spain between 2006 and 2010

Nocardia species (no. of strains)	Genes (bp)[a]	No. of haplotypes (HGDI, S^2, SD)[b]	No. of SNPs (divergence rate)[c]	SNPs per strain (avg, mode)
N. abscessus (29)	16S rRNA (409)	1	0	0
	gyrB (653)	14 (0.889, 0.002, 0.046)	24 (0.0–1.7)	0–9 (3, 2)
	rpoB (354)	18 (0.948, 0.00058, 0.024)	65 (0.0–16)	6–49 (42, 44)
N. cyriacigeorgica (30)	16S rRNA (514)	1	0	0
	gyrB (727)	17 (0.94, 0.00055, 0.023)	77 (0.0–5.6)	0–38 (28, 36/38)
	rpoB (354)	12 (0.88, 0.00135, 0.037)	24 (0.0–4.4)	4–15 (9, 13)
N. farcinica (31)	16S rRNA (507)	1	0	0
	gyrB (739)	11 (0.854, 0.0014, 0.038)	11 (0.0–0.8)	0–6 (3, 2/4)
	rpoB (351)	5 (0.716, 0.001, 0.036)	9 (0.0–1.4)	4–6 (4, 4)
N. nova (29)	16S rRNA (529)	8 (0.672, 0.00739, 0.086)	8 (0.0–1.1)	0–4 (1, 2)
	gyrB (698)	26 (0.985, 0.0003, 0.017)	51 (0.0–5.2)	0–36 (22, 34)
	rpoB (285)	29 (1.0, 0.00008, 0.009)	49 (0.7–12.6)	16–44 (32, 31)

[a] Analyzed size in number of base pairs.
[b] HGDI, Hunter-Gaston discriminatory index; S^2, variance.
[c] SNP, single nucleotide polymorphism; the divergence rate is expressed as a percentage among strains of each group.

16S, *gyrB*, and *rpoB* sequencing. DNA was extracted by the boiling method. Amplification was performed using PuReTaq Ready-To-Go PCR beads (Amersham Biosciences, Buchinghamshire, United Kingdom) under the following conditions: 30 cycles at 95°C for 1 min, primer annealing (5 pM) at 55°C for 1 min, and extension at 72°C for 1.5 min. The primers used for 16S amplification and sequencing were 5′-GCTTAACACATGC AAGTCG-3′ and 5′-GAATTCCAGTCTCCCCTG-3′ (8), for *gyrB* they were 5′-GAGGTCGTCATGACCCAGCTGCA-3′ and 5′-GTCTTGGTC TGGCCCTCGAACTG-3′ (9), and for *rpoB* they were 5′-CGACCACTT CGGCAACCG-3′ and 5′-TCGATCGGGCACATCCGG-3′ (12). The amplification products were electrophoresed and purified using ExoSAP-IT reagent (GE Healthcare, NJ, USA) and sequenced by capillary electrophoresis in an ABI Prism 3100 apparatus (Applied Biosystems, Foster City, CA, USA).

16S, *gyrB*, and *rpoB* analyses. Sequences were assembled using SeqMan software (DNAStar, Inc., Madison, WI). The sequence lengths were adjusted to match the length of the shortest sequence of each species and aligned using the ClustalW algorithm (see http://www.ebi.ac.uk/Tools/clustalw2/index.html). The Hunter-Gaston discrimination index (HGDI) (13), single nucleotide polymorphisms (SNPs), haplotype numbers, and other variables (Table 1) were assessed using DnaSP software (14). A phylogenetic assessment of each species was undertaken using MEGA 4.1 software (15). Phylogenetic trees were constructed using the neighbor-joining (16), maximum-parsimony (17), and maximum-likelihood methods (18), with bootstrap analyses based on 1,000 resamplings. Branches corresponding to partitions that were reproduced in <50% of bootstrap replicates were collapsed. The evolutionary distances between the nucleotide and amino acid sequences of the *rpoB* and *gyrB* genes were determined using the Kimura 2-parameter model and the Poisson correction model (19). *N. farcinica* strain DSM 43665^T (GenBank accession no. NC_006361) was used as an outgroup (see http://nocardia.nih.go.jp/), except in *N. farcinica* analyses, in which *N. abscessus* strain DSM 44432^T (GenBank accession no. JN041489 for 16S, AB447398 for *gyrB*, and JN215593 for *rpoB*) was employed.

