Announcement

Brunswick, Germany – March 3, 2016 – A new web service is now available via the GGDC web site that allows for the calculation of pairwise similarities between, e.g., 16S rRNA gene sequences under the recommended settings detailed in the publication by Meier-Kolthoff et al. (2013) found below. Moreover, the service allows users to infer both maximum-likelihood and maximum-parsimony phylogenies for a set of uploaded sequences. The service is available under the following URL:

http://ggdc.dsmz.de/phylo_form.php

When should a DDH experiment be mandatory in microbial taxonomy?

Jan P. Meier-Kolthoff [*], **Markus Göker** [*] [†], **Cathrin Spröer** [*] **and Hans-Peter Klenk** [*]

[*]Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, c/o Markus Göker, Inhoffenstr. 7B, 38124 Braunschweig, Germany, and [†]Correspondence: markus.goeker@dsmz.de

Submitted to Archives of Microbiology – Received: 25 February 2013 / Revised: 27 March 2013 / Accepted: 30 March 2013 / Published online: 17 April 2013

DNA–DNA hybridizations (DDH) play a key role in microbial species discrimination in cases when 16S rRNA gene sequence similarities are 97 % or higher. Using real-world 16S rRNA gene sequences and DDH data, we here re-investigate whether or not, and in which situations, this threshold value might be too conservative. Statistical estimates of these thresholds are calculated in general as well as more specifically for a number of phyla that are frequently subjected to DDH. Among several methods to infer 16S gene sequence similarities investigated, most of those routinely applied by taxonomists appear well suited for the task. The effects of using distinct DDH methods also seem to be insignificant. Depending on the investigated taxonomic group, a threshold between 98.2 and 99.0 % appears reasonable. In that way, up to half of the currently conducted DDH experiments could safely be omitted without a significant risk for wrongly differentiated species.

DNA-DNA hybridization | Species concept | 16S rRNA gene | Generalized linear model | BLAST | Smith-Waterman | Substitution model | Microbial taxonomy

Abbreviations: DDH, DNA-DNA hybridization

Introduction

Within the realm of the artificial species concept in microbiology, the question whether two strains belong to the same species is one of the key questions for microbial taxonomists and most often decided by the determination of the degree of DNA-DNA hybridization (DDH) (Wayne et al., 1987). Because the various established techniques for the determination of DDH values are tedious and performed in only few specialized labs internationally, it is of relevance in practice to avoid unnecessary DDH measurements if evidence suggests that a value of 70 % (=both strains belong to the same species) is unexpected. Methods have been established to replace wet-lab DDH by in silico methods based on genome sequencing (Auch et al., 2010b,a; Konstantinidis and Tiedje, 2005; Meier-Kolthoff et al., 2013; Richter and Rosselló-Móra, 2009), but at the moment, a 16S rRNA gene sequence is still easier to obtain than a (partial) genome sequence. In 1994, Stackebrandt and Goebel recommended a 97 % 16S rRNA gene sequence similarity threshold up to that an additional DDH determination need not be conducted for confirming that two strains do not belong to the same species (Stackebrandt and Goebel, 1994). This guideline was based on the visual interpretation of an empirical dataset in which only pairs of strains with a 16S rRNA gene sequence similarity >97 % resulted in DDH values ⩾70 % and was widely accepted by the scientific community (Tindall et al., 2006), as demonstrated by, at the time of writing (February 2013), 3,361 citations according to Google Scholar alone.

Sequencing errors lead to an underestimation of pairwise similarities (Stackebrandt and Ebers, 2006). Improvements in sequencing technology during the last decade yielded more accurate sequences and similarity values, thus allowing taxonomists, in principle, to significantly reduce the number of tedious determinations of DDH values in wetlab experiments to be conducted. In 2006, Stackebrandt and Ebers hence suggested an updated, higher threshold (98.7 - 99.0 %) again based on an empirical dataset of DDH values and respective 16S rRNA gene sequence similarities collected from the taxonomic literature (Stackebrandt and Ebers, 2006). However, editors of the International Journal of Systematic and Evolutionary Microbiology (IJSEM) seemed reluctant to accept this shift toward a higher threshold value, probably because it did not appear to be sufficiently conservative.

