

**Table S1.** Comparison of the main typing characteristics (16S rDNA, *gyrB*, and GyrB) between the *N. cyriacigeorgica* soil strains from Lara State (Venezuela) and clinical samples from Spanish patients.

| Target<br>Source<br>(no. of strains)                 | 16S rRNA (514bp*)      |                            | <i>gyrB</i> (726bp*)   |                            | GyrB (242aa†)          |                            |
|------------------------------------------------------|------------------------|----------------------------|------------------------|----------------------------|------------------------|----------------------------|
|                                                      | Soil samples<br>(n=29) | Clinical samples<br>(n=30) | Soil samples<br>(n=29) | Clinical samples<br>(n=30) | Soil samples<br>(n=29) | Clinical samples<br>(n=30) |
| Haplotypes number                                    | 3                      | 1                          | 13                     | 17                         | 12                     | 9                          |
| HGDI‡                                                | 0.537                  | 0                          | 0.761                  | 0.940                      | 0.726                  | 0.786                      |
| SNPs number <sup>§</sup><br>(divergence rate)        | 2 (0-1)                | 0                          | 212 (0.0-16.4)         | 77 (0.0-5.6)               | 121 (0.0-1.8)          | 12 (0.0-0.4)               |
| SNP range per strain <sup>§</sup><br>(average, mode) | 0-1 (0.1, 0)           | 0                          | 1-109 (43, 50)         | 0-38 (28, 36/38)           | 2-72 (30, 37)          | 0-10 (6, 7)                |

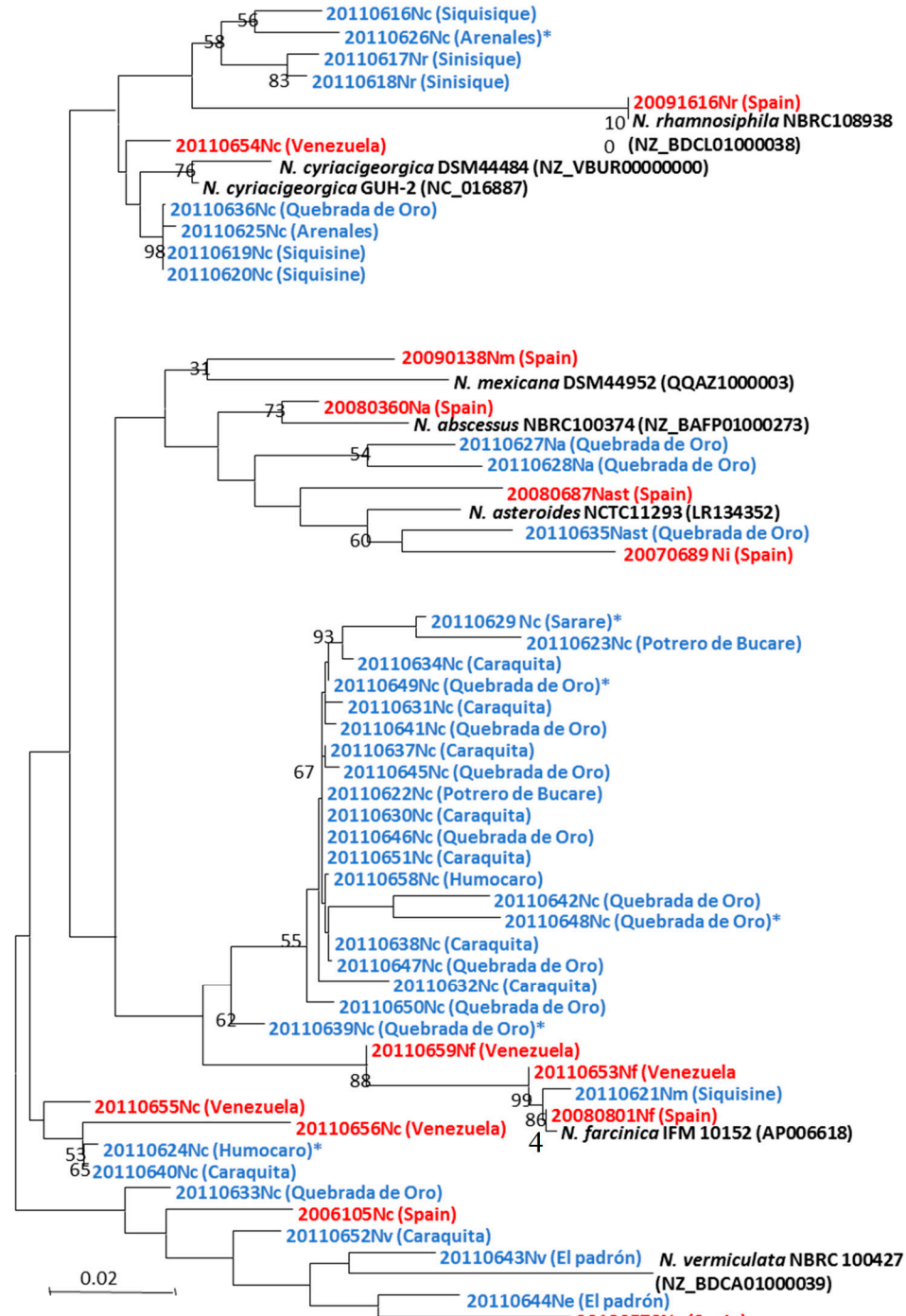
\*Analyzed sequence size in numbers of base pairs; †Analyzed sequence size in numbers of amino acids; ‡HGDI correspond to the Hunter and Gaston discrimination index; §With respect to *N. cyriacigeorgica* DSM 44484 (GenBank accession number NR\_041857 for 16S rRNA, and JN041323 for *gyrB* and GyrB).