Species assignation and assignment of detected polymorphisms. The 16S, *gyrB*, and *rpoB* fragments sequenced for each *Nocardia* strain were compared to sequences in the GenBank database and identified using BLAST (version 2.2.10; see http://www.ncbi.nlm.nih.gov/BLAST) and the Bioinformatics Bacteria Identification (BIBI) version 0.2 software (see http://umr5558-sud-str1.univ-lyon1.fr/lebibi/lebibi.cgi). A similarity score of ≥99.0% between the 16S rRNA sequence and database sequence(s) was deemed to indicate that strains belonged to the same species (20). Species assignment was confirmed via the greatest similarity with respect to the 16S, *gyrB*, and *rpoB* sequences. To detect polymorphisms and their genome positions, the following type strains were used: *N. abscessus* DSM 44432^T (GenBank accession no. JN041489 for 16S, AB447398 for *gyrB*, and JN215593 for *rpoB*), *N. cyriacigeorgica* DSM 44484^T (GenBank accession no. GQ376180 for 16S, GQ496121 for *gyrB*, and JN215664 for *rpoB*), *N. farcinica* DSM 43665^T (GenBank accession no. GQ217499 for 16S, GQ496115 for *gyrB*, and DQ085117 for *rpoB*), and *N. nova* DSM 44481^T (GenBank accession no. GQ376190 for 16S, GQ496102 for *gyrB*, and JN215754 for *rpoB*).

RESULTS

Between 2006 and 2010, 698 clinical strains of *Nocardia* spp. were submitted to our center for identification by 16S analysis. The most prevalent *Nocardia* spp. responsible for severe clinical conditions were *N. abscessus*/*N. asiatica* (15.9%), *N. cyriacigeorgica* (24.5%), *N. farcinica* (13.6%), and the *N. nova* complex (13.0%). For each species, some 30 strains of different geographical origins and clinical backgrounds were selected (see Table S1 in the supplemental material).

Table 1 shows the corresponding HGDI and SNP variables (number, arithmetic mean and mode, etc.) for the 16S, *gyrB*, and *rpoB* genes for each species.

Analysis of 16S polymorphisms. Twenty-nine *N. abscessus* strains shared a 16S rRNA haplotype that was also identical to that of *N. abscessus* strain DSM 44432^T (GenBank accession no. JN041489) and *N. asiatica* DSM 44668^T (GenBank accession no. DQ659897). This full similarity between *N. abscessus* and *N. asiatica* is a consequence of the fact that the current 16S fragment used in routine clinical testing does not include the area of difference (which extends from the middle of the gene to the 3′-end).

Thirty *N. cyriacigeorgica* strains had a common 16S haplotype, which was also that of *N. cyriacigeorgica* DSM 44484^T (GenBank accession no. GQ376180). All 31 *N. farcinica* strains had the same 16S haplotype, which was identical to that of *N. farcinica* DSM 43665^T (GenBank accession no. GQ217499). Eight 16S haplotypes were found in 29 *N. nova* complex strains with respect to the reference strain *N. nova* DSM 44481^T (GenBank accession no. GQ376190) (similarity, ≥99%).

TABLE 2 GyrB amino acid polymorphisms identified in the *N. abscessus* ($n = 29$), *N. cyriacigeorgica* ($n = 30$), and *N. nova* ($n = 29$) strains

Species	Amino acid (nucleotides) at the indicated positions (1–4)[a]								
N. farcinica DSM 43665[b]	160	161	164	181	183	185	187	188	233
N. abscessus strains[c]									
DSM 44432[b] ($n = 11$)	10 Glu (GAG)							38 Lys (AAG)	
Mutated[b] ($n = 18$)	Lys (AAG)							Gln (CAG)	
Frequency (%)	3.4							3.4	
N. cyriacigeorgica strains[d]									
DSM 44484[c] ($n = 8$)		29 Val (GTC)	32 Ala (GCC)	49 Ile (ATC)		53 Pro (CCG)	55 Thr (ACG)		
Mutated[c] ($n = 22$)		Ile (ATC)	Asp (GAC)	Val (GTT)		Ala (GCG)	Ser (TCC), Lys (AAG)		
Frequency (%)		3.3	3.3	3.3		43.3	36.7 6.6		
N. nova strains[e]									
DSM 44481[d] ($n = 8$)					51 Gly (GGT)				101 Glu (GAG)
Mutated[d] ($n = 21$)					Asp (GAC), Ser (AGC)				Gln (CAG)
Frequency (%)					3.4, 3.4				62.0

(Continued on next page)

Analysis of *gyrB* polymorphisms. Table 2 shows the identified GyrB amino acid substitutions, their types, and frequencies, along with the HGDI values and the percent divergence between the studied strains. Figure S1 in the supplemental material shows the inferred phylogenetic tree for the 4 studied species. The topologies of the trees produced by the neighbor-joining, maximum-parsimony, and maximum-likelihood methods were almost identical (data not shown).

