Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR* genes

Javier Pascual,^{1,2} M. Carmen Macián,^{2,3} David R. Arahal,^{2,3}
Esperanza Garay^{2,3} and María J. Pujalte^{2,3}

Correspondence

María J. Pujalte
maria.j.pujalte@uv.es

¹Instituto Cavanilles de Biodiversidad y Biología Evolutiva (ICBiBE), Universidad de Valencia, Spain

²Departamento de Microbiología y Ecología, Universidad de Valencia, Spain

³Colección Española de Cultivos Tipo (CECT), Universidad de Valencia, Spain

The central clade of the genus *Vibrio*, also called the *Vibrio* core group, comprises six species that are tightly related (DNA–DNA reassociation values are very close to 70 % for most species pairs). Identification of novel strains to the species level within this group is troublesome and results are quite often dependent on the methodology employed. Therefore, this group represents an excellent framework to test the robustness of multilocus sequence analysis (MLSA) not only for inferring phylogeny but also as an identification tool without the need for DNA–DNA hybridization assays. The genes selected, 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR*, were amplified by direct PCR from 44 *Vibrio* core-group strains. Subsequent analysis allowed us to recognize *toxR* and *rpoD* as the most resolving individual genes and showed that concatenated sequences of *rpoD*, *rctB* and *toxR* were more useful than concatenated sequences of all seven genes. To validate our conclusions, MLSA similarities have been correlated with DNA–DNA relatedness values obtained in this study and values taken from the literature. Although the seven concatenated genes gave the best correlation, the concatenated sequences of *rpoD*, *rctB* and *toxR* have the practical advantage of showing a considerable gap between the maximal interspecies similarity and the minimal intraspecies similarity recorded, meaning that they can be used quite conveniently for species identification of vibrios.

INTRODUCTION

MLSA (multilocus sequence analysis) is a recently developed technique derived from multilocus sequence typing that has been used successfully in epidemiology studies. The use of MLSA as an alternative method for species delineation in bacteriology is currently being evaluated, following a recommendation of the ad hoc committee for the re-evaluation of the species definition in bacteriology (Stackebrandt *et al.*, 2002). In MLSA studies, several housekeeping genes (more than five) are analysed and relationships between taxa are established. One important advantage of this kind of approach is the

availability of genomic sequences from any laboratory, avoiding the problems of lack of comparability that are associated with DNA–DNA reassociation data. However, the usefulness of MLSA for describing bacterial species has to be demonstrated for each of the taxa or group of species studied and for the selected housekeeping genes. Authors should prove that there is a sufficient degree of congruence between the alternative technique (i.e. MLSA) and DNA–DNA reassociation data. To date, the technique has been applied to discrete bacterial taxa, mainly genera of lactic acid bacteria (Naser *et al.*, 2006), *Borrelia* (Richter *et al.*, 2006), mycobacteria (Mignard & Flandrois, 2008), pseudomonads and relatives (Young & Park, 2007), *Ensifer* (Martens *et al.*, 2007, 2008) and the *Vibrionaceae* (Goarant *et al.*, 2006; Urbanczyk *et al.*, 2007; Thompson *et al.*, 2008; Rameshkumar *et al.*, 2008).

In the family *Vibrionaceae*, sequence similarities for the 16S rRNA gene are ≥ 97.6 % among members of the so-called *Vibrio* core group, which is formed by six species, *Vibrio harveyi*, *V. campbellii*, *V. rotiferianus*, *V. parahaemolyticus*, *V. alginolyticus* and *V. natriegens* (Dorsch *et al.*, 1992). DNA–DNA relatedness values are also very high, in some

Abbreviations: ML, maximum-likelihood; MLSA, multilocus sequence analysis; MP, maximum-parsimony; NJ, neighbour-joining.

The GenBank/EMBL/DBJ accession numbers of newly reported sequences are provided in Table 1.

Details of genes examined in this study and DNA–DNA hybridization results taken from the literature, single-gene maximum-likelihood and maximum-parsimony trees and additional trees based on concatenated sequences and details of detected recombinations are available as supplementary material with the online version of this paper.

cases approaching 70 %, as is the case for *V. harveyi* and *V. campbellii* (Reichelt *et al.*, 1976; Brenner *et al.*, 1983). Sawabe *et al.* (2007) estimated the divergence of this pair of species at only 39 million years ago, one of the more recent divergence events within the genus. Accordingly, the phenotypic and genotypic resemblance among the members of the group is so high that almost no traits are available for absolute discrimination between species, making identification to the species level very difficult. As several species in the group are pathogens of humans and/or fish, accurate differentiation among them is particularly relevant. In this context, the study of Thompson *et al.* (2007) represented a major contribution to the differentiation of *V. harveyi* and *V. campbellii* through MLSA; however, only a few strains of their sister species *V. rotiferianus* and no other members of the clade were included.

The aim of this study was to investigate the usefulness of the MLSA approach for the resolution of the central clade of *Vibrio*, using six protein-encoding genes and the 16S rRNA gene from 44 representatives of all species and to compare the output of this technique with DNA–DNA relatedness within the species complex.

METHODS

Strains and culture conditions. Forty-four *Vibrio* strains were used in this study (Table 1), including type, reference and environmental strains. The six species of the central clade of *Vibrio* together with the type strain of *Vibrio cholerae*, CECT 514^T, were studied. Type and reference strains were obtained from CECT (Colección Española de Cultivos Tipo, Spain) and CAIM (Collection of Aquaculture Important Microorganisms, Mexico). Some environmental strains were obtained during surveys on bacteria associated with cultured fishes in several Spanish Mediterranean fish farms and tentatively identified as *V. harveyi* by Pujalte *et al.* (2003). Some other strains were received as *V. rotiferianus* according to (GTG)₅ analysis. The origin of the strains is reported in Table 1. Strains were grown aerobically on marine agar (MA) and in marine broth (MB) (Difco 2260) overnight at room temperature. Cultures were maintained on semi-solid MA stabs (prepared with 0.2 % w/v agar) at room temperature and as suspensions in MB plus 20 % glycerol at –80 °C.

