Application of sliding-window discretization and minimization of stochastic complexity for the analysis of fAFLP genotyping fingerprint patterns of *Vibrionaceae*

Peter Dawyndt,¹ Fabiano L. Thompson,¹ Brian Austin,² Jean Swings,¹ Timo Koski³ and Mats Gyllenberg^{4†}

Correspondence

Peter Dawyndt

Peter.Dawyndt@ugent.be

¹Laboratorium voor Microbiologie, Universiteit Gent, B-9000 Gent, Belgium

²School of Life Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

³Department of Mathematics, University of Linköping, S-58183 Linköping, Sweden

⁴Department of Mathematics, University of Turku, FIN-20014 Turku, Finland

Minimization of stochastic complexity (SC) was used as a method for classification of genotypic fingerprints. The method was applied to fluorescent amplified fragment length polymorphism (fAFLP) fingerprint patterns of 507 *Vibrionaceae* representatives. As the current BinClass implementation of the optimization algorithm for classification only works on binary vectors, the original fingerprints were discretized in a preliminary step using the sliding-window band-matching method, in order to maximally preserve the information content of the original band patterns. The novel classification generated using the BinClass software package was subjected to an in-depth comparison with a hierarchical classification of the same dataset, in order to acknowledge the applicability of the new classification method as a more objective algorithm for the classification of genotyping fingerprint patterns. Recent DNA–DNA hybridization and 16S rRNA gene sequence experiments proved that the classification based on SC-minimization forms separate clusters that contain the fAFLP patterns for all representatives of the species *Enterovibrio norvegicus*, *Vibrio fortis*, *Vibrio diazotrophicus* or *Vibrio campbellii*, while previous hierarchical cluster analysis had suggested more heterogeneity within the fAFLP patterns by splitting the representatives of the above-mentioned species into multiple distant clusters. As a result, the new classification methodology has highlighted some previously unseen relationships within the biodiversity of the family *Vibrionaceae*.

INTRODUCTION

Bacterial taxonomy has a long tradition of using hierarchical cluster algorithms to establish taxonomies based on phenotypic and genotypic characteristics. Notwithstanding this, there has been concern about the subjective nature of the process (Priest & Austin, 1993). It is appreciated that when using a hierarchical method of cluster analysis early

decisions in the construction process may preclude certain meaningful groupings at later stages (Sneath & Sokal, 1973). Moreover, for a given dataset there might be many meaningful groupings that each reflect different aspects of the underlying relationships. Therefore, a single classification may give a distorted view of the multifaceted set of patterns. Consequently, if there are several meaningful groupings, a variety of cluster analysis techniques will be needed to reveal them all (Anderberg, 1973).

Inspired by these observations, the family of classification methods that optimize a given expression in information theory (such as entropy or stochastic complexity) was proven to be a complementary alternative to the hierarchical classifications (Gyllenberg *et al.*, 1997b, 1998, 1999, 2002) used in bacterial taxonomy. However, the application of this kind of method for the classification of genotypic fingerprint patterns is hindered by the prerequisite that they only work on a vectorized data representation. Austin

Published online ahead of print on 19 July 2004 as DOI 10.1099/ijs.0.63136-0.

[†]Present address: Department of Mathematics and Statistics, Rdf Nevanlinna Institute, University of Helsinki, FIN-00014, Finland.

Abbreviations: fAFLP, fluorescent amplified fragment length polymorphism; HMO, hypothetical median organism; SC, stochastic complexity.

An example of a stochastic-complexity-minimizing dendrogram for the fAFLP dataset of *Vibrionaceae* and the source code and a user manual for the BinClass software package are available as supplementary material in IJSEM Online.

et al. (2004) demonstrated the impact that alternative discretization methods may have on the final classification result, and showed that sliding-window discretization results in a better conservation of the information content of the original fingerprint patterns, in comparison with other data discretization methods.

Thompson *et al.* (2001) analysed a set of 507 fluorescent amplified fragment length polymorphism (fAFLP) fingerprint patterns of *Vibrionaceae* strains based on a hierarchical classification using Ward's hierarchical cluster algorithm (Ward, 1963). This algorithm does not make a direct classification of the fingerprint patterns, but works on an intermediate similarity or dissimilarity matrix. In their report, Thompson *et al.* (2001) calculated this matrix using the Dice similarity coefficient s_D (Dice, 1945). From the hierarchical clustering a plain classification was derived by selecting a rather arbitrary α -cut based on the intuition of the authors and the distribution of type and reference strains included in the dataset.

The current study has sought to reclassify the same set of fingerprint patterns based on more objective foundations and without taking into account prior knowledge about the bacterial strains. The classification method based on minimizing the stochastic complexity (SC) of a binary vector representation of the data (Gyllenberg *et al.*, 1997a) was chosen as a representative method from the family of optimization algorithms for classification. This method has been implemented in the BinClass software package (Gyllenberg *et al.*, 2001), which not only enables the calculation of an optimal classification in the sense of SC, but also allows the construction of an SC-driven hierarchy built on top of the optimal classification (Gyllenberg *et al.*, 1998). An example of an SC-minimizing dendrogram for the fAFLP dataset of *Vibrionaceae* is available as supplementary material in IJSEM Online (Fig. S1).

The two classifications have been compared in depth, which has shown a global correlation between the major parts of the two groupings. However, differences between the classifications have stimulated the discovery of new relationships within the taxonomy of the *Vibrionaceae* strains, which have been confirmed by recent DNA–DNA hybridization and 16S rRNA gene sequence experiments (Thompson *et al.*, 2003a, b, c, d). These results prove the value of the new methodology as an alternative classification strategy in its own right, and the need to inspect a given dataset from different angles using different mathematical models in order to get a complete picture of all the relationships present.

METHODS

Bacterial strains. In this study we have analysed a total of 507 fAFLP (Janssen *et al.*, 1996) fingerprint patterns from isolates of the family *Vibrionaceae*. The dataset was identical to that studied by Thompson *et al.* (2001), where the number of bands of the fAFLP profiles ranged from 46 to 164, with a global mean of 107 bands and a standard deviation of 23. More detailed band statistics for each class resulting from the classification generated in the framework of

this study are included in Table 1. In their original report, Thompson *et al.* (2001) classified the fAFLP fingerprint patterns using Ward's hierarchical clustering algorithm (Ward, 1963). For pairwise fragment comparison, the Dice similarity coefficient s_D (Dice, 1945) was applied with a band position tolerance of 0.5% to compensate for misalignment of homologous bands due to technical imperfections. These calculations were performed using the BioNumerics software package (Applied Maths). After cluster delineation based on the results from previous studies and prior knowledge about the distribution of the type strains within the dataset, this classification strategy resulted in a set of 69 clusters (labelled A1 to A69) and four singleton fingerprint patterns (labelled U1 to U4).

