Defining the Species *Micromonospora saelicesensis* and *Micromonospora noduli* Under the Framework of Genomics

Raúl Riesco 1, Lorena Carro 1, Brenda Román-Ponce 1, Carlos Prieto 2, Jochen Blom 3, Hans-Peter Klenk 4, Philippe Normand 5 and Martha E. Trujillo 1*

1 Departament of Microbiology and Genetics, Edificio Departamental, University of Salamanca, Salamanca, Spain; 2 Servicio de Bioinformática, NUCLEUS, Edificio I-D+i, University of Salamanca, Salamanca, Spain; 3 Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany; 4 School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; 5 Centre National de la Recherche Scientifique-UMR5557 Ecologie Microbienne, Université de Lyon, Université Lyon1, Villeurbanne, France

The type isolates of species *Micromonospora saelicesensis* and *Micromonospora noduli* are Gram-stain positive actinobacteria that were originally isolated from nitrogen fixing nodules of the legumes *Lupinus angustifolius* and *Pisum sativum*, respectively. These two species are very closely related and questions arise as to whether they should be merged into a single species. To better delineate the relationship of *M. saelicesensis* and *M. noduli*, 10 strains isolated from plant tissue (nodules and leaves) and identified by their 16S rRNA gene sequences as either *M. saelicesensis* or *M. noduli*, based on a cut-off value of ≥99.5% were selected for whole-genome sequencing and compared with the type strains of *M. saelicesensis* Lupac 09T and *M. noduli* GUI43T using overall genome relatedness indices (OGRI) which included ANI, OrthoANI and digital DNA-DNA hybridization. Whole- and core-genome phylogenomic analyses were also carried out. These results were compared with the topologies of the 16S rRNA and *gyrB* gene phylogenies. Good correlation was found between all trees except for the 16S rRNA gene. Overall results also supported the current classification of *M. saelicesensis* and *M. noduli* as separate species. Especially useful was the core-genome phylogenetic analyses based on 92 genes and the dDDH results which were highly correlated. The importance of using more than one strain for a better definition of a species was also shown. A series of *in vitro* phenotypic assays performed at different times were compared with *in silico* predictions based on genomic data. *In vitro* phenotypic tests showed discrepancies among the independent studies, confirming the lack of reproducibility even when tests were performed in the same laboratory. On the other hand, the use of *in silico* predictions proved useful for defining a stable phenotype profile among the strains analyzed. These results provide a working framework for defining *Micromonospora* species at the genomic and phenotypic level.

Keywords: *Micromonospora*, genome sequencing, phylogenomic analysis, nitrogen-fixing nodule, taxonomy, species delimitation
INTRODUCTION

To define a species, current prokaryotic taxonomy integrates multiple aspects of a microorganism that include phenotypic and genotypic data (Chun and Rainey, 2014). This approach, known as polyphasic taxonomy (Colwell, 1970; Vandamme et al., 1996) has contributed for several decades to improve classification and identification schemes, however, its limitations and pitfalls, particularly in relation to reproducibility of some methods and/or the difficulty of data storage have been timely addressed (Sutcliffe et al., 2012; Vandamme and Peeters, 2014; Thompson et al., 2015).

The introduction and improvement of cost-effective whole-genome sequencing methods provide a new working framework. Unlike DNA-DNA hybridization, that was heralded as the “golden standard” for defining genomic species in 1987 (Wayne et al., 1987), genomic data can be stored and made available to the scientific community for subsequent comparisons (Chun and Rainey, 2014). Furthermore, genomic data can also be used to predict phenotypic traits which can then be tested in the laboratory, reducing the need to perform labor-intensive and non-reproducible tests (Sutcliffe et al., 2012; Amaral et al., 2014).

Another problem is the definition of species based on single-strain representatives. This approach does not allow the recognition of intra-species diversity and limits the proposal for a sound and testable definition of a prokaryotic species. While the number of genomes representing bacterial species, in most cases, include only the type strain, it is also necessary to study several members within the same species to better understand intraspecies variation. Unfortunately, for most species, only one strain (the type strain) has been described and single-strain descriptions hinder the possibility for such studies.

The genus Micromonospora represented by Gram-stain positive, filamentous and sporulating actinobacteria, belongs to the family Micromonosporaceae of the order Micromonosporales in the phylum Actinobacteria (Genilloud, 2015a,b,c). The type species of the genus is Micromonospora chalcea and currently includes 81 species with validly published names http://www.bacterio.net/micromonospora.html (Parte, 2014). Most of these species have been described in the past 10 years with representative strains isolated from diverse habitats such as soil (Li and Hong, 2016; Lee and Whang, 2017), aquatic habitats (Trujillo et al., 2005; de Menezes et al., 2012), plant tissues (Carro et al., 2012, 2013; Kittiwongwattana et al., 2015; Trujillo et al., 2015; Kaewkla et al., 2017) and other environments (Hirsch et al., 2004; Nimaichand et al., 2013; Lin et al., 2015). Recently, a revised classification of the genus Micromonospora based on genome sequence data has been proposed (Carro et al., 2018).

In 2007, three strains recovered from internal nodular tissue of the plant Lupinus angustifolius were formally described as Micromonospora saelicesensis (Trujillo et al., 2007). Recent studies have shown that this species is widely distributed in legumes (e.g., Trifolium, Lupinus, Pisum, etc.), especially in nodules (Trujillo et al., 2015). Micromonospora noduli described as a single representative, strain GU143T, isolated from the nodular tissue of Pisum sativum was found to be closely related to M. saelicesensis (Carro et al., 2016). While the DNA-DNA hybridization value of 63.4% (62.3 reciprocal) is below the accepted threshold of 70% (Wayne et al., 1987) these two species share many features, and the question arises whether they should be merged into a single species. Therefore, this study was designed to determine the level of taxonomic relationship between ten strains initially identified as M. saelicesensis or M. noduli using a 16S rRNA gene sequence comparison with a similarity threshold of $\geq$99.5%. Draft whole-genome sequences were obtained for all strains, including the type strain M. noduli GU143T and data was analyzed using a combination of overall genome related indices (OGRI) and phylogenomic analyses. Furthermore, to obtain information about intra-species variation, especially at the phenotypic level, these studies were complemented with physiological and biochemical data. The integration of all these studies support the current status of M. saelicesensis and M. noduli as different species and the approach presented in this work provides a good method for the definition of species in the genus Micromonospora.