A higher range of threshold values has been supported by the studies of (Konstantinidis and Tiedje, 2005, 2007), but based upon average nucleotide identities (ANI) of complete and partial genome sequences rather than wet-lab DDH values. Such an indirect approach is apparently based on a high correlation between ANI and wet-lab DDH values (Goris et al., 2007). The slightly older genome BLAST distance phylogeny (GBDP) approach (Henz et al., 2005), if appropriately adapted, yielded somewhat higher correlations with wet-lab DDH than ANI (Auch et al., 2010b; Meier-Kolthoff et al., 2013). Moreover, the linear-regression model used in the ANI and the earlier GDBP studies is not optimal for this kind of data (Meier-Kolthoff et al., 2013). But even if more suitable models (and larger empirical training datasets) are used, model parameters for predicting DDH from ANI or GBDP cannot be estimated without uncertainty (Meier-Kolthoff et al., 2013). In an indirect approach, the limiting factor is thus not the (impressive) size of the dataset relating 16S rRNA to ANI or GBDP (Konstantinidis and Tiedje, 2005, 2007), but the (smaller) size of the dataset relating ANI or GBDP to wetlab DDH (Meier-Kolthoff et al., 2013). For this reason, a direct approach for relating DDH and 16S rRNA gene similarities (Stackebrandt and Ebers, 2006) seems preferable to tackle this crucial taxonomic problem. Furthermore, an appropriate statistical assessment is necessary (Motulsky and Christopoulos, 2004, pp. 58-79), including the quantification of the effect of the taxonomic group (Keswani and Whitman, 2001).

Here, the situation is re-evaluated based on two premises. Firstly, predictions from empirical data (which are always limited) are always prone to a certain degree of failure. For instance, even a 16S rRNA gene similarity below 97 % does, in theory, not fully exclude the possibility that a DDH value

Communicated by Erko Stackebrandt.

Electronic supplementary material The original online version of this article contains supplementary material, which is available to authorized users. It is found at the following address: http://dx.doi.org/10.1007/s00203-013-0888-4

⩾70 % could be obtained between two strains, but the probability of such an event is regarded as negligible in practice. Secondly, calculating the probabilities of estimation errors is an empirical and statistical question, but defining the maximum acceptable probability of failure is a task for the editorial board of journals covering taxonomic issues. This study thus statistically estimates probabilities of failure for a series of 16S rRNA gene similarity thresholds, yielding a decision table that could serve as a guidance for editors and authors to accept or reject the proposal to create new species.

Materials and Methods

The dataset used in (Stackebrandt and Ebers, 2006) comprised accession numbers and DDH values for 376 distinct pairs of strains collected from volume 55 of the IJSEM. For the present study, 195 additional pairs were obtained from taxonomic journals (mostly IJSEM, too). These included 177 pairs of meanwhile publicly available DDH values and 16S rRNA gene sequences obtained at the German Collection of Microorganisms and Cell Cultures (see Table S1, Supplementary material) as assessed in the last couple of years. All in all 45 of these 571 pairs were associated with a DDH value above 70 %. Since this work studies models with dichotomous outcomes, this number of so-called "events" is the crucial factor for the necessary size of the dataset to avoid overfitting (Harrell et al., 1984). Rules of thumb (Peduzzi et al., 1996; Steyerberg et al., 2000) suggest at least a 1:10 relationship between the number of predictor variables and the number of events. Here, all models are at most based on two predictor variables (SSU similarity and phylum affiliation; see below) and thus are clearly satisfying this requirement.