Discretization of fAFLP fingerprint patterns. A classification strategy based on the optimization of an expression in information theory such as SC, which is a quantitative criterion for evaluation of the global goodness of a given classification with respect to the given dataset, requires that the fAFLP fingerprint patterns of the *Vibrio* isolates are transformed into a binary vector representation. Austin *et al.* (2004) have shown that the choice of discretization method used may have a major impact on the final classification. They proved that, among other algorithms, sliding-window discretization resulted in the binary vector representation with the highest conservation of the original information content of the fAFLP fingerprint patterns included in this study.

The sliding-window procedure is an extension of the well-known equal-width discretization method, controlled by two parameters: (i) the position tolerance parameter ε , which defines a window through which bands that are measured at different physical locations are logically considered as similar regardless of their position within the same fingerprint profile or within different profiles, thus compensating for misalignment of homologous bands due to technical imperfections; and (ii) the resolution of the method δ , which controls the step size by which the window shifts over the continuous interval that must be sampled in a discrete manner. If the resolution of the method is smaller than the position tolerance, the different sampling windows used by the sliding-window method will overlap, in contrast to the equal-width method, where sampling windows never overlap with each other. This implies a fuzzyfication of the window boundaries in the resulting vector representation.

For the classification within the current study, the sliding-window discretization method was applied with ε set to 0.007 and δ set to 0.001, in order to maximize the Pearson product-moment correlation r with the s_D similarity matrix from the study of Thompson *et al.* (2001). After all, to enable an objective comparison of the different classification strategies applied to the same *Vibrio*/fAFLP dataset, it is important to mimic the same band position tolerance behaviour and reduce the effect of fragment comparison of the fingerprint patterns within the classification process. As a result, the fAFLP fingerprint patterns were transformed by the sliding-window discretization method into binary vectors of length 994.

Classification of binary vectors. The binary vectors resulting from the sliding-window discretization procedure were classified using minimization of SC (Gyllenberg *et al.*, 1997a), as it was implemented in the BinClass software package (Gyllenberg *et al.*, 2001). This algorithm is an example of an unsupervised non-hierarchical classification method, insofar as it does not make use of prior knowledge or assumptions on the dataset other than the binary vector representation of its characteristics and its final result is a plain partitioning of the dataset into non-overlapping classes. The default parameter settings were chosen, except for the $-F$ parameter (safety value), which was set to 50, and the $-S$ parameter, which was set to 20. This has resulted in the classification with 64 classes (labelled BC1 to BC64) that is summarized in Table 1. The optimal SC found for the dataset was 739.92. The BinClass software package

Table 1. BinClass classification based on data discretized by the sliding-window discretization method with the position tolerance parameter ε set to 0.007 and the resolution of the method δ set to 0.001 (so that vector length $d=994$)

BinClass run performed with command line settings (–F50 –S20), resulting in a classification with 64 classes and an SC of 739.92280. 1, Class identifier. 2, Size of class (number of strains n). 3, Mean number of bands (standard deviation) over all profiles in the class. 4, Minimal and maximal number of bands of all profiles in the class. 5, fAFLP cluster in classification by Thompson *et al.* (2001). 6, Name of fAFLP cluster as given by Thompson *et al.* (2001). 7, Frequency of original fAFLP cluster within class. 8, Mean Shannon code length of the class. 9, Class distortion. 10, Nearest class. 11, Hamming distance to nearest class. 12, Farthest class. 13, Hamming distance to farthest class. 14, Hamming distance between type strain and hypothetical median organism. 15, Shannon code length of type strain.

1	2	3	4	5	6*	7	8	9	10	11	12	13	14	15
BC1	24	101 (8)	90–114	A67	<i>V. haliotocoli</i> †	24/24	475	141	BC9	329	BC63	468	206	658
BC2	24	116 (20)	75–147	A49	<i>V. diabolicus</i>	2/6	541	171	BC36	312	BC63	467		
				A62	<i>V. alginolyticus</i> †	22/22							151	494
BC3	22	76 (9)	58–100	A68	<i>E. norvegicus</i>	18/18	418	131	BC44	392	BC63	487		
				A69	<i>E. norvegicus</i> †	4/4							122	412
BC4	21	126 (14)	96–146	A5	<i>V. neptunius</i> †	21/21	240	78	BC61	257	BC63	488	140	494
BC5	18	93 (13)	60–114	A38	<i>V. anguillarum</i> †	8/8	523	168	BC42	371	BC63	500	291	759
				A39	<i>V. ordalii</i> †	10/10							74	366
BC6	16	116 (13)	79–135	A9	<i>V. fortis</i> †	1/8	482	151	BC11	328	BC63	493	351	1014
				A61	<i>V. cyclitrophicus</i>	15/15								
BC7	14	66 (11)	46–91	A11	<i>V. logei</i> †	1/7	691	243	BC25	356	BC63	491	348	924
				A21	<i>V. gazogenes</i> †	1/2							325	897
				A27	<i>V. fluvialis</i> †	4/4							239	697
				A28	<i>V. fluvialis</i>	2/2								
				A29	<i>V. furnisii</i> †	5/5							212	581
				U4	<i>Vibrio</i> sp. R-3681	1/1								
BC8	13	134 (19)	106–164	A9	<i>V. fortis</i>	5/8	417	133	BC57	340	BC60	478		
				A60	<i>V. fortis</i>	8/8								
BC9	13	79 (9)	66–93	A10	<i>V. pelagius</i>	1/2	610	206	BC1	329	BC63	483		
				A12	<i>V. cincinnatiensis</i>	1/6								
				A64	<i>'V. neonatus'</i> †	9/9							167	547
				A65	<i>Vibrio</i> sp.	2/2								
BC10	13	120 (12)	105–141	A30	<i>V. harveyi</i>	12/14	570	187	BC56	254	BC63	479		
				A59	<i>V. tubiashii</i>	1/18								
BC11	12	126 (10)	109–143	A50	<i>V. splendidus</i> †	11/16	630	218	BC47	322	BC3	463	190	560
				A59	<i>V. tubiashii</i>	1/18								
BC12	12	115 (14)	94–147	A36	<i>V. harveyi</i> †	12/12	520	179	BC6	363	BC63	503	124	435
				A36	<i>V. trachuri</i> †								208	554
BC13	12	101 (8)	90–112	A66	<i>'V. ezurae'</i> †	12/12	388	129	BC2	381	BC63	478	83	321
BC14	12	111 (9)	98–126	A63	<i>V. ichthyenteri</i> †	12/12	376	130	BC53	398	BC48	489	248	572
BC15	11	137 (9)	124–149	A4	<i>V. coralliilyticus</i>	10/10	271	77	BC16	264	BC58	483		
				A59	<i>V. tubiashii</i>	1/18								
BC16	10	134 (9)	116–148	A1	<i>V. coralliilyticus</i>	4/10	412	141	BC15	264	BC60	472		
				A2	<i>V. coralliilyticus</i>	6/6								
BC17	10	121 (15)	94–149	A52	<i>V. chagasii</i>	4/7	588	201	BC36	285	BC63	469		
				A53	<i>V. chagasii</i> †	5/5							130	470
				A59	<i>V. tubiashii</i>	1/18								
BC18	10	109 (15)	72–128	A8	<i>V. brasiliensis</i> †	7/7	546	175	BC2	368	BC60	473	26	314
				A11	<i>V. logei</i>	1/7								
				A49	<i>V. diabolicus</i> †	2/6							374	836
BC19	9	119 (11)	105–144	A11	<i>V. logei</i>	2/7	6	178	BC11	346	BC60	472		
				A55	<i>V. lentus</i>	7/7								
BC20	9	116 (17)	95–145	A46	<i>V. kanaloae</i> †	5/5	621	213	BC36	344	BC63	458	102	440
				A56	<i>P. damsela</i> subsp. <i>damsela</i> †	2/8							259	661
				A56	<i>P. histaminum</i> †								292	741
				A59	<i>V. tubiashii</i>	2/18								