DNA preparation, PCR amplification and sequencing of the genes. Bacterial genomic DNAs were extracted according to a method described previously (Pitcher *et al.*, 1989). The following genes were studied: 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR*. PCR primers used for amplification and sequencing of these genes are listed in Supplementary Table S1, available in IJSEM Online. PCR mixtures were composed of 5.0 µl PCR buffer (10 ×), 0.75 µl MgCl₂ (100 mM), 1.0 µl dNTPs (10 mM each), 1.0 µl each forward and reverse primers (50 µM), 0.5 µl *Taq* polymerase (5 U µl^{–1}; Genotag) and 5.0 µl template DNA (50 ng µl^{–1}) in a total volume of 50 µl. PCR amplifications were performed using a PTC-100 MJ Research ThermoCycler. For the protein-encoding genes, the thermal program consisted of (i) 5 min at 95 °C; (ii) 3 cycles of 1 min at 95 °C, 2 min 15 s at 55 °C and 1 min 15 s at 72 °C; (iii) 30 cycles of 35 s at 95 °C, 1 min 15 s at 55 °C and 1 min 15 s at 72 °C; and (iv) a final extension at 72 °C for 10 min. The conditions for 16S rRNA gene amplification were (i) 4 min at 94 °C; (ii) 30 cycles of 1 min at 94 °C, 1 min 30 s at 52 °C and 2 min at 72 °C; and (iii) a final elongation

step of 10 min at 72 °C. Amplified products were examined by agarose gel electrophoresis (1.2 %) and ethidium bromide staining. Purified amplicons (Mo Bio Laboratories) were sequenced by the dideoxy method using the BigDye Terminator v. 3.0 Ready Reaction cycle sequencing kit and analysed in an ABI PRISM 3700 sequencer (Applied Biosystems). Sequencing primers were the same ones used in the amplification reaction but diluted tenfold (5 pmol).

Recombination analyses. Putative recombination events were detected using RDP3 Beta 27 (Martin *et al.*, 2005). This program identifies recombinant sequences and recombination breakpoints using two phylogenetic methods, RDP and Bootscan, which infer recombination when different parts of the genome result in discordant topologies, and four nucleotide substitution methods, MaxChi, Chimera, Geneconv and SiScan, which examine the sequences either for significant clustering of substitutions or for a fit to an expected statistical distribution. Recombination events were considered only when they were identified by at least two methods. Statistical significance was set at the $P \leq 0.01$ level.

Phylogenetic data analysis. Nucleotide sequence alignments were revised visually to identify positions with uncertain alignment, mainly at the ends of the sequences, to be corrected or omitted from further analyses. Multiple sequence alignments were obtained using CLUSTAL_X (Thompson *et al.*, 1997), taking into account the corresponding amino acid alignment for protein-coding genes. Phylogenetic analysis was performed using the program PAUP* version 4.0b10 (Swofford, 2002). Neighbour-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML) analyses were obtained for each gene. NJ analyses were performed using Kimura's two-parameter model. MP analyses were carried out using a heuristic search option. For ML analysis of each gene, the optimal models of nucleotide substitution were estimated through the program MODELTEST 3.7 (Posada & Crandall, 1998; Posada & Buckley, 2004) using the Akaike information criterion (AIC) (Supplementary Table S1). Bootstrap analyses were performed using 1000 replications for NJ, MP and ML.

DNA–DNA hybridization and correlation studies. DNA–DNA hybridization experiments were done with the hydroxyapatite method using microtitre plates as described by Ziemke *et al.* (1998) with a hybridization temperature of 60 °C. These experiments were performed in triplicate on 11 strains selected as representatives of each subclade of the *V. harveyi*, *V. campbellii* and *V. rotiferianus* clades obtained in the seven-gene concatenated analysis.

DNA–DNA hybridization data obtained in this study (Table 2) as well as values for type and reference strains retrieved from the literature (Supplementary Table S2) were compared to the distance matrix data for each gene individually and the distance matrix corresponding to the concatenated gene sequences. Correlation between values was calculated using Pearson's product–moment correlation coefficient.

RESULTS AND DISCUSSION

In this study, we have analysed the sequences of seven genes (six protein-encoding genes and the 16S rRNA gene) from 44 *Vibrio* strains representing the six species of the core group (Table 1). Among the environmental strains, eight were reidentified according to the results obtained in this study (see footnote to Table 1), and this identity is used in the discussion that follows. All sequences were obtained in our laboratory except for six 16S rRNA gene sequences that were retrieved from public databases (Table 1). The length of the alignments used were:

Table 1. Origin and identity of strains used in this study and sequence accession numbersSequences represented by accession numbers in *italics* were retrieved from public databases.