MATERIALS AND METHODS

Isolation of Strains

The list of Micromonospora strains used in this study is given in Table 1. All strains, except for LAH08, were isolated from nitrogen fixing nodules of five different legumes between 2003 and 2015 as described previously (Trujillo et al., 2010). Isolation of strain LAH08 from the leaves of L. angustifolius was done after surface sterilization of the plant material by immersing in 70% ethanol (v/v) for 1 min, transferred to 3.5% w/v sodium hypochlorite solution for 2 min and rinsed five times with sterile distilled water. Sample was crushed with a sterile homogenizing pestle in a microtube and the resulting slurry plated onto yeast extract-humic acid agar (de la Vega, 2010).

16S rRNA Gene Sequencing and Analysis

DNA extraction (REDEExtract-N-Amp Plant PCR kit [Sigma]), 16S rRNA gene amplification and sequencing were carried out as previously explained (Trujillo et al., 2010). Assembled sequences were compared against the Ezbiocloud Database (Yoon et al., 2017) and other public platforms (Genbank, EMBL, etc.), and aligned with ClustalX2 (Thompson et al., 1997). Catellatospora citrea IMSNU 22008T, a member of the family Micromonosporaceae was used as outgroup.

Phylogenetic analyses were performed using MEGA (v 7.0.14) (Kumar et al., 2016); distances were calculated with the Kimura 2-parameter and tree topologies were based on the Maximum Likelihood algorithm (Felsenstein, 1981). Total analysis included 1318 positions and a bootstrap (Felsenstein, 1985) sampling of 1,000.

Whole-Genome Sequencing, Assembly, and Annotation

DNA was isolated from 1 g of bacterial cultures grown in ISP 2 broth (Shirling and Gottlieb, 1966) at 28°C for 5–7 days. Cell lysis was done in 5 ml EC buffer containing 60 µl of lysozyme (300 mg/ml, Sigma-Aldrich, USA) and 50 µl of mutanolysin (1,000 U/ml), and incubated at 37°C for 90 min. Five ml of 2% SDS
(w/v) and 200 µl of proteinase k (10 mg/ml) were added with gentle mixing for protein precipitation and incubated at 55°C for 3 h. Samples were extracted with phenol:chloroform:i soamyl alcohol (25:24:1 v/v), treated with 35 µl of RNase A (10 mg/ml) and precipitated with 70% ethanol. Draft genome sequences were determined by MiSeq (300 bp paired end) (Chunlab, Inc.). Libraries were prepared using TruSeq DNA LT Sample Prep kit (Illumina, San Diego, CA, USA) for the Illumina system (Chunlab, Inc.). Illumina sequencing data were assembled with SPAdes 3.10.1 (Algorithmic Biology Lab, St. Petersburg Academic University of the Russian Academy of Sciences). Protein-coding sequences (CDSs) were predicted by Prodigal 2.6.2 (Hyatt et al., 2010). Genes coding for tRNA were searched using tRNAscan-SE 1.3.1 (Schattner et al., 2005). The rRNA and other non-coding RNAs were searched by a covariance model search with CRISPR-finder, available in http://crispr.i2bc.paris-saclay.fr (Grissa et al., 2007) using default parameters. EDGAR 2.0 platform (Blom et al., 2016) was used to calculate the core genome, dispensable genome and singleton genes.

### gyrB Gene Phylogeny
GyrB nucleotide gene sequences extracted from whole genome sequence data or downloaded from the public databases were used to construct a Maximum-Likelihood phylogenetic tree based on Kimura 2-parameter, using 1001 nucleotide positions and a bootstrap value of 1000. Catellatospora citrea DSM 44097T, a member of the family Micromonosporaceae was used as outgroup.

### OGRI Analyses
Average Nucleotide Identity (ANI) (Goris et al., 2007) and OrthoANI (Lee et al., 2016) comparisons were made with the Orthologous Average Nucleotide Identity Tool (OAT) v0.93 https://www.ezbiocloud.net/tools/orthoani. Digital DNA-DNA hybridizations (dDDH) and G+C content differences were obtained with Genome to Genome Distance Calculator (GGDC) v2.0 available at https://ggdc.dsmz.de/ggdc.php#. Using the recommended settings (Meier-Kolthoff et al., 2013b). A dDDH heatmap was constructed using ComplexHeatmap R package v 1.17.1 (Gu et al., 2016).
Whole-Genome Phylogenomic Analyses

Genome Blast Distance Phylogeny (GBDP) was used to calculate the intergenomic distances based on whole proteomes (Meier-Kolthoff et al., 2013b). Calculation of a distance matrix was done using the on-line GGDC server, with BLAST+ and recommended formula 2 (optimized for draft genome sequences) (Meier-Kolthoff et al., 2013b). Phylogenetic trees were constructed with FastMe tool (Lefort et al., 2015). Genome sequence accession numbers are provided in Table S1.

UBCG Phylogenomic Analysis

Ninety-two bacterial core genes based on the Up to date Bacterial Core Gene (UBCG) tool, https://www.ezbiocloud.net/tools/ubcg were used for phylogenomic tree reconstruction using default parameters (Na et al., 2018). The selection of the representative genes was based on 1429 complete genome sequences, covering 28 phyla and providing a set of genes present in the majority of the genomes or highly conserved single copy genes (Na et al., 2018).