**Table S2.** Interpretations of the analysis of the genomes of the soil *N. cyriacigeorgica* strains and the NCBI-available *N. cyriacigeorgica* genomes in terms of *gyrB*, ANI, AAI, *in silico* genome-to-genome distance similarity (GGDH; DDH-estimate), and differences in G+C content.

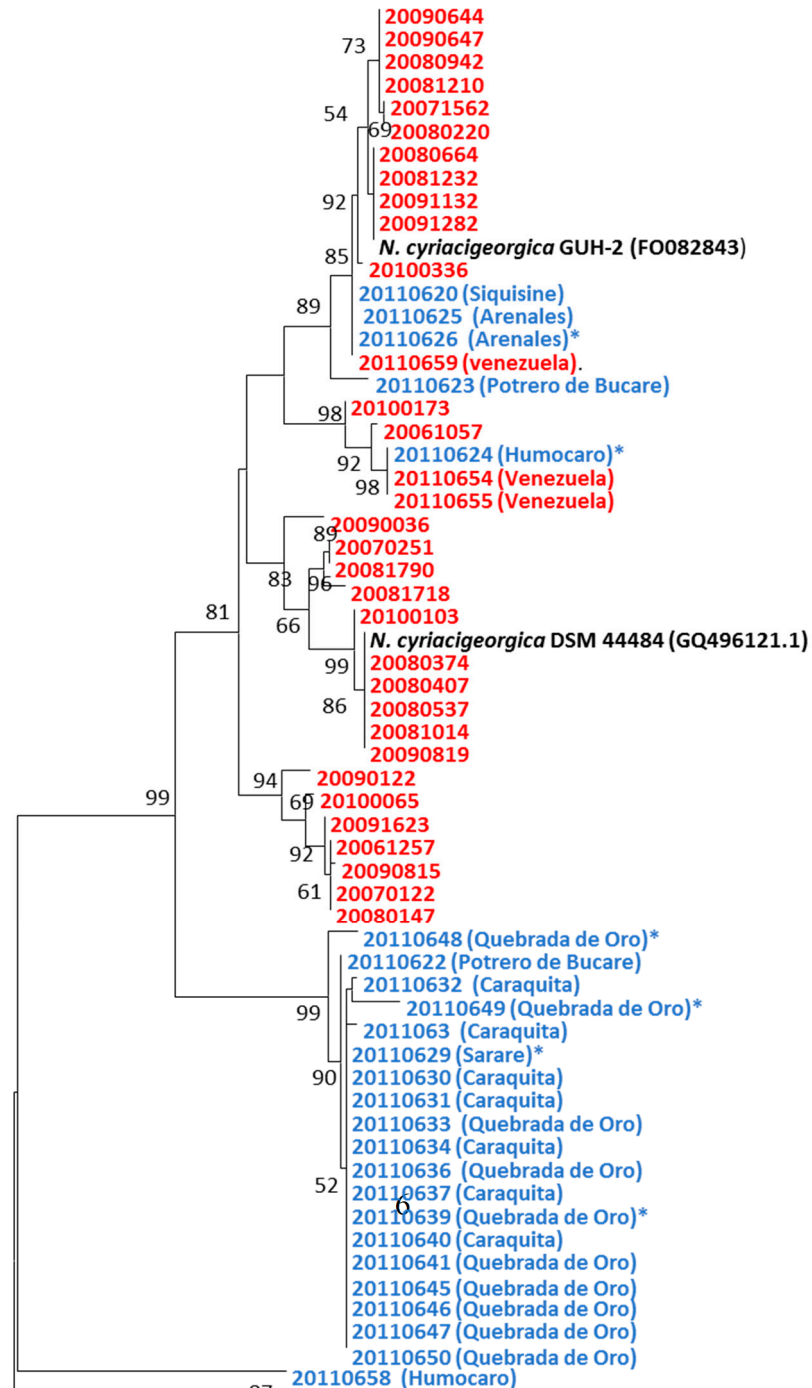
| Strain<br>(ID no./refSeq)                          | <i>gyrB</i><br>(≥93.5%) <sup>1,2</sup> |                        | TrueBac™<br>ANI <sup>3</sup>                                          | AAI profiler <sup>4</sup><br>(average<br>identity) | Type (Strain) Genome Server <sup>5</sup><br>(estimate, difference of <1% G+C) <sup>1,2</sup> and interpretation |                                                                  |
|----------------------------------------------------|----------------------------------------|------------------------|-----------------------------------------------------------------------|----------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
|                                                    | GUH-2                                  | DSM 44484 <sup>T</sup> | GUH-2                                                                 | GUH-2                                              | GUH-2                                                                                                           | NBRC 100375 (alternative designation<br>DSM 44484 <sup>T</sup> ) |
| GUH-2 ( <a href="#">NC_016887</a> )                | --                                     | 94.77                  | Not determined <sup>6</sup>                                           | --                                                 | --                                                                                                              | 0.17<br>Potential new species                                    |
| DSM 44484 <sup>T</sup>                             | 94.77                                  | --                     | Not determined <sup>6</sup>                                           | 0.93                                               | 0.15                                                                                                            | 0.03<br>Belongs to <i>N. cyriacigeorgica</i>                     |
| 20110624<br>(JAAGVC000000000)                      | 95.32                                  | 95.98                  | <i>Nocardia</i> sp. nov. genomic<br>evidence                          | 0.80                                               | 0.05                                                                                                            | 90.26<br>65.9 - 73.5, 0.13<br>Potential new species              |
| 20110626<br>(JAAGVB000000000)                      | 99.45                                  | 95.34                  | 98.34<br><i>Nocardia</i> sp. nov.                                     | 0.85                                               | 0.11                                                                                                            | 87.76<br>61.2 - 68.6, 0.06<br>Potential new species              |
| 20110629<br>(JAAGVA000000000)                      | 93.12                                  | 92.37                  | <i>Nocardia</i> sp. nov<br><i>cyriacigeorgica</i> GUH-2<br>(FO082843) | 0.76                                               | 1.75                                                                                                            | 44.8 - 51.7, 1.58<br>Potential new species                       |