Strain	Species	Origin	16S rRNA	<i>recA</i>	<i>pyrH</i>	<i>rpoD</i>	<i>gyrB</i>	<i>rctB</i>	<i>toxR</i>
CECT 525 ^T	<i>V. harveyi</i>	Dead luminescent amphipod <i>Talorchestia</i> sp.; USA	AY750575	FM204794	FM202541	FM202498	FM202583	FM202625	FM202678
SA F3s 40	<i>V. harveyi</i>	<i>Sparus aurata</i> ; Spain, 1999	FM204836	FM204795	FM202542	FM202499	FM202584	FM202626	FM202679
H2	<i>V. harveyi</i>	<i>Sparus aurata</i> ; Spain, 2004	FM204837	FM204796	FM202543	FM202500	FM202585	FM202627	FM202680
DL F5 57	<i>V. harveyi</i>	<i>Dicentrarchus labrax</i> ; Spain, 1999	FM204838	FM204797	FM202544	FM202501	FM202586	FM202628	FM202681
SA F1s 10	<i>V. harveyi</i>	<i>Sparus aurata</i> ; Spain, 1999	FM204839	FM204798	FM202545	FM202502	FM202587	FM202629	FM202682
SA F4s 17	<i>V. harveyi</i>	<i>Sparus aurata</i> ; Spain, 1999	FM204840	FM204799	FM202546	FM202503	FM202588	FM202630	FM202683
LPD 1-1-10	<i>V. harveyi</i>	<i>Dicentrarchus labrax</i> ; Spain, 1999	FM204841	FM204800	FM202547	FM202504	FM202589	FM202631	FM202684
LPD 1-3-1	<i>V. harveyi</i>	<i>Dicentrarchus labrax</i> ; Spain, 2000	FM204842	FM204801	FM202548	FM202505	FM202590	FM202632	FM202685
LPD 1-3-27	<i>V. harveyi</i>	<i>Dicentrarchus labrax</i> ; Spain, 2000	FM204843	FM204802	FM202549	FM202506	FM202591	FM202633	FM202686
LPD 1-3-35	<i>V. harveyi</i>	<i>Dicentrarchus labrax</i> ; Spain, 2000	FM204844	FM204803	FM202550	FM202507	FM202592	FM202634	FM202687
LPD 1-1-11	<i>V. rotiferianus</i> *	<i>Dicentrarchus labrax</i> ; Spain, 1999	FM204861	FM204822	FM202569	FM202526	FM202611	FM202653	FM202706
LPD 1-1-24	<i>V. rotiferianus</i> *	<i>Dicentrarchus labrax</i> ; Spain, 1999	FM204862	FM204823	FM202570	FM202527	FM202612	FM202654	FM202707
LPD 1-1-84	<i>V. rotiferianus</i> *	<i>Dicentrarchus labrax</i> ; Spain, 1999	FM204863	FM204824	FM202571	FM202528	FM202613	FM202655	FM202708
LPD 1-1-86	<i>V. rotiferianus</i> *	<i>Dicentrarchus labrax</i> ; Spain, 1999	FM204864	FM204825	FM202572	FM202529	FM202614	FM202656	FM202709
CECT 523 ^T	<i>V. campbellii</i>	Seawater; USA	X74692	FM204804	FM202551	FM202508	FM202593	FM202635	FM202688
CAIM 9	<i>V. campbellii</i>	<i>Penaeus monodon</i> (hepatopancreas); Philippines	FM204845	FM204805	FM202552	FM202509	FM202594	FM202636	FM202689
CAIM 128	<i>V. campbellii</i>	Diseased shrimp (<i>Penaeus</i> sp.), haemolymph; Mexico, 2000	FM204846	FM204806	FM202553	FM202510	FM202595	FM202637	FM202690
CAIM 333	<i>V. campbellii</i>	Seawater from shrimp (<i>Litopenaeus stylirostris</i>) broodstock tank; Mexico, 1999	FM204847	FM204807	FM202554	FM202511	FM202596	FM202638	FM202691
CAIM 392	<i>V. campbellii</i>	Diseased Z3 shrimp (<i>Penaeus chinensis</i>) larvae; China	FM204848	FM204808	FM202555	FM202512	FM202597	FM202639	FM202692
CAIM 415	<i>V. campbellii</i>	Seawater from shrimp (<i>Litopenaeus stylirostris</i>) broodstock tank; Mexico, 1999	FM204849	FM204809	FM202556	FM202513	FM202598	FM202640	FM202693
CAIM 419	<i>V. campbellii</i>	Shrimp (<i>Litopenaeus</i> sp.), haemolymph; Mexico, 2001	FM204850	FM204810	FM202557	FM202514	FM202599	FM202641	FM202694
CAIM 521	<i>V. campbellii</i>	Seawater; USA	FM204851	FM204811	FM202558	FM202515	FM202600	FM202642	FM202695
CAIM 780	<i>V. campbellii</i>	Shrimp (<i>Litopenaeus vannamei</i>), haemolymph; Mexico, 2003	FM204852	FM204812	FM202559	FM202516	FM202601	FM202643	FM202696
R379	<i>V. campbellii</i> †	Oyster (<i>Crassostrea gigas</i>); Bay of Alfacos, Spain, 2005	FM204854	FM204814	FM202561	FM202518	FM202603	FM202645	FM202698
R1024	<i>V. campbellii</i> †	Coral (<i>Mussismilia hispida</i>); Brazil, 2005	FM204853	FM204813	FM202560	FM202517	FM202602	FM202644	FM202697
R1117	<i>V. campbellii</i> †	Coral mucus (<i>Mussismilia hispida</i>); Brazil, 2005	FM204855	FM204815	FM202562	FM202519	FM202604	FM202646	FM202699
R1311	<i>V. campbellii</i> †	Spider crab (<i>Maja brachidactyla</i>); Spain	FM204856	FM204816	FM202563	FM202520	FM202605	FM202647	FM202700
CAIM 573	<i>V. rotiferianus</i>	Rotifer (<i>Brachionus plicatilis</i>), water; Belgium, 1999	FM204857	FM204817	FM202564	FM202521	FM202606	FM202648	FM202701
CAIM 574	<i>V. rotiferianus</i>	Rotifer (<i>Brachionus plicatilis</i>), water; Belgium, 1999	FM204858	FM204818	FM202565	FM202522	FM202607	FM202649	FM202702
CAIM 575	<i>V. rotiferianus</i>	Rotifer (<i>Brachionus plicatilis</i>), water; Belgium, 1999	FM204859	FM204819	FM202566	FM202523	FM202608	FM202650	FM202703
CAIM 576	<i>V. rotiferianus</i>	Rotifer (<i>Brachionus plicatilis</i>), water; Belgium, 1999	FM204860	FM204820	FM202567	FM202524	FM202609	FM202651	FM202704
CAIM 577 ^T	<i>V. rotiferianus</i>	Rotifer (<i>Brachionus plicatilis</i>), water; Belgium, 1999	AJ316187	FM204821	FM202568	FM202525	FM202610	FM202652	FM202705
CECT 526 ^T	<i>V. natriegens</i>	Mud from a salt marsh; USA	X74714	FM204826	FM202573	FM202530	FM202615	FM202657	FM202710
CECT 7465	<i>V. natriegens</i>	Oysters; USA	FM999824	FM999812	FM999814	FM999816	FM999818	FM999820	FM999822
CECT 7466	<i>V. natriegens</i>	Seawater; USA	FM999825	FM999813	FM999815	FM999817	FM999819	FM999821	FM999823
CECT 511 ^T	<i>V. parahaemolyticus</i>	Patient suffering from 'Shirashu' food poisoning; Japan, 1951	X74720	FM204827	FM202574	FM202531	FM202616	FM202658	FM202711
CECT 5305	<i>V. parahaemolyticus</i>	Diseased <i>Seriola dumerili</i> ; Spain, 1998	FM204865	FM204828	FM202575	FM202532	FM202617	FM202659	FM202712
CECT 5271	<i>V. parahaemolyticus</i>	Cockles causing fatal food poisoning; UK	FM204866	FM204829	FM202576	FM202533	FM202618	FM202660	FM202713
CECT 611	<i>V. parahaemolyticus</i>	Seawater; Spain, 1981	FM204867	FM204830	FM202577	FM202534	FM202619	FM202661	FM202714
CECT 521 ^T	<i>V. alginolyticus</i>	Spoiled <i>Trachurus trachurus</i> (horse mackerel) causing food poisoning; Japan	X74690	FM204831	FM202578	FM202535	FM202620	FM202662	FM202715

Table 1. cont.