The combined dataset included strains affiliated to the phyla Actinobacteria (#140), Bacteroidetes (#26), Deinococcus-Thermus (#25), Euryarcheaota (#14), Firmicutes (#100), and Proteobacteria (#266) according to the List of Prokaryotic names with Standing in Nomenclature (Euzéby, 1997). Since a vast majority of samples from the denoted journal volume lacked sufficient information on how 16S rRNA gene similarities were calculated, and because comparable 16S rRNA gene sequence similarity data were needed for the old and novel part of the dataset, all 376 16S rRNA gene sequence similarities from reference (Stackebrandt and Ebers, 2006) were recalculated, which was possible because both parts of the dataset contained mostly type strains, whose accession numbers were collected from reference (Euzéby, 1997).

Pairwise similarities of 16S rRNA gene sequences were calculated from exact pairwise sequence alignments using the Smith-Waterman algorithm as implemented in EMBOSS version 6.3.1 (Rice et al., 2000). Additionally, pairwise similarities (recalculated from distances by subtraction from 1.0) were inferred from the alignments using PAUP* version 4b10 (Swofford, 2003) and uncorrected (p distances) as well as the substitution models JC (Jukes and Cantor, 1969, pp. 21-132), K2P (Kimura, 1980), and HKY85 (Hasegawa et al., 1985) with and without applying a gamma distribution (Yang, 1993) with an alpha parameter of 0.5 for modeling site heterogeneity. Alternatively, local alignments [high-scoring segment pairs (HSPs)] were generated with NCBI BLAST (Altschul et al., 1990) (blastall version 2.2.25) under default settings, overlapping ones, if any, corrected by keeping the longer HSP (Henz et al., 2005), and pairwise similarities calculated by adding up the numbers of identical positions within HSPs and the overall HSP lengths. For BLAST hits that comprise only a single HSP (which is usually the case for highly similar sequences), this procedure is equivalent to the usually reported BLAST similarity, but potentially more accurate in all other cases because additional non-overlapping HSPs are also considered.

A generalized linear model (GLM) is a flexible generalization of standard linear regression that allows for response variables that need not be normally distributed (Nelder and Wedderburn, 1972). Here, a GLM was constructed with DDH encoded as a binary response variable (0 if DDH <70 %, indicating distinct species; 1 otherwise, indicating a pair of strains belongs to the same species) and the respective 16S rRNA gene similarity values as predictor variable (Crawley, 2007, pp. 593-609). Such a model yields for any given 16S rRNA gene similarity value the probability that it corresponds to a DDH value ⩾70 %. For a given similarity threshold, a special kind of error can occur when it is decided to not conduct a DDH experiment because a 16S rRNA gene similarity was found below that threshold. Such a decision apparently only caused an erroneous result if the strains belonged to the same species (because then the low 16S rRNA gene similarity would directly be treated as indicative of distinct species). Thus, the GLM yields, for each possible similarity value, the maximum probability of such an error if this similarity value is accepted as threshold up to which a 16S rRNA gene similarity is regarded as sufficiently low to allow one to omit DDH (Distinct usage examples are provided below).

A GLM was inferred for each of the aforementioned procedures for the inference of 16S rRNA gene sequence similarities, and the best one(s) chosen with respect to the following statistical criteria: (a) highest Kendall correlation coefficient between 16S rRNA gene similarity and DDH values, (b) lowest Akaike information criterion (Akaike, 1974) of the model, and (c) lowest minimum standard error in tenfold cross-validation (Stone, 1974) of the model. The latter is indicative of how well the results of a statistical analysis could be generalized to other datasets. To assess whether affiliation to a certain phylum or the use of a specific method for the determination of DDH values (Ezaki et al., 1989; De Ley et al., 1970; Tourova and Antonov, 1988) was specifically affecting the error probability per 16S rRNA gene sequence similarity threshold in a statistically significant way and thus would need to be considered in the final model, a multiple regression analysis was conducted and the significance of the overall effects of both the phyla and the specific wet-lab method used for the determination of DDH values assessed. The model was thus as above, but with the affiliations to both a phylum and DDH method as additional predictor variables. The overall prediction accuracy of the models was measured as the number of pairs with an actual DDH ⩾70 % that would have been omitted for the determination of DDH values given the respective 16S rRNA gene sequence similarity thresholds, thus falsely assuming distinct species. All statistical computations were done in R (R Development Core Team, 2011).