Table 1. cont.

1	2	3	4	5	6*	7	8	9	10	11	12	13	14	15
BC21	9	115 (23)	75–141	A06	<i>V. mediterraneit</i> / <i>V. shiloniit</i> †	9/10	468	162	BC55	309	BC58	472	113/208	423/578
BC22	9	121 (10)	106–143	A45	<i>V. tasmaniensis</i> †	6/6	461	154	BC47	330	BC25	463	65	308
				A56	<i>P. angustum</i> / <i>P. damsela</i> subsp. <i>damsela</i>	1/8								
				A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> / <i>V. tapetis</i>	2/17								
BC23	9	106 (5)	96–115	A32	<i>V. harveyi</i>	4/9	408	137	BC51	271	BC60	469		
				A33	<i>V. rotiferianus</i> †	5/5							77	324
BC24	8	104 (18)	85–139	A1	<i>V. coralliilyticus</i>	4/10	617	213	BC15	290	BC63	497		
				A9	<i>V. fortis</i>	2/8								
				A10	<i>V. pelagius</i> †	1/2							336	903
				A12	<i>V. cinnamatisensis</i> †	1/6							340	838
BC25	8	64 (13)	50–97	A12	<i>V. cinnamatisensis</i>	1/6	599	217	BC10	329	BC63	502		
				A19	<i>P. phosphoreum</i> †	1/2							341	916
				A25	<i>V. proteolyticus</i> †	3/3							248	568
				A31	<i>V. harveyi</i>	3/3								
BC26	8	85 (9)	73–98	A11	<i>V. logei</i>	3/7	498	174	BC20	363	BC63	474		
				A24	<i>V. vulnificus</i> †	5/5							179	481
BC27	8	103 (9)	92–119	A12	<i>V. cinnamatisensis</i>	1/6	658	231	BC32	294	BC3	460		
				A14	<i>V. campbellii</i>	1/4								
				A37	<i>V. campbellii</i>	4/10								
				A57	<i>V. diazotrophicus</i>	1/4								
				A59	<i>V. tubiashii</i> †	1/18							275	717
BC28	7	109 (13)	96–128	A42	<i>V. parahaemolyticus</i> †	6/6	535	182	BC2	362	BC63	488	103	393
				A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> / <i>V. tapetis</i>	1/17								
BC29	7	112 (17)	84–129	A13	<i>V. nigripulchritudo</i> / <i>V. orientalis</i>	1/7	419	128	BC40	362	BC63	475		
				A19	<i>P. phosphoreum</i>	1/2								
				A43	<i>V. pectenocida</i> †	5/5							54	288
BC30	7	96 (12)	85–120	A12	<i>V. cinnamatisensis</i>	1/6	554	193	BC27	370	BC63	481		
				A34	<i>V. diazotrophicus</i>	5/5								
				A57	<i>V. diazotrophicus</i> †	1/4							247	632
BC31	7	115 (7)	103–122	A3	<i>V. coralliilyticus</i>	7/7	158	56	BC16	288	BC60	500		
BC32	7	95 (12)	79–117	A18	<i>V. navarrensis</i>	1/2	514	167	BC27	294	BC63	477		
				A37	<i>V. campbellii</i>	6/10								
BC33	6	115 (16)	84–131	A14	<i>V. campbellii</i> †	3/4	598	224	BC10	303	BC3	471	191	484
				A32	<i>V. harveyi</i>	1/9								
				A59	<i>V. tubiashii</i>	2/18								
BC34	6	115 (15)	101–146	A21	<i>V. gazogenes</i> / <i>V. salmonicida</i> †	1/2	705	253	BC61	380	BC63	479	364	896
				A54	<i>V. mytili</i> / <i>P. leiognathi</i> †	3/5							156	523
				A56	<i>P. angustum</i> †/ <i>P. damsela</i> subsp. <i>damsela</i>	1/8							302	778
				A59	<i>V. tubiashii</i>	1/18								
BC35	6	123 (17)	103–146	A16	<i>V. hispanicus</i> †	3/3	559	197	BC17	332	BC3	474	153	461
				A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> / <i>V. tapetis</i>	1/17								
				A59	<i>V. tubiashii</i>	2/18								
BC36	6	124 (12)	107–141	A49	<i>V. diabolicus</i>	1/6	663	237	BC17	285	BC60	458		
				A52	<i>V. chagasii</i>	3/7								
				A59	<i>V. tubiashii</i>	2/18								
BC37	6	91 (15)	66–111	A7	<i>V. wodanis</i> †	6/6	374	129	BC51	365	BC3	485	127	384
BC38	6	93 (15)	67–109	A13	<i>V. nigripulchritudo</i> / <i>V. orientalis</i> †	1/7	496	170	BC20	362	BC63	497	310	758
				A48	<i>V. natriegens</i> †	5/5							142	453
BC39	5	98 (20)	66–118	A15	<i>V. fischeri</i> †/ <i>P. iliopiscarius</i> †	4/4	652	234	BC19	370	BC63	481	129/270	523/715

Table 1. cont.