Strain	Species	Origin	16S rRNA	recA	pyrH	rpoD	gyrB	rctB	toxR
CECT 436	<i>V. alginolyticus</i>	Food; 1962	FM204868	FM204832	FM202579	FM202536	FM202621	FM202663	FM202716
CECT 586	<i>V. alginolyticus</i>	Mussels; Netherlands	FM204869	FM204833	FM202580	FM202537	FM202622	FM202664	FM202717
CECT 609	<i>V. alginolyticus</i>	Marine plankton; Spain	FM204870	FM204834	FM202581	FM202538	FM202623	FM202665	FM202718
CECT 514 ^T	<i>V. cholerae</i>	Human faeces	X76337	FM204835	FM202582	FM202539	FM202624	FM202666	FM202719

*Identified phenotypically as *V. harveyi* in a previous study (Pujalte *et al.*, 2003).

[†]Received as *V. rotiferianus* based on (GTG)₃ genotyping (B. Gómez-Gil, personal communication).

Table 2. DNA–DNA hybridization among 11 representative strains of *V. harveyi*, *V. campbellii* and *V. rotiferianus*

Values are means of at least three independent replicates. –, Not done.

Source of unlabelled DNA	Reassociation (%) with labelled DNA from:		
	1	2	3
<i>V. harveyi</i>			
1. CECT 525 ^T	100	50	–
SA F1s 10	96	49	59
SA F4s 17	100	54	69
DL F5 57	99	51	59
<i>V. campbellii</i>			
2. CECT 523 ^T	–	100	–
CAIM 333	58	71	58
CAIM 392	64	80	68
R1024	65	78	69
<i>V. rotiferianus</i>			
3. CAIM 577 ^T	55	52	100
LPD 1-1-84	53	46	86
LPD 1-1-86	64	57	78

768 bp for *recA*, 552 bp for *pyrH*, 780 bp for *rpoD*, 939 bp for *gyrB*, 591 bp for *rctB*, 477 bp for *toxR* and 1385 bp for the 16S rRNA gene (Supplementary Table S1).

Individual gene sequences

For each gene, three phylogenetic trees were obtained, based on the NJ, MP and ML methods. NJ trees are shown in Fig. 1. Other trees are supplied as Supplementary Figs S1 and S2. The range of lengths of the amplicons was from 477 nt for *toxR* to 1385 nt for the 16S rRNA gene. There was no direct relationship between the number of nucleotides and the number of parsimony-informative sites in the nucleotide sequence. The amplicon with the largest number of nucleotides (16S rRNA gene) has the smallest number of parsimony-informative sites and, vice versa, the shortest amplicon (*toxR*) has the largest number of parsimony-informative sites (Supplementary Table S1).

In order to check for possible recombination events in the seven genes studied, we analysed the 44 sequences of each gene with the RDP3 Beta 27 program. Several recombination events were noted in *recA*, *rpoD*, *gyrB*, *rctB* and *toxR* (Supplementary Table S3). Among them, only the events involving the *gyrB* and *rctB* genes affect tree topology. Starting with *gyrB*, all *V. harveyi* strains as well as strains CECT 523^T, CAIM 128, CAIM 419 and CAIM 521 of *V. campbellii* have suffered a recombination event of 280 nt with an unknown strain; therefore, all of them cluster together and separately from the remaining *V. campbellii* strains (*gyrB* trees; Fig. 1 and Supplementary Figs S1 and S2). For the same gene, another strain of *V. campbellii*, R379, and the four strains of *V. alginolyticus* show a recombination event of 256 nt. This event explains the position of strain R379 within the *V. alginolyticus* clade (*gyrB* trees; Fig. 1 and Supplementary Figs S1 and S2).

(a) 16S rRNA gene

(b) *recA* (c) *pyrH* (d) *rpoD*

Fig. 1. Phylogenetic reconstructions based on individual analyses of the 16S rRNA (a), *recA* (b), *pyrH* (c), *rhoD* (d), *gyrB* (e), *rctB* (f) and *toxR* (g) genes using the neighbour-joining method. Bars, 0.01 (a), 0.05 (b–e) and 0.1 (f, g) expected nucleotide substitutions per site. Only bootstrap values above 75 % are shown (1000 resamplings) at branching points.

Finally, *V. alginolyticus* CECT 521^T shows a recombination affecting 169 nt of the *rctB* gene with *Vibrio parahaemolyticus* CECT 5305. This explains the position of the type strain of *V. alginolyticus* close to *V. parahaemolyticus* when using the *rctB* gene (*rctB* trees; Fig. 1 and Supplementary Figs S1 and S2) and increases the within-species distances for this gene (Fig. 2).

The ranges of intraspecific sequence similarity were 98.8–100 % (16S rRNA gene), 92.7–100 % (*recA*), 93.7–100 % (*pyrH*), 95.6–100 % (*rpoD*), 86.8–100 % (*gyrB*), 85.6–100 % (*rctB*) and 77.2–100 % (*toxR*) (Table 3), whereas the ranges of interspecific sequence similarity were 97.6–99.9 % (16S rRNA gene), 87.9–99.9 % (*recA*), 86.4–97.8 % (*pyrH*), 79.1–96.0 % (*rpoD*), 83.1–99.5 % (*gyrB*), 74.3–92.7 % (*rctB*) and 33.8–72.5 % (*toxR*) (Table 3).

For the protein-encoding genes, we also examined the amino acid sequences, applying MP and NJ analyses using the Poisson correction. As expected, these sequences resulted in a less clear separation between species because of the more conservative character of amino acid sequences.