Physiology

A set of physiological and biochemical tests reported to differentiate between M. saelicesensis and M. noduli were carried out; these included carbon source utilization, determination of enzymatic activity, NaCl and pH tolerance, temperature range growth and degradation of starch, Tween 20, Tween 80, tyrosine, and urea (Carro et al., 2016). All tests were done in triplicate.

Several carbon sources (19) were tested in vitro at different times in the laboratory (2016 and 2017) and compared with the results of the original description of M. saelicesensis to check for reproducibility (Trujillo et al., 2007). Draft-genome data was screened for genes coding for proteins for carbon metabolism of the carbon sources assayed.

Biolog Characterization

To generate phenotypic fingerprints of 71 carbon source utilization and 23 chemical sensitivity assays, the strains were tested at 28°C using GEN III Microplates in an Omnilog device (BIOLOG Inc., Hayward, CA, USA). The reference strains Micromonospora saelicesensis Lupac 09T and Micromonospora noduli GUI43T were included for parallel comparison. One week old cells were suspended in an inoculating fluid (IF C) provided by the manufacturer and inoculated in the GEN III Microplates at a cell density of 80% transmittance. Phenotype microarray mode was used to measure respiration rates yielding a total running time of 7 days using two independent replicates for each strain. Data were recovered and analyzed using the opm package for R, v.1.0.6 (Vaas et al., 2012, 2013). Clustering analyses of the phenotypic microarrays were constructed using the pvclust package for R v.1.2.2 (Suzuki and Shimodaira, 2015). Distinct behaviors between the two repetitions in the reactions were regarded as ambiguous.

RESULTS

16S rRNA Gene Sequence Analysis

The 16S rRNA gene sequences were used to determine the nearest phylogenetic neighbors based on overall sequence similarity in relation to currently described Micromonospora species. In all cases, the closest species were M. saelicesensis and M. noduli with similarity values of 99.2–100% (Table 1).

A phylogenetic tree constructed with the new sequences and those of 81 Micromonospora species described to date, distributed the 10 strains into two clusters: Group I contained the type strain M. saelicesensis Lupac 09T and the isolates PSN13, GAR06, PSN01, Lupac 06, GAR05, and Lupac 07. Group II was formed with ONO 86, ONO23, LAH08, and MED15 and M. noduli GUI43T (Figure 1). The topology clearly showed the close relationship between the two groups as visualized by the branch lengths which were almost inexistent. Group II (M. noduli) also showed a close relationship with the type strains of Micromonospora profundi isolated from a deep marine sediment (Veyisoglu et al., 2016) and Micromonospora ureilytica isolated from Pisum sativum (Carro et al., 2016), also recovered in this group. Reported DDH values between M. saelicesensis and M. ureilytica and M. profundi (Veyisoglu et al., 2016) were 28.4 and 56.9%, respectively. A DDH value of 50.9% was found between M. noduli G143T and M. ureilytica GUI23T (Carro et al., 2016).

gyrB Gene Phylogeny

The phylogenetic tree constructed with the gyrB gene sequences showed a similar topology to the 16S rRNA gene tree with respect to the study strains (Figure 2). These were recovered in their respective groups defined in the 16S rRNA gene tree. The exception was strain Lupac 07, recovered in the M. noduli cluster (Group II) and this rearrangement was supported by a bootstrap value of 99%. This strain was originally classified as M. saelicesensis (Trujillo et al., 2007). The positions of M. profundi DS 3010T and M. ureilytica GUI23T also changed and moved out of the M. noduli cluster. As previously noted (Garcia et al., 2010; Carro et al., 2012), the gyrB gene phylogeny yielded a better resolution as observed by slightly larger distance branches, however, the topology of the remaining type strains was very different from that obtained using the 16S rRNA gene. This phylogeny was very similar to a tree constructed using a concatenated set of five housekeeping genes as proposed previously (Carro et al., 2012) (data not shown).

Comparative Genomic Characteristics

Eleven high quality draft genomes (depth >100X), including that of the type strain M. noduli GUI43T were obtained. General genome characteristics of the sequenced strains are provided in Table 2. Genome sizes varied from 6.8 to 7.4 Mb, the largest genome being that of strain PSN13. The G+C mol% among all strains was very homogeneous. The values of the M. saelicesensis group ranged from 71.1 to 71.2% while the strains in the M. noduli cluster, including Lupac 07, varied from 70.9 to 71.1%. As observed, the G+C mol% values between the two species was less than 1%.
FIGURE 1 | Maximum-likelihood phylogenetic tree based on 165 rRNA gene sequences showing the relationships between 81 Micromonospora type and the study strains. Distances were calculated with the Kimura 2-parameter. The tree is based on 1,318 nt. Bootstrap percentages ≥50% (1,000 samplings) are shown at nodes. Bar, 0.02 substitutions per nucleotide.
FIGURE 2 | Maximum-likelihood phylogenetic tree based on gyrB gene sequences. A total of 76 Micromonospora type strains and 10 non-type strains have been used for the analysis. Distances were calculated with the Kimura 2-parameter, using 999 nucleotide positions and a bootstrap of 1,000. Bar, 0.05 substitutions per nucleotide.
The number of coding DNA sequences (CDS) was also very similar between both species representatives, strain GAR06 showed the lowest number with 6410 and strain PSN13 had the highest count with 6823, a difference of 413 CDS. A slightly larger difference was observed between the number of CDS in the group of *M. saelicesensis* (410) with respect to the *M. noduli* group (127). The number of rRNA operons was also higher in *M. saelicesensis*, with the type strain Lupac 09T accounting for the highest number (9 operons), followed by GAR06 (8 operons) and GAR05 (6 operons). The number of rRNA operons in the *M. noduli* was lower (3–5 operons) with the type strain, GUI43T having only 3 and strain ONO23 accounting for 5. In the case of tRNAs, it was observed that the number of these molecules was higher in all strains identified as *M. saelicesensis* (64–68) than in the *M. noduli* strains (51–57). A high number of tRNAs (77) was also reported for *Micromonospora lupini* Lupac 08T also isolated from nitrogen fixing nodules (Trujillo et al., 2014), while tRNAs reported for available *Micromonospora* genomes ranged from 48 to 87 (Carro et al., 2018).