(i) *N. abscessus* strains. To confirm the above identifications that were made using the partial 16S sequence, several reference strains were used in *gyrB* analysis: *N. abscessus* DSM 44432^T (GenBank accession no. AB447398) and *N. asiatica* DSM 44668^T (GenBank accession no. GQ217495). All 29 studied strains showed a *gyrB* sequence that was similar (98.3 to 99.8%) to that of the reference strain *N. abscessus* DSM 44332^T. None showed the sequence seen in *N. asiatica* DSM 44668^T (similarity, 94.9 to 95.8%), which lacks three nucleotides possessed by *N. abscessus* DSM 44432^T (^{109}Ala [GCC]-^{110}Ala [GCG] is modified to ^{109}Ala [GCG]). The 14 *gyrB* haplotypes found fell into one SNP cluster group (SCG) along with *N. abscessus* DSM 44332^T, and the *gyrB* haplotype of *N. asiatica* DSM 44668^T lay outside this cluster (Table 1; see also Fig. S1 in the supplemental material). A total of 24 SNPs were detected.

The GyrB protein of all 29 strains harbored ^{109}Ala-^{110}Ala. Eighteen strains showed four amino acid substitutions with respect to the GyrB protein of *N. abscessus* DSM 44432^T (accession no. BAH10710). The most common change was the appearance of ^{87}Glu (seen as the only substitution in 13 strains but appearing in 58.6% of the strains in some combination with other changes). ^{10}Lys, ^{38}Gln, and ^{109}Ser appeared in one, one, and three strains, respectively.

(ii) *N. cyriacigeorgica* strains. Seventeen *gyrB* haplotypes were identified for *N. cyriacigeorgica* DSM 44484^T (GenBank accession no. GQ496121) (similarity values, 95.6 to 99.9%), which fell into two SCGs. Seventy-seven SNPs were detected (Table 1; see also Fig. S1 in the supplemental material).

In comparison with the *N. cyriacigeorgica* DSM 44484^T (GenBank accession no. ACV89678) GyrB reference sequence, 22 strains (73.3%) showed a total of 14 amino acid replacements in 12 locations, resulting in 9 GyrB protein types (Table 2). The amino acid changes accumulated in individual strains ranged from 1 to 7.

(iii) *N. farcinica* strains. Eleven *gyrB* haplotypes were identified for *N. farcinica* DSM 43665^T (GenBank accession no. GQ217499) (similarity, 99.2 to 99.9%) (Table 1), all of which fell into one SCG. Eleven SNPs were detected, all of which were silent.

(iv) *N. nova* complex strains. In *gyrB* analysis, several reference sequences were taken into account due to the closeness of the *N. nova* complex members (21): *N. nova* DSM 44481^T (GenBank accession no. GQ496102), *N. aobensis* DSM 44805^T (GenBank accession no. JN041378), *N. cerradoensis* DSM 44546^T (GenBank accession no. GQ496123), *N. kruczakiae* DSM 44877^T (GenBank accession no. AB450793), *N. elegans* DSM 44890^T (GenBank accession no. AB450785), and *N. africana* DSM 44491^T (GenBank accession no. JN041368). Twenty-six *gyrB* haplotypes were found with respect to *N. nova* DSM 44481^T (GenBank accession no. GQ496102) (83.1 to 99.9%), which fell into 2 SCGs. Fifty-one *gyrB* SNPs were detected. Almost 59% of the strains harbored ≥22 SNPs, with a maximum of 36 SNPs in a strain.

The GyrB protein of 21 strains showed a ^{129}Pro insertion compared to the reference strain *N. nova* DSM 44481^T (GenBank accession no. ACV89659) and 10 of the other studied strains. This difference thus divided them into two separate clades. Eleven GyrB protein types were seen. Eight amino acid changes were detected with respect to the sequence of the *N. nova* DSM 44481^T protein (Table 2). The substitutions accumulated per strain ranged from 0 to 7 (mode, 6). The combination ^{101}Gln-^{105}Glu-$^{129'}$Pro-^{131}Pro/Arg-^{157}Ala was seen in 17 strains with the ^{129}Pro insertion, and $^{129'}$Pro-^{131}Pro/Arg-^{157}Ala was seen in two strains. The triple amino acid association $^{129'}$Pro-^{131}Pro-^{157}Ala was found in the *N. aobensis*, *N. cerradoensis*, and *N. kruczakiae* reference strains. In addition, the ^{126}AAAPEH motif was seen in three GyrB protein types (possessed in 10 strains). However, in the strains with the ^{129}Pro insertion (underlined below), this motif was changed to AAAPPEP in 1 GyrB protein type (two strains), to SAAPPEP in 2 GyrB protein types (3 strains) (SAAPPEP was also seen in the *N. aobensis* DSM 44805^T and *N. kruczakiae* DSM 44877^T reference strains), to TAAPPEP in 3 GyrB protein types (12 strains) (similar to the motif TATPPEP seen in the *N. cerradoensis* DSM 44546^T reference strain), to TAAPPER in one GyrB protein type (one strain), and finally to NAAPPEP in 1 GyrB protein type (one strain) (see Fig. S3 in the supplemental material). Strains with the $^{129'}$Pro insertion and GyrB changes showed a range of similarity scores of 16S with respect to the *N. nova* DSM 44481^T of 99.2% to

