Infection, Genetics and Evolution 13 (2013) 67–75

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Review

Highlights on molecular identification of closely related species

Lígia A. Almeida [a,b], Ricardo Araujo [a,c,*]

[a] IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
[b] ICBAS, Instituto de Ciências Biomédicas Abel Salazar, University of Porto, Rua de Jorge Viterbo Ferreira n.° 228, 4050-313 Porto, Portugal
[c] Faculty of Sciences, University of Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal

ARTICLE INFO

Article history:
Received 15 May 2012
Received in revised form 6 August 2012
Accepted 8 August 2012
Available online 12 September 2012

Keywords:
Bacterial complex
DNA–DNA hybridization
Molecular identification
Multilocus sequence typing
Species definition
Taxonomy

ABSTRACT

The term "complex" emerged in the literature at the beginning of the genomic era associated to taxonomy and grouping organisms that belong to different species but exhibited similar patterns according to their morphological, physiological and/or other phenotypic features. DNA–DNA hybridization values ⩾70% and high identity on 16S rRNA gene sequences were recommended for species delineation. Electrophoretic methods showed in some cases to be useful for species identification and population structure but the reproducibility was questionable. Later, the implementation of polyphasic approaches involving phenotypic and molecular methods brought new insights into the analysis of population structure and phylogeny of several "species complexes", allowing the identification of new closely related species. Likewise, the introduction of multilocus sequence typing and sequencing analysis of several genes offered an evolutionary perspective to the term "species complex". Several centres worldwide have recently released increasing genetic information on distinct microbial species. A brief review will be presented to highlight the definition of "species complex" for selected microorganisms, mainly the prokaryotic *Acinetobacter calcoaceticus – Acinetobacter baumannii*, *Borrelia burgdorferi* sensu lato, *Burkholderia cepacia*, *Mycobacterium tuberculosis* and *Nocardia asteroides* complexes, and the eukaryotic *Aspergillus fumigatus*, *Leishmania donovani* and *Saccharomyces* sensu stricto complexes. The members of these complexes may show distinct epidemiology, pathogenicity and susceptibility, turning critical their correct identification. Dynamics of prokaryotic and eukaryotic genomes can be very distinct and the term "species complex" should be carefully extended.

Contents

* Corresponding author at: IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal. Tel.: +351 225570700; fax: +351 225570799.
E-mail addresses: almeida.ligia.mca@gmail.com (L.A. Almeida), ricjparaujo@yahoo.com (R. Araujo).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.08.011

1. Introduction

The first classification of microscopic organisms relied on morphological and physiological observations. The term "complex" emerged in the literature at the beginning of the genomic era associated to taxonomy in a wide variety of scenarios. The intended meaning was almost the same: grouping organisms belonging to different species but exhibited similar patterns according to their morphological, physiological or other phenotypic characteristics. The word "complex" was employed in different groups of microorganisms, such as bacteria, fungi, parasites, or higher organisms (Beaman and Beaman, 1994; Edwards-Ingram et al., 2004; Vandamme et al., 1997). With bacteria, constant changes in taxonomy were evident and, while some species were transferred from one genus to the other, new bacterial species and genera were also proposed. Wide consensus was sometimes difficult to achieve (Vandamme et al., 1997; Yabuuchi et al., 1992).

Polyphasic taxonomy of prokaryotes was first proposed 40 years ago by Colwell (1970) and aimed to integrate different types of data. This approach proposed a taxonomic classification of any isolate according to a set of criteria (Colwell, 1970). The development of nucleic acid hybridization methodologies and their application to prokaryotes in the 1960's allowed the first measurement of whole genome and gene sequence identities between strains. Complementary interactions between DNA–DNA and DNA–mRNA employing gel structures or membrane filters and radioactive measurements provided quantitative information on the genetic relatedness of several species (Brenner et al., 1967; McCarthy and Bolton, 1963). Since the 1970's, DNA–DNA hybridization (DDH) has been used to compare nucleotide sequences and to delineate bacterial taxonomies (Colwell, 1970; Johnson et al., 1970). In 1987, DDH values $\geqslant$70% was finally recommended as criteria for standard species delineation by the *ad hoc* Committee of the International Committee for Systematic Bacteriology (Wayne et al., 1987). The 16S rRNA sequence-based methodologies, although only applied to prokaryote characterization by the late 1980s, were believed to be a valuable tool for establishing relationships between microorganisms through phylogenetic analysis (Collins et al., 1989). The 16S rRNA gene is present in the genome of all prokaryotes and its primary structure is highly conserved within organisms of the same genus. However, 16S rRNA gene analysis do not allow distinguishing many genetically related species, especially those presenting over 98% 16S rRNA sequence identity, such as the species complexes included in this review.