Results

Cross-validation yielded a uniformly low minimum standard error (0.1) for all models, indicating the general suitability of the underlying dataset for such modeling, but no distinction between the models. The Akaike information criterion, however, preferred 16S rRNA gene similarities calculated with PAUP* (184) over Smith-Waterman (203) and BLAST (201) similarities, whereas the Kendall correlations (0.32-0.33) showed little variation. There was no performance difference between the distinct distance formulas implemented in PAUP*; even the resulting similarity thresholds showed little deviation (Table 1). The multiple regression did not confirm an overall significant effect of the methods used for the determination of the DDH values (p value: 0.12), but a significant

one for the phyla (p value: 0.009). The DDH determination method was thus neglected in the final model and the error probabilities were determined for general as well as phylum-specific 16S rRNA gene sequence similarity thresholds, shown in Table 1 for critical error-probability values; a short usage example is given in the corresponding caption. One of the optimal models is depicted in Fig. 1 together with the underlying data points. File S2 (Supplementary material) contains details on all inferred models.

All models revealed 100 % prediction accuracy under all permitted error probabilities ≤5 %, except for the phylum-specific model for the Proteobacteria that falsely assumed distinct species for a low number of data points (see Table 1). But that was the case only if 16S rRNA gene sequence similarity thresholds were applied in conjunction with an allowed error probability ⩾0.5 %.

Discussion

Regression analysis of the enlarged dataset confirms the previously posited thresholds (Konstantinidis and Tiedje, 2007; Stackebrandt and Ebers, 2006) for the three largest phyla (Actinobacteria, Firmicutes and Proteobacteria) considering a maximum accepted probability of failure of 1 %. The results differ, however, for other phyla (as this aspect of the taxonomic affiliation was shown to have a significant effect) and when other maximally accepted error probabilities are chosen. This effect of the phyla is in accordance with observations made earlier (Keswani and Whitman, 2001), and in contrast to earlier studies (Konstantinidis and Tiedje, 2005; Stackebrandt and Ebers, 2006), we here provide phylum-specific thresholds. Such distinct thresholds might be due to accelerated or retarded evolution of rRNA genes in some phyla. A preliminary assessment did not reveal a correlation between the phylum-specific thresholds and the average GC content usually reported for the phyla present in our dataset. For instance, the low-GC Firmicutes and the high-GC Actinobacteria (Woese, 1987) showed the same trend regarding their thresholds (Table 1). The investigated dataset may well be regarded as prone to a taxonomically biased composition, as only 61 data points were selected from the other three phyla. On the other hand, it is obvious that the relative abundances of the contained phyla do represent their proportion in the current application of DDHs in taxonomy, confirming that the values depicted in Table 1 are indeed of practical relevance. For phyla not covered here we suggest to use the thresholds calculated from the entire dataset (leftmost column) in conjunction with a conservative maximally accepted probability of failure, which could without a significant loss in confidence be set distinctly below the 1:10,000 failure rate (0.01 %) that is currently in use. Indeed, Table 1 shows that the 97 % sequence similarity threshold (Stackebrandt and Goebel, 1994) was chosen so conservative that only 1 in 10,000 pairs of strains with a 16S rRNA gene sequence similarity of 97 % and above will result in a DDH value higher than 70 %. Given that microbiologists currently know only about 10,000 validly named species of bacteria and archaea, this means that we so far are paying a very high price (for DDH determinations) to achieve a level of confidence that is exorbitantly high.

