DNA–DNA hybridization values and their relationship to whole-genome sequence similarities

Johan Goris,^{1†} Konstantinos T. Konstantinidis,^{1‡} Joel A. Klappenbach,¹ Tom Coenye,² Peter Vandamme² and James M. Tiedje¹

Correspondence

Johan Goris

johan_goris@applied-maths.com

¹Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824, USA

²Laboratory for Microbiology, Gent University, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

DNA–DNA hybridization (DDH) values have been used by bacterial taxonomists since the 1960s to determine relatedness between strains and are still the most important criterion in the delineation of bacterial species. Since the extent of hybridization between a pair of strains is ultimately governed by their respective genomic sequences, we examined the quantitative relationship between DDH values and genome sequence-derived parameters, such as the average nucleotide identity (ANI) of common genes and the percentage of conserved DNA. A total of 124 DDH values were determined for 28 strains for which genome sequences were available. The strains belong to six important and diverse groups of bacteria for which the intra-group 16S rRNA gene sequence identity was greater than 94 %. The results revealed a close relationship between DDH values and ANI and between DNA–DNA hybridization and the percentage of conserved DNA for each pair of strains. The recommended cut-off point of 70 % DDH for species delineation corresponded to 95 % ANI and 69 % conserved DNA. When the analysis was restricted to the protein-coding portion of the genome, 70 % DDH corresponded to 85 % conserved genes for a pair of strains. These results reveal extensive gene diversity within the current concept of 'species'. Examination of reciprocal values indicated that the level of experimental error associated with the DDH method is too high to reveal the subtle differences in genome size among the strains sampled. It is concluded that ANI can accurately replace DDH values for strains for which genome sequences are available.

INTRODUCTION

There is a general consensus among taxonomists that all taxonomic information about a bacterium is incorporated in the complete nucleotide sequence of its genome (Stackebrandt *et al.*, 2002; Wayne *et al.*, 1987). As whole-genome sequencing did not become available until recent years and, even then, only for a limited number of organisms, other parameters by which organisms could be classified into practical categories were needed. Since the 1960s, DNA–DNA hybridization (DDH) experiments have been performed to determine relatedness between bacteria, as this was one of the few universally applicable techniques available that could offer truly genome-wide comparisons between organisms. A value of 70 % DDH was proposed by Wayne *et al.* (1987) as a recommended standard for delineating species. Although this recommendation is not a

strict standard and, in fact, several studies have used more stringent DDH cut-off values or have employed no DDH experiments at all, we focused our analysis on 70 % DDH because it is the best-known standard. Several principally different methods for the measurement of DDH values have been described (Brenner *et al.*, 1969; Crosa *et al.*, 1973; De Ley *et al.*, 1970; Ezaki *et al.*, 1989), and the use of DDH in bacterial taxonomy has recently been reviewed in detail (Rosselló-Mora, 2006). While the technique has the above-mentioned advantages, it also has several important drawbacks. Because relatively large quantities of DNA (in comparison with PCR-based techniques) of a high quality are required, the whole process of performing DDHs often becomes rather time-consuming and labour-intensive. Also, the diverse methods that are available can yield different results, especially for lower reassociation values (Grimont *et al.*, 1980; Huß *et al.*, 1983). Its main disadvantage, however, is that because of the comparative nature of the technique no incremental databases can be built, in contrast to sequence information, for example (Gevers *et al.*, 2005; Stackebrandt, 2003). Because of these drawbacks, bacterial taxonomists are actively searching for alternative methods that can replace DDH experiments (Cho & Tiedje, 2001; Coenye *et al.*, 2005; Gevers *et al.*, 2005).

[†]Present address: Applied Maths NV, Keistraat 120, B-9830 Sint-Martens-Latem, Belgium.

[‡]Present address: 15 Vassar Street, Room 48-336, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Abbreviations: ANI, average nucleotide identity; DDH, DNA–DNA hybridization.

We have recently shown that the average nucleotide identity (ANI) of conserved genes present in two sequenced strains represents a robust measure of the genetic and evolutionary distance between them, because it shows a strong correlation with 16S rRNA gene sequence similarity and the mutation rate of the genome, it is not affected by lateral transfer or variable recombination rates of single (or a few) genes and it offers resolution at the subspecies level (Konstantinidis & Tiedje, 2005). Previously, ANI was compared with DDH values using a limited number of published data often obtained with different hybridization methods. When DDH values were not available for the sequenced strains, mean DDH values for other strains of the same species were used in the calculations (Konstantinidis & Tiedje, 2005). However, it is important to account for strain differences within species and to perform all DDH experiments with a single, well-established method under identical experimental conditions.

The goal of the present study was to examine more accurately the relationship between DDH values and (genomic) sequence-derived parameters, such as ANI. For this purpose, we determined a large number of DDH values among related strains for which the whole genome had been sequenced. Furthermore, we evaluated whether genome size differences can explain differences in reciprocal reactions that are often observed with DDH.

METHODS

Strain selection. Strains in this study were chosen on the basis of the availability of their complete genome sequences (either fully closed or high draft sequences). Hybridization groups were defined as groups of strains sharing more than 94% 16S rRNA gene sequence identity (Table 1). The following strains were generously provided by the researchers indicated: *Shigella sonnei* 53G (Dr Ian Henderson, Department of Microbiology and Immunology, Queen's University of Belfast School of Medicine, UK), *Pseudomonas fluorescens* SBW25 (Dr Andrew J. Spiers, Department of Plant Sciences, University of Oxford, UK), *P. fluorescens* PfO-1 (Dr Mark W. Silby, Department of Molecular Biology and Microbiology, Tufts University School of Medicine, MA, USA), *Pseudomonas syringae* pv. *Syringae* B728a (Dr Steve E. Lindow, Department of Plant and Microbial Biology, University of California, CA, USA) and *P. syringae* pv. *Tomato* DC300 (Dr Cheng Yang He, Department of Plant Biology, Michigan State University, MI, USA). All six *Escherichia coli* strains and *Shigella flexneri* 2a 2457^T were generously provided by Dr Thomas S. Whittam (National Food Safety and Toxicology Center, Michigan State University, MI, USA). Dr Margaret F. Romine (Pacific Northwest National Laboratory, Richland, WA, USA) provided the *Shewanella* strains. *Bacillus cereus* ATCC 10987, *P. fluorescens* Pf-5 (=ATCC BAA-477), *Pseudomonas aeruginosa* PAO1 (=ATCC BAA-47), *Streptococcus agalactiae* 2603 V/R (=ATCC BAA-611) and *Streptococcus agalactiae* NEM 316 (=ATCC 12403) were obtained from the American Type Culture Collection (Manassas, VA, USA), and *Bacillus cereus* ATCC 14579^T (=LMG 6923^T) and *Streptococcus agalactiae* A909 (=LMG 15083) were provided by the BCCM/LMG Bacteria Collection (Gent University). All strains were grown aerobically for 24 h (see Table 1 for details of the growth media and incubation temperatures).