| 20110639<br>(JAAGUZ000000000)                      | 93.12                                  | 92.6                   | 87.78<br><i>Nocardia</i> sp. nov.                                     | 0.76                                               | 1.43                                                                                                            | 45.2 - 52.1, 1.26<br>Potential new species                       |
| 20110648<br>(JAAGUY000000000)                      | 92.85                                  | 92.85                  | 87.78<br><i>Nocardia</i> sp. nov.                                     | 0.75                                               | 1.42                                                                                                            | 45.0 - 51.9, 1.25<br>Potential new species                       |
| 20110649<br>(JAAGUX000000000)                      | 92.16                                  | 91.20                  | Not determined <sup>6</sup>                                           | 0.75                                               | 1.46                                                                                                            | 44.7 - 51.5, 1.28<br>Potential new species                       |
| 3012STDY6756504<br>( <a href="#">NZ_LR215973</a> ) | 96.84                                  | 97.93                  | Not determined <sup>6</sup>                                           | 0.82                                               | 0.13                                                                                                            | 80.9 - 87.8, 0.04<br>Potential new species                       |
| EML 446<br>( <a href="#">NZ_VBUT000000000</a> )    | 97.08                                  | 97.03                  | Not determined <sup>6</sup>                                           | 0.80                                               | 0.14                                                                                                            | 66.0 - 73.5, 0.03<br>Potential new species                       |
| EML 1456<br>( <a href="#">NZ_VBUU000000000</a> )   | 96.94                                  | 96.95                  | Not determined <sup>6</sup>                                           | 0.75                                               | 0.34                                                                                                            | 61.1 - 68.5, 0.17<br>Potential new species                       |
| MDA3349 ( <a href="#">NZ_CP026746</a> )            | 96.55                                  | 99.84                  | Not determined <sup>6</sup>                                           | 0.82                                               |                                                                                                                 | 85.4 - 91.5, 0.09<br>Belongs to <i>Nocardia cyriacigeorgica</i>  |
| MDA3732<br>( <a href="#">NZ_PSZF000000000</a> )    | 94.32                                  | 96.97                  | Not determined <sup>6</sup>                                           | 0.82                                               |                                                                                                                 | 83.9 - 90.3, 0.03<br>Belongs to <i>Nocardia cyriacigeorgica</i>  |