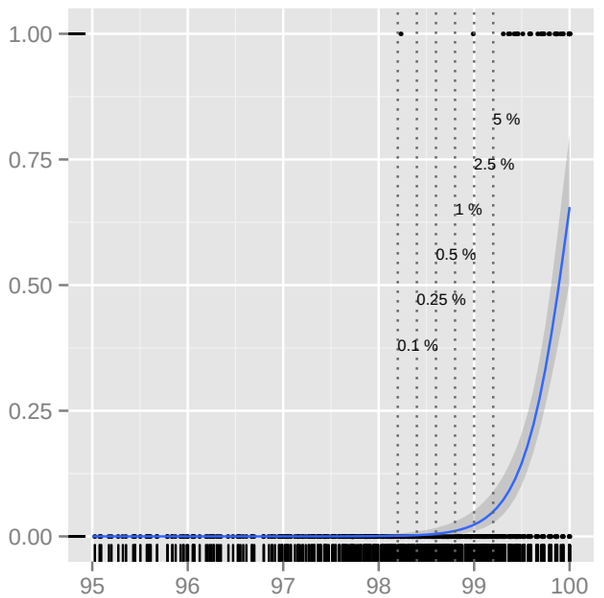

Fig. 1. Results of DDH experiments as inferred from 16S rRNA gene similarities. Values on the x-axis are similarities calculated from uncorrected (p) 16S rRNA gene distances inferred with PAUP*, and those on the y-axis wet-lab DDH results coded as 1 if ⩾70 % and 0 if <70 %. The scale on the y-axis is thus the probability for DDH ⩾70 %. The curve depicts a GLM for inferring such probabilities from 16S rRNA gene similarities, without considering the affiliation to a phylum as additional independent variable (but see Table 1). The dotted lines indicate similarity thresholds for some critical maximum allowed error probabilities. For instance, for 16S rRNA gene similarities of at most 99.0 %, the probability of a DDH experiment yielding ⩾70 % is at most 2.5 % (which is reached, of course, at a similarity of exactly 99.0 %). Thus, if wet-lab DDH experiments were omitted for all pairs of strains whose 16S rRNA gene similarity was 99 % or below, one would falsely regard them as distinct species in up to 2.5 % of all cases. This maximum error probability would drop to 1 error in 100 cases if a similarity threshold of 98.8 % was used (see also Table 1)

Table 1. General and phylum-specific 16S rRNA thresholds for critical maximum error probabilities ranging from 0.01 to 10 %. The 16S rRNA gene thresholds are equally applicable to uncorrected (p), JC, K2P, and HKY85 distances, with or without a gamma distribution (= 0.5), because they did not differ after rounding. Within parentheses, the number of samples with an actual DDH ⩾70 % that would have been overlooked given the respective 16S rRNA threshold is given if greater than zero. A usage example is as follows. Assume a taxonomic journal specifies an error probability of at most 1 in 100 cases. Then, for Actinobacteria the sequence similarity could be at most 99.0 % for allowing authors to omit a DDH experiment; for Firmicutes, 98.8 %; and for Proteobacteria, 98.7 %. If at most one error per 1,000 cases deemed appropriate according to the policy of the journal, these values were 98.2, 98.2, and 98.0 %, respectively, yielding a higher number of mandatory DDH experiments, but still much less than the currently frequently applied general threshold of 97 %. Act., Actinobacteria; Firm., Firmicutes; Prot., Proteobacteria; Dein., Deinococcus-Thermus; Eury., Euryarchaeota; Bact., Bacteroidetes

Max. probability of error (%)	General thresholds	Phylum-specific thresholds					
		Act.	Firm.	Prot.	Dein.	Eury.	Bact.
0.01	97.6	97.5	97.5	97.2	97.2	96.8	96.8
0.025	97.8	97.8	97.8	97.5	97.5	97.0	97.0
0.05	98.0	98.0	98.0	97.8	97.7	97.2	97.2
0.10	98.2	98.2	98.2	98.0	97.8	97.5	97.5
0.25	98.4	98.5	98.5	98.2	98.0	97.8	97.8
0.50	98.6	98.8	98.7	98.3 (1)	98.2	98.0	98.0
1.00	98.8	99.0	98.8	98.7 (2)	98.5	98.2	98.2
2.50	99.0	99.2	99.2	98.8 (3)	98.8	98.3	98.5
5.00	99.2	99.3	99.3	99.0 (5)	99.0	98.7	98.7
10.0	99.4	99.5 (1)	99.5 (2)	99.2 (7)	99.2	98.8	98.8