DNA–DNA hybridizations. DNA from Gram-negative organisms was prepared as described by Marmur (1961), with reagent volumes adapted to the use of 50 ml centrifuge tubes. For Gram-positive

organisms, DNA was prepared according to De Clerck *et al.* (2004). DDH reactions were as described by Ezaki *et al.* (1989), with slight modifications. Briefly, DNA was non-covalently adsorbed to polystyrene microplates (black MaxiSorp, FluoroNunc; Nunc) by incubating 100 µl portions of a denatured DNA solution [10 ng DNA per µl phosphate-buffered saline (PBS)/MgCl₂ (8 mM NaH₂PO₄, 1.5 mM KH₂PO₄, pH 7.2, 137 mM NaCl, 2.7 mM KCl, 0.1 M MgCl₂)] per well at 30 °C for 4 h in a hybridization oven. Before incubation, plates were sealed with self-adhesive vinyl tape (Nunc). The plates were then washed once with 300 µl PBS per well with the aid of a multichannel pipette, dried at 45 °C for 15 min and stored in a desiccator at 4 °C. Probe DNA was labelled by mixing 10 µl DNA solution [0.5 µg µl⁻¹ in 0.1 × SSC (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7.0 ± 0.2)] plus 10 µl photobiotin solution (Sigma) (0.5 µg µl⁻¹ in water) in a 1.5 ml Eppendorf tube and illuminating the mixture for 30 min under a 400 W mercury-vapour lamp while the open tube was kept upright in a cooling block on ice. The labelled probe DNA was diluted by adding 185 µl 0.1 M Tris/HCl (pH 9.0) and the remaining free photobiotin was removed by extracting twice with 200 µl water-saturated 1-butanol. The probe DNA was then fragmented with 30 ultrasonic pulses at 70% output (model W-385 sonicator; Heat Systems-Ultrasonics), denatured at 100 °C for 10 min and immediately cooled on ice. A pre-hybridization step was performed by adding 200 µl pre-hybridization solution (2 × SSC, 5 × Denhardt's solution, 50% formamide, 100 µg denatured salmon sperm DNA ml⁻¹) per well, sealing the microplate with vinyl tape and incubating it for 30 min at the appropriate hybridization temperature (Table 1) in the hybridization oven. For the actual hybridization, the pre-hybridization solution was removed and 100 µl hybridization solution (pre-hybridization solution plus 2.5% dextran sulfate and 1 µg probe DNA ml⁻¹) was added per well. The microplate was sealed again with vinyl tape and incubated for 3 h at the appropriate hybridization temperature. Hybridization temperatures (given in Table 1) chosen were about 5 °C higher (stringent conditions) than the optimal renaturation temperature calculated as $[0.51 \times (G+C \text{ mol}\%) + 47] - 36^\circ\text{C}$, where 36 °C is the correction for the presence of 50% formamide (De Ley, 1970; McConaughy *et al.*, 1969). The microplate was then washed three times with 300 µl 1 × SSC per well. For the enzymic development, 100 µl streptavidin-β-D-galactosidase (Gibco BRL) solution was added per well (0.5 U ml⁻¹ in PBS plus 0.5% BSA) and the microplate was covered with a preheated empty microplate and incubated for 10 min at 37 °C. Subsequently, the plate was washed three times with 300 µl 1 × SSC per well, using the microplate washer. Finally, the substrate for β-D-galactosidase, 4-methylumbelliferyl β-D-galactopyranoside (Sigma), was added (100 µl per well, 0.1 mg ml⁻¹ in PBS plus 1 mM MgCl₂) and the plate was incubated at 37 °C. The reaction product, 4-methylumbelliferone (excitation max., 360 nm; emission max., 465 nm) was quantified using a SpectraMax M2 microplate reader (Molecular Devices) at 0, 15, 30 and 45 min and data were immediately transferred to a personal computer. DDH values were calculated using the fluorescence measurements at 30 min; a homologous reaction was regarded as representing 100% reassociation.

Unless otherwise stated, the DDH values reported here are the means of at least two independent experiments (i.e. DNA immobilization and actual hybridizations performed in different batches on different days). In each of these experiments, all hybridization reactions were done in quadruplicate and calculations were based on the mean fluorescence values (clearly aberrant fluorescence values were omitted). All reciprocal hybridizations (different hybridizations using the same DNAs, A and B, but once with A as the immobilized DNA and once with B as the immobilized DNA) were carried out.

