

The new bacterial species description: the genus *Mycobacterium* as an example*

Descrição de novas espécies de bactérias: o gênero *Mycobacterium* como exemplo

Enrico Tortoli

IRCCS Istituto Scientifico San Raffaele, Milano, Italia

Description of a new bacterial species is a long and onerous route; yet it remains a major aspiration for many microbiologists.

The basis of the process relies on a polyphasic approach¹, combining a number of phenotypic and genotypic investigations, performed by the proponent, to support the diversity of the organism from the most closely-related taxa known. Though minimal rules for such identifications do exist, not all are followed; however, determination of the sequence of 16S rRNA is practically the only requirement that remains mandatory. The International Journal of Systematic and Evolutionary Microbiology is the official journal of record for novel prokaryotic taxa.

In absence of an explicit position against the formal description of a new species based on the identification of a single strain, the number of such cases is steadily increasing. The concept of species, for prokaryotes in particular, is hardly defined, but it can definitely be thought of as a 'group'. Description of a species, based on a single isolate, actually is just a description of the isolate, practically excluding the biodiversity. This is indirectly confirmed by the high percentage of such taxa, reported in the literature, without any further isolation. For most species, the type strains deposited in the World Federation for Culture Collections remain the only proof of their existence. It has been calculated that 25 strains are necessary for an accurate description of a species, and the lowest tolerable limit being 10², but practically such numbers can rarely be achieved. The wait for encountering that many strains may be endless and frustrating. An acceptable compromise for the reliable description of a new species cannot be lower than two, provided each is proven spatially and temporarily independent.

Another critical point for proposing a new species is the certainty that it does not overlap with an already-accepted taxon. The DNA-DNA Hybridization test (DDH)³ represents the reference for species

circumscription. To verify if two strains belong to different species, the mixture of their denatured DNAs is left to re-associate forming hybrid molecules (hetero-duplex) under suitable conditions. The degree of similarity is analyzed by comparing the results obtained with the mixed DNAs to that with pure DNA (forming homo-duplex only). DDH provides a clear and objective numerical threshold: values < 70% warrant the assignment of the two strains to different species. The DDH test is labor-intensive and error-prone⁴, hence limiting its implementation to very few cases for supporting the description of new species. Nowadays, bioinformatic algorithms are available, validated by multiple studies, which can be used to infer the DDH from genomic data; the best known are the Average Nucleotide Identity (ANI)⁵ and the Genome to Genome Distance (GGD)^{6,7}. The ANI represents a mean of identity between the homologous genomic regions shared by two genomes. Two strains characterized by pairwise ANI value < 95% belong to different species, while when the value is > 96%, they are members of a single species; a confident attribution is not possible for values between 95 and 96%. The GGD algorithm is the *in silico* equivalent of the DDH and produces values directly convertible to DDH%, consequently subjected to the threshold of 70% for species demarcation.

In a recent study, we investigated the whole genome of 144 out of the 180 species included in the genus *Mycobacterium*. The analyses conducted with ANI and GGD led us to detect 10 illegitimate species, and to create: four subspecies in *M. intracellulare*; three each in *M. farcinogenes* and *M. abscessus*; and two each in *M. austroafricanum*, *M. marinum*, and *M. pyrenivorans* (Table 1).

The need of assembled genomic data has so far hampered the exploitation of ANI and GGD in the description of new species. In recent years, however, availability of reliable genomic data has become increasingly easy and cost-effective, and in fact, modern taxonomy owes its growth to the pivotal contribution of

* Opinion article written by Speaker for the II International Scientific Meeting of Instituto Evandro Chagas, held from October 25 to 27, 2017, in Ananindeua, Pará, Brazil. All the articles of this modality were analyzed by the Scientific Committee of the Event and, later, by the Editors of RPAS.

Correspondence / Correspondência:

Enrico Tortoli

San Raffaele Scientific Institute

San Michele Building, via Olgettina 60, 20132 – Milan, Italy – Phone #: +39 02 26435684

E-mail: tortoli.enrico@hsr.it

genomic information. Therefore, determination of the whole genome, of every newly described species, is a primary requisite. If not already available in public repository, the 16S rRNA sequence, closest in similarity with the proposed species, should be determined. The pairwise ANI or GGD with every closely related genome should be calculated to avoid the risk of duplicating an already existing species. The availability of exhaustive

genomic characterization will reduce the need of detailed characterization of the weakly informative properties such as phenotypic traits. The availability of genomes of a large majority of *Mycobacterium* species in GenBank could potentially establish a tree based on whole genomes⁸, besides the traditional phylogenetic trees, based on 16S rRNA or on concatenated housekeeping genes.

Table 1 – Synonymous species of genus *Mycobacterium**

				ANI%	GGD [†]
<i>M. conceptionense</i>	<i>M. farcinogenes</i>	<i>M. senegalense</i>		98.3–99.4	83–86
<i>M. abscessus</i>	<i>M. bolletii</i>	<i>M. massiliense</i>		97.2–97.4	85–88
<i>M. chimaera</i>	<i>M. intracellulare</i>	<i>M. paraintracellulare</i>	<i>M. yongonense</i>	97.6–98.7	77–90
<i>M. austroafricanum</i>	<i>M. vanbaalenii</i>			98.7	80
<i>M. marinum</i>	<i>M. pseudoshottsii</i>			98.2	82
<i>M. monacense</i>	<i>M. pyrenivorans</i>			97.5	84

* In each line, the prior name, basing on the publication year, is in bold type and becomes the new species name; Names in normal face become names of subspecies and follow the new species name; [†] DDH% equivalent of GGD values.



REFERENCES

- Vandamme P, Pot B, Gillis M, De Vos P, Kersters K, Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev.* 1996 Jun;60(2):407-38.
- Sneath PHA. The maintenance of large numbers of strains of microorganisms, and the implications for culture collections. *FEMS Microbiol Lett.* 1977 Jun;1(6):333-4.
- Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, et al. Report of the Ad Hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol.* 1987 Oct;37(4):463-4.
- Rosselló-Mora R, Amann R. The species concept for prokaryotes. *FEMS Microbiol Rev.* 2001 Jan;25(1):39-67.
- Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol.* 2014 Feb;64(Pt 2):346-51.
- Auch AF, von Jan M, Klenk HP, Goker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci.* 2010 Jan;2(1):117-34.
- Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics.* 2013 Feb;14:60.
- Fedrizzi T, Meehan CJ, Grottola A, Giacobazzi E, Serpini GF, Tagliazucchi S, et al. Genomic characterization of nontuberculous mycobacteria. *Sci Rep.* 2017 Mar;7:45258.

Received / Recebido em: 20/9/2017
Accepted / Aceito em: 11/10/2017