1	2	3	4	5	6*	7	8	9	10	11	12	13	14	15
				A56	<i>P. angustum</i> / <i>P. damsela</i> subsp. <i>damsela</i>	1/8								
BC40	5	133 (13)	111–147	A51	<i>V. pomeroyi</i> †	4/6	480	162	BC42	345	BC63	495	175	458
				A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> / <i>V. tapetis</i>	1/17								
BC41	5	94 (30)	67–149	A12	<i>V. cincinnatiensis</i>	1/6	520	175	BC62	318	BC48	454		
				A26	<i>V. hepatarius</i> †	3/4							28	325
				A59	<i>V. tubiashii</i>	1/18								
BC42	5	131 (12)	117–147	A51	<i>V. pomeroyi</i>	2/6	626	216	BC40	345	BC3	476		
				A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> / <i>V. tapetis</i>	2/17								
				A59	<i>V. tubiashii</i>	1/18								
BC43	5	76 (12)	61–94	A22	<i>S. costicola</i> †	2/2	573	209	BC41	380	BC63	475	266	568
				A23	<i>V. xuii</i> †	3/3							143	534
BC44	5	81 (17)	53–98	A17	<i>V. scopthalmi</i> †	3/3	611	216	BC32	355	BC63	496	72	430
				A18	<i>V. navarrensis</i> †	1/2							365	791
				U2	<i>V. hollisae</i> †	1/1							388	819
BC45	5	103 (13)	83–120	A47	<i>V. pacinii</i> †	3/3	503	182	BC20	368	BC3	475	109	441
				A54	<i>V. mytili</i> †	2/5							274	560
BC46	5	145 (9)	135–156	A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> / <i>V. tapetis</i> †	5/17	492	168	BC47	327	BC7	490	42	337
BC47	5	119 (13)	100–136	A50	<i>V. splendidus</i>	5/16	440	153	BC11	322	BC13	469		
BC48	4	100 (8)	90–110	A44	<i>V. nereis</i> †	4/4	425	167	BC61	383	BC14	489	219	514
BC49	4	102 (23)	71–130	A20	<i>V. metschnikovii</i> †	3/3	353	130	BC5	376	BC63	504	98	255
				A59	<i>V. tubiashii</i>	1/18								
BC50	4	96 (6)	87–102	A35	<i>V. aestuarianus</i> †	4/4	321	115	BC30	374	BC63	490	166	400
BC51	4	99 (14)	84–120	A32	<i>V. harveyi</i>	4/9	387	149	BC23	271	BC49	477		
BC52	4	104 (13)	82–113	A40	<i>V. cholerae</i> †	4/4	489	181	BC56	383	BC63	490	229	594
BC53	3	127 (6)	121–135	A1	<i>V. coralliilyticus</i>	2/10	414	150	BC15	272	BC48	471		
				A13	<i>V. nigrapulchritudo</i> / <i>V. orientalis</i>	1/7								
BC54	3	126 (6)	117–131	A58	<i>V. penaeicida</i> / <i>V. rumoiensis</i> †/ <i>V. tapetis</i>	3/17	406	147	BC2	355	BC63	475	71	330
BC55	3	68 (3)	64–71	A6	<i>V. mediterranei</i> / <i>V. shilonii</i>	1/10	519	188	BC21	309	BC63	522		
				A49	<i>V. diabolicus</i>	1/6								
				U1	<i>V. aerogenes</i> †	1/1							181	512
BC56	3	111 (11)	96–122	A13	<i>V. nigrapulchritudo</i> / <i>V. orientalis</i>	1/7	444	161	BC10	254	BC48	476		
				A30	<i>V. harveyi</i>	2/14								
BC57	3	128 (22)	98–150	A13	<i>V. nigrapulchritudo</i> / <i>V. orientalis</i>	1/7	508	184	BC8	340	BC59	472		
				A58	<i>V. penaeicida</i> †/ <i>V. rumoiensis</i> / <i>V. tapetis</i>	2/17							210	534
BC58	2	103 (5)	98–107	A41	<i>V. mimicus</i> †	2/2	257	128	BC27	387	BC63	500	99	257
BC59	2	85 (33)	52–118	A56	<i>V. tubiashii</i>	1/8	504	252	BC20	356	BC48	481		
				U3	<i>Vibrio</i> sp. R-1586	1/1								
BC60	2	103 (9)	94–111	A57	<i>V. diazotrophicus</i>	2/4	218	109	BC13	385	BC31	500		
BC61	2	109 (10)	99–118	A56	<i>P. damsela</i> subsp. <i>damsela</i>	2/8	159	79	BC4	257	BC63	495		
BC62	2	75 (11)	64–86	A13	<i>V. nigrapulchritudo</i> †/ <i>V. orientalis</i>	1/7	390	195	BC41	318	BC63	490	148	390
				A26	<i>V. hepatarius</i>	1/4								
BC63	1	127 (0)	127–127	A59	<i>V. tubiashii</i>	1/18	1	0	BC16	423	BC55	522		
BC64	1	100 (0)	100–100	A13	<i>V. nigrapulchritudo</i> / <i>V. orientalis</i>	1/7	1	0	BC38	374	BC59	471		

**E.*, *Enterovibrio*; *P.*, *Photobacterium*; *S.*, *Salinivibrio*; *V.*, *Vibrio*. Bold type indicates revised name since publication of the report by Thompson *et al.* (2001).

†Type strain position.

automatically accommodated for monomorphic bands by discarding vector indexes that had the same binary value before performing the classification, hence taking only into account the bands that were polymorphic within the dataset.