Sequence-based comparisons. All pairwise, whole-genome sequence comparisons were performed as follows. The genomic sequence from one of the genomes in a pair ('the query') was cut into

Table 1. Strains used in this study

Strain	Sequenced by:	Genome size (Mb)	G + C content (mol%)	Hybridization temperature (°C)	Growth conditions*
<i>Bacillus</i> hybridization group					
<i>Bacillus cereus</i> ATCC 10987	TIGR	5.22	35.6	35	NA; 30 °C
<i>Bacillus cereus</i> ATCC 14579 ^T	TIGR	5.41	35.3	35	NA; 30 °C
<i>Burkholderia</i> hybridization group					
<i>Burkholderia</i> sp. ATCC 17760	JGI	8.7	66.3	50	LB; 30 °C
<i>Burkholderia cenocepacia</i> AU1054	JGI	7.25	66.8	50	LB; 30 °C
<i>Burkholderia cenocepacia</i> HI2424	JGI	8.14	66.5	50	LB; 30 °C
<i>Burkholderia cenocepacia</i> J2315 ^T	Sanger	8.06	66.9	50	LB; 30 °C
<i>Burkholderia xenovorans</i> LB400 ^T	JGI	9.70	62.6	50	LB; 30 °C
<i>Burkholderia vietnamiensis</i> G4	JGI	8.4	65.5	50	LB; 30 °C
<i>Escherichia/Shigella</i> hybridization group					
<i>Escherichia coli</i> CFT073	University of Wisconsin	5.23	50.4	42	LB; 37 °C
<i>Escherichia coli</i> E2348/69	Sanger	5.07	50.5	42	LB; 37 °C
<i>Escherichia coli</i> K-12 MG1655	University of Wisconsin	4.60	50.7	42	LB; 37 °C
<i>Escherichia coli</i> O157:H7 EDL933	University of Wisconsin	5.60	50.3	42	LB; 37 °C
<i>Escherichia coli</i> O157:H7 Sakai	Japanese Consortium	5.50	50.5	42	LB; 37 °C
<i>Escherichia coli</i> O42	Sanger	5.36	50.5	42	LB; 37 °C
<i>Shigella flexneri</i> 2a 2457 ^T	University of Wisconsin	4.60	50.8	42	LB; 37 °C
<i>Shigella sonnei</i> 53G	Sanger	5.23	50.7	42	LB; 37 °C
<i>Pseudomonas</i> hybridization group					
<i>Pseudomonas aeruginosa</i> PAO1	University of Washington/ PathoGenesis Corp.	6.30	66.4	50	TSA; 37 °C
<i>Pseudomonas fluorescens</i> Pf-5	TIGR	7.10	63.3	50	TSA; 30 °C
<i>Pseudomonas fluorescens</i> PfO-1	JGI	6.44	60.6	50	TSA; 30 °C
<i>Pseudomonas fluorescens</i> SBW25	Sanger	6.72	60.5	50	TSA; 30 °C
<i>Pseudomonas syringae</i> pv. <i>Syringae</i> B728a	JGI	6.09	59.2	50	TSA; 30 °C
<i>Pseudomonas syringae</i> pv. <i>Tomato</i> DC3000	TIGR/Cornell	6.40	58.3	50	TSA; 30 °C
<i>Shewanella</i> hybridization group					
<i>Shewanella baltica</i> OS155	JGI	5.09	46.2	37	MB; 22 °C
<i>Shewanella oneidensis</i> MR-1 ^T	TIGR	4.97	43.7	37	TSB; 30 °C
<i>Shewanella putrefaciens</i> CN-32	JGI	4.58	44.4	37	TSB; 30 °C
<i>Streptococcus</i> hybridization group					
<i>Streptococcus agalactiae</i> 2603 V/R	TIGR	2.20	35.5	35	TSAB; 37 °C
<i>Streptococcus agalactiae</i> A909	TIGR	2.10	35.7	35	TSAB; 37 °C
<i>Streptococcus agalactiae</i> NEM 316	Institut Pasteur	2.21	35.5	35	TSAB; 37 °C

*LB, Luria-Bertani agar (1 l contains 10 g tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar); MB, marine broth 2216 (279110; Difco); NA, nutrient agar (CM0003; Oxoid); TSA, tryptone soya agar (CM0131; Oxoid); TSAB, TSA with 5 % defibrinated sheep blood; TSB, tryptic soy broth (211825; Difco).

consecutive 1020 nt fragments. The 1020 nt cut-off was used to correspond with the fragmentation of the genomic DNA to approximately 1 kb fragments during the DDH experiments. The use of different cut-offs (e.g. smaller fragments) did not notably modify our results (data not shown). The 1020 nt fragments were then used to search against the whole genomic sequence of the other genome in the pair ('the reference') by using the BLASTN algorithm (Altschul *et al.*, 1997); the best BLASTN match was saved for further analysis. The BLASTN algorithm was run using the following settings: X=150 (where X is the drop-off value for gapped alignment), q = -1 (where q is the penalty for nucleotide mismatch) and F=F (where F is the filter for repeated sequences); the rest of the parameters were used at the default settings. These settings give better sensitivity than the default settings when more distantly related genomes are being compared, as the latter target sequences that are more similar to each other.

To calculate the percentage of conserved DNA between a query and a reference, only those BLASTN matches reaching values above a cut-off point of 90 % nucleotide sequence identity were considered, regardless of the extent of the alignable region. The lengths of the alignable regions for all such matches were summed and the sum was divided by the total length of the genomic DNA of the query genome to provide a genome size-independent measurement of the percentage of the query's DNA that was conserved in the reference genome.

The ANI between the query genome and the reference genome was calculated as the mean identity of all BLASTN matches that showed more than 30 % overall sequence identity (recalculated to an identity along the entire sequence) over an alignable region of at least 70 % of their length. This cut-off is above the 'twilight zone' of similarity searches in which an inference of homology is error prone because of low levels of

similarity between aligned sequences (Rost, 1999; Sander & Schneider, 1991). Therefore we can assume that only homologous DNA fragments were considered in our calculations.

Reverse searching, i.e. in which the reference genome is used as the query, was also performed to provide reciprocal values. Perl scripts were used to extract 1020 nt fragments from whole-genome sequence files, formatting databases for BLAST searches and automatically parsing BLAST outputs. These scripts are available upon request.