TABLE 2 (Continued)

Amino acid (nucleotides) at the indicated positions (1–4)[a]										
237	238	240	257	258	259	260	262	265	292	349
87 Asp (GAC) Glu (GAA) 58.6						109 Ala (GCC) Ser (TCC) 10.3				
	106 Thr (ACC) Ala (GCC) 6.6	108 Val (GTC) Ile (ATC) 20.0	125 Ala (GCC) Glu (GAG) 23.3	126 Glu (GAG) Gln (CAG) 23.3	127 Ala (GCG) Thr (ACG) 23.3	128 Gln (CAG) Pro (CCG), Ala(GCG) 66.6 3.3	130 Thr (ACC) Ala (GCC) 50.0			
105 Asp (GAT) Glu (GAG) 62.0				126 Ala (GCG) Thr (ACC), Ser (AGC), Asn (AAC) 44.8, 10.34, 3.4			129′ Pro (CCG) 65.5	131 Hys (CAG) Pro (CCG), Arg (CGG) 62, 3.4	157 Pro (CCG) Ala (GCG) 68.9	215 Val (GTG) Phe (TTC) 3.4

[a] Amino acid polymorphism positions (modified nucleotides are underlined) with respect to the complete genome sequence of reference strain *N. farcinica* DSM 43665^T (accession no. NC_006361), *N. abscessus* DSM 44432^T (accession no. AB447398), *N. cyriacigeorgica* DSM 44484^T (accession no. GQ496121), and *N. nova* DSM 44481^T (accession no. GQ496102).

[b] All the *N. farcinica* strains harbored the same GyrB as *N. farcinica* DSM 43665^T (accession no. NC_006361).

[c] No. of GyrB types, 5; HGDI, 0.94; divergence, 0.5 to 1.4%.

[d] No. of GyrB types, 9; HGDI, 0.94; divergence, 0.4 to 4.3%.

[e] No. of GyrB types, 11; HGDI, 0.81; divergence, 0.4 to 2.6%.

100%. These percentage scores decreased (~1 to 2 percentage points) when the comparison was done with respect to the *N. aobensis*, *N. cerradoensis*, and *N. kruczakiae* reference strains.

Analysis of *rpoB* polymorphisms. Figure S2 in the supplemental material shows the phylogenetic trees produced according to the neighbor-joining method that were based on the *rpoB* gene.

(i) *N. abscessus* strains. Eighteen haplotypes were found for the 29 studied strains with respect to the *rpoB* gene of the reference strain *N. abscessus* DSM 44432^T (GenBank accession no. JN215593) (similarity range, 84.0 to 99.7%). Sixty-five SNPs were detected. Thirty-three nucleotides were changed in >85% of the strains; these SNPs divided the 29 strains into two clades, with 26 strains in the larger clade. Nineteen of these changes resulted in 15 RpoB protein types (Table 3).

(ii) *N. cyriacigeorgica* strains. Twelve *rpoB* haplotypes were identified for *N. cyriacigeorgica* DSM 44484^T (GenBank accession no. DQ085116) (95.6 to 99.7% similarity). Twenty-four SNPs were identified (Table 1; see also Fig. S2 in the supplemental material). All these SNPs were synonymous, i.e., they did not affect the amino acid sequence of the RpoB protein.

(iii) *N. farcinica* strains. Five *rpoB* haplotypes were detected with respect to the reference strain *N. farcinica* DSM 43665^T (GenBank accession no. DQ085117) (96.5 to 99.7% similarity). Nine SNPs were seen. A common RpoB protein type was seen for the entire population.

(iv) *N. nova* complex strains. Each strain showed a different *rpoB* haplotype compared to the reference strain *N. nova* DMS 44481^T (GenBank accession no. JN215754) (87.4 to 99.3% similarity). Forty-nine SNPs were detected. Twenty amino acid changes in 14 locations led to 22 RpoB protein types (Table 3).

Nucleotide sequence accession numbers. The new variants of 16S, *gyrB*, and *rpoB* sequences of each studied species were assigned the GenBank accession no. KC662119 to KC662126 and KC631324 to KC631385.

