2 DNA–DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation

Ramon Rosselló-Mora

2.1 Introduction

DNA–DNA reassociation techniques are used for many purposes, but in the field of microbial systematics they are in most cases linked to the circumscription of prokaryotic species. Actually, as we will see, the use of whole genome hybridizations in the definition of prokaryotic species has had an enormous influence since the origin of the polythetic classification system (Rosselló-Mora and Kämpfer 2004). The importance of morphology in the middle of the eighteenth century was substituted for that of biochemical properties at the beginning of the nineteenth century; and subsequently the emerging "modern spectrum" techniques emphasized the importance of genetic measurements, such as DNA–DNA reassociation experiments. However, after almost 50 years of the application of these techniques to circumscribe species, there is increasing reluctance to use them because of the intrinsic pitfalls in the methods (e. g. Stackebrandt 2003; Stackebrandt et al. 2002). Consequently, the question that arises is: if DNA reassociation techniques are to be substituted, what will take their place? However, in my opinion, it is still too soon to substitute these techniques because of several reasons: (a) the use of such parameters in the definition of species has been of paramount influence and has actually determined the size and shape of what we call 'species', (b) there are almost 5,000 species described (Garrity et al. 2004), many of them based on reassociation experiments, and the legitimacy of new circumscription methods should be validated and (c) the alternatives proposed are not yet standardized and tested sufficiently enough to offer a reliable, pragmatic and easy to use circumscription tool. Any new technique with the potential to act as a substitute for DNA–DNA reassociation experiments should demonstrate that: (a) it is more reliable, workable and pragmatic, (b) it does not radically change the present classification system and (c) it leads to results that fit into a genomically based perspective without losing sight of the organisms themselves. Any

Ramon Rosselló-Mora: Grup d'Oceanografia Interdisciplinar, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), C/Miquel Marqués 21, 07190 Esporles, Illes Balears, Spain, E-mail: rossello-mora@uib.es

Molecular Identification, Systematics, and Population Structure of Prokaryotes
E. Stackebrandt (Ed.)

intended substitution of a technique that has implications for the circumscription parameters that have served as a basis for the establishment of the current taxonomic system should also take into account the purpose of taxonomy. The end result itself is to provide a system that is operative and predictive; and the information behind a name should be more than a mere set of genes with no meaning. What has hitherto been constructed is a classification system based on the circumscription of taxa when the overall information collected indicated that such circumscription would be enough to recognize them as unique and identifiable. Behind a species name there is more than a binomial, there is a collection of data that allows identification from several independent sources that gives a prediction of how an organism may be and might behave. Our system is perhaps not perfect and deserves improvement, but as already noted "it is the envy of those who wish to implement similar systems in botany or zoology" (Euzéby and Tindall 2004).

DNA–DNA reassociation techniques, also known as DNA–DNA hybridization techniques, are based on an attempt to make raw comparisons of whole genomes between different organisms in order to calculate their overall genetic similarities. Just after the discovery of the intrinsic properties of DNA (i. e. information content and secondary structure resilience), a good number of techniques were developed and applied to microbial taxonomy in order to circumscribe its basic unit, the species. At that time, it was believed that such genetic comparisons would render more stable classifications than those simply based on phenotypic similarities (Krieg 1988). There is no doubt that the first attempt to elucidate taxonomic relationships based on single-stranded DNA reassociation conducted by Schildkraut et al. (1961) was a breakthrough for microbial systematics and for the construction of the current microbial classification system. They demonstrated that duplex formation between the denatured DNA of one organism and that of another organism would only occur if the overall DNA base compositions were similar and if the organisms from which the DNA was extracted were genetically related. At the time when a monothetic classification was abandoned in favour of a polythetic (or phenetic; Rosselló-Mora and Amann 2001) classification, these developments in DNA techniques led to microbial taxonomists extending the definition of the species by using reassociation results and by determining the GC mole percentage of each individual genome. The great practical advantage seen in DNA–DNA hybridization experiments was that the results did not show the continua often observed between groups defined by phenotypic characteristics, but instead the genomes appeared clustered in discrete groups, whether organisms tended to be closely related or not (Krieg 1988). Since then, such techniques have routinely been applied in most of the new species characterizations, especially those that involved new taxa in already existing