RESULTS

The DDH values measured, the ANI data and the percentages of conserved DNA among strains within the six bacterial groups are presented in Table 2. For the DDH values, the mean values of replicate reactions as well as the standard deviations are given. We evaluated the correlation between DDH values, ANI data and the percentages of conserved DNA using different regression models (Figs 1 and 2). Regression analysis between DDH values and ANI data performed using linear, exponential, power and logarithmic models gave comparably high r^2 correlation values (0.94, 0.94, 0.95 and 0.94, respectively). The recommended cut-off point of 70 % DDH for species delineation thereby corresponded to an ANI of 95 ± 0.5 %, depending on the specific regression model used (Fig. 1a). When the analysis was restricted to the protein-coding portion of the genome, the 70 % DDH recommendation corresponded to a mean value of 85 % gene conservation for a pair of strains. In an effort to obtain a more conservative estimation of functional similarity, a reciprocal best-match approach was used to determine the orthologous fraction of the conserved genes. In this case, 70 % DDH corresponded to 79 % conserved genes. For the percentage of conserved DNA, we initially evaluated different sequence identity cut-off points to test which of them corresponded best to the DDH values. The 60, 70, 80, 90 and 95 % sequence identity cut-off points evaluated resulted in r^2 values of 0.85, 0.87, 0.87, 0.96 and 0.78, respectively. As the best correlation was obtained with a 90 % cut-off point, only those values (for percentage of conserved DNA) calculated with this cut-off are given in Table 2 and used in Fig. 1(b). In comparison with the linear model ($r^2=0.95$), significantly lower r^2 values were calculated with the exponential, power and logarithmic models (0.87, 0.82 and 0.71, respectively). With the linear model, a DDH value of 70 % corresponded to 69 % conserved DNA (Fig. 1b). The genomic sequence-derived parameters ANI and the percentage of conserved DNA correlated well at cut-off points above about 80 % ANI ($r^2=0.96$ for the linear model; Fig. 2). Given the 90 % cut-off point used for the calculation of the percentage of conserved DNA, no correlation should be expected at ANI values lower than 80 %. As ANI and the percentage of conserved DNA could not be considered as independent parameters, no multiple regression analysis with DDH could be performed.

Every DDH reaction was run in the reverse direction as well, i.e. in which the immobilized genome of the forward

reaction is used as the probe genome in the reverse reaction. We evaluated whether DNA content differences account for differences between values for the reciprocal reactions. For instance, if the amount of DNA conserved in two genomes A and B (where A has a larger genome size than B), is x Mb, then this conserved DNA represents a larger percentage of the genomic DNA of strain B compared with strain A. Therefore, the reassociation reaction with B as the immobilized genome should have a higher value than the reverse reassociation reaction (i.e. proportional to the genome size difference between the two strains). We found a weakly positive ($r^2=0.19$) but significant ($P<0.001$) linear correlation between the difference in the reciprocal DDH values and the difference in the percentage of conserved DNA between two genomes (Fig. 3). The weakness of the correlation is probably attributable to experimental error, which, although reasonably small (the mean standard deviation was 2.7 %), is relatively large with respect to the typically small differences between the reciprocal DDH values (the mean difference was 4.3 %). The ANI between a pair of strains typically shows <0.1 % difference between the reciprocal searches, i.e. when the query genome of the forward search was used as the reference in the reverse (reciprocal) search. Therefore, ANI should have only a minimal effect on differences between reciprocal DDH values, and therefore we did not investigate this correlation.

Finally, we looked for correlations between the mean G + C content of the genome, the percentage of genomic fragments deviating from the mean genomic G + C content, genomic duplication and the extent of variation between replicates of the same DDH hybridization (i.e. the experimental error). Genomic duplication was defined here as the fraction of the 1020 nt genomic fragments that have another match within their genome with more than 80 % identity. We found no significant correlations between any of these parameters and the experimental error. Therefore, the experimental error associated with DDH values is presumably attributable exclusively to technical issues concerning DDH experiments.

DISCUSSION

We determined 124 mean DDH values between 28 genome sequenced strains belonging to six phylogenetically distinct groups. Strains within a group were closely related, i.e. showing >70 % ANI and >94 % 16S rRNA gene sequence identity. The six bacterial groups were *Bacillus cereus*, *Burkholderia* species, *E. coli/Shigella* species, *Pseudomonas* species, *Shewanella* species and *Streptococcus agalactiae* (Table 2). These groups represent opportunistic pathogens as well as environmental species, they include both Gram-positive and Gram-negative bacteria and they comprise a wide range of genome sizes (Table 1), from 2 Mb (*Streptococcus agalactiae*) to mean genome sizes of 4–6 Mb (*Escherichia*, *Pseudomonas* and *Shewanella*) and up to 8 Mb (*Burkholderia*), which allowed for robust

Table 2. DDH values, ANI and percentage of conserved DNA among the strains studied

Values for DDH are means $\pm$ SD for independent repetitions of the same hybridization reaction; values without a standard deviation were determined only once. The percentage of conserved DNA was determined at a cut-off point of 90 % nucleotide identity (as calculated by BLASTN).