DISCUSSION

In Spain, the estimated incidence of nocardial infection in the population as a whole is 0.55/100,000 (22), although solid organ recipients, people with chronic lung disease, diabetes mellitus, or an immunodeficiency, and those undergoing corticosteroid treatment are more susceptible. Over the last 10 years, the main etiological agents of human nocardiosis in the country have been *N. abscessus* (prevalence, 13.8%), *N. cyriacigeorgica* (25.5%), *N. farcinica* (12.5%), and *N. nova* complex (14.5%) (figures were obtained from 1,024 *Nocardia* isolates from across the country). When these rates are compared with those reported in studies that examined >90 isolates, differences in the prevalences of the four species can be seen; e.g., in Guipuzcoa (northern Spain), the prevalences of the four species are 12.4, 15.0, 23.1, and 29.6 (23), in Belgium, 6.5, 14.0, 44.1, and 21.5 (24), in Taiwan, 2.0, 16.0, 8.0, and 5.3 (25), and in the United States, 5.6, 13.2, 14.0, and 28.0, respectively (26).

The classification of species and subgroups of the genus *Nocardia* has traditionally been based on the evaluation of growth characteristics, antibiotic resistance, and biochemical testing (1). However, these methods are laborious and time-consuming (6 to 8 weeks). Molecular methods provide crucial insights into identification, epidemiology, and intraspecies variability. 16S analysis is the gold standard and allows for rapid identification, but the conservation of the studied sequence means that it cannot be used to study intrageneric relationships. Therefore, other targets need to be explored (11). Typing techniques, such as pulsed-field gel electrophoresis and amplified fragment length polymorphism, have been shown to be unsatisfactory for distinguishing between *N. farcinica* strains involved in outbreaks (27).

Housekeeping genes, such as *secA1*, *hsp65*, *rpoB*, and *gyrB* (2, 7–9), may be better molecular markers than 16S for such analyses. When *secA1* was examined, several disagreements with 16S-based identifications were seen for very closely related species (e.g., *N. abscessus* and *N. asiatica*) (8). *hsp65* shows better discriminatory power at the species level but not as much as *gyrB* and *rpoB* (6, 9). Even in species with nearly identical 16S sequences, *rpoB* and *gyrB* polymorphisms are highly discriminatory (6) and show close correlations with the identification results of DNA-DNA hybridization tests (11). *gyrB* has even been shown to be useful in the description of new *Nocardia* species (2, 3).

TABLE 3 RpoB amino acid polymorphisms identified in the *N. abscessus* ($n = 29$) and *N. nova* ($n = 29$) strains

Species	RpoB amino acid (nucleotides) at the indicated positions (1–6)[a]										
N. farcinica DSM 43665	357	358	361	365	367	376	378	405	413	414	420
N. abscessus strains[b]											
DSM 44432 ($n = 11$)[c]	3 Leu (CTC)	3 Asp (GAC)	3 Gly (GGC)			13 Ile (ATC)		42 Ile (ATC)	50 Ala (GCG)		
Mutated ($n = 18$)	Ile (ATC)	Ala (GCC)	Arg (CGC)			Met (ATG)		Met (ATG)	Gly (GGG)		
Frequency (%)	3.4	3.4	3.4			10.3		89.6	86.2		
N. nova strains[d]											
DSM 44481 ($n = 0$)				2 Leu (CTG)	4 Thr (ACG)	13 Leu (CTG)	15 Val (GTC)			51 Ile (ATC)	57 Thr (ACC)
Mutated($n = 29$)				Val (GTC), Gly (GGC)	Pro (CCC)	Ile (ATC), Met (ATG)	Ile (ATC)			Val (GTC)	Ser (ACT)
Frequency (%)				93.1, 6.9	10.3	93.1, 6.9	100			96.5	34.5

(Continued on next page)

In two recent multilocus sequence typing studies involving reference strains for each species examined, five and 14 housekeeping genes were tested as markers of diversity in *Nocardia* (6, 11). However, if smaller numbers of housekeeping genes could be used in identification procedures, these analyses could be simpler and cheaper. Certainly, the analyses of *gyrB* and *rpoB* appear to identify the strains of the four main *Nocardia* species, type them at the inter- and intraspecies levels, and provide information on nocardial community diversity, a relevant event in clinical settings (28).