genera and/or those where more than a single isolate was used to circumscribe the taxon. The application of these techniques to circumscribe species was reinforced by a recommendation from an ad hoc committee on systematics (Wayne et al. 1987). In fact, the committee (using ΔT_m to indicate melting temperature increment) stated that "the phylogenetic definition of a species generally would include strains with approximately 70% or greater DNA–DNA relatedness and with 5 °C or less ΔT_m. Both values must be considered. Phenotypic characteristics should agree with this definition and would be allowed to override the phylogenetic concept of species only in a few exceptional cases". In addition, they reinforced that "it is recommended that a distinct genospecies that cannot be differentiated from another genospecies on the basis of any known phenotypic property not be named until they can be differentiated by some phenotypic property". That recommendation had two main effects. On the one hand, it forced descriptions based on both genomic and phenotypic properties but, on the other hand, it unwittingly created the belief that a rigid boundary of 70% genome similarity would be sufficient for the recognition of species. Both aspects have had an enormous influence on prokaryotic taxonomy.

Emerging techniques at the end of the twentieth century, such as rRNA gene sequencing and phylogenetic reconstructions, were expected to help in the replacement of DNA–DNA reassociation experiments. However, it was soon realized that, due to the length and information of the molecule, the resolution power needed to discriminate different species within a genus was not always adequate (e. g. Amann et al. 1992; Fox et al. 1992; Martínez-Murcia et al. 1992). For these reasons, it was accepted at that time that no other methodology could replace genome similarity analysis (Stackebrandt and Goebel 1994). It has always been clear that the best way to understand similarities would be to truly compare whole genome sequences (e. g. Owen and Pitcher 1985), a fact that has nowadays almost become possible. The increasing number of completely sequenced genomes allows such comparisons and the first speculations on how species can be circumscribed by this newly emerging information (Konstantinidis and Tiedje 2005; Santos and Ochman 2004; Stackebrandt et al. 2002; Zeigler 2003). However, all these new circumscription attempts should be previously validated by contrasting them with the criteria used to construct the current taxonomic schema.

DNA–DNA reassociation experiments have often been criticized due to their high experimental error and their failure at generating cumulative databases (e. g. Sneath 1989; Stackebrandt 2003). However, their use has never been abandoned because no other alternative has been either found or tested. In order to illustrate how often DNA–DNA reassociation experiments are still used to circumscribe species, a survey on all the publications that appeared in 'Int. J. Syst. Evol. Microbiol.' during 2004 has been under-

Table 2.1. 'Int. J. Syst. Evol. Microbiol.' survey: absolute numbers and percentages of articles or new descriptions that were published in the six issues of vol 54 of the journal during 2004

Articles with new descriptions	305	
Articles with reassociation experiments	199	65%[a]
Articles without reassociation experiments	106	35%[a]
Spectrophotometric reassociation experiments	67	34%[b]
Non-radioactive microtitre-plate hybridizations	96	48%[b]
Non-radioactive filter methods (chemiluminescence)	9	5%[b]
Radioactive filter, S1, or hydroxyapatite methods	27	14%[b]
New species	351	
New species with a single isolate	191	54%[c]
New genera	65	
New 'candidatus'	17	

[a] percentages refer to the 305 articles with new descriptions

[b] percentages refer to the 199 articles where reassociation experiments were performed

[c] percentages refer to the total number of 351 new species classifications

taken (Table 2.1). In that year, around 305 articles appeared that compiled the description of about 351 new species, 65 new genera, and 17 new 'candidatus'. Among all these new species descriptions, about 65% of them used DNA–DNA reassociation experiments. From the 35% of the remaining descriptions where no reassociation was used, more than 75% were based on a single isolate and more than half corresponded to new genera. In such cases, the rationale for taxa descriptions were mainly based on 16S rDNA sequence dissimilarities. However, it is also worth noting that among all the descriptions where DNA–DNA reassociation was used, nearly 60% of them were also based on a single isolate. In these cases, the use of hybridizations was to show enough dissimilarity to their closest relative species.

There is a desire to replace DNA–DNA reassociation for other more accurate techniques (Stackebrandt et al. 2002) but its use still cannot be avoided. Consequently, this is a timely review concerning existing techniques, their pitfalls and the meaning of their results. In addition, the possibility to replace them will also be discussed.