Immobilized DNA/reference genome	Probe DNA/tester genome	DDH (%)	ANI (%)	Percentage conserved DNA
<i>Bacillus</i> hybridization group				
ATCC 10987	ATCC 14579 ^T	66.4 $\pm$ 0.8	90.93	49.7
ATCC 14579 ^T	ATCC 10987	61.2 $\pm$ 3.0	90.87	48.3
<i>Burkholderia</i> hybridization group				
ATCC 17760	HI2424	57.4 $\pm$ 4.1	89.77	39.6
	J2315 ^T	54.5 $\pm$ 2.4	89.88	42.7
	LB400 ^T	13.8 $\pm$ 1.8	76.10	0.4
	G4	32.4 $\pm$ 4.7	86.31	18.1
AU1054	LB400 ^T	14.8 $\pm$ 0.8	76.80	0.5
	G4	36.5 $\pm$ 3.6	87.35	24.6
HI2424	ATCC 17760	45.4 $\pm$ 10.7	89.89	43.3
	J2315 ^T	66.1 $\pm$ 0.8	93.84	72.6
	LB400 ^T	12.5 $\pm$ 1.1	76.50	0.5
	G4	37.7 $\pm$ 6.0	87.16	23.1
J2315 ^T	ATCC 17760	55.8 $\pm$ 10.8	90.28	46.9
	HI2424	68.8 $\pm$ 0.1	93.86	72.1
	LB400 ^T	13.0 $\pm$ 3.1	76.60	0.6
	G4	37.0 $\pm$ 4.7	87.40	23.3
LB400 ^T	ATCC 17760	12.6 $\pm$ 1.5	75.95	0.5
	AU1054	11.8 $\pm$ 4.1	76.24	0.4
	HI2424	12.7 $\pm$ 1.1	76.21	0.4
	J2315 ^T	13.7 $\pm$ 2.2	76.04	0.4
	G4	10.1 $\pm$ 1.2	76.15	0.5
G4	ATCC 17760	29.9 $\pm$ 9.0	86.94	18.6
	AU1054	26.3 $\pm$ 8.0	87.56	21.6
	HI2424	26.2 $\pm$ 1.9	87.52	21.8
	J2315 ^T	34.2 $\pm$ 5.2	87.54	22.4
	LB400 ^T	9.0 $\pm$ 1.2	76.60	0.5
<i>Escherichia/Shigella</i> hybridization group				
CFT073	E2348/69	74.8 $\pm$ 3.6	97.62	79.5
	K-12 MG1655	89.0	96.07	73.6
	O157:H7 EDL933	78.2 $\pm$ 1.3	96.01	74.6
	O157:H7 Sakai	85.5 $\pm$ 3.0	96.00	74.9
	O42	81.4 $\pm$ 7.1	95.90	77.9
	2a 2457 ^T	66.9 $\pm$ 5.5	95.90	69.0
	53G	72.8	95.93	70.4
E2348/69	CFT073	78.1 $\pm$ 6.3	97.85	82.8
	K-12 MG1655	75.7 $\pm$ 6.8	96.39	74.8
	O157:H7 EDL933	74.2 $\pm$ 3.6	96.18	78.4
	O157:H7 Sakai	81.8 $\pm$ 3.4	96.23	78.8
	O42	78.7 $\pm$ 3.5	96.14	77.2
	2a 2457 ^T	65.0 $\pm$ 0.7	96.09	70.7
	53G	64.3 $\pm$ 3.4	96.01	72.6
K-12 MG1655	CFT073	95.7 $\pm$ 1.5	96.38	83.1
	E2348/69	75.5 $\pm$ 4.1	96.48	81.5
	O157:H7 EDL933	78.8 $\pm$ 2.2	97.50	88.0
	O157:H7 Sakai	84.8 $\pm$ 3.7	97.53	88.0
	O42	75.5 $\pm$ 1.8	96.83	87.5
	2a 2457 ^T	63.4 $\pm$ 0.6	97.61	81.2
	53G	70.4	97.80	84.1

Table 2. cont.

Immobilized DNA/reference genome	Probe DNA/tester genome	DDH (%)	ANI (%)	Percentage conserved DNA
O157:H7 EDL933	CFT073	83.0 ± 7.1	96.04	72.6
	E2348/69	78.2 ± 2.7	95.90	73.8
	K-12 MG1655	79.7 ± 4.7	97.28	74.2
	O157:H7 Sakai	104.1 ± 3.1	99.68	99.6
	O42	73.3 ± 1.7	95.91	77.3
	2a 2457 ^T	65.6 ± 1.6	96.66	68.5
	53G	75.1 ± 4.8	96.77	72.4
O157:H7 Sakai	CFT073	83.8 ± 3.5	95.85	71.4
	E2348/69	77.1 ± 3.9	95.76	72.8
	K-12 MG1655	75.9 ± 6.5	97.25	73.2
	O157:H7 EDL933	97.7 ± 1.1	99.63	97.3
	O42	78.2 ± 3.6	95.84	76.8
	2a 2457 ^T	61.3 ± 0.5	96.54	67.8
	53G	71.2 ± 2.5	96.57	71.9
O42	CFT073	84.5 ± 6.0	96.04	76.5
	E2348/69	71.4 ± 3.1	96.10	73.1
	K-12 MG1655	71.7 ± 2.2	96.75	76.1
	O157:H7 EDL933	76.7 ± 1.6	96.26	79.1
	O157:H7 Sakai	77.1 ± 4.2	96.27	79.5
	2a 2457 ^T	64.4 ± 0.7	96.39	70.3
	53G	66.7	96.10	72.4
2a 2457 ^T	CFT073	78.1 ± 7.5	96.40	82.0
	E2348/69	67.5 ± 1.7	96.31	79.3
	K-12 MG1655	72.8 ± 3.4	97.75	84.6
	O157:H7 EDL933	70.0 ± 6.5	97.30	84.6
	O157:H7 Sakai	70.5 ± 0.0	97.36	84.8
	O42	69.2 ± 1.7	96.70	84.7
	53G	78.4 ± 1.1	97.89	85.3
53G	CFT073	83.3	96.20	73.6
	E2348/69	77.6 ± 2.0	96.20	71.6
	K-12 MG1655	79.2	97.80	77.7
	O157:H7 EDL933	73.6 ± 6.0	97.15	77.7
	O157:H7 Sakai	76.8 ± 3.3	97.09	78.3
	O42	76.1	96.55	77.3
	2a 2457 ^T	81.0 ± 4.7	97.68	75.8
<i>Pseudomonas</i> hybridization group				
PAO1	Pf-5	14.4 ± 0.6	75.75	0.003
	PfO-1	13.9 ± 0.2	74.86	0.002
	SBW25	14.7 ± 1.5	74.47	0.002
	B728a	8.5 ± 0.5	73.92	0.003
	DC3000	6.3 ± 0.3	73.57	0.002
Pf-5	PAO1	13.6 ± 0.3	75.50	0.0001
	PfO-1	25.8 ± 0.3	80.63	2.6
	SBW25	29.6 ± 2.6	80.02	2.3
	B728a	14.1 ± 1.3	76.89	0.006
	DC3000	11.2 ± 0.0	76.63	0.006
PfO-1	PAO1	13.0 ± 0.0	74.73	0.001
	Pf-5	29.4 ± 0.1	80.64	2.9
	SBW25	32.4 ± 2.1	80.18	2.6
	B728a	19.1 ± 0.8	77.01	0.007
	DC3000	15.3 ± 0.3	76.95	0.006
SBW25	PAO1	10.5 ± 7.0	74.27	0.003
	Pf-5	24.9 ± 3.0	79.97	2.0

Table 2. cont.