The Hamming distance between two binary vectors is defined as the number of bits that are different, from which the distance between two classes can be defined as the mean pairwise Hamming distance between members of the two classes (Gyllenberg *et al.*, 1999). The centroid of a class is by definition the vector giving the frequencies of 1's for the different attributes. By rounding off each component of the centroid to the nearest binary value (0 or 1) one obtains the hypothetical median organism (HMO; Liston *et al.*, 1963). The distortion of a class, defined as the mean number of bits by which the members of the class differ from the HMO (mean Hamming distance), can be regarded as a measure of the heterogeneity of a class. Shannon code length (Cover & Thomas, 1991) between a class member and the centroid of the class was used as an alternative distance function, where the mean Shannon code length of all class members with respect to the centroid of the class gives an alternative quantifier for describing the heterogeneity of the class. For both the distortion and the mean Shannon code length it holds that classes with lower values for these parameters are more homogeneous than classes with higher values.

From a taxonomic viewpoint the fingerprint profiles of the type strains that were included in the dataset are fairly neatly distributed over the different classes that result from the BinClass classification (see Table 1), although this was not forced by any subjective decision-making within the classification strategy and the type strain information was not regarded as prior knowledge by the classification scheme. This proves that the partitioning of the fAFLP patterns from the current study by minimization of the SC generated classes that generally correspond well with the species delineated within the family *Vibrionaceae*. In this context, both the centroid and the HMO can be regarded as estimations of fictitious representatives for each class and thus, by extension, for the species that are represented by these classes. The Hamming distance to the HMO and the Shannon code length may then be employed as measures for evaluating the typicality of a pattern for the class it belongs to or, stated differently, they measure how typical a strain is of the species to which it is identified by a given classification procedure applied on a chosen set of characters of the strain. Since a type strain has to be designated when a species is first described and named, the nomenclatural type strain is nothing more than the name bearer of the species and is usually the first strain known (Buchanan, 1925). Hence, at the time of type strain selection so little information has yet been found out about the constellation of the species which will be represented by that strain that there is not enough statistical evidence in order to ensure that the type strain is indeed a typical strain (Sneath, 1984). Moreover, some of the type strains might be comparatively old and have lost useful characters due to gene loss caused by the long preservation time of the strains.

The source code and a user manual of the BinClass software package are available as supplementary materials in IJSEM Online. Due to the command line interface and the ANSI C compliance of the source code, the software easily compiles on most operating systems (e.g. Win32, UNIX and Linux). For the convenience of the readership, all BinClass-formatted input files and all the output files generated by the software package in the framework of this study are also included as supplementary materials, together with an executable version of the program that has been compiled to run on all Win32 platforms. One of the output files generated by the BinClass software package contains a complete description of the classification results, in which each of the 507 *Vibrionaceae* strains is accounted for. This information is far too extensive to be reported in detail within this publication.

Comparison of different classifications. In order to rate the value of minimizing SC for the classification of bacterial genotyping fingerprint patterns, the BinClass classification of the *Vibrio*/AFLP dataset was compared with the classification of the same dataset as described by Thompson *et al.* (2001). Fig. 1 shows a graphical representation of the comparison between these two classifications. In this representation, each row represents a class from the classification described previously by Thompson *et al.* (2001), with the assigned class identifier in the first column and the number of strains in the last column. Each column represents a class from the BinClass classification, with the assigned class identifier in the first row and the number of strains in the last row. The values in the row/column intersections represent the number of strains that the two corresponding classes have in common. The following section gives a detailed discussion about the similarities and differences found between the two classifications, together with the taxonomic implications for the *Vibrio* dataset.

RESULTS AND DISCUSSION

The 507 strains examined in this study formed 64 classes (BC1 to BC64), several of which (i.e. BC1, BC4, BC12, BC13, BC14, BC31, BC37, BC48, BC50, BC52 and BC58) corresponded exactly with classes obtained from the clustering of the same dataset using s_D /Ward and an arbitrary cluster cut off value of 45 % (Thompson *et al.*, 2001). In addition, new relationships among former fAFLP clusters have been disclosed, many of which are in agreement with recent DNA–DNA hybridization and 16S rRNA gene sequence experiments.

Class BC1 contained 24 *Vibrio haliotocoli* strains, with low pattern distortion (heterogeneity), and was most closely related to BC9, which constituted a new *Vibrio* species, '*Vibrio neonatus*', phylogenetically related to *V. haliotocoli* (Sawabe *et al.*, 2002, 2004). Class BC2 had 24 strains, including the type strain of *Vibrio alginolyticus* and two *Vibrio diabolicus* strains, according to the clustering obtained by Thompson *et al.* (2001). Interestingly, these two *V. diabolicus* strains were originally identified by Vandenberghe *et al.* (1999), using phenotypic and genotypic techniques, as *V. alginolyticus*. The nearest class of BC2 was BC36, which included one *V. diabolicus* strain.

Class BC3 contained 22 *Enterovibrio norvegicus* strains. The combination of the two fAFLP clusters A68 and A69 in BC3 is in complete agreement with more recent analyses based on DNA hybridization and 16S rRNA gene sequences which proved that these two clusters are in fact a single species, *E. norvegicus* (Thompson *et al.*, 2002). The nearest class to BC3 was BC44, which contains *Vibrio hollisae*. This is also the closest phylogenetic neighbour of *Enterovibrio* based on 16S rRNA gene analysis, having about 95 % sequence similarity (Thompson *et al.*, 2002). Class BC4 consisted of 21 *Vibrio neptunius* strains of remarkably low pattern diversity, while class BC5 merged all *Vibrio anguillarum* and *Vibrio ordalii* strains analysed. *V. ordalii* was described by Schiewe *et al.* (1981) to encompass biotype II of *V. anguillarum*. It is well known that these species are highly related, having nearly 100 % 16S rRNA gene similarity and 70 % DNA–DNA similarity. Class BC6 contained the type

Fig. 1. Comparison of the classification described by Thompson *et al.* (2001) and the BinClass classification based on data discretized by the sliding-window method with the position tolerance parameter ε set to 0.007 and the resolution of the method δ set to 0.001 (so that the vector length $d=994$). BinClass was run with command line settings (`-F50 -S20`), resulting in a classification with 64 classes and an SC of 739.92280. Each row represents a class from the classification described previously by Thompson *et al.* (2001), with the assigned class identifier in the first column and the number of strains in the last column. Each column represents a class from the BinClass classification, with the assigned class identifier in the first row and the number of strains in the last row. The values in the row/column intersections represent the number of strains that the two corresponding classes have in common.

strain of *Vibrio fortis* and 15 *Vibrio cyclitrophicus* strains, while most of the other members of *V. fortis* appeared in BC8 and BC24. *V. fortis* was proposed to encompass strains of the former fAFLP clusters A9 and A60 (Thompson *et al.*, 2003a). The fact that the type strain of this species clustered apart from all other species members may suggest that the type strain of *V. fortis* is a species on its own and that this heterogeneous species may be split into new species in the future. Class BC8 merged five strains of the former fAFLP group A9 and all eight strains of A60, which correspond to the newly described *V. fortis*. Clearly BC8 differs from the type strain of *V. fortis* (in BC6).