Because of the sequenced size and the criterion of 99% sequence similarity, the results of 16S analysis were not conclusive for *N. abscessus* and the *N. nova* complex strains. The study of the whole-gene 16S may distinguish *N. abscessus* from *N. asiatica* and allow for a more accurate species assignment of the *N. nova* complex. *Nocardia* species were previously clustered based on *gyrB* sequence similarity values of ≥93.5% (9). Here, the lower scores for the similarities of each studied species were as follows: *N. abscessus*, 98.3%; *N. cyriacigeorgica*, 94.4%; *N. farcinica*, 99.2%; and *N. nova* complex, 94.8% (higher values than the one previously proposed). In the case of *rpoB*, the recommended limit for a wide number of bacterial genera was ≥94.0% (29). *N. cyriacigeorgica* and *N. farcinica* strains showed higher minimum similarity scores, 95.6% and 98.6%, respectively. Meanwhile, *N. abscessus* and the *N. nova* complex displayed values lower than the suggested one (84.0% and 87.4%). The resulting scores are with respect to the considered type reference strains. Taking into account these criteria for the three genes, 65 strains simultaneously fulfilled them: 3 *N. abscessus* strains, 30 *N. cyriacigeorgica* strains, 31 *N. farcinica* strains, and 1 *N. nova* strain. Meanwhile, 54 strains fulfilled the 16S and *gyrB* breakpoints: 26 *N. abscessus* strains and 28 *N. nova* complex strains. Because of this dual behavior, *gyrB* and *rpoB* are complements of 16S identification, each with their usefulness and limits.

GyrB analysis was, however, very helpful in this respect. The GyrB of the strains from *N. abscessus* possessed ^{109}Ala-^{110}Ala, as did the *N. abscessus* reference strains, while the reference strains of *N. asiatica* possessed ^{109}Ala. 16S analysis was sufficient to identify the *N. nova* complex. The *gyrB* gene of the *N. nova* complex strains, however, showed great variability, as revealed by the different motifs seen in the corresponding GyrB proteins. Similar heterogeneity was previously reported in the concatenated *gyrB-16S-secA1-hsp65-rpoB* phylogenetic tree constructed in multilocus sequence analysis studies of species grouped in the *N. nova* complex (6).

Unfortunately, 16S analysis is unable to distinguish between the species within certain genera. *gyrB* and *rpoB* have also been used as identification tools, but the genetic similarity values at which identities are confirmed vary depending on the species and the target gene (30, 31). To make identifications based on *gyrB* or *rpoB*, it is necessary to analyze the sequences from large populations of the same species (30). Owing to the faster evolution of these genes, the ranges of similarity are wider than those reported for 16S (31) (in the present case, 83.4 to 99.2 for *gyrB* and 84.0 to 98.6 for *rpoB*). To date, no consensus divergence breakpoint has been fully established for the majority of bacterial species.

Typing was successful using just one of the housekeeping genes. For *gyrB* and *rpoB*, the HGDI values ranged from 0.88 to 1 in all species, except for *N. farcinica* (HGDI, 0.85 and 0.72 for *gyrB* and *rpoB*, respectively). One gene of these two was usually better, depending on the species. The abilities of *gyrB* and *rpoB* to discriminate between strains of the most prevalent species are important in confirming person-to-person transmission and for identifying new and recurrent infections. Thus, *gyrB* and *rpoB* are promising genes for use in molecular typing studies. The large number of *gyrB* and *rpoB* SNPs detected, both synonymous and nonsynonymous, provided insight into the diversity of these clinical strains (with the exception of *N. farcinica*).

After describing the genetic diversities of *gyrB* and *rpoB* in populations made up of different *Nocardia* species (7, 10, 12), the second step was to assess the intraspecies variation in a large number of strains. To our knowledge, this is the first work to explore this in a single-species scenario. Three levels of intraspecies diversity were seen: low level, which was shown by the *N. farcinica* strains with their 11 *gyrB* and 4 *rpoB* haplotypes, and single GyrB and RpoB protein types; mid-level, which was shown by the *N. abscessus* and *N. cyriacigeorgica* strains, whose large numbers of SNPs resulted in 14 *gyrB* and 18 *rpoB* and 17 *gyrB* and 12 *rpoB* haplotypes, respectively, along with 5 GyrB and 15 RpoB protein types among the *N. abscessus* strains and 9 GyrB and 1 RpoB protein types among the *N. cyriacigeorgica* strains; and high level, which was shown by the *N. nova* complex strains with their 26 *gyrB* and 29 *rpoB* haplotypes and 11

TABLE 3 (Continued)