Immobilized DNA/reference genome	Probe DNA/tester genome	DDH (%)	ANI (%)	Percentage conserved DNA
B728a	PfO-1	26.3 ± 3.0	80.14	2.8
	B728a	14.8 ± 1.2	76.59	0.009
	DC3000	11.7 ± 1.5	76.33	0.008
	PAO1	7.6 ± 0.7	73.64	0.001
	Pf-5	12.7 ± 0.1	76.62	0.008
	PfO-1	16.7 ± 0.2	76.66	0.008
	SBW25	18.0 ± 1.2	76.47	0.007
DC3000	DC3000	37.8 ± 2.3	85.46	13.1
	PAO1	6.1 ± 0.2	73.61	0.003
	Pf-5	10.5 ± 0.5	76.84	0.008
	PfO-1	13.0 ± 0.2	76.91	0.009
	SBW25	15.7 ± 1.6	76.65	0.009
	B728a	38.5 ± 2.8	85.73	12.4
<i>Shewanella</i> hybridization group				
OS155	MR-1 ^T	25.7	78.32	2.31
MR-1 ^T	CN-32	26.1	81.88	6.85
	OS155	26.2	78.52	3.61
CN-32	CN-32	21.1 ± 4.9	78.63	2.90
	OS155	27.8	81.90	7.67
	MR-1 ^T	22.1 ± 8.1	78.46	2.20
<i>Streptococcus</i> hybridization group				
2603 V/R	A909	95.5 ± 4.0	98.78	87.6
	NEM 316	89.6 ± 4.8	98.36	87.3
A909	2603 V/R	91.4 ± 5.3	98.74	86.9
	NEM 316	92.8 ± 0.1	98.61	86.2
NEM 316	2603 V/R	83.1 ± 4.3	98.34	84.9
	A909	88.5 ± 3.3	98.61	84.7

interpretations. The microplate hybridization method of Ezaki *et al.* (1989) is a well-established and frequently used method in bacterial taxonomy (Rosselló-Mora, 2006). A mean standard deviation of 2.7 % was calculated between replicate DDH experiments. This deviation is comparable to or smaller than those reported in previous studies (Christensen *et al.*, 2000; Goris *et al.*, 1998; Huß *et al.*, 1983; Johnson, 1991), indicating that the experimental error was consistently small.

Several reviews (Rosselló-Mora, 2006; Rosselló-Mora & Amann, 2001; Stackebrandt & Goebel, 1994; Stackebrandt & Liesack, 1993) mention that DNA fragments must share at least 80 % identity in order to hybridize during DDH experiments. However, this statement is based on early studies of the hybridization kinetics of unmodified and alkali-deaminated DNA (Ullmann & McCarthy, 1973) or synthetic polyribonucleotides (Bautz & Bautz, 1964). We found that a cut-off point of 90 % nucleotide identity gave a significantly better correlation between the percentages of conserved DNA and the DDH values than the 80 % cut-off point ($r^2 = 0.96$ versus 0.87, respectively). Our analysis does not preclude the possibility that some genomic fragments of 80 % (or less) identity cross-hybridize, but it shows that fragments of greater identity are more important during the

genome-scale hybridizations; consequently, a 90 % cut-off point was used in the remaining analysis to determine the percentage of conserved DNA between two strains.

Our results revealed a close relationship between DDH values and ANI (Fig. 1a) and between DDH values and the percentage of conserved DNA (Fig. 1b) for each pair of strains. Because of the very small differences between different models (linear, exponential, power and logarithmic) in terms of their ability to describe the relationship between DDH and ANI, no assumptions can be made about the mechanisms underlying this relationship based on these comparisons. The relationship between DDH values and percentage of conserved DNA was best described by a linear model. The model's divergence from the ideal situation ($y = x$) at lower values for the percentage of conserved DNA could be explained by a small contribution from DNA fragments with less than 90 % identity but that still hybridize (Fig. 1b). The finding that ANI and the percentage of conserved DNA are strongly correlated is consistent with previous results (Konstantinidis & Tiedje, 2005). Therefore, only one of these two genome-derived parameters is necessary for a fairly accurate prediction of the expected DDH values between two strains. According to our dataset, the classical cut-off point of 70 % DDH similarity for species

Fig. 1. Relationship between DDH values and genomic sequence identity and conservation. Each filled circle represents the value for DDH between two strains (y -axis), plotted against the ANI of the conserved genes between the strains (a) and the percentage of conserved DNA between the strains (b). The standard deviations for the DDH values, omitted from (a) for simplicity, are shown in (b). A linear trend line is shown, but other regression models were evaluated as well (see text). The horizontal broken lines denote the 70% DDH recommendation for species delineation, while the vertical broken lines denote the corresponding ANI (a) and percentage of conserved DNA (b) values for linear regression.

delineation corresponds to 95% ANI and 69% conserved DNA. With the analysis restricted to the protein-coding portion of the genome, 70% DDH corresponds to 85 or 79% conserved genes between a pair of strains when, respectively, a one-way or a reciprocal best-match approach was used to determine the orthologous fraction of the conserved genes. These results reveal that the 70% DDH recommendation encompasses relatively homogeneous strains at the genomic level, which is consistent with previous studies on the phenotypic similarity of the strains (Stackebrandt & Goebel, 1994; Wayne *et al.*, 1987). Nonetheless, a difference of up to 21% in gene content between strains showing $\geq 70\%$ DDH represents a large genetic and (presumably) phenotypic difference, e.g. up to

Fig. 2. Relationship between genomic sequence identity and conservation. Each filled circle represents the percentage of conserved DNA shared between two strains (determined at 90% nucleotide identity), plotted against the ANIs of their common genes.