Class BC7 consisted of 14 strains, including the type strains of *Vibrio gazogenes*, *Vibrio fluvialis*, *Vibrio furnissii* and *Vibrio logei*. A remarkable feature of this diverse class is the large distance from all the type strains to the HMO (Liston *et al.*, 1963). It is reasonable that *V. fluvialis* and *V. furnissii* (former biotype of *V. fluvialis*) group together, although one would not expect the inclusion of the psychrophilic *V. logei* in this class. A 16S rRNA gene-based phylogenetic analysis of *V. logei* revealed that this organism is more related to the psychrophilic vibrios (e.g. *Vibrio fischeri*, *Vibrio salmonicida*, *Vibrio wodanis*) than to any of the species within BC7. Class BC9 ($n=13$) contained nine strains of a new *Vibrio* species, *Vibrio neonatus* (Sawabe *et al.*, 2004). This class included two other vibrios named *Vibrio pelagius* and *Vibrio cincinnatiensis*, which are most probably representatives of this new species. Classes BC10, BC12, BC23, BC25, BC33, BC51 and BC56 contained *Vibrio harveyi* strains, with the type strain of *V. harveyi* being in BC12, suggesting that this is a very diverse species. In contrast, BC11 ($n=12$) included the type strain and most reference strains of *Vibrio splendidus* and one strain of *Vibrio tubiashii*. This class was related to BC47, which contained all the remaining *V. splendidus* strains; these may be a variant of *V. splendidus* or a new species. Two strains of BC47 were found in the so-called ribotype cluster C described by Mácian *et al.* (2000). Strains originally allocated to *V. tubiashii* (Thompson *et al.*, 2001) were repartitioned into different BC classes (BC10, BC11, BC15, BC17, BC20, BC27, BC33, BC34, BC35, BC36, BC41, BC42, BC49, BC59 and BC63), with the type strain allocated to BC27. This suggests that the original *V. tubiashii* group as delineated by Thompson *et al.* (2001) was quite artificial.

Class BC13 contained a new *Vibrio* species, '*Vibrio ezurae*' (Sawabe *et al.*, 2002, 2004), while BC14 contained 12 *Vibrio ichthyocyteri* strains. Classes BC15 and BC16 included most *Vibrio coralliilyticus* strains (Ben-Haim *et al.*, 2002), although BC24, BC31 and BC53 included four, seven and two strains of this species, respectively. It is quite remarkable that BC31 made a (homogeneous) cluster on its own, suggesting it is a variant of *V. coralliilyticus*. Strains of this class appear to be specialized in causing disease in *Nodipecten nodosus* bivalve larvae, while other *V. coralliilyticus* are known coral pathogens (Ben-Haim *et al.*, 2002). Class BC17 consisted of 10 strains including the type strain

of *Vibrio chagasii* and most of the reference strains of this new species (Thompson *et al.*, 2003c). The nearest neighbour of BC17 was BC36, which contained three *V. chagasii* strains (LMG 13220, LMG 13222 and LMG 13239), suggesting that they may be yet another new species. In fact, a closer examination of the fAFLP patterns of *V. chagasii* strains and the DNA–DNA hybridization data indicates a large diversity within this species in support of the new grouping obtained here. In class BC18 two species, *Vibrio brasiliensis* and *V. diabolicus*, were merged. This is not an ideal situation, but the grouping might be just a reflection of the phylogenetic relatedness between the two species as they share approximately 98% 16S rRNA gene similarity. BC19 included all *Vibrio lentus* strains and two *V. logei* strains. Interestingly, the nearest neighbour class to BC19 was BC11. It is well known that *V. lentus* and *V. splendidus* are highly related species. The strains allocated to *V. logei* in the previous fAFLP analysis (Thompson *et al.*, 2001) were quite heterogeneous, as can be seen from the new partitioning of the strains. The type strain of *V. logei* appeared in BC7, but another one, two and three strains appeared in BC18, BC19 and BC26, respectively. These strains were supposed to be *V. logei*, however the results presented here undermine this assumption.

BC20 merged together three type strains (i.e. *P. histaminum*, *Photobacterium damsela* and *Vibrio kanaloae*), which might undermine the value of the new classification. However, *P. damsela* was originally clustered with *Photobacterium angustum* (Thompson *et al.*, 2001). *P. angustum* appears now in BC34 along with the type strains of *V. salmonicida* and *Photobacterium leiognathi*. BC21 consisted of nine *Vibrio mediterranei* strains including the former type strain of *Vibrio shilonii* (Thompson *et al.*, 2001). BC22 comprised six strains of a newly described species, *Vibrio tasmaniensis* (Thompson *et al.*, 2003b), and three strains from two very heterogeneous fAFLP clusters, A56 and A58. Thompson *et al.* (2001) highlighted that the precise taxonomic allocation of isolates clustering with more than one type strain was unclear, requiring further investigation. Here we demonstrate the partitioning of such isolates by using minimization of SC, indicating the usefulness of this new approach.

BC23 contained *Vibrio rotiferianus* (Gomez-Gil *et al.*, 2003) and four *V. harveyi* strains. According to Gomez-Gil *et al.* (2003) both species are highly related, and the results presented here may suggest that those four strains identified as *V. harveyi* are in fact *V. rotiferianus*. BC23 was related to BC51, which comprised four diverse *V. harveyi* isolates. Class BC24 included the type strains of *V. cincinnatiensis* and *V. pelagius*, whereas BC25 included those of *Photobacterium phosphoreum* and *Vibrio proteolyticus*. All these species were clearly separated in the clustering of Thompson *et al.* (2001), but with the SC-minimizing classification one may expect that certain species be grouped together. This fact may be just a reflection of the limitation of band patterns, which happen to give similar fingerprints between

completely unrelated species (e.g. *P. phosphoreum* and *V. proteolyticus*). We inspected the original patterns of the species within BC25 and found large gaps, which will turn out as zeros in the binarized patterns used for comparisons.