RpoB amino acid (nucleotides) at the indicated positions (1–6)[a]												
423	425	427	429	430	434	435	438	451	456	464	465	471
60 Leu (CTG)	62 Gln (CAG)	64 Met (ATG)	66 Gln (CAG)	67 Asn (AAC)	71 Ser (TCG)	72 Gly (GGC)	75 Hys (CAC)	88 Ser (TCC)		101 Pro (CCG)	102 Ser (TCG)	108 Cys (TGC)
Met (ATG)	Val (GTG), Glu (GAG)	Leu (CTG)	Glu (GAG)	Arg (CGC)	Ala (GCC)	Ser (AGC)	Gln (CAG)	Thr (ACC)		Tyr (TAC)	Thr (ACC)	Ala (GCC)
96.5	86.2, 3.4	86.2	89.6	89.6	89.6	89.6	86.2	89.6		89.6	10.3	6.8
		64 Met (ATG) Leu (CTG)	66 Gln (CAG) Glu (GAA)	67 Asn (AAC) Arg (CGC), Ser (AGC), Hys (CAC), Cys (TGC)	71 Ser (TCG) Ala (GCG)	72 Gly (GGT) Ser (AGT)	75 Hys (CAC) Asn (AAC)	88 Ser (TCC) Cys (TGC), Thr (ACC)	93 Gly (GGC) Ala (GCT)			
		41.4	24.1	24.1, 20.7, 3.4, 3.4	80.7	89.6	27.6	10.4, 3.4	6.8			

[a] Amino acid polymorphism positions (modified nucleotides are underlined) with respect to the complete genome sequence of reference strain *N. farcinica* DSM 43665^T (accession no. NC_006361), *N. abscessus* DSM 44432^T (accession no. JN215593), and *N. nova* DMS 44481^T (accession no. JN215754). All the *N. cyriacigeorgica* strains harbored the same RpoB as *N. cyriacigeorgica* DSM 44484^T (accession no. JN215664).
[b] No. of RpoB types, 15; HGDI, 0.93; divergence, 0.9 to 12.6%.
[c] A fragment of RpoB was not indicated in the GenBank-deposited *N. abscessus* DSM 44432^T (accession no. JN215593) data; here, a consensus sequence is described.
[d] No. of RpoB types, 22; HGDI, 0.97; divergence, 1.1 to 11.1%.

GyrB and 22 RpoB protein types. The absence of the ^{129}Pro insertion differentiates *N. nova sensu stricto* from other members of the *N. nova* complex.

N. farcinica, with its few and always monomorphic SNPs, might belong to a younger lineage than the other species. In the *N. nova* strains identified as such by 16S analysis, variations in the GyrB motif ^{126}AAAPEH suggest different sublineages of recent appearance.

In conclusion, the present work shows that analysis of the *gyrB* or *rpoB* genes offers a rapid and relatively cheap means of studying strains of clinically important *Nocardia* species, in combination with 16S analysis.

ACKNOWLEDGMENTS

This work was funded by a grant to N.G. from the Instituto de Salud Carlos III (MPY-1446/11).

We are grateful to the CNM Biopolymers Unit for assistance in sequencing, to Adrian Burton for linguistic assistance in the preparation of the manuscript, and to the laboratories that submitted the *Nocardia* strains to our center for identification.

REFERENCES

1. **Brown-Elliott BA, Brown JM, Conville PS, Wallace RJ, Jr.** 2006. Clinical and laboratory features of the *Nocardia* spp. based on current molecular taxonomy. Clin. Microbiol. Rev. **19:**259–282.
2. **Jannat-Khah D, Kroppenstedt RM, Klenk HP, Spröer C, Schumann P, Lasker BA, Steigerwalt AG, Hinrikson HP, Brown JM.** 2010. *Nocardia mikamii* sp. nov., isolated from human pulmonary infections in the USA. Int. J. Syst. Evol. Microbiol. **60:**2272–2276.
3. **Moser BD, Klenk HP, Schumann P, Pötter G, Lasker BA, Steigerwalt AG, Hinrikson HP, Brown JM.** 2011. *Nocardia niwae* sp. nov., isolated from human pulmonary sources. Int. J. Syst. Evol. Microbiol. **61:**438–442.
4. **Saubolle MA, Sussland D.** 2003. Nocardiosis: review of clinical and laboratory experience. J. Clin. Microbiol. **41:**4497–4501.
5. **Woo PC, Lau SK, Teng JL, Tse H, Yuen KY.** 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin. Microbiol. Infect. **14:**908–934.
6. **McTaggart LR, Richardson SE, Witkowska M, Zhang SX.** 2010. Phylogeny and identification of *Nocardia* species on the basis of multilocus sequence analysis. J. Clin. Microbiol. **48:**4525–4533.
7. **Rodriguez-Nava V, Couble A, Devulder G, Flandrois JP, Boiron P, Laurent F.** 2006. Use of PCR-restriction enzyme pattern analysis and sequencing database for *hsp65* gene-based identification of *Nocardia* species. J. Clin. Microbiol. **44:**536–546.
8. **Kong F, Wang H, Zhang E, Sintchenko V, Xiao M, Sorrell TC, Chen X, Chen SC.** 2010. *secA1* gene sequence polymorphisms for species identification of *Nocardia* species and recognition of intraspecies genetic diversity. J. Clin. Microbiol. **48:**3928–3934.
9. **Takeda K, Kang Y, Yazawa K, Gonoi T, Mikami Y.** 2010. Phylogenetic studies of *Nocardia* species based on *gyrB* gene analyses. J. Med. Microbiol. **59:**165–171.
10. **Case RJ, Boucher Y, Dahllöf I, Holmström C, Ford-Doolittle W, Kjelleberg S.** 2007. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. Appl. Environ. Microbiol. **73:**278–288.
11. **Tamura T, Matsuzawa T, Oji S, Ichikawa N, Hosoyama A, Katsumata H, Yamazoe A, Hamada M, Suzuki K, Gonoi T, Fujita N.** 2012. A genome sequence-based approach to taxonomy of the genus *Nocardia*. Antonie Van Leeuwenhoek **102:**481–491.
12. **Devulder G, Pérouse de Montclos M, Flandrois JP.** 2005. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. Int. J. Syst. Evol. Microbiol. **55:**293–302.
13. **Hunter PR, Gaston MA.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465–2466.
14. **Librado P, Rozas J.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:**1451–1452.
15. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
16. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for constructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
17. **Fitch WM.** 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20:**406–416.
18. **Felsenstein J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:**368–376.
19. **Kimura M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
20. **Kong F, Chen SCA, Chen X, Sintchenko V, Halliday C, Cai L, Tong Z, Lee OC, Sorrell TC.** 2009. Assignment of reference 5′-end 16S rDNA sequences and species-specific sequence polymorphisms improves species identification of *Nocardia*. Open Microbiol. J. **3:**97–105.
21. **Conville PS, Brown JM, Steigerwalt AG, Lee JW, Anderson VL, Fishbain JT, Holland SM, Witebsky FG.** 2004. *Nocardia kruczakiae* sp. nov., a pathogen in immunocompromised patients and a member of the "*N. nova* complex." J. Clin. Microbiol. **42:**5139–5145.