1000 genes may differ between two strains with a 5 Mb genome (approximately the mean genome size). Such a large difference in gene content would probably be responsible for a suite of important phenotypes, which could justify the description of such strains as separate species or, at least,

Fig. 3. Reciprocal DDH values versus DNA-content differences. Each filled circle represents the difference between the reciprocal DDH reactions (y -axis) for a pair of strains (e.g. strain A versus strain B, with A being the query and immobilized strain in the sequence comparisons and DDH experiments, respectively), plotted against the difference in the percentage of conserved DNA between the two strains. A negative value on the y -axis indicates that the reciprocal reaction, i.e. using B as the query and immobilized strain, gave a higher DDH value than the forward reaction. A negative value on the x -axis indicates that the conserved DNA between the two strains represents a larger fraction of strain B (or, alternatively, that strain B has a smaller genome size than strain A).

ecotypes. It is possible, however, that these phenotypes would only be important under natural conditions, i.e. they might not be apparent under laboratory conditions because of technological limitations. If the results based on the six bacterial groups considered here are more universally applicable in the prokaryotic world, then our results suggest that the 70 % DDH criterion can only serve as a first (coarse) level of screening for species. Higher resolution should then be adopted, as necessary, for particular groups of organism.

Theoretically, differences in DNA content (or genome size) could be expected to lead to differences in reciprocal DDH values. However, the low level of correlation found between differences in reciprocal DDH values and the difference in the percentage of conserved DNA between two genomes (Fig. 3) indicates that DDH is too coarse a method (i.e. the experimental error is too high) to reveal subtle differences in genome size between strains. Experimental error could not be explained by (deviations of) the mean genomic G+C content or genomic duplications and is therefore probably solely attributable to technical errors such as DNA impurities and fragmentation affecting the efficiencies of the immobilization, hybridization and enzymic reactions.

Our results (Fig. 1), together with data from other studies (Rademaker *et al.*, 2000; Vauterin *et al.*, 1995), suggest that DDH values are continuous, i.e. theoretically, every value between 0 and 100 % could be obtained in DDH experiments. These data are supportive of a continuous gradient of genetic relatedness rather than discrete species boundaries. Although relatively few of the 28 strains studied appeared to be moderately related (i.e. showing 80–90 % ANI or 30–60 % DDH) (Fig. 1), this result is probably attributable to a bias in the collection of sequenced strains rather than to species boundaries. We recently reported on the occurrence of species-specific diagnostic genetic signatures among sequenced representatives of *E. coli/Shigella* and *Salmonella* (Konstantinidis & Tiedje, 2005). The recent description of *Escherichia albertii* (Huys *et al.*, 2003), a novel species that probably spans the genetic gap between *E. coli* and *Salmonella* (Hyma *et al.*, 2005), as well as the analysis of environmental *E. coli* isolates (Byappanahalli *et al.*, 2006; Ishii *et al.*, 2006), indicates, however, that a genetic continuum may indeed be present for this group of bacteria as well. Besides, because of the pronounced decrease in the percentage of conserved DNA shown with increasing evolutionary distance (Fig. 2), discontinuities in the DDH values should be expected every time distantly related groups are compared (e.g. <80–85 % ANI) such as *E. coli* versus *Salmonella* (~80 % ANI). Shorter evolutionary scales, e.g. corresponding to 85–100 % ANI, are the most important – and at the same time the most under-investigated – areas for investigation with respect to species boundaries. The current dataset is simply too small for either validation or rejection of the existence of ‘condensed nodes in a cloudy and confluent taxonomic space’

(Vandamme *et al.*, 1996). As has been stated previously by other authors (e.g. Rosselló-Mora, 2003), species delineation through the rigid application of any standard for DNA–DNA relatedness (such as the 70 % cut-off point) is purely arbitrary. Our data further validate this statement, as the 70 % cut-off point does not necessarily correlate with clear genomic clusters within the set of strains investigated.

A note should be made here regarding DDH values and species designation. Despite the fact that they are classified within separate genera, it is known that in the context of population genetics the four *Shigella* species belong to the diverse species *E. coli* (Lan & Reeves, 2002). This is clearly reflected in our hybridization results (Table 2): reassociation values between 61.3 and 83.3 % were found between *Shigella sonnei* 53G or *Shigella flexneri* 2a 2457^T and the six *E. coli* strains. These values are comparable with those found among the *E. coli* strains (71.4–100 %, highest level of similarity being between the two O157 serotypes).

Remarkably low DDH values were found between *Pseudomonas* strains that are reported to belong to the same species: *P. fluorescens* strains Pf-5, SBW25 and PfO-1 yielded DDH values between 25 and 32 %, whereas the two *P. syringae* strains, B728a and DC3000, yielded DDH values of 38–39 % (Table 2). These low reassociation values demonstrate that these strains cannot belong to the same species. Consistent with this, the ANI values among these genomes are much lower than the 95 % ANI value corresponding to the 70 % DDH recommendation.

While the first example might be common knowledge among microbiologists, the second illustrates that caution should be exercised when drawing conclusions in genome-comparison studies based on the reported species name for some sequenced strains.