Focusing on the species and their appearance as monophyletic, paraphyletic or polyphyletic groups in the trees generated, we can conclude that *V. parahaemolyticus* and *V. natriegens* were monophyletic in all instances, followed by *V. harveyi* (monophyletic with all genes except *gyrB*, where it was paraphyletic to *V. campbellii* CAIM 128). *V. alginolyticus* and *V. rotiferianus* were monophyletic with four genes: *rctB*, *rpoD* and *toxR* for both species, in addition to *pyrH* and *gyrB*, respectively. Finally, *V. campbellii* was monophyletic only with *rctB*, *rpoD* and *toxR* (Fig. 1 and Supplementary Figs S1 and S2). Therefore, in terms of tree topology, it can be concluded that the 16S rRNA gene, *recA*, *gyrB* and *pyrH* are not adequate for resolution of the six species of the core group of the genus *Vibrio*. In contrast, the other three genes, *rctB*, *rpoD* and *toxR*, can individually resolve each of the six species of the core group, although their relative positions can vary

depending on the gene and the treeing method applied (Fig. 1 and Supplementary Figs S1 and S2).

Following a representation used in a previous MLSA study (Martens *et al.*, 2007), Fig. 2 shows the taxonomic resolution of each individual gene, defined as the overlap or gap between the maximum intraspecific distance and the minimum interspecific distance, depicted as percentage similarity. For the 16S rRNA gene sequence, the overlap between the intraspecific and interspecific distances is very large, from which almost no discrimination ability can be expected. Genes *recA*, *gyrB*, *pyrH* and *rctB* also show some degree of overlap. In the case of *gyrB* and *rctB*, the detected recombination events are the cause of this overlap. For *pyrH*, the low resolution affects only the species *V. campbellii* and *V. rotiferianus*, since several strains among them cause them to be polyphyletic. For *recA*, the resolution is even lower, making it less suitable for MLSA in this group of organisms. Finally, *rpoD* and *toxR*, with a slight overlap and a large gap, respectively, are the genes that wield the highest discriminatory power.

Concatenated gene sequences

Analysis of concatenated gene sequences enhances the quality of the phylogenetic reconstruction and optimizes the taxonomic resolution by adding more informative data and minimizing the weight of recombination events.

Concatenation of the seven genes yielded a sequence of 5492 nt. Trees resulting from NJ, MP and ML analyses are shown in Fig. 3 and Supplementary Fig. S3. All six *Vibrio* species appear as monophyletic, well-defined clades in all trees. The relative branching order of the six species obtained with NJ seems to be more in agreement with DNA–DNA hybridization data from the classical studies on the genus *Vibrio* (Reichelt *et al.*, 1976; Brenner *et al.*, 1983) than the relationships established with MP and ML analyses. In those studies, the *V. harveyi* hybridization group was first established, containing two pairs of sister species, *V. harveyi*–*V. campbellii* and *V.*

Fig. 2. Taxonomic resolution based on the ranges of intraspecific (open bars) and interspecific (filled bars) similarity (%).

Table 3. Distances as percentages of differences in total nucleotide sequence positions

Intraspecific (in bold) and interspecific variability applies for *V. harveyi* (10 strains), *V. campbellii* (13 strains), *V. rotiferianus* (nine strains), *V. parahaemolyticus* (four strains), *V. alginolyticus* (four strains) and *V. natriegens* (three strains).

Gene	1	2	3	4	5	6
16S rRNA						
1. <i>V. harveyi</i>	0.0–0.1					
2. <i>V. campbellii</i>	0.8–2.2	0.0–1.2				
3. <i>V. rotiferianus</i>	0.2–1.3	0.1–1.7	0.0–0.6			
4. <i>V. parahaemolyticus</i>	1.2–2.4	0.4–1.4	0.4–1.2	0.0–0.1		
5. <i>V. alginolyticus</i>	1.0–2.4	0.4–1.2	0.4–1.2	0.1–0.5	0.0–0.2	
6. <i>V. natriegens</i>	1.2–2.4	0.6–1.3	0.7–1.3	0.4–0.6	0.3–0.6	0.0–0.1
recA						
1. <i>V. harveyi</i>	0.0–2.0					
2. <i>V. campbellii</i>	2.1–4.4	0.0–4.0				
3. <i>V. rotiferianus</i>	2.5–4.2	0.10–4.1	0.0–3.3			
4. <i>V. parahaemolyticus</i>	8.6–10.9	7.9–10.8	8.2–10.9	0.4–3.1		
5. <i>V. alginolyticus</i>	5.7–7.4	4.6–7.6	4.8–7.4	9.0–12.0	0.6–7.3	
6. <i>V. natriegens</i>	9.0–10.4	8.9–10.6	9.6–10.4	9.7–12.1	9.9–11	1.2–1.9
gyrB						
1. <i>V. harveyi</i>	0.0–0.7					
2. <i>V. campbellii</i>	0.5–12.2	0.0–13.2				
3. <i>V. rotiferianus</i>	8.5–9.2	4.7–13.3	0.0–0.9			
4. <i>V. parahaemolyticus</i>	14.0–14.8	13.6–14.9	13.7–14.8	0.5–1.3		
5. <i>V. alginolyticus</i>	11.0–12.0	1.4–13.5	12.2–13.7	13.1–14.0	0.5–2.4	
6. <i>V. natriegens</i>	16.4–17.2	14.7–16.9	14.7–15.6	11.9–12.4	14.6–14.9	1.1–1.4
pyrH						
1. <i>V. harveyi</i>	0.0–0.4					
2. <i>V. campbellii</i>	4.3–6.6	0.0–4.2				
3. <i>V. rotiferianus</i>	6.8–8.6	2.2–7.6	0.0–6.3			
4. <i>V. parahaemolyticus</i>	8.1–9.0	8.9–11.2	10.9–11.6	0.2–0.7		
5. <i>V. alginolyticus</i>	12.0–13.5	12.4–13.6	11.7–13.4	7.6–8.5	0.2–2.7	
6. <i>V. natriegens</i>	10.2–11.2	10.0–11.4	9.8–10.3	4.6–5.5	7.9–9.2	0.6–0.9
rctB						
1. <i>V. harveyi</i>	0.0–0.8					
2. <i>V. campbellii</i>	18.1–19.9	0.0–9.0				
3. <i>V. rotiferianus</i>	22.4–23.3	13.7–17.3	0.0–0.2			
4. <i>V. parahaemolyticus</i>	20.4–21.1	16.2–19.2	19.0–19.9	0.8–1.5		
5. <i>V. alginolyticus</i>	22.5–26.1	16.5–22.7	20.4–22.5	7.3–20.6	1.7–14.4	
6. <i>V. natriegens</i>	23.9–24.6	21.1–23.6	24.8–25.7	20.3–21.7	19.9–23.0	0.5–0.8
rpoD						
1. <i>V. harveyi</i>	0.0–0.4					
2. <i>V. campbellii</i>	4.0–4.9	0.1–4.4				
3. <i>V. rotiferianus</i>	13.1–13.9	13.6–14.9	0.0–2.3			
4. <i>V. parahaemolyticus</i>	10.4–11.0	10.3–11.3	15.4–17.2	0.0–0.1		
5. <i>V. alginolyticus</i>	11.3–12.4	10.5–11.5	15.7–17.1	13.2–13.7	0.1–1.0	
6. <i>V. natriegens</i>	13.3–14.2	13.5–15.4	18.0–20.1	14.4–15.5	13.9–15.1	1.1–1.9
toxR						
1. <i>V. harveyi</i>	0.0–0.6					
2. <i>V. campbellii</i>	28.9–32.2	0.0–22.8				
3. <i>V. rotiferianus</i>	57.1–58.5	56.6–63.1	0.0–1.6			
4. <i>V. parahaemolyticus</i>	43.5–44.9	41.9–46.2	60.6–64.4	0.2–1.1		
5. <i>V. alginolyticus</i>	43.0–45.0	39.2–47.4	56.1–58.2	27.5–32.9	0.5–20.4	
6. <i>V. natriegens</i>	55.7–65.9	51.1–64.8	58.3–66.2	46.1–51.7	39.6–46.1	0.6–0.8
Concatenated seven genes						
1. <i>V. harveyi</i>	0.0–0.5					
2. <i>V. campbellii</i>	6.0–8.5	0.1–6.4				
3. <i>V. rotiferianus</i>	10.8–11.4	9.0–11.3	0.0–1.7			
4. <i>V. parahaemolyticus</i>	11.8–12.4	11.2–11.9	13.3–14.0	0.5–0.9		
5. <i>V. alginolyticus</i>	11.6–12.2	9.1–11.8	12.7–13.2	9.5–11.2	0.9–4.7	
6. <i>V. natriegens</i>	13.8–14.7	13.1–14.2	14.4–14.8	11.8–12.3	11.9–12.4	0.8–0.8
Concatenated three genes						
1. <i>V. harveyi</i>	0.0–0.4					
2. <i>V. campbellii</i>	14.6–15.8	0.1–9.9				
3. <i>V. rotiferianus</i>	26.0–26.7	23.6–25.1	0.0–1.4			
4. <i>V. parahaemolyticus</i>	21.4–21.9	20.0–21.0	26.9–27.6	0.4–0.6		
5. <i>V. alginolyticus</i>	22.5–24.2	19.9–22.9	26.5–27.9	15.2–20.3	1.5–9.6	
6. <i>V. natriegens</i>	26.6–28.8	25.2–27.5	30.1–30.9	24.4–25.5	22.7–24.2	0.8–1.1