Class BC26 consisted of five *Vibrio vulnificus* strains. This class also included three so-called *V. logei* strains. Two of these strains, i.e. VIB 523 and STD3-996, are clearly *V. vulnificus* representatives misidentified by Thompson *et al.* (2001). Arias *et al.* (1997) identified VIB 523 to the species *V. vulnificus* using phenotypic and genotypic techniques, while our 16S rRNA gene sequence of STD3-996 revealed 100 % similarity to *V. vulnificus*. BC27 consisted of eight strains, including the type strain of *V. tubiashii*, while BC28 had seven strains, including the type strain of *Vibrio parahaemolyticus*. BC28 was closely related to BC2, which contained *V. alginolyticus* (a former variant of *V. parahaemolyticus*). BC29 comprised the species *Vibrio pectenida* and two other strains which were probably misidentified in the former analysis by Thompson *et al.* (2001). Surprisingly, class BC30 put together the former fAFLP cluster A34 and the type strain of *Vibrio diazotrophicus*. Originally A34 was thought to be a new species; however, recent DNA–DNA hybridization data have proven that A34 belongs to *V. diazotrophicus*. Class BC33 contained six strains, including the type strain of *Vibrio campbellii*.

Class BC35 accommodated *Vibrio hispanicus* and three other strains. BC38 merged the type strains of *Vibrio orientalis* and *Vibrio natriegens*. Originally the type strains of *V. orientalis* and *Vibrio nigripulchritudo* were grouped together with another five strains of uncertain taxonomic position (Thompson *et al.*, 2001). BC39 grouped *V. fischeri* and *Photobacterium iliopiscarius* and so did the previous clustering (Thompson *et al.*, 2001). BC40 consisted of four *Vibrio pomeroyi* strains including the type strain, but two further strains of this species were found in BC42. This is interesting in that these two strains (LMG 21351 and LMG 21352) differ in fAFLP patterns and the DNA–DNA hybridization data, suggesting that they are in fact at the outskirts of the species *V. pomeroyi* (Thompson *et al.*, 2003c). Classes BC43 to BC45 all suffered from the same problem, namely merging of more than one type strain. Oddly enough, *Salinivibrio costicola* and *Vibrio xuii*, which are such different species, were included in the same class, BC43. On the other hand, BC46 consisted of three *Vibrio tapetis* strains which were previously merged with *Vibrio penaeicida* and *Vibrio rumoiensis* in the so-called fAFLP cluster A58. In the present analysis *V. rumoiensis* appeared in BC54 and *V. penaeicida* in BC57. Classes BC48, BC49, BC50, BC52 and BC58 contained *Vibrio nereis*, *Vibrio metschnikovii*, *Vibrio aestuarianus*, *Vibrio cholerae* and *Vibrio mimicus*, respectively. All these species had previously formed clusters on their own (Thompson *et al.*, 2001).

For the *Vibrio*/AFLP dataset, there was good overall agreement between the classification based on the minimization of SC and the classification as described by Thompson *et al.* (2001) based on Ward's hierarchical clustering algorithm.

However, it is quite clear that certain heterogeneous fAFLP groups, e.g. A12, A13, A56, A58 and 59, were totally repartitioned with the new classification. Obviously, many of the repartitioned strains had been given only tentative names. For instance, A59 contained the type strain of *V. tubiashii* and thus we assumed that all strains clustering together with this type strain at the level of as low as 45 % similarity would belong to the species *V. tubiashii*. Apparently this assumption goes beyond the discrimination of fAFLP and the mathematical algorithms used for fAFLP pattern analysis. Additionally, in other cases (e.g. BC46, BC54 and BC57) the new classification has repartitioned into separate classes type strains which were grouped together in a former analysis (Thompson *et al.*, 2001). Some 'hidden' relationships have been disclosed with the help of the BinClass classification. This is the case for the former fAFLP groups A68 and A69, which appear in class BC3 (*E. norvegicus*). Likewise clusters A9 and A60 are merged in BC8 (*V. fortis*), clusters A34 and A57 in BC30 (*V. diazotrophicus*), and also clusters A14 and A37 in BC27 (*V. campbellii*). All these four classes have been validated by DNA–DNA hybridization data which showed that these classes include strains from the same species (Thompson *et al.*, 2002, 2003a).

Conclusion

In this report we have demonstrated the applicability of sliding-window discretization in combination with the minimization of SC for the classification of genotypic fingerprint patterns. As such, this new classification strategy forms a complementary alternative to the hierarchical cluster algorithms that are commonly used in bacterial taxonomy. Comparison of the classifications for the same set of fAFLP fingerprint patterns by different classification strategies has revealed that there was a good overall correlation between the alternative groupings, but that no single classification managed to reflect all the taxonomic relationships within the *Vibrionaceae*.

ACKNOWLEDGEMENTS

F. L. Thompson worked on a PhD scholarship (no. 2008361/98-6) from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. J. Swings acknowledges grants from the Fund for Scientific Research (FWO), Belgium. The research of T. Koski has been supported by the Swedish Research Council (VR/NT). The research of M. Gyllenberg was supported by the Academy of Finland.

REFERENCES

- Anderberg, M. R. (1973). *Cluster Analysis for Applications*. New York: Academic Press.
- Arias, C. R., Verdonck, L., Swings, J., Garay, E. & Aznar, R. (1997). Intraspecific differentiation of *Vibrio vulnificus* biotypes by amplified fragment length polymorphism and ribotyping. *Appl Environ Microbiol* 63, 2600–2606.