In conclusion, we have shown that DDH values correlate well with the genome sequence-derived parameters ANI and the percentage of conserved DNA. A value of 70 % DDH thereby corresponds to about 95 % ANI and 69 % conserved DNA. Previously published DDH values could be used to give a rough approximation of the ANI values and gene-content differences between the strains evaluated, using the equations described here (Fig. 1). For more accurate measurements, however, alternative methods are needed. At present, only a relatively small fraction of all available strains can be fully sequenced, but multilocus sequencing analysis using appropriate genetic markers might have potential application in this area. Further investigation is required, however, to determine whether multilocus sequencing analysis correlates as well as DDH with ANI. Despite its drawbacks, DDH remains valuable in bacterial taxonomy. It is the accepted standard, and, to date, no other universally applicable and cost-effective technique offers genome-wide comparison. However, the steadily decreasing cost of DNA sequencing means that DDH is likely to be replaced by sequence-based techniques in the not-too-distant future.

ACKNOWLEDGEMENTS

J. G. acknowledges the support he received from the Belgian American Educational Foundation, in the form of a BAEF post-doctoral fellowship. The authors thank the researchers (named in the text) that generously provided strains.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Bautz, E. K. F. & Bautz, F. A. (1964). The influence of non complementary bases on the stability of ordered polynucleotides. *Proc Natl Acad Sci U S A* **52**, 1476–1481.
- Brenner, D. J., Fanning, G. R., Rake, A. V. & Johnson, K. E. (1969). Batch procedure for thermal elution of DNA from hydroxyapatite. *Anal Biochem* **28**, 447–459.
- Byappanahalli, M. N., Whitman, R. L., Shively, D. A., Sadowsky, M. J. & Ishii, S. (2006). Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ Microbiol* **8**, 504–513.
- Cho, J. C. & Tiedje, J. M. (2001). Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl Environ Microbiol* **67**, 3677–3682.
- Christensen, H., Angen, Ø., Møtters, R., Olsen, J. E. & Bisgaard, M. (2000). DNA-DNA hybridization determined in micro-wells using covalent attachment of DNA. *Int J Syst Evol Microbiol* **50**, 1095–1102.
- Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147–167.
- Crosa, J. H., Brenner, D. J. & Falkow, S. (1973). Use of a single-strand specific nuclease for analysis of bacterial and plasmid deoxyribonucleic acid homo- and heteroduplexes. *J Bacteriol* **115**, 904–911.
- De Clerck, E., Rodriguez-Diaz, M., Vanhoutte, T., Heyman, J., Logan, N. A. & De Vos, P. (2004). *Anoxybacillus contaminans* sp. nov. and *Bacillus gelatini* sp. nov., isolated from contaminated gelatin batches. *Int J Syst Evol Microbiol* **54**, 941–946.
- De Ley, J. (1970). Re-examination of the association between melting point, buoyant density, and chemical base composition of deoxyribonucleic acid. *J Bacteriol* **101**, 738–754.
- De Ley, J., Cattoir, H. & Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**, 133–142.
- Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane-filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39**, 224–229.
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P. & other authors (2005). Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733–739.
- Goris, J., Suzuki, K., De Vos, P., Nakase, T. & Kersters, K. (1998). Evaluation of a microplate DNA-DNA hybridization method compared with the initial renaturation method. *Can J Microbiol* **44**, 1148–1153.
- Grimont, P. A. D., Popoff, M. Y., Grimont, F., Coynault, C. & Lemelin, M. (1980). Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Curr Microbiol* **4**, 325–330.
- HuB, V. A. R., Festl, H. & Schleifer, K. H. (1983). Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Syst Appl Microbiol* **4**, 184–192.
- Huys, G., Cnockaert, M., Janda, J. M. & Swings, J. (2003). *Escherichia albertii* sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. *Int J Syst Evol Microbiol* **53**, 807–810.
- Hyma, K. E., Lacher, D. W., Nelson, A. M., Bumbaugh, A. C., Janda, J. M., Strockbine, N. A., Young, V. B. & Whittam, T. S. (2005). Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. *J Bacteriol* **187**, 619–628.
- Ishii, S., Ksoll, W. B., Hicks, R. E. & Sadowsky, M. J. (2006). Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl Environ Microbiol* **72**, 612–621.
- Johnson, J. L. (1991). DNA reassociation experiments. In *Nucleic Acid Techniques in Bacterial Systematics*, pp. 21–44. Edited by E. Stackebrandt & M. Goodfellow. Chichester: Wiley.
- Konstantinidis, K. T. & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**, 2567–2572.
- Lan, R. & Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect* **4**, 1125–1132.
- Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J Mol Biol* **3**, 208–218.
- McConaughy, B. L., Laird, C. D. & McCarthy, B. J. (1969). Nucleic acid reassociation in formamide. *Biochemistry* **8**, 3289–3295.
- Rademaker, J. L., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P. & de Bruijn, F. J. (2000). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int J Syst Evol Microbiol* **50**, 665–677.
- Rosselló-Mora, R. (2003). Opinion: the species problem, can we achieve a universal concept? *Syst Appl Microbiol* **26**, 323–326.
- Rosselló-Mora, R. (2006). DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In *Molecular Identification, Systematics and Population Structure of Prokaryotes*, pp. 23–50. Edited by E. Stackebrandt. Berlin: Springer.
- Rosselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* **25**, 39–67.
- Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng* **12**, 85–94.
- Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**, 56–68.
- Stackebrandt, E. (2003). The richness of prokaryotic diversity: there must be a species somewhere. *Food Technol Biotechnol* **41**, 17–22.
- Stackebrandt, E. & Goebel, B. M. (1994). A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.
- Stackebrandt, E. & Liesack, W. (1993). Nucleic acids and classification. In *Handbook of New Bacterial Systematics*, pp. 151–194. Edited by M. Goodfellow & A. G. O'Donnell. London: Academic Press.
- Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kämpfer, P., Maiden, M. C. J., Nesme, X., Rosselló-Mora, R., Swings, J. & other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1047.
- Ullmann, J. S. & McCarthy, B. J. (1973). The relationship between mismatched base pairs and the thermal stability of DNA duplexes. *Biochim Biophys Acta* **294**, 416–424.

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**, 407–438.

Vauterin, L., Hoste, B., Kersters, K. & Swings, J. (1995). Reclassification of *Xanthomonas*. *Int J Syst Bacteriol* **45**, 472–489.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E. & other authors (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.