The core genome of the six strains identified as *M. saelicesensis* (Group I) was calculated to be 5313 genes (81.3%) considering an average genome of 6531 genes. The number of singletons ranged from 94 for GAR06 to 706 for PSN13. In the case of the *M. noduli* strains (group II), the core genome included 5759 genes (88.05%) for an average genome of 6540 genes. In this group, strain Lupac 07 had the lowest number of singletons, 84, while strain ONO86 showed the largest variation with 369 genes. The calculation of a core genome based on all strains dropped to 74.72% and contained 4884 genes (Table S2). The calculated pangenomes were 8405, 7857, and 9867 genes for *M. saelicesensis* (Group I), *M. noduli* (Group II) and the combination of both species, respectively (Figure S1). As expected, an increase in the number of genes in the global pangenome was observed when all strains were combined, suggesting an important degree of variation between the genomes. The progression of the pan- and core genome can be seen in Figure S2.

Over 85% of the CDS for each species group were classified into Clusters of Orthologous Groups (COGs). COG profiles were very similar in all strains and were assigned into 22 categories ranging from 94 for GAR06 and strain PSN13 had the highest number (9 operons), followed by GAR06 (8 operons) and GAR05 (6 operons). The number of rRNA operons in the *M. noduli* was lower (3–5 operons) with the type strain, GUI43T having only 3 and strain ONO23 accounting for 5. In the case of tRNAs, it was observed that the number of these molecules was higher in all strains identified as *M. saelicesensis* (64–68) than in the *M. noduli* strains (51–57). A high number of tRNAs (77) was also reported for *Micromonospora lupini* Lupac 08T also isolated from nitrogen fixing nodules (Trujillo et al., 2014), while tRNAs reported for available *Micromonospora* genomes ranged from 48 to 87 (Carro et al., 2018).

The core genome of the six strains identified as *M. saelicesensis* (Group I) was calculated to be 5313 genes (81.3%) considering an average genome of 6531 genes. The number of singletons ranged from 94 for GAR06 to 706 for PSN13. In the case of the *M. noduli* strains (group II), the core genome included 5759 genes (88.05%) for an average genome of 6540 genes. In this group, strain Lupac 07 had the lowest number of singletons, 84, while strain ONO86 showed the largest variation with 369 genes. The calculation of a core genome based on all strains dropped to 74.72% and contained 4884 genes (Table S2). The calculated pangenomes were 8405, 7857, and 9867 genes for *M. saelicesensis* (Group I), *M. noduli* (Group II) and the combination of both species, respectively (Figure S1). As expected, an increase in the number of genes in the global pangenome was observed when all strains were combined, suggesting an important degree of variation between the genomes. The progression of the pan- and core genome can be seen in Figure S2.

**OGRI Indices**

Overall genomic relatedness indices (Chun and Rainey, 2014) were used to determine the relatedness between each pair of genomes used in the present study. In *M. saelicesensis* (Group I) ANI and OrthoANI values ranged from 97.82 to 99.13% and 97.96 to 99.19%, respectively, between the type and study strains. The *M. noduli* group (Group II) had ANI and OrthoANI values from 99.05 to 99.12% and 99.14% respectively (Table 3). In both cases, these values were above the recommended cut-off value of 96% for species recognition (Richter and Rosselló-Móra, 2009).

The ANI and OrthoANI values between the type strains *M. saelicesensis* and *M. noduli* were 96.64 and 96.82 respectively. Overall, pairwise comparison between the two groups showed the highest ANI and OrthoANI values corresponded to strains GAR05 and GUI43T (96.68%, ANI) and PSN01 and ONO23 (96.90%, OrthoANI) (Table S3). Both results are slightly above the border line of 95–96% for the delineation of species. However, these results are comparable to OrthoANI values obtained for the genome pairs of *M. carbonacea* and *M. haikouensis* (95.16%), and *M. inyonensis* and *M. sagamiensis* (96.5%).

---

**TABLE 2 | General genomic characteristics of *M. saelicesensis* and *M. noduli* strains.**

| Strain           | Genome size (Mb) | G+C ratio (mol%) | Number of Contigs | CDS  | rRNA operon | tRNA | Genes in COGs (%) | Genes in COGs (%) ** | Confirmed CRISPR count |
|------------------|------------------|------------------|-------------------|------|-------------|------|------------------|----------------------|------------------------|
| *M. saelicesensis* DSM 44871T | 7.1              | 71.1             | 11                | 6668 | 9           | 67   | 87.13            | 86.13                | 4                     |
| GAR 05           | 7.0              | 71.2             | 75                | 6526 | 6           | 68   | 87.07            | 86.07                | 0                     |
| GAR 06           | 6.9              | 71.2             | 60                | 6410 | 8           | 67   | 89.58            | 88.63                | 4                     |
| Lupac 06         | 7.0              | 71.2             | 59                | 6495 | 4           | 65   | 87.62            | 86.61                | 1                     |
| PSN 01           | 6.8              | 71.1             | 456               | 6458 | 4           | 67   | 85.71            | 84.5                 | 4                     |
| PSN 13           | 7.4              | 71.1             | 41                | 6823 | 3           | 64   | 86.91            | 85.90                | 1                     |
| *M. noduli* GUI43T | 7.2              | 70.9             | 224               | 6393 | 3           | 57   | 87.81            | 86.88                | 0                     |
| LAH 08           | 7.2              | 71.1             | 62                | 6627 | 4           | 56   | 88.13            | 87.14                | 3                     |
| Lupac 07         | 7.1              | 71.1             | 56                | 6500 | 4           | 51   | 86.95            | 86.06                | 2                     |
| MED 15           | 7.2              | 71.1             | 43                | 6554 | 4           | 56   | 87.97            | 87.00                | 1                     |
| ONO 23           | 7.2              | 71.0             | 125               | 6565 | 5           | 56   | 87.39            | 86.43                | 2                     |
| ONO 08           | 7.1              | 70.9             | 407               | 6591 | 4           | 55   | 85.48            | 84.60                | 2                     |