<sup>1</sup>The reference breakpoints (for *gyrB*, ANI, AAI, and DDH-estimate and differences in G+C content) for assigning membership to a specific species are indicated in brackets. <sup>2</sup>Values lower than the reference breakpoints, suggestive of a distinct species, are indicated in italics. <sup>3</sup>Average nucleotide identity (ANI) determined using the TrueBac™ system (<https://www.truebacid.com/genome/>). <sup>4</sup>Average amino acid identity (AAI) determined by the AAI-profiler system

(<http://ekhidna2.biocenter.helsinki.fi/AAI>). <sup>5</sup>DDH-estimate and the difference in genomic GC content determined by the Type Strain Genome Server (TYGS) method (<https://tygs.dsmz.de>). <sup>6</sup>TrueBacTM, is a commercial web server (only the above-mentioned strains could be examined without cost).



**Figure S1.** Phylogenetic ML tree based on the MLSA analysis (*gyrB*-16S rRNA-*secA*-*hsp65* genes) of the 38 *N. cyriacigeorgica* strains from soil (in blue), 5 *Nocardia* clinical strains from Venezuelan patients and nine Spanish clinical strains representing each species present in soil (in red), plus the type strains (in black). The asterisk indicates the strains selected for WGS. Na stands for *N. abscessus*, so on Nast for *N. asteroides*, Nc for *N. cyriacigeorgica*, Ne for *N. elegans*, Nf for *N. farcinica*, Ni for *N. ignorata*, Nm for *N. mexicana*, Nn for *N. nova*, Nr for *N. rhamnosiphila*, and Nv for *N. vermiculata*. The reliability of the topologies was assessed by the bootstrap method with 1000 replicates.