Only few cases of significant intragenomic rRNA gene polymorphisms are known that could bias taxonomic decisions. For instance, some Haloarcula species express distinct rRNA genes under distinct environmental conditions (Cui et al., 2009). The copies from the genomes of Haloarcula marismortui (NCBI accession numbers AY596297 and AY596298) and Haloarcula vallismortis (AOLQ01000000) yielding intergenomic and intragenomic similarity scores of about 94 % and about 99 %, respectively. But the smaller of these values would lead to taxonomic bias under both the 97 % and the here proposed thresholds, whereas the larger similarity value would suggest a DDH experiment in either case.

It is well known that DDH and 16S rRNA gene sequence similarities are not linear and the DDH values obtained for a given 16S rRNA gene sequence similarity value may differ significantly (Keswani and Whitman, 2001). DDH values are influenced by several physicochemical parameters, many of which are not recorded in the original citation. Repetition of DDH by another research group may therefore yield different values, which can potentially lead to taxonomic rearrangement in those cases where the new determination is above or below the threshold value of 70 %. An example is the reclassification of Lactobacillus arizonensis (Swezey et al., 2000) as a later heterotypic synonym of Lactobacillus plantarum (Kostinek et al., 2005) because the originally determined DDH value of 42 % was redetermined to be as high as 73 %. Nevertheless, a statistically significant effect of the main kind of DDH method used was not found in the present study, indicating that the currently predominating DDH approaches are, on average, comparable and reliable.

It is also good news that only negligible differences were observed between the 16S rRNA gene sequence similaritycalculation approaches investigated (except for BLAST and the direct use of Smith-Waterman similarities), because it is seldom exactly stated in the taxonomic literature which one was chosen (see the dataset compiled by (Stackebrandt and Ebers, 2006). This result is not unexpected, however, because correction formulas for pairwise distances have a strong effect only for large distances (Felsenstein, 2004, pp. 156-158). Similarly, sequence alignment also appears to play a negligible role for highly similar 16S rRNA gene sequences, as confirmed by a high Kendall correlation (0.8) between the similarities collected by (Stackebrandt and Ebers, 2006) and those recalculated here. The Smith-Waterman algorithm was chosen because it solves pairwise alignments exactly, but other programs might as well be used without risk. The poor performance of the Smith-Waterman similarities themselves is most likely caused by their treatment of gaps as real (not missing) information, which fails if 16S rRNA gene sequences have large insertions (e.g., AY639871 vs. U46145). BLAST similarities also cannot be recommended (even though here all non-overlapping HSPs were considered, in contrast to the naïve approach), most likely because they are only based on sequence parts within HSPs, in contrast to a full pairwise alignment. This conclusion is in agreement with recommendations published in taxonomic journals (Tindall et al., 2010).

With this exception, the software routinely applied by microbial taxonomist appears well suited to safely estimate 16S rRNA gene sequence similarities, which could now be used in conjunction with Table 1 for a reasonable decision on whether or not the DDH value between two strains should be determined for the discrimination of species in a taxonomic analysis. As an example of the proportion of tedious experiments that could be omitted, we estimated that the following percentage of DDHs covered by our dataset could have been avoided if a maximum probability of failure of 1 % was accepted: Actinobacteria, 51 %; Bacteroidetes, 35 %; Deinococcus-Thermus, 28 %; Euryarchaeota, 50 %; Firmicutes, 43 %; Proteobacteria, 38 %. Table 1 could serve as a general guideline for researchers even if favored other maximum possible error probabilities, but based on our experience at DSMZ, a value of 1 % appears to be reasonable. We hope that this study helps taxonomists to avoid tedious wet-lab work without sacrificing accuracy until finally a situation will be reached in which genome sequencing is routinely applied in species descriptions anyway (Lagier et al., 2013), thus generating the data that can be used by bioinformatics approaches for species delimitation (Auch et al., 2010b,a; Goris et al., 2007; Klenk and Göker, 2010; Meier-Kolthoff et al., 2013; Richter and Rosselló-Móra, 2009).