alginolyticus–*V. parahaemolyticus*, with *V. natriegens* as a less-related member of the group. Although *V. rotiferianus* was not described at the time of those studies, DNA–DNA relatedness data obtained later (Gomez-Gil *et al.*, 2003, 2004) suggest a closer relation to the pair *V. harveyi*–*V. campbellii*. All these phylogenetic relationships are depicted in Fig. 3. Therefore, concatenation of all seven genes is shown to be a very good tool for identification at the species level as well as for inferring phylogeny in the core group (Fig. 3). In addition, this combination of genes improves the phylogenetic reconstruction obtained by Thompson *et al.* (2007), which resolved *V. harveyi* as a monophyletic cluster but yielded several paraphyletic clades for *V. campbellii* and *V. rotiferianus* strains.

A second approach to combine sequence data used the three most resolving genes, *toxR*, *rpoD* and *rctB* (Fig. 2). The concatenated sequence was 1848 nt long. As shown in Supplementary Fig. S4, the phylogeny inferred is essentially the same as that obtained with the seven concatenated genes. Nonetheless, the resolution, in terms of interspecies sequences similarity, was superior for the three concatenated sequences to that obtained with the seven concatenated sequences (Fig. 2).

The six protein-encoding genes used in the present study were chosen in light of the fact that all of them had been used previously in studies of species of the *Vibrionaceae* (Le Roux *et al.*, 2005; Thompson *et al.*, 2005, 2007; Urbanczyk *et al.*, 2007; Rameshkumar *et al.*, 2008). All of them fit the following requirements: suitable size to be sequenced, enough informative sites and an evolutionary rate faster than that of the 16S rRNA gene but slow enough to preserve phylogenetic information.

Thompson *et al.* (2005) used *rpoA*, *recA* and *pyrH* to elucidate the phylogeny of the *Vibrionaceae*, but the trees including the core group species contained only the sequences of the type strains and therefore did not show intraspecific variability. Later, Thompson *et al.* (2007) used *pyrH*, *recA* and *gyrB*, among others, in an MLSA of *V. harveyi* and *V. campbellii* and found that *pyrH* resolved both species satisfactorily, while *recA* and *gyrB* did not clearly resolve them. We agree with the lack of resolution of *recA* and *gyrB* in the *V. harveyi* group, but our results show that *pyrH* is unable to resolve *V. campbellii* and *V. rotiferianus*, and therefore it also should not be recommended for discriminating closely related species.

Recent taxonomic studies sustain the contribution of MLSA data to the description of novel taxa (Le Roux *et al.*, 2005; Urbanczyk *et al.*, 2007; Rameshkumar *et al.*, 2008). Le Roux *et al.* (2005) described *Vibrio gigantis* on the basis of individual gene trees, comparing the four strains of the species with type strains of the *Vibrio splendidus* clade. This clade contains eight species that are related genomically and phylogenetically as closely as the core group. Of the four protein-encoding genes used, *gyrB* was the least resolving, while the others, *rpoD*, *rctB* and

toxR, gave the best discrimination between species. These results agree completely with our own results for the core group.

Urbanczyk *et al.* (2007) proposed a new genus, *Aliivibrio*, to accommodate *Vibrio fischeri* and closely related species. They sequenced several genes (*recA*, *rpoA*, *pyrH*, *gyrB* and 16S rRNA), and the concatenated sequence was analysed phylogenetically, supporting the proposal of a novel taxon. Rameshkumar *et al.* (2008) reported the phylogenetic analysis of the concatenated sequence of four genes, the 16S rRNA gene, *rpoA*, *recA* and *pyrH*, in the description of a novel species, *Vibrio porteresiae*. However, in these two studies, the role of MLSA is rather superfluous, since the phylogenetic inferences they present can be sustained by 16S rRNA gene sequences alone and, in any case, none of them serves as an example of discrimination between closely related species.