*mapped against bact HMM database included in eggNOG 4.5.1.*
**mapped against actNOG HMM database included in eggNOG 4.5.1.*
Species delineation based on dDDH values ranged from 81.0 to 93.7% for the six strains in *M. saelicesensis* (Group I) and 92.3–93.8% for the 5 components of *M. noduli* (Group II); all values clearly above the 70% recommended threshold (Table S3). dDDH values between the two species groups ranged from 71.0 to 71.8% (Table 3). Similar to ANI and OrthoANI results, dDDH values between the two groups were slightly above the border limit threshold value of 70% (68.1–74.5%). Overall pairwise comparisons of the study genomes and 48 additional *Micromonospora* type strains show the close relationship of the strains but clearly delineate each group within this 70–71% dDDH radius (Figure 4). A similar situation is observed between the species *Micromonospora sagamiensis* and *Micromonospora inyonensis* which share a dDDH value close to 70% (69.8%, dDDH; 61.3%, experimental DDH) (Kroppenstedt et al., 2005). Meier-Kolthoff and colleagues (Meier-Kolthoff et al., 2014), recently proposed the delineation of subspecies using genomic data. Specifically, these authors recommend a threshold of 79–80% to define subspecies in prokaryotic taxonomy. In the present study, the values obtained for the *Micromonospora* strains are much lower than this range and these strains are better classified as different species rather than subspecies.

Whole-Genome Phylogenomic Analysis

Phylogenomic tree reconstruction based on whole-genome distances calculated with the GBDP tool is presented in Figure 5. This tree included the 10 study genomes, all *Micromonospora* strains (type and non-type) published previously (Carro et al., 2018) and the genome sequences of the type strains *M. noduli* GU143T (this work), *M. aviciniae* DSM 45748T, *M. pisi* DSM 45175T, *M. pattaloongensis* DSM 45245T, *M. rosaria* DSM 803T and *M. wenchangensis* CCTCC AA 2012002T. The composition of the *M. saelicesensis* and *M. noduli* groups defined in the *gyrB* gene tree were identical, including the position of Lupac 07 as a member of *M. noduli* (Group II). The overall topology of this tree and the one published by Carro et al. (2018) was very similar, however, the inclusion of 17 additional genomes, as expected, influenced the distribution of the type strains, especially the
TABLE 3 | OGRID indices: ANI, OrthoANI, and dDDH percentage values calculated between the type and study strains.

| Strain                | ANI                  | OrthoANI              | dDDH                  |
|-----------------------|----------------------|-----------------------|-----------------------|
|                       | M. saelicesensis DSM 44871^T/M. noduli GUI43T | M. saelicesensis DSM 44871^T/M. noduli GUI43T | M. saelicesensis DSM 44871^T/M. noduli GUI43T |
| M. saelicesensis DSM 44871^T | 100/96.64          | 100/96.82             | 100/71.20             |
| GAR 05                | 99.11/96.68         | 99.16/96.82           | 99.50/71.50           |
| GAR 06                | 99.09/96.61         | 99.18/96.79           | 99.70/71.20           |
| Lupac 06              | 99.13/96.63         | 99.19/96.82           | 99.70/71.20           |
| PSN 01                | 99.12/96.61         | 99.19/96.82           | 99.80/71.60           |
| PSN 13                | 97.82/96.55         | 97.96/96.79           | 91.20/71.30           |
| M. noduli GUI43T/M. saelicesensis DSM 44871^T | 100/96.64          | 100/96.82             | 100/71.20             |
| M. noduli GUI43T/M. noduli GUI43T | 100/96.82 | 100/71.20             | 100/71.20             |
| M. noduli GUI43T/M. noduli GUI43T | 100/96.82 | 100/71.20             | 100/71.20             |
| M. noduli GUI43T/M. saelicesensis DSM 44871^T | 100/96.82          | 100/71.20             | 100/71.20             |
| M. noduli GUI43T/M. noduli GUI43T | 100/96.82 | 100/71.20             | 100/71.20             |
| M. noduli GUI43T/M. noduli GUI43T | 100/96.82 | 100/71.20             | 100/71.20             |
| M. noduli GUI43T/M. noduli GUI43T | 100/96.82 | 100/71.20             | 100/71.20             |

inclusion of the six additional type strains. Nevertheless, three out of Carro’s five defined groups (I, IV, and V) were almost completely recovered in the present phylogenetic analysis, the major rearrangements were observed in Carro’s groups II and III. In the present phylogenomic analysis, the strains in group II (M. purpureochromogenes, M. coxensis and M. halophytica) fused with M. rifamycinica and M. matsumotoense (group III) and were joined by M. wenchangensis (new to the analysis). The instability of group III was already highlighted (Carro et al., 2018). This rearrangement reduced group III to M. olivasterospora, M. carbonacea, and M. haikouensis.