**Figure S2.** Phylogenetic relationships of the 29 Venezuelan *N. cyriacigeorgica* soil strains (in blue), three Venezuelan and 30 Spanish *N. cyriacigeorgica* clinical strains (in red), as revealed by their *gyrB* genes. The reliability of the ML topologies was assessed by the bootstrap method (1000 replications). The asterisk indicates the strains selected for WGS.

|          |                                                                                |                        |       |       |       |       |       |       |       |       |
|----------|--------------------------------------------------------------------------------|------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
|          | 10                                                                             | 20                     | 30    | 40    | 50    | 60    | 70    | 80    | 90    | 100   |
| GUH-2    | MTETTLPPNGGAGDRIEPVDIQNEMQSSYIDYAMSVIVGRALPDVRDGLKPVHRRVLYAMYDNGYRPDRGYVKSARPV | ETMGNYPHGHGDASIYDTLVRM |       |       |       |       |       |       |       |       |
| DSM44484 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110626 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110624 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110629 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | S     | ..... |
| 20110639 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | S     | ..... |
| 20110648 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | S     | ..... |
| 20110649 | .....                                                                          | .....                  | ..... | ..... | ..... | ..... | ..... | ..... | S     | ..... |

|          |                                                                                    |            |         |       |       |       |       |       |       |       |
|----------|------------------------------------------------------------------------------------|------------|---------|-------|-------|-------|-------|-------|-------|-------|
|          | 110                                                                                | 120        | 130     | 140   | 150   | 160   | 170   | 180   | 190   | 200   |
| GUH-2    | AQPWSLRYPLVDGQGNFGSRGNDGAAAMRYTECRITPLAMELLREIDHETVDFTPNYDGRSQEPVVLPSRVPNLLMNGSNGI | AVGMATNIPP | HNLELAE |       |       |       |       |       |       |       |
| DSM44484 | .....                                                                              | .....      | .....   | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110626 | .....                                                                              | .....      | .....   | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110624 | .....                                                                              | .....      | .....   | ..... | ..... | ..... | ..... | ..... | ..... | D     |
| 20110629 | .....                                                                              | .....      | .....   | M     | ..... | V     | ..... | T     | ..... | A     |
| 20110639 | .....                                                                              | .....      | .....   | M     | ..... | V     | ..... | T     | ..... | A     |
| 20110648 | .....                                                                              | .....      | .....   | M     | ..... | V     | ..... | T     | ..... | A     |
| 20110649 | .....                                                                              | .....      | .....   | M     | ..... | V     | ..... | T     | ..... | A     |