DNA–DNA hybridization

The eleven strains used for DNA–DNA hybridization experiments included the type strains of *V. harveyi*, *V. campbellii* and *V. rotiferianus*, along with additional representative strains of each subclade formed in the concatenated analysis (Table 2). Intraspecific DNA–DNA reassociation was always above the 70 % threshold value, with *V. harveyi* showing the highest values (96–100 %) and *V. campbellii* the lowest (71–80 %). These results are in agreement with the lowest intraspecific genetic diversity of the species *V. harveyi* obtained in our study and noted previously by Thompson *et al.* (2007). Interspecies reassociation values were always above 45 %, reaching values very close to 70 % in the case of some environmental strains of *V. campbellii* and *V. harveyi* with respect to the type strain of *V. rotiferianus*. Such high DNA–DNA reassociation values have been reported already for the pair *V. harveyi*–*V. campbellii* by Reichelt *et al.* (1976) and Brenner *et al.* (1983) and for *V. rotiferianus* with the other two species by Gomez-Gil *et al.* (2003, 2004). Therefore, within the core group of *Vibrio*, the species *V. harveyi*, *V. campbellii* and *V. rotiferianus* form a subgroup of more recent speciation, making species identification difficult even by complete genome comparison.

Correlation of DNA–DNA hybridization values with MLSA data

In order to validate the usefulness of MLSA for the resolution of the *Vibrio* species studied, we have calculated the correlation of the MLSA data against DNA–DNA hybridization. In addition to the DNA–DNA hybridization data obtained in this study (Table 2), we have also used data retrieved from the literature (Supplementary Table S2; Reichelt *et al.*, 1976; Gomez-Gil *et al.*, 2003, 2004), provided that they corresponded to strains used in our MLSA study or known equivalents.

The linear Pearson's product–moment correlation coefficient was calculated between the DNA–DNA relatedness

Fig. 3. Phylogenetic reconstruction based on concatenated 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR* gene sequences. Analysis was done using the neighbour-joining method. Bar, 0.05 expected nucleotide substitutions per site. Only bootstrap values above 75 % are shown (1000 resamplings) at branching points.

matrix and the corresponding similarity matrices of the seven individual genes, the concatenation of the seven genes and the concatenation of the three most resolving genes (*rpoD*, *rctB* and *toxR*). Each comparison correlated 62 pairs of values. The coefficients obtained were as low as 0.46 for the 16S rRNA gene, followed by 0.62 for *recA*, 0.65 for *rpoD*, 0.68 for *pyrH*, 0.69 for *gyrB*, 0.74 for *toxR* and 0.81 for *rctB*. The corresponding value for the seven concatenated genes increased to 0.84, which is the best value obtained, while the concatenation of the three most discriminative genes gave a figure of 0.79. Therefore, the only good correlation values are those for *rctB* and the seven concatenated genes, which are above 0.80. Adékambi *et al.* (2008) found similar values when they examined the correlation of DNA–DNA hybridization values with *rpoB*

gene sequence similarities. These authors stressed the difficulties found in correlating DNA–DNA hybridization values and sequence similarities, mainly resulting from the multiple factors that affect hybridization experiments. In our study, another reason for these low correlation values may be the inclusion of strains with a borderline position for species definition according to DNA–DNA reassociation values.

We also represented DNA–DNA hybridization values against similarities of concatenated sequences of the three most resolving genes (*rpoD*, *rctB* and *toxR*) (Fig. 4). Two threshold values can be defined: above 90.3 % sequence similarity, two strains can be considered to be members of the same species and, below 85.2 %, they can be considered to represent different species. These results are satisfactory because they allow reliable differentiation of *Vibrio* species even within the closely related species of the core group by nucleotide sequence comparisons without the need to perform DNA–DNA hybridization and make interlaboratory data comparison easier.

Conclusions

MLSA has been widely demonstrated to be a good substitute for DNA–DNA hybridization in studies of the *Vibrionaceae*. However, as discussed above, the choice of the protein-encoding genes is of great importance in this method, since not all genes to which it has been applied so far are really useful, particularly if the strains being analysed belong to tightly related species. This study also shows the importance of including a reasonable number of strains for each species, and not only type strains, so that intraspecies variability can be measured and compared with the interspecies variability. Failure to do so can result in an overestimation of the discriminatory power of the genes employed. Finally, if, in addition to phylogenetic analysis, researchers intend to propose MLSA as a replacement for DNA–DNA hybridization assays, it is very important to validate the method by comparing both. To the best of our knowledge, this is the first report of this kind employing *Vibrio* species.

Our study demonstrates that, of six protein-encoding genes, only two – *toxR* and *rpoD* – served the purpose of reliable species identification based on sequence similarity. Of the other genes, *recA*, *pyrH* and *gyrB* gave poor resolution and, in some aspects, they did not represent a real advantage over the 16S rRNA gene.

For the aim of phylogenetic inference, three genes – *rpoD*, *rctB* and *toxR* – succeeded in presenting all species as monophyletic groups but offered different interpretations about their relative order of evolution. The concatenated sequences of these three genes harmonized and stabilized the resulting topologies (Supplementary Fig. S4) to the same extent as the concatenated sequences of all seven genes (Fig. 3 and Supplementary Fig. S3). This latter concatenation gave the best correlation (0.84) between

Fig. 4. Relationship between DNA–DNA reassociation values and sequence similarities of concatenated *rpoD*, *rctB* and *toxR* genes ($n=62$).

sequence similarities and DNA–DNA hybridization values. Nevertheless, its taxonomic resolution in terms of the gap between intraspecific and interspecific sequence similarities was not as good as that of the concatenated sequences of *rpoD*, *rctB* and *toxR* (Fig. 2).

Therefore, and depending on the purpose, future studies on *Vibrio* species can be better designed, minimizing the number of genes that need to be studied by taking into account these conclusions.

ACKNOWLEDGEMENTS

Thanks are due to Dr Bruno Gómez-Gil for providing some strains. This work was supported by the Spanish Ministerio de Educación y Ciencia (projects AGL-2002-0475-C02 to E. G. and CGL-2005-02292 to M. J. P.). J. P. is the recipient of a PhD fellowship from the Spanish Ministry of Education and Science. D.R.A. and M.C.M. are the recipients of contracts granted by the Ramón y Cajal and ‘support personnel’ programmes, respectively (Spanish Ministry of Education and Science).