Multilocus enzyme electrophoresis (MLEE), initially employed to study eukaryotes, was applied to prokaryotes by Milkman (1973). Milkman was a pioneer in the analysis of the electrophoretic motility of five loci of an extensive number of *Escherichia coli* strains. The MLEE methods enabled the detection of small alterations on nucleotide sequences of genes encoding enzymes and the electrophoretic patterns that were then correlated with the alleles of each locus. MLEE, by establishing a correlation between electrophoretic types (ETs) and alleles of housekeeping loci, allowed the evaluation of bacterial evolution and the inference of relationships between strains, proving in some cases to be more useful for species identification and analysis of population structure than DNA hybridization and 16S rRNA sequence analysis (Stackebrandt and Goebel, 1994). Methods based on enzymatic digestion of DNA molecules in combination with electrophoresis separation were developed in the 1980's and turned out as relevant instruments for the analysis of larger DNA molecules (10 kb – 10 Mb). Pulsed field gel electrophoresis (PFGE) and restriction fragment length polymorphism (RFLP) were the most widely used and often considered as "gold standard" methods for identifying varieties of bacteria, sometimes at strain level (Li et al., 2009; Olive and Bean, 1999).

The term "species complex" was proposed by Ursing et al. (1995) for grouping strains with distinct genomic characteristics, therefore, probably representing new distinct species. Each group would be named genomovar, followed by a roman number, being the first number attributed to the type strain of the species. The genomovars were, thus, transitory attributes of the putative species while they were waiting further confirmation of their phenotypic distinctiveness (Ursing et al., 1995). Subsequently, Vandamme et al. (1996) revised the concept proposing the normalization of taxonomic classification and ascertained coherency, reproducibility and uniformity of criteria. In addition, they applied the polyphasic approach, comprising a biochemical profile analysis, a whole cell protein profile, fatty acid analysis, sequencing of 16S rRNA and *recA* genes and DDH to the identification of *Burkholderia cepacia* complex (BCC) species (Vandamme et al., 1997).

Multilocus typing methodologies, such as the multilocus sequence typing (MLST; http://www.mlst.net/), have significantly improved the accuracy of species characterization. MLST introduced by Maiden et al. (1998) to study *Neisseria meningitis* population, was based on the same principles of MLEE. But, instead of enzymes, MLST assigned the alleles of each locus of housekeeping genes by sequencing conserved fragments. The number of alleles identified was higher than in MLEE and this methodology was approved for strain genotyping, species identification and analysis of several bacterial populations (Maiden et al., 1998).

With the introduction of the MLST and sequencing analysis of complete genomes, the term "species complex" gained a new perspective within the epidemiologic and evolutionary settings of bacterial biology. Closely related species may show distinct epidemiology, pathogenicity and susceptibility to the antibiotics and it is critical that correct identification of microbes be performed, particularly at clinical laboratories targeting pathogenic and relevant infectious agents. Now we have incoming genetic information available from several centres worldwide and it urges an overview of the findings that have been described by researchers in distinct microorganisms.

Is taxonomic classification based on a set of criteria among bacteria or eukaryotic populations? How difficult is it to define a microbial species in prokaryotes and eukaryotes? A brief review of a few selected microorganisms is presented below to highlight the observations regarding the definition of "species complex". We selected organisms that currently show the most solid information regarding the definition of closely related species.

2. Prokaryotic "species complex"

2.1. Burkholderia cepacia complex

B. cepacia complex (BCC) is a group of strictly aerobic, gram-negative, motile bacilli (Palleroni, 1984; Palleroni and Holmes, 1981). BCC also known as "*B. cepacia* like bacteria" have their taxonomic origin on *Pseudomonas cepacia*, first described in 1950 as phytopathogen by Burkholder (1950). Presently, the BCC includes 17 species: *B. ambifaria*, *B. anthina*, *B. arboris*, *B. cepacia*, *B. cenocepacia*, *B. contaminans*, *B. diffusa*, *B. dolosa*, *B. lata*, *B. latens*, *B. metallica*, *B. multivorans*, *B. pyrrocinia*, *B. seminalis*, *B. stabilis*, *B. ubonensis*, *B. vietnamiensis* (Vanlaere et al., 2009).

The BCC was initially defined as a group of closely related species that were phenotypically similar and exhibited intermediate level of DDH values (30–60%), 98–100% identity on their 16S rRNA sequences and high *recA* sequence identity (94–95%). Lower values of DDH (below 30%) were observed between BCC strains and representatives of other species of the genus *Burkholderia* (Vanlaere et al., 2009).