UBCG Phylogenomic Analysis
The same dataset as above was used to construct a phylogenetic tree based on a core genome set of 92 genes using the UBCG tool (Na et al., 2018). Most of the selected genes (67/92) fall in the translation COG category (J), coding for ribosomal proteins (25/92, 50S and 18/92, 30S), aminoacid-tRNA ligases (10/92) and elongation an initiator factors (4/92) (Table S4). Again, the ten strains were distributed in two groups of identical composition as that of the gyrB gene and whole-genome phylogenomic analyses with significant branch support as indicated by the bootstrap values and gene support indices (GSI) (Figure 6). GSI values indicate the reliability of the branches on the phylogenomic tree based on the total number of genes used to construct the tree (92 genes) (Na et al., 2018). The topology of this tree with respect to the composition of the two groups was the same as the whole-genome and gyrB gene trees, including the position of strain Lupac 07, recovered in the M. noduli group. The topology of the UBCG tree highly correlated to the topology of the whole-genome phylogenomic tree of this study. Especially interesting was the fact that the new redefined groups II and III were recovered in entirety together with groups I, IV and V. In this analysis, a new group that contained strains from Carro’s groups I (M. mirobrigensis and M. siamensis), III (M. yangpuensis) and IV (M. krabienis), in addition to the newly included type strains M. aviciniae and M. rosaria was formed (Figure 6). Another important difference between the whole- and UBCG trees of this study was the position of Salinispora pacifica and Salinispora arenicola which in the latter tree was found associated to group IV. In this case, the up-to-date bacterial core gene analysis was not resolutive.

Phenotypic Profiles
Thirty-one phenotypic tests reported previously to be useful for the differentiation of the species M. saelicesensis and M. noduli (Carro et al., 2016) were carried out with all test strains. The number of characteristics that phenotypically differentiated between the two species was significantly reduced to one test when the number of strains compared increased (Table S5). Specifically, the use of rhamnose as a carbon source substrate was positive for all strains in the M. noduli group while the results were all negative for M. saelicesensis strains except for isolate PSN13 which was positive. The results of the remaining tests varied at the strain level and did not relate to their species identification.

Intra-species variability within each species group ranged from 0 to 33.3%. Range of pH growth and lipase production were the most variable tests in M. saelicesensis; utilization of serine as carbon source or the 23 biochemical tests served to differentiate between the two species, given the variability observed among the duplicate tests (Table S6). Strain Lupac 06 was the most variable with 35.1% discrepancies recorded. Overall intraspecies variability for M. saelicesensis and M noduli was 25.5 and 26.6% respectively.
Nineteen carbon sources were also assayed at different times (2007, 2016, and 2017) to check for reproducibility. Nine of the eleven strains tested expressed discrepant results over the different testing times. Three strains (Lupac 09T, Lupac 06, and Lupac 07) showed the highest variation with 26% of the tests yielding conflicting results while MED15, LAH08, and PSN13 had the lowest variation (5.2%). The use of D-serine as carbon source was the least reproducible test with seven strains yielding conflicting results (Table S7).

Draft genomes of the test strains were screened for genes involved in the carbon metabolism of the corresponding 19 substrates assayed in vitro. The predicted phenotypes correlated 100% with the results obtained in the laboratory for 11 tests. However, in the case of L-alanine, L-arginine, L-histidine, L-lysine, myo-inositol, L-rhamnose, D-serine, and D-trehalose, discrepant results were found between wet lab and in silico predictions (Figure S3). In most cases, the genes were localized in the genome but the experimental results varied (+/−) suggesting that even when the tests were carried out in the same laboratory and using the same method, they were not 100% reproducible.

In the case of L-rhamnose, in vitro tests for strain GU143T were positive but the genes related to the metabolism of this compound were not located. This is probably explained by the
DISCUSSION

The genus Micromonospora is highly relevant in biotechnological applications in areas such as medicine, agriculture and biofuels (Hirsch and Valdés, 2010; Trujillo et al., 2015; Carro et al., 2018). At present, this taxon holds 81 species with validly published names (LPNS), most of them described based on a polyphasic approach (Colwell, 1970; Vandamme et al., 1996). Within this framework, DNA-DNA hybridization (DDH) has been considered the key test to decide if a new strain represents a new species, despite its well spelled limitations (Gevers et al., 2005).
Given the drawbacks of DDH, it is not always straightforward to delineate the species limits, especially when DDH values close to the threshold. Therefore, the development of whole genome sequencing seems more appropriate to deduce relatedness by comparing genome sequences rather than performing DDH experiments (Vandamme and Peeters, 2014). Genomic data was recently used as the backbone to revisit the classification of the genus Micromonospora using a set of 45 draft genomes providing a useful dataset for comparison (Carro et al., 2018). While 16S rRNA is limited in resolving phylogenetic relationships at the species level (Katayama et al., 2007; Hahnke et al., 2016; Carro et al., 2018; Na et al., 2018), it has provided a good starting point for taxonomic studies. In this work,
16S rRNA gene sequencing was used to identify the closest neighbors of ten Micromonospora strains isolated from various legumes (Table 1). The sequence similarity values indicated that M. saelicesensis or M. noduli were the two most closely related species although in some cases, similarity values were identical between the test and both type strains (e.g., GAR06 and LAH08). The 16S rRNA gene tree topology yielded two very tight groups which could be interpreted as a single one when the branch lengths from these clusters were compared against the lengths of the remaining 79 Micromonospora type strains included in the analysis.

The use of gyrB gene sequences to resolve phylogenetic relationships in the genus Micromonospora has been recommended by several authors (Kasai et al., 2000; Garcia et al., 2010; Carro et al., 2012) given its higher resolution when compared to 16S rRNA gene phylogeny. In this study, the gyrB gene tree topology showed a similar arrangement to the 16S rRNA gene tree with respect to the test strains, however several differences were observed. The branch lengths were slightly longer, but still very small when compared to the rest of the Micromonospora species included in the tree. The most relevant change was the position of strain Lupac 07, which, together with strains Lupac 06 and Lupac 09T were originally classified as M. saelicesensis (Trujillo et al., 2007). The latter strains remained in the M. saelicesensis cluster but Lupac 07 moved to the M. noduli group. As expected, topologies of both trees in relation to the type strains were very different confirming that phylogenies based on single genes are very limited and unstable, making identification of nearest phylogenetic neighbors difficult.