ACKNOWLEDGMENTS. We are grateful to Prof. Erko Stackebrandt for providing data and for helpful comments.

CONFLICT OF INTEREST. The authors declare that they have no conflict of interest.

References

H. Akaike. A new look at the statistical model identification. *IEEE T Automat Contr*, 19(6):716–723, 1974. ISSN 0018-9286. doi: 10.1109/TAC.1974.1100705.

S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, 1990.

A. F. Auch, H.-P. Klenk, and M. Göker. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci*, 2(1):142–148, jan 2010a. ISSN 1944-3277. doi: 10.4056/sigs.541628. URL .

A. F. Auch, M. von Jan, H.-P. Klenk, and M. Göker. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci*, 2(1):117–134, jan 2010b. ISSN 1944-3277. doi: 10.4056/sigs.531120.

M. J. Crawley. *The R book*. Wiley Publishing, Chichester, 1st edition, 2007. ISBN 0470510242, 9780470510247.

H.-L. Cui, P.-J. Zhou, A. Oren, and S.-J. Liu. Intraspecific polymorphism of 16S rRNA genes in two halophilic archaeal genera, Haloarcula and Halomicrobium. *Extremophiles*, 13 (1):31–37, 2009. ISSN 1431-0651. doi: 10.1007/s00792-008-0194-2.

J. De Ley, H. Cattoir, and A. Reynaerts. The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem*, 12(1):133–142, jan 1970. ISSN 1432-1033. doi: 10.1111/j.1432-1033.1970.tb00830.x. URL .

J. P. Euzéby. List of bacterial names with standing in nomenclature: a folder available on the internet. *Int J Syst Bacteriol*, 47(2):590–592, 1997. doi: 10.1099/00207713-47-2-590.

T. Ezaki, Y. Hashimoto, and E. Yabuuchi. Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in microdilution 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(3):224–229, 1989. doi: 10.1099/00207713-39-3-224.

J. Felsenstein. *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts, 1st edition, jan 2004.

J. Goris, K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme, and J. M. Tiedje. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*, 57(1):81–91, 2007. doi: 10.1099/ijs.0.64483-0.

F. E. Harrell, K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati. Regression modelling strategies for improved prognostic prediction. *Stat Med*, 3(2):143–152, 1984. ISSN 1097-0258. doi: 10.1002/sim.4780030207.

M. Hasegawa, H. Kishino, and T. Yano. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*, 22:160–174, 1985.

S. R. Henz, D. H. Huson, A. F. Auch, K. Nieselt-Struwe, and S. C. Schuster. Whole-genome prokaryotic phylogeny. *Bioinformatics*, 21(10):2329–2335, 2005. ISSN 13674803. doi: 10.1093/bioinformatics/bth324.

T. Jukes and C. Cantor. *Evolution of protein molecules*. Academic Press, New York, 1969.

J. Keswani and W. B. Whitman. Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes. *Int J Syst Evol Microbiol*, 51(Pt 2):667–78, mar 2001. ISSN 1466-5026. URL .

M. Kimura. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16:111–120, 1980.

H.-P. Klenk and M. Göker. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol*, 33(4):175–182, jul 2010. ISSN 1618-0984. doi: 10.1016/j.syapm.2010.03.003.

K. T. Konstantinidis and J. M. Tiedje. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA*, 102(7):2567–2572, 2005. doi: 10.1073/pnas.0409727102.

K. T. Konstantinidis and J. M. Tiedje. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol*, 10(5):504–509, oct 2007. ISSN 1369-5274. doi: 10.1016/j.mib.2007.08.006.