REFERENCES

- Adékambi, T., Shinnick, T. M., Raoult, D. & Drancourt, M. (2008). Complete *rpoB* gene sequencing as a suitable supplement to DNA–DNA hybridization for bacterial species and genus delineation. *Int J Syst Evol Microbiol* **58**, 1807–1814.
- Brenner, D. J., Fanning, G. R., Hickman-Brenner, F. W., Steigerwalt, A. G., Davis, B. R. & Farmer, J. J., III (1983). DNA relatedness among Vibrionaceae, with emphasis on the *Vibrio* species associated with human infection. *Colloq INSERM* **114**, 175–184.
- Dorsch, M., Lane, D. & Stackebrandt, E. (1992). Towards a phylogeny of the genus *Vibrio* based on 16S rRNA sequences. *Int J Syst Bacteriol* **42**, 58–63.
- Goarant, C., Reynaud, Y., Ansquer, D., Decker, S. D., Saulnier, D. & Roux, F. I. (2006). Molecular epidemiology of *Vibrio nigrripulchritudo*, a pathogen of cultured penaeid shrimp (*Litopenaeus stylirostris*) in New Caledonia. *Syst Appl Microbiol* **29**, 570–580.
- Gomez-Gil, B., Thompson, F. L., Thompson, C. C. & Swings, J. (2003). *Vibrio rotiferianus* sp. nov., isolated from cultures of the rotifer *Brachionus plicatilis*. *Int J Syst Evol Microbiol* **53**, 239–243.
- Gomez-Gil, B., Soto-Rodríguez, S., García-Gasca, A., Roque, A., Vazquez-Juarez, R., Thompson, F. L. & Swings, J. (2004). Molecular identification of *Vibrio harveyi*-related isolates associated with diseased aquatic organisms. *Microbiology* **150**, 1769–1777.
- Le Roux, F., Goubet, A., Thompson, F. L., Faury, N., Gay, M., Swings, J. & Saulnier, D. (2005). *Vibrio gigantis* sp. nov., isolated from the haemolymph of cultured oysters (*Crassostrea gigas*). *Int J Syst Evol Microbiol* **55**, 2251–2255.
- Martens, M., Delaere, M., Coopman, R., De Vos, P., Gillis, M. & Willems, A. (2007). Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* **57**, 489–503.
- Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P. & Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *Int J Syst Evol Microbiol* **58**, 200–214.
- Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.
- Mignard, S. & Flandrois, J. P. (2008). A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. *Int J Syst Evol Microbiol* **58**, 1432–1441.
- Naser, S. M., Hagen, K. E., Vancanneyt, M., Cleenwerck, I., Swings, J. & Tompkins, T. A. (2006). *Lactobacillus suntoryeus* Cachat and Priest 2005 is a later synonym of *Lactobacillus helveticus* (Orla-Jensen 1919) Bergey et al. 1925 (Approved Lists 1980). *Int J Syst Evol Microbiol* **56**, 355–360.
- Pitcher, D. G., Saunders, N. A. & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* **8**, 151–156.
- Posada, D. & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst Biol* **53**, 793–808.
- Posada, D. & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Pujalte, M. J., Sitjà-Bobadilla, A., Macián, M. C., Belloch, C., Álvarez-Pellitero, P., Pérez-Sánchez, J., Uruburu, F. & Garay, E. (2003). Virulence and molecular typing of *Vibrio harveyi* strains isolated from cultured dentex, gilthead sea bream and European sea bass. *Syst Appl Microbiol* **26**, 284–292.
- Rameshkumar, N., Fukui, Y., Sawabe, T. & Nair, S. (2008). *Vibrio porteresiae* sp. nov., a diazotrophic bacterium isolated from a

- mangrove-associated wild rice (*Porteresia coarctata* Tateoka). *Int J Syst Evol Microbiol* **58**, 1608–1615.
- Reichelt, J. L., Baumann, P. & Baumann, L. (1976). Study of genetic relationships among marine species of the genera *Beneckeia* and *Photobacterium* by means of in vitro DNA/DNA hybridization. *Arch Microbiol* **110**, 101–120.
- Richter, D., Postic, D., Sertour, N., Livey, I., Matuschka, F. R. & Baranton, G. (2006). Delineation of *Borrelia burgdorferi sensu lato* species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp. nov. *Int J Syst Evol Microbiol* **56**, 873–881.
- Sawabe, T., Kita-Tsukamoto, K. & Thompson, F. L. (2007). Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. *J Bacteriol* **189**, 7932–7936.
- 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.
- Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), version 4. Sunderland, MA: Sinauer Associates.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., Munn, C. B. & Swings, J. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl Environ Microbiol* **71**, 5107–5115.
- Thompson, F. L., Gomez-Gil, B., Ribeiro-Vasconcelos, A. T. & Sawabe, T. (2007). Multilocus sequence analysis reveals that *Vibrio harveyi* and *Vibrio campbellii* form distinct species. *Appl Environ Microbiol* **73**, 4279–4285.
- Thompson, C. C., Thompson, F. L. & Vicente, A. C. P. (2008). Identification of *Vibrio cholerae* and *Vibrio mimicus* by multilocus sequence analysis (MLSA). *Int J Syst Evol Microbiol* **58**, 617–621.
- Urbanczyk, H., Ast, J. C., Higgins, M. J., Carson, J. & Dunlap, P. V. (2007). Reclassification of *Vibrio fischeri*, *Vibrio logei*, *Vibrio salmonicida* and *Vibrio wodanis* as *Aliivibrio fischeri* gen. nov., comb. nov., *Aliivibrio logei* comb. nov., *Aliivibrio salmonicida* comb. nov. and *Aliivibrio wodanis* comb. nov. *Int J Syst Evol Microbiol* **57**, 2823–2829.
- Young, J. M. & Park, D. C. (2007). Probable synonymy of the nitrogen-fixing genus *Azotobacter* and the genus *Pseudomonas*. *Int J Syst Evol Microbiol* **57**, 2894–2901.
- Ziemke, F., Höfle, M. G., Lalucat, J. & Rosselló-Mora, R. (1998). Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. *Int J Syst Bacteriol* **48**, 179–186.