The tree topologies based on the phylogenomic analyses of the UBCG (92 genes) and the whole draft genomes were similar. In both trees, strain Lupac 07 was recovered in the M. noduli group, strongly suggesting that this strain should be reclassified as a member of this species. The remaining 9 strains were recovered in the same species groups throughout all analyses.

In this study, both phylogenomic analyses contained a total of 70 genomes, including six additional Micromonospora type strains (see above). Overall, good agreement was found between the two phylogenies of this work and recently published data. In all cases, groups I, IV, and V previously defined (Carro et al., 2018) were recovered in their entirety with M. avinum (this analysis) joining group IV. The main difference between the three phylogenies was the composition of Carro’s groups II and III which were clearly influenced by the addition of M. rosaria DSM 803T and M. wenchangensis CCTCC AA 2012002T, producing a new group recovered in both phylogenies of the present work. Nevertheless, the groups I, IV, and V remained very stable considering that 11 new genomes (M. noduli GU143T and 10 test strains) were added and these were assigned to group IV where M. saelicesensis Lupac 09 was originally assigned. These rearrangements reinforce the argument that classification and identification systems are data dependent and constant rearrangement should be expected as more data are added and alternative methods are applied (Carro et al., 2018).

The new analysis tool UBCG proved useful for the construction of phylogenomic analysis, showing good correlation with trees using whole-draft genome data even though it did not resolve well the position of the Salinispora representatives, however, this may be due to the small number of representatives in the data set. An advantage of this pipeline is the use of bootstrap and GSI values to support the phylogenetic branches. It is also expected that as more genome sequences are added to the database, the more resolutive it should become.

Genome relatedness indices (ANI, Ortho-ANI, and dDDH) were calculated to complement the phylogenomic analyses for species demarcation. Overall, the three methods showed good agreement and the two species groups defined in the gyrB core-genome and whole-genome phylogenetic analyses supported the recognition of the 10 strains in two species.

Furthermore, these studies served to highlight the close relationship between the species M. saelicesensis and M. noduli. ANI values proposed for species delineation have been set to 95–96% as this range has been found to be correlated with the experimental DDH threshold of 70% (Goris et al., 2007; Richter and Rosselló-Móra, 2009). An alternative means to measure relatedness between two genomes is the calculation of dDDH using the GBDP method which appears to show a better correlation than ANI to the data derived from DDH experiments (Auch et al., 2010; Meier-Kolthoff et al., 2013b; Peeters et al., 2016).

In this work, the OGRI values were slightly above the recommended threshold for species delineation, if strictly applied, the study strains should be recognized as members of the same species. However, the consideration of other results in this work support the recognition of the strains as two separate species, M. saelicesensis and M. noduli. As previously expressed, thresholds are necessary for guidance but these should be applied in a flexible manner and considering other biological properties (Li et al., 2015). The present work is a good example for the interpretation and application of these values.

The use of phenotypic traits to identify and differentiate species in prokaryotic systematics is of limited value as previously discussed (Sutcliffe et al., 2013; Amaral et al., 2014; Vandamme and Peeters, 2014). In this work, several strains identified as one species, expressed different phenotypes, highlighting the problem of using diagnostic tables based on single strains to list differential characteristics between species. Information about intra-species variation is crucial for the development of stable diagnostic characteristics and the convenience of using more than a single isolate have been previously discussed (Sutcliffe et al., 2012; Oren and Garrity, 2014).

Our results confirm that the use of phenotypic tests, even when performed under the same conditions are not reliable for species differentiation due to the high variability observed within several members of the same species (Kumar et al., 2015). Instead, phenotypic studies should be regarded as complementary information to understand the biology of a microorganism and they should be restricted to strain characterization. Understandably, the inclusion of additional strains for the description of a taxon is often regarded as a burden because a lot of extra work is needed, especially when looking for differential phenotypic tests with questionable taxonomic value (Sutcliffe et al., 2012; Vandamme and Peeters, 2014).
Genomic information can be used to determine the intrinsic variability between a set of strains based on the core and pangenome profiles (Coenye et al., 2005; Sutcliffe et al., 2012; Oren and Garrity, 2014). In this work the calculation of these parameters has pointed out an important degree of variation between the species M. saelicesensis and M. noduli supporting their recognition as separate taxa. The complete elucidation of the gene functions within each group may provide an initial set of stable differential characteristics for each species, some of which may be phenotypically expressed.

CONCLUDING REMARKS

As additional data is generated, genome-based classifications should become more stable and provide a new working frame for the systematics of prokaryotes. The present study illustrates the advantage of using a diverse array of methods for the correct identification of new strains and the importance of using more than one isolate for a better characterization and definition of a species. OGRI values and especially dDDH values seem very appropriate for the delineation of prokaryotic species, but threshold numbers should be applied with a sufficient level of flexibility and considering other features inherent to a microorganism such as ecology, physiology, etc. (Li et al., 2015). There is no doubt that phenotypic information is useful for the good characterization of strains, but these studies should aim to provide information on the biology of a microorganism and not necessarily and not only to fill out a table with results of questionable value.

AUTHOR CONTRIBUTIONS

RR, LC and BR-P carried out experiments and bioinformatic analyses. CP and JB performed bioinformatic analyses. H-PK, PN, and MT designed the study and wrote the manuscript. All authors read the manuscript.

FUNDING

MT received financial support from the Ministerio de Economía y Competitividad (MICINN) under project CGL2014-52735-P. RR received a PhD scholarship from the University of Salamanca. BR-P and LC received a postdoctoral fellowship from CONACYT, México and University of Salamanca, respectively.

ACKNOWLEDGMENTS

Markus Göker is acknowledged for providing the iTOL annotation script.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01360/full#supplementary-material

Table S1 | Genome sequence accession numbers of strains used in this work.

Table S2 | Number of orthologous genes that conform the pan genome, core genome and singletons of Micromonospora saelicesensis (Group I) and Micromonospora noduli (Group II). In parenthesis, values expressed as percentages based on an average genome of 6531 genes for M. saelicesensis and 6540 genes for M. noduli.