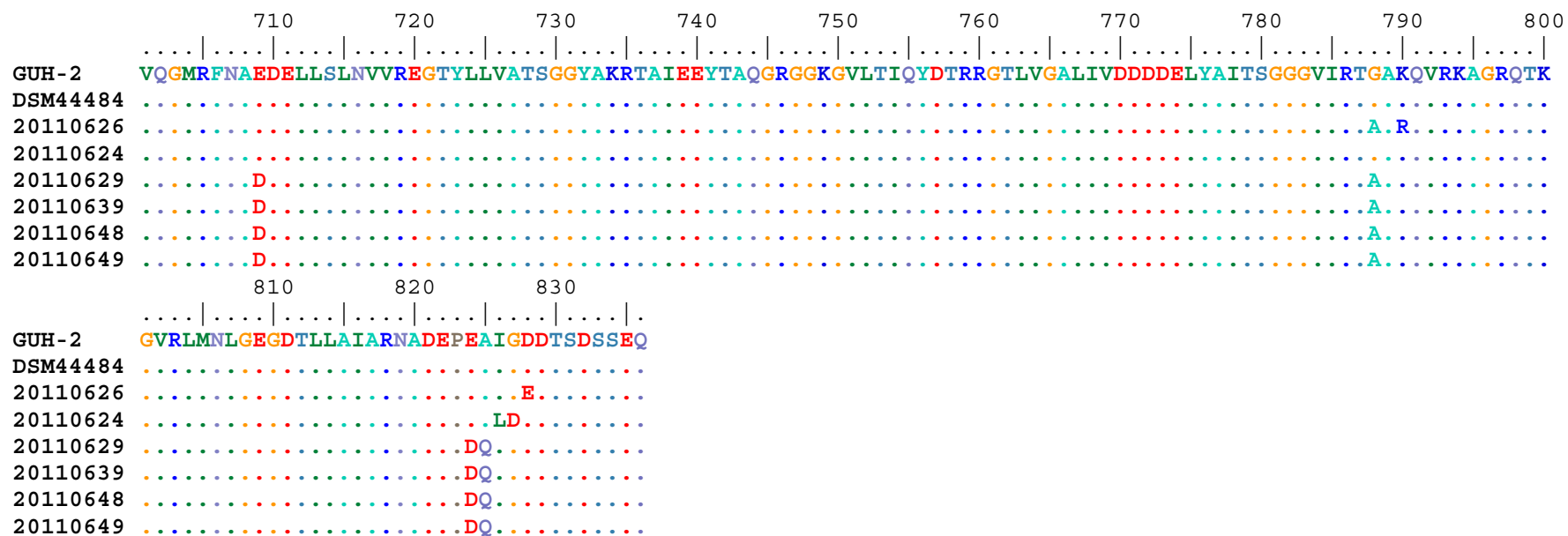
|          |                                                                                                      |       |       |       |       |       |       |       |       |       |
|----------|------------------------------------------------------------------------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|          | 210                                                                                                  | 220   | 230   | 240   | 250   | 260   | 270   | 280   | 290   | 300   |
| GUH-2    | AIYWALENYDADEEATLAACMERVKGPDFPTAGLIVGGQGIHDAYTTGRGSIRMRGVVEIEEDTRGRTTIVITELPYQVNTDNFINSIAEQVKDGKIAGI |       |       |       |       |       |       |       |       |       |
| DSM44484 | .....                                                                                                | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110626 | .....                                                                                                | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110624 | .....                                                                                                | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 20110629 | .....                                                                                                | D     | E     | ..... | S     | ..... | ..... | S     | ..... | ..... |
| 20110639 | .....                                                                                                | D     | E     | ..... | S     | ..... | ..... | S     | ..... | ..... |
| 20110648 | .....                                                                                                | D     | E     | ..... | S     | ..... | ..... | S     | ..... | ..... |
| 20110649 | .....                                                                                                | D     | E     | ..... | S     | ..... | ..... | S     | ..... | ..... |

|       |                                                                                                      |     |     |     |     |     |     |     |     |     |
|-------|------------------------------------------------------------------------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|       | 310                                                                                                  | 320 | 330 | 340 | 350 | 360 | 370 | 380 | 390 | 400 |
| GUH-2 | SDIHDESSDRAGMRIVVTVKRDAVAKVVLNNLYKHTQLQTSFGANMLSIVDGVPRTLRLDQMIRYVVEHQLEVIVRRTRYLLRKAEERAHILRGLVKALD |     |     |     |     |     |     |     |     |     |



DSM44484 .....  
 20110626 .....  
 20110624 .....  
 20110629 .....F.....  
 20110639 .....F.....  
 20110648 .....F.....  
 20110649 .....F.....  
 .....410.....420.....430.....440.....450.....460.....470.....480.....490.....500  
 GUH-2 ALDEVIALIRRSANTDTARTGLMQLLDIDEIQATAILDMQLRRLSALERQKIIDELAKIEAEIADLKDILDKPERQRAIVRDELAIEIVEKYGDERRTKII  
 DSM44484 .....V.....G.....  
 20110626 .....  
 20110624 .....V.....  
 20110629 .....D.E.....V.....E.....  
 20110639 .....D.E.....V.....E.....  
 20110648 .....D.E.....V.....E.....  
 20110649 .....D.E.....V.....E.....  
 .....510.....520.....530.....540.....550.....560.....570.....580.....590.....600  
 GUH-2 AADGDVADEDLIAREDVVVTITETGYAKRTKTDLYRSQKRGGKGVQGAGLKQDDLVLKHFFITSTHDWLLFFTNGKRVYRAKAYELPEANRTARGQHVANL  
 DSM44484 .....  
 20110626 .....  
 20110624 .....  
 20110629 .....S.....S.....  
 20110639 .....S.....S.....  
 20110648 .....S.....S.....  
 20110649 .....S.....S.....  
 .....610.....620.....630.....640.....650.....660.....670.....680.....690.....700  
 GUH-2 LAFQPDEKIAQIIRIKSYEDAPYLVLATKNGLVKKSRLSDFDSNRSGGIVAVNLRDNDELVGAVLCSADDDLLVSAQGQSIRFAATDEALRPMGRATSG  
 DSM44484 .....  
 20110626 .....  
 20110624 .....  
 20110629 .....S.....  
 20110639 .....S.....  
 20110648 .....S.....  
 20110649 .....S.....