22. **Minero MV, Marín M, Cercenado E, Rabadán PM, Bouza E, Muñoz P.** 2009. Nocardiosis at the turn of the century. Medicine (Baltimore) **88:** 250–261.
23. **Larruskain J, Idigoras P, Marimón JM, Pérez-Trallero E.** 2011. Susceptibility of 186 *Nocardia* sp. isolates to 20 antimicrobial agents. Antimicrob. Agents Chemother. **55:**2995–2998.
24. **Glupczynski Y, Berhin C, Janssens M, Wauters G.** 2006. Determination of antimicrobial susceptibility patterns of *Nocardia* spp. from clinical specimens by Etest. Clin. Microbiol. Infect. **12:**905–912.
25. **Lai CC, Liu WL, Ko WC, Chen YH, Tang HJ, Huang YT, Hsueh PR.** 2011. Antimicrobial-resistant *Nocardia* isolates, Taiwan, 1998–2009. Clin. Infect. Dis. **52:**833–835.
26. **Uhde KB, Pathak S, McCullum I, Jr, Jannat-Khah DP, Shadomy SV, Dykewicz CA, Clark TA, Smith TL, Brown JM.** 2010. Antimicrobial-resistant *Nocardia* isolates, United States, 1995–2004. Clin. Infect. Dis. **51:**1445–1448.
27. **Kalpoe JS, Templeton KE, Horrevorts AM, Endtz HP, Kuijper EJ, Bernards AT, Klaassen CH.** 2007. Molecular typing of a suspected cluster of *Nocardia farcinica* infections by use of randomly amplified polymorphic DNA, pulsed-field gel electrophoresis, and amplified fragment length polymorphism analyses. J. Clin. Microbiol. **45:**4048–4050.
28. **Muñoz J, Mirelis B, Aragón LM, Gutiérrez N, Sánchez F, Español M, Esparcia O, Gurguí M, Domingo P, Coll P.** 2007. Clinical and microbiological features of nocardiosis 1997–2003. J. Med. Microbiol. **56:**545–550.
29. **Adékambi T, Drancourt M, Raoult D.** 2008. The *rpoB* gene as a tool for clinical microbiologist. Trends Microbiol. **17:**37–45.
30. **Tayeb LA, Lefevre M, Passet V, Diancourt L, Brisse S, Grimont PA.** 2008. Comparative phylogenies of *Burkholderia*, *Ralstonia*, *Comamonas*, *Brevundimonas* and related organisms derived from *rpoB*, *gyrB* and *rrs* gene sequences. Res. Microbiol. **159:**169–177.
31. **Küpfer M, Kuhnert P, Korczak BM, Peduzzi R, Demarta A.** 2006. Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, *gyrB* and *rpoB* gene sequences. Int. J. Syst. Evol. Microbiol. **56:**2743–2751.