Table S3 | Pair-wise OGRI values for ANI, OrthoANI and dDDH between M. saelicesensis and M. noduli strains. The orange and blue lines show the progression in the pan- and core genomes as more genomes are added.

Table S4 | List of genes used in core-genome phylogenomic analysis based on UBCG (Na et al., 2018).

Table S5 | Differential phenotypic characteristics between M. saelicesensis and M. noduli as reported by Carro and colleagues (Carro et al., 2016). +, Positive; –, Negative; w, Weak.

Table S6 | BIOLOG phenotypic profiles of M. saelicesensis and M. noduli strains. +, positive; –, negative; c, conflicting.

Table S7 | Carbon source substrates tested at different times using the same laboratory conditions. +, positive; –, negative; w, weak.

REFERENCES

Amaral, G. R. S., Dias, G. M., Wellington-Oguri, M., Chimetto, L., Campeão, M. E., Thompson, F. L., et al. (2014). Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. Int. J. Syst. Evol. Microbiol. 64, 357–365. doi: 10.1099/ijs.0.057927-0

Auch, A. F., von Jan, M., Klöppel, H. P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. 2, 117–134. doi: 10.4056/sigs.531120

Bateman, A., Martin, M. J., O’Donovan, C., Magrane, M., Apweiler, R., Alpi, E., et al. (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212. doi: 10.1093/nar/gku989

Blom, J., Kreis, J., Spanig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene context analyses. Nucleic Acids Res. 44, W22–W28. doi: 10.1093/nar/gkw255

Carro, L., Nouioui, I., Sangal, V., Meier-Kolthoff, J. P., Trujillo, M. E., Montero-Calasanz, M. D. C., et al. (2018). Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential. Sci. Rep. 8, 52925. doi: 10.1038/s41598-017-17392-0

Carro, L., Pujic, P., Trujillo, M. E., and Normand, P. (2013). Micromonospora is a normal occupant of actinorhizal nodules. J. Bacteriol. 38, 685–693. doi: 10.1128/jb.02383-10

Carro, L., Riesco, R., Spröer, C., and Trujillo, M. E. (2016). Micromonospora urelytica sp. nov., Micromonospora noduli sp. nov. and Micromonospora vinacea sp. nov., isolated from Pisum sativum nodules. Int. J. Syst. Evol. Microbiol. 66, 3509–3514. doi: 10.1099/ijsem.0.001231

Carro, L., Spröer, C., Alonso, P., and Trujillo, M. E. (2012). Diversity of Micromonospora strains isolated from nitrogen fixing nodules and rhizosphere of Pisum sativum analyzed by multilocus sequence analysis. Syst. Appl. Microbiol. 35, 73–80. doi: 10.1016/j.syapm.2011.11.003
Nawrocki, E. P., Eddy, S. R., Nawrocki, E. P., and Eddy, S. R. (2016). Computational identification of functional RNA homologs in metagenomic data. *RNA Biol.* 6286, 1170–1179. doi: 10.4161/rna.25038

Nimaichand, S., Zhang, Y. G., Cheng, J., Li, L., Zhang, D. F., Zhou, E. M., et al. (2013). *Micromonospora kangleipakensis* sp. nov., isolated from a sample of limestone quarry. *Int. J. Syst. Evol. Microbiol.* 63, 4546–4551. doi: 10.1099/ijs.0.052746-0

Oren, A., and Garrity, G. M. (2014). Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie van Leeuwenhoek* 106, 43–56. doi: 10.1007/s10482-013-0084-1

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuan, H. Y., Cohoon, J. K., Shirling, E., and Gottlieb, D. (1966). Methods for characterization of *B. subtilis* in the laboratory. *Methods for characterization of Prokaryotes.* 2nd ed., p. 88–99. Washington, DC: American Society for Microbiology.

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S., Lin, S., Sutcliffe, I. C., Trujillo, M. E., Whitman, W. B., and Goodfellow, M. (2013). A call to action for the International Committee on Systematics of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 63, 5961–5702. doi: 10.1093/ijs/iqt086

Parte, A. C. (2014). *LPSN - List of prokaryotic names with standing in nomenclature.* *Nucleic Acids Res.* 42, 613–616. doi: 10.1093/nar/gkt1111

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S., Lin, S., Sutcliffe, I. C., Trujillo, M. E., Whitman, W. B., and Goodfellow, M. (2013). A call to action for the International Committee on Systematics of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 63, 5961–5702. doi: 10.1093/ijs/iqt086

Richard, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W185–W191. doi: 10.1093/nar/gki866

Suzuki, R., and Shimodaira, H. (2015). Phylogenetic trees from multiple alignments: inferDNA. *Bioinformatics* 31, 1355–1361. doi: 10.1093/bioinformatics/btv465

Vaas, L. A. I., Sikorski, J., Hofner, B., Fiébig, A., Buddruhs, N., Klenk, H. P., et al. (2013). Opm: an R package for analysing OmniLog® phenotype microarray data. *Bioinformatics* 29, 1823–1824. doi: 10.1093/bioinformatics/btq291

Wang, X., Tang, Y., Horikoshi, M., and Li, W. (2016). ggfortify: unified interface to visualize complex R objects. *Bioinformatics* 32, 10–20. doi: 10.1093/bioinformatics/btw706

Yang, D., Xia, Y., Wang, Z., and Zhang, L. (2016). ClustalX 2.1: improving the sensitivity of multiple sequence alignment methods. *Nucleic Acids Res.* 44, 338–341. doi: 10.1093/nar/gkw960

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., and Seo, H., et al. (2017). Bioinformatics for microbiome. *PLoS ONE* 12, e0175378. doi: 10.1371/journal.pone.0175378

Copyright © 2018 Riesco, Carro, Román-Ponce, Prieto, Blom, Kleek, Normand and Trujillo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.