**Figure S3.** Amino acid sequences of GyrA alignment from *N. cyriacigeorgica* genome reference strain GUH-2, type strain DSM44484<sup>T</sup>, CNM20110626, and CNM20110624 with major allele GyrA1, and the strains of the soil-only cluster (CNM20110629, CNM20110639, CNM20110648, and CNM20110649) with major allele GyrA2.

## Material and methods. Soil *Nocardia* spp. culture and presumptive identification

The collection, culture, and presumptive identification were carried out by microbiologists of the Microbiology and Parasitology Department of the Faculty of Pharmacy and Bioanalysis (Los Andes University, Mérida-Venezuela).

This appendix includes the method previously published in Spanish by Ramírez A, Blanco M and García E in 2003 titled "Biogeografía de *Nocardia*: Estudio de la población edáfica de *Nocardia* en diversas zonas climáticas del Estado de Lara". *Rev Soc Ven Microbiol.* 23:1-7

Soil samples each weighing  $\approx 300$  g were collected randomly from a depth of 10 cm, after removing the surface. The samples were transferred aseptically to sterile bags and maintained at 4°C after recollection. The loam and the sandy-clay loam were sifted through (2mm/10 mesh), handily homogenized for 15 minutes and dried at 105° C for 3 days. Each 50g of this sample was added to 450ml of charcoal-free broth [1] and gently shaken for 20 minutes. Later, the sediment of 10 minutes conforming to the  $10^{-1}$  dilution was serially diluted to  $10^{-8}$  [2].

From each  $10^{-8}$  dilution, five aliquots of (1 ml each tube) were incubated at 34 °C for five weeks including a paraffin dipstick and nine milliliters of CFB. As previously is described, the paraffin batting technique is based on the ability of the *Nocardia* genus to use paraffin as an only Carbon source [3]. The emergent colonies were grown in nutrient agar and identified as *Nocardia* genus by: phenotypic analysis of macro- and microscopic characteristics, Gram stain and degree of fragmentation of the hyphae (microculture), lysozyme resistance, presence of methyl esters of mycolic acids by thin-layer chromatography on silica gel [4]; and by the analysis of hydrolysis or decomposition of casein, xanthine, hypoxanthine, tyrosine and urea [5].

## References

1. McClung N. Isolation of *Nocardia asteroides* from soils. *Mycology.* **1960.** 52: 154-156.
2. Ramirez, A.; Blanco, M.; Garcia, E. Biogeografía de *Nocardia*: Estudio de la población edáfica de *Nocardia* en diversas zonas climáticas del estado Lara, Venezuela. *Rev Soc Ven Microbiol.* **2003.**23:142-147
3. Shawar, R.M; Moore, D.G.; LaRocco, M.T. Cultivation of *Nocardia* spp. on chemically defined media for selective recovery of isolates from clinical specimens. *J Clin Microbiol.* **1990.** 28:508-512.
4. Hamid, M.E.; Minnikin, D.E.; Goodfellow, M.A. Simple chemical test to distinguish mycobacteria from other mycolic-acid-containing actinomycetes. *J Gen Microbiol.* **1993.** 139:2203-2213.
5. Brown-Elliott BA, Brown JM, Conville PS, Wallace RJ. Clinical and laboratory features of the *Nocardia*. **2006**