M. Kostinek, R. Pukall, A. P. Rooney, U. Schillinger, C. Hertel, W. H. Holzapfel, and C. M. A. P. Franz. ¡i¿Lactobacillus arizonensis¡/i¿ is a later heterotypic synonym of ¡i¿Lactobacillus plantarum¡/i¿. *Int J Syst Evol Micr*, 55(6):2485–2489, 2005. doi: 10.1099/ijs.0.63880-0.

J.-C. Lagier, K. E. Karkouri, R. Rivet, C. Couderc, D. Raoult, and P.-E. Fournier. Non contiguous-finished genome sequence and description of Senegalemassilia anaerobia gen. nov., sp. nov. *Stand Genomic Sci*, 7(3), 2013. ISSN 1944-3277. doi: 10.4056/sigs.3246665.

J. P. Meier-Kolthoff, A. F. Auch, H.-P. Klenk, and M. Göker. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14(1):60, 2013. ISSN 1471-2105. doi: 10.1186/1471-2105-14-60. URL .

H. Motulsky and A. Christopoulos. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press, Oxford, UK, 1st edition, 2004. ISBN 9780195171792.

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *J R Stat Soc*, 135(3):370–384, 1972.

P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*, 49 (12):1373–1379, dec 1996.

R Development Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL .

P. Rice, I. Longden, and A. Bleasby. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16: 276–277, 2000.

M. Richter and R. Rosselló-Móra. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA*, 106(45):19126–19131, 2009. doi: 10.1073/pnas.0906412106.

E. Stackebrandt and J. Ebers. Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today*, 33:152–155, 2006.

E. Stackebrandt and B. M. Goebel. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol*, 44(4):846–849, 1994. doi: 10.1099/00207713-44-4-846. URL .

E. W. Steyerberg, M. J. C. Eijkemans, F. E. Harrell, and J. D. F. Habbema. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*, 19(8):1059–1079, 2000. ISSN 1097-0258. doi: 10.1002/(SICI)1097-0258(20000430)19:8¡1059::AID-SIM412¿3.0.CO;2-0.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc*, 36(2):111–147, 1974. ISSN 00359246. doi: 10.2307/2984809.

J. L. Swezey, L. K. Nakamura, T. P. Abbott, and R. E. Peterson. ¡i¿Lactobacillus arizonensis¡/i¿ sp. nov., isolated from jojoba meal. *Int J Syst Evol Microbiol*, 50(5):1803–1809, 2000. doi: 10.1099/00207713-50-5-1803.

D. L. Swofford. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. *Sinauer Associates, Sunderland, Massachusetts.*, 2003.

B. J. Tindall, P. Kämpfer, J. P. Euzéby, and A. Oren. Valid publication of names of prokaryotes according to the rules of nomenclature: past history and current practice. *Int J Syst Evol Microbiol*, 56(11):2715–2720, nov 2006. ISSN 1466-5026. doi: 10.1099/ijs.0.64780-0. URL .

B. J. Tindall, R. Rosselló-Móra, H.-J. Busse, W. Ludwig, and P. Kämpfer. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol*, 60(1):249–266, 2010. doi: 10.1099/ijs.0.016949-0. URL .

T. P. Tourova and A. S. Antonov. Identification of microorganisms by rapid DNA-DNA hybridization. In R. R. Colwell and R. Grigorova, editors, *Method Microbiol*, volume 19 of *Methods in Microbiology*, pages 333–355. Academic Press, 1988. doi: 10.1016/S0580-9517(08)70414-8.

L. G. Wayne, D. J. Brenner, R. R. Colwell, P. a. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Truper. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol*, 37(4):463–464, oct 1987. ISSN 0020-7713. doi: 10.1099/00207713-37-4-463.

C. R. Woese. Bacterial evolution. *Microbiol Rev*, 51(2):221–271, 1987.

Z. Yang. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol*, 10(6):1396–1401, nov 1993.