2.2 Semantic Considerations

Prokaryotic taxonomy, like eukaryotic taxonomy, is filled with semantic misuses. There are several examples that in some respect are responsible for the so-called 'species problem': (a) the use of homology as a synonym of similarity, (b) the persistent homonymy of the term species and (c) the

synonymy between concept and definition. Although these issues will be thoroughly discussed elsewhere, it is worth providing some clarifications at this point:

1. Homology vs similarity: since the early days of the interpretation of DNA–DNA reassociation results, homology and similarity have been used as synonyms. However, it was soon noted that the use of the term homology would not be appropriate for interpreting hybridization results, because there was no certainty that bound stretches of DNA from different organisms would contain identical nucleotide sequences and the use of terms such as relatedness or DNA binding would be more accurate (Brenner and Cowie 1968; De Ley et al. 1970). However, these recommendations were not taken into account and for decades the term homology has been used to express DNA–DNA reassociation results. Later, there was again the temptation to abandon the term homology (Stackebrandt and Liesack 1993) by arguing that the values observed were not linearly correlated with sequence identity. Homology is not a measurable parameter: either two characters (in this case sequences or DNA fragments) are homologous or not, which means that either they have the same evolutionary origin or not (Fitch 2000; Mindell and Meyer 2001; Tindall 2002). Homology basically has an evolutionary meaning and thus cannot be applied either as a synonym for sequence identity or to express DNA–DNA reassociation results. The term similarity is perhaps the best choice because it does not imply any evolutionary nor phylogenetic meaning. Despite the reiterated recommendations, there are still quite a few publications that wrongly use the term homology.

2. Homonymy of the term species: perhaps the most important cause of the 'species problem' is the persistent homonymy (Reydon 2004). This means that different scientific disciplines adopt different concepts to embrace their devised units, but the same term 'species' is given to all of them. This has always been regarded as a clear case of pluralism (Brigandt 2002; Ereshefsky 1998; Mishler and Donoghue 1982; Reydon 2004). For some, it would be better to eliminate the term species and each scientific discipline should instead adopt a unique and specially tailored basic unit, such as 'biospecies', 'ecospecies' or 'phylospecies' (Ereshefsky 1998). However, for others, pluralism is still an adequate choice, with the term 'species' being kept for general-purpose classification, which should retain binomials as a property of the taxonomic system (Brigandt 2002). These problems, which have been thoroughly discussed in eukaryotic taxonomies, are well represented when classifying prokaryotes. Actually, what taxonomists mean by a species does not satisfy, for instance, microbial ecologists or population geneticists, although it would probably not be possible for these groups to come to any mutual agreement on terminology. It is also important to note that, for example "evolution was inferred from the classification, not vice

versa" (Sneath 1988) and thus the ultimate concept of 'species' is a property of taxonomy. These disagreements are the basis for most of the discussions on the adequacy of the current species concept in use (Rosselló-Mora and Kämpfer 2004) and, therefore, most probably it would be recommendable to adopt a clear pluralistic approach. Taking into account that the term and idea of 'species' is the basal taxonomic unit originally devised to support a universal hierarchic system (Ereshefsky 1994), the main arguments expressed here are within the framework of taxonomy and refer to the species concept currently applied to the classification of prokaryotes. Perhaps the most updated version of the prokaryotic species concept is "a category that circumscribes a (preferably) genomically coherent group of individual isolates/strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardized conditions" (Stackebrandt et al. 2002). The whole critical viewpoint here revolves around the adequacy of DNA–DNA reassociation experiments to circumscribe genomically coherent groups.

3. Concept and definition: another exponent example of semantic misunderstanding is the confusion between concept and definition. Both terms are often used as synonyms, but it is important to take into account that distinguishing them may very much help in clarifying our prokaryotic species 'problem'. The species concept is the idea that explains and circumscribes the patterns of recurrence observed in nature. It is the essence of what we think is the basic unit for constructing an operative and predictive classification. Within the concept, we should find the reasons for including or excluding naturally occurring individuals within a category. However, the species definition is the way we recognize that individuals belong to a category. The definition provides a set of parameters that are sufficient to recognize that a certain group of individuals belong to a recurrent pattern in nature. Actually, this responds in the most pragmatic way to identify what we think is a unit. Our reductionistic approach to understanding nature allows us to formulate the simplest way to recognize units (Rosselló-Mora 2003). For example, in this chapter, 'genomic coherency' applies to the concept, whereas the relaxed (or not) results or values of DNA–DNA reassociation experiments would apply to the definition. For example, changing the method and parameters to recognize coherent genomic groups, such as substituting DNA–DNA reassociation experiments (e. g. MLST), would result in a change in how we define species but not how we conceive them. The concept remains the same.