- Austin, B., Dawyndt, P., Gyllenberg, M., Koski, T., Lund, T., Swings, J. & Thompson, F. L. (2004). Sliding window discretization: a new method for multiple band matching of bacterial genotyping fingerprints. *Bull Math Biol* **66**, 1575–1596.
- Ben-Haim, Y., Thompson, F. L., Thompson, C. C., Cnockaert, M. C., Hoste, B., Swings, J. & Rosenberg, E. (2003). *Vibrio coralliilyticus* sp. nov., a temperature-dependent pathogen of the coral *Pocillopora damicornis*. *Int J Syst Evol Microbiol* **53**, 309–315.
- Buchanan, R. E. (1925). *General Systematic Bacteriology*. Baltimore: Williams & Wilkins.
- Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley.
- Dice, L. R. (1945). Measures of the amount of ecological association between species. *Ecology* **26**, 297–302.
- Gomez-Gil, B., Thompson, F. L., Thompson, C. C. & Swings, J. (2003). *Vibrio pacinii* sp. nov., from cultured aquatic organisms. *Int J Syst Evol Microbiol* **53**, 1569–1573.
- Gyllenberg, M., Koski, T. & Verlaan, M. (1997a). Classification of binary vectors by stochastic complexity. *J Multivariate Anal* **63**, 47–72.
- Gyllenberg, H. G., Gyllenberg, M., Koski, T., Lund, T., Schindler, J. & Verlaan, M. (1997b). Classification of *Enterobacteriaceae* by minimization of stochastic complexity. *Microbiology* **143**, 721–732.
- Gyllenberg, H. G., Gyllenberg, M., Koski, T. & Lund, T. (1998). Stochastic complexity as a taxonomic tool. *Comput Methods Programs Biomed* **56**, 11–22.
- Gyllenberg, H. G., Gyllenberg, M., Koski, T., Lund, T. & Schindler, J. (1999). *Enterobacteriaceae* taxonomy approached by minimization of stochastic complexity. *Quantitative Microbiol* **1**, 157–170.
- Gyllenberg, M., Koski, T. & Lund, T. (2001). BinClass: a software package for classifying binary vectors. User's guide. *TUCS Technical Report* **411** <http://www.tucs.fi/publications/techreports/TR411.php>.
- Gyllenberg, M., Dawyndt, P., Koski, T., Lund, T., Thompson, F., Austin, B. & Swings, J. (2002). New methods for the analysis of binarized BIOLOG GN data of *Vibrio* species: minimization of stochastic complexity and cumulative classification. *Syst Appl Microbiol* **25**, 403–415.
- Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M. & Kersters, K. (1996). Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**, 1881–1893.
- Liston, J., Wiebe, W. J. & Colwell, R. R. (1963). Quantitative approach to the study of bacterial species. *J Bacteriol* **85**, 1061–1070.
- Mácian, M. C., Garay, E., Gonzalez-Candelas, F., Pujalte, M. J. & Aznar, R. (2000). Ribotyping of *Vibrio* populations associated with cultured oysters (*Ostrea edulis*). *Syst Appl Microbiol* **23**, 409–417.
- Priest, F. & Austin, B. (1993). *Modern Bacterial Taxonomy*, 2nd edn. London: Chapman & Hall.
- Sawabe, T., Thompson, F. L., Heyrman, J. & 7 other authors (2002). Fluorescent amplified fragment length polymorphism and repetitive extragenic palindrome-PCR fingerprinting reveal host-specific genetic diversity of *Vibrio halotictoli*-like strains isolated from the gut of Japanese Abalone. *Appl Environ Microbiol* **68**, 4140–4144.
- Sawabe, T., Hayashi, K., Moriwaki, J., Thompson, F. L., Swings, J. & Christen, R. (2004). *Vibrio neonatus* sp. nov. and *Vibrio ezuriae* sp. nov. isolated from the gut of Japanese abalones. *Syst Appl Microbiol* **27**, 527–534.
- Schiewe, M. H., Trust, T. J. & Crosa, J. H. (1981). *Vibrio ordalii* sp. nov.: a causative agent of vibriosis in fish. *Curr Microbiol* **6**, 343–348.
- Sneath, P. H. A. (1984). *Bacterial Nomenclature*. In *Bergey's Manual of Systematic Bacteriology*, vol. 1, pp. 19–23. Edited by N. R. Krieg & J. G. Holt. Baltimore: Williams & Wilkins.
- Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. San Francisco: W. H. Freeman.
- Thompson, F. L., Hoste, B., Vandemeulebroecke, K. & Swings, J. (2001). Genomic diversity amongst *Vibrio* isolates from different sources determined by fluorescent amplified fragment length polymorphism. *Syst Appl Microbiol* **24**, 520–538.
- Thompson, F. L., Hoste, B., Thompson, C. C., Goris, J., Gomez-Gil, B., Huys, L. & Swings, J. (2002). *Enterovibrio norvegicus* gen. nov., sp. nov., isolated from the gut of turbot (*Scophthalmus maximus*) larvae: a new member of the family *Vibrionaceae*. *Int J Syst Evol Microbiol* **52**, 2015–2022.
- Thompson, F. L., Thompson, C. C., Hoste, B., Vandemeulebroecke, K., Gullian, M. & Swings, J. (2003a). *Vibrio fortis* sp. nov. and *Vibrio hepatarius* sp. nov., isolated from aquatic animals and the marine environment. *Int J Syst Evol Microbiol* **53**, 1495–1501.
- Thompson, F. L., Thompson, C. C. & Swings, J. (2003b). *Vibrio tasmaniensis* sp. nov., isolated from atlantic salmon (*Salmo salar* L.). *Syst Appl Microbiol* **26**, 65–69.
- Thompson, F. L., Thompson, C. C., Li, Y., Gomez-Gil, B., Vandenberghe, J. & Swings, J. (2003c). *Vibrio kanaloae* sp. nov., *Vibrio pomeroyi* sp. nov. and *Vibrio chagasii* sp. nov., from sea water and marine animals. *Int J Syst Evol Microbiol* **53**, 753–759.
- Thompson, F. L., Li, Y., Gomez-Gil, B. & 8 other authors (2003d). *Vibrio neptunius* sp. nov., *Vibrio brasiliensis* sp. nov. and *Vibrio xuii* sp. nov., isolated from the marine aquaculture environment (bivalves, fish, rotifers and shrimps). *Int J Syst Evol Microbiol* **53**, 245–252.
- Vandenberghe, J., Verdonck, L., Robles-Arozarena, R. & 7 other authors (1999). *Vibriosis* associated with *Litopenaeus vannamei* larvae, postlarvae, broodstock and hatchery probionts. *Appl Environ Microbiol* **65**, 2592–2597.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* **58**, 236–244.