Taxonomic revision of Harveyi clade bacteria (family Vibrionaceae) based on analysis of whole genome sequences

Henryk Urbanczyk,1 Yoshitoshi Ogura2,3 and Tetsuya Hayashi2,3

1Interdisciplinary Research Organization, University of Miyazaki, Miyazaki, Japan
2Division of Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, Miyazaki, Japan
3Division of Microbiology, Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

Use of inadequate methods for classification of bacteria in the so-called Harveyi clade (family Vibrionaceae, Gammaproteobacteria) has led to incorrect assignment of strains and proliferation of synonymous species. In order to resolve taxonomic ambiguities within the Harveyi clade and to test usefulness of whole genome sequence data for classification of Vibrionaceae, draft genome sequences of 12 strains were determined and analysed. The sequencing included type strains of seven species: Vibrio sagamiensis NBRC 104589T, Vibrio azureus NBRC 104587T, Vibrio harveyi NBRC 15634T, Vibrio rotiferianus LMG 21460T, Vibrio campbellii NBRC 15631T, Vibrio jasicida LMG 25398T, and Vibrio owensii LMG 25443T. Draft genome sequences of strain LMG 25430, previously designated the type strain of Vibrio beijerinckii, and two strains (MWB 21 and 090810c) from the ‘beijerinckii’ lineage were also determined. Whole genomes of two additional strains (ATCC 25919 and 200612B) that previously could not be assigned to any Harveyi clade species were also sequenced. Analysis of the genome sequence data revealed a clear case of synonymy between V. owensii and V. communis, confirming an earlier proposal to synonymize both species. Both strains from the ‘beijerinckii’ lineage were classified as V. jasicida, while the strains ATCC 25919 and 200612B were classified as V. owensii and V. campbellii, respectively. We also found that two strains, AND4 and Ex25, are closely related to Harveyi clade bacteria, but could not be assigned to any species of the family Vibrionaceae. The use of whole genome sequence data for the taxonomic classification of the Harveyi clade bacteria and other members of the family Vibrionaceae is also discussed.

INTRODUCTION

Vibrio harveyi (Vibrionaceae, Gammaproteobacteria) and related species are colloquially referred to as the 'Harveyi clade' (Sawabe et al., 2007). There is no clear consensus on which species comprise the Harveyi clade, and up to 11 species have been included in the clade: Vibrio rotiferianus, Vibrio parahaemolyticus, Vibrio harveyi, Vibrio azureus, Vibrio sagamiensis, Vibrio campbellii, Vibrio owensii, Vibrio jasicida, Vibrio alginolyticus, Vibrio natriegens and Vibrio mytilli (Sawabe et al., 2007; Cano-Gomez et al., 2011; Hoffman et al., 2012; Yoshizawa et al., 2012). Other species related to V. harveyi have also been proposed, but they were either not verified, or were found to be synonyms of previously described species. Examples of these species include Vibrio carchariae (Grimes et al., 1984) and Vibrio trachuri (Iwamoto et al., 1995), which were found to be later heterotypic synonyms of V. harveyi (Pedersen et al., 1998; Thompson et al., 2002). Recently, two novel species in the clade were proposed, V. owensii (Cano-Gomez et al., 2010) and [Vibrio communis] (Chimetto et al., 2011). However, a thorough comparison of strains from both species revealed significant similarities between them, which led Cano-Gomez et al., (2011) to propose that [V. communis] should be recognized as a later heterotypic synonym of V. owensii.
Use of multi-locus sequence analysis (MLSA) helped to resolve some controversies surrounding the taxonomy of the Harveyi clade (Gomez-Gil et al., 2004; Sawabe et al., 2007; Thompson et al., 2007; Cano-Gomez et al., 2011; Hoffman et al., 2011). MLSA also revealed significant, previously unrecognized diversity within the clade. For example, Figge et al., (2011) used MLSA to reveal substantial phylogenetic diversity within the Harveyi clade, and described so-called ‘beijerinckii’ lineage. The ‘beijerinckii’ lineage encompasses marine luminous bacteria that were phylogenetically distinct from other species of the family Vibrionaceae, but were apparently closely related to Harveyi clade bacteria (Figge et al., 2011).

Robust and reliable taxonomy of the Harveyi clade is of great importance. Bacteria in the clade represent some of the major pathogens of marine shrimp, fish and molluscs (Thompson et al., 2004; Austin & Zhang, 2006), V. parahaemolyticus is an important human pathogen (Thompson et al., 2004) and V. owensii was linked to coral diseases (Ushijima et al., 2012). In addition to pathogenicity, the bacteria are widely used as model organisms in microbiology. Studies of Harveyi clade bacteria can be traced to the early days of microbiology (Figge et al., 2011; Dunlap & Urbanczyk, 2013), and they are used as models in studies of bacterial luminescence (Lin & Bassler, 2004), biofilm formation (Yildiz & Visick, 2009) and multi-chromosomal genome organization (Okada et al., 2005; Kirkup et al., 2010). However, the use of Harveyi clade bacteria as model organisms is hampered by difficulties in assigning model strains to species. For example, strain ATCC BAA-1116 that was used as a model for studies of V. harveyi quorum sensing mechanism for more than 15 years, was recently revealed to represent V. campbellii (Lin et al., 2010).

Difficulties in assigning bacterial strains to species and frequent taxonomic revisions are not unique to the Harveyi clade or to the family Vibrionaceae. In search for new and reliable methodologies, modern bacterial taxonomy has been strongly influenced by the developments in genomics. Low costs of whole genome sequencing makes it now possible to routinely use genome sequence data during taxonomic classification of bacteria. One of the most promising methods that uses bacterial whole genome sequences for taxonomic classification is the average nucleotide identity (ANI) of conserved genes present in two strains (Konstantinidis & Tiedje, 2005; Richter & Rosselló-Móra, 2009). ANI provides a robust measurement of genetic distance between two bacterial genomes and was shown to have a very good correlation with the results of DNA–DNA hybridization (DDH) (Konstantinidis & Tiedje, 2005; Goris et al., 2007). ANI was suggested as a suitable replacement for DDH, with an ANI of 95–96% corresponding to approximately 70% DNA–DNA reassociation value (Goris et al., 2007; Richter & Rosselló-Móra, 2009).

To resolve ambiguities in the Harveyi clade taxonomy and to test usefulness of whole genome sequence data for the taxonomic studies of the family Vibrionaceae, in this study we determined and analysed whole genome sequences of 12 strains from the Harveyi clade. The sequencing included seven type strains from species in the clade, and five additional strains. Results of the study were used to discuss the use of whole genome sequence data for taxonomic classification of Harveyi clade bacteria and other members of the family Vibrionaceae.

METHODS

Strains and culture conditions. Strains NBRC 104589T, NBRC 104587T, NBRC 15634T and NBRC 15631T were obtained from the Biological Resource Center (NBRC), while strains LMG 21460T, LMG 25398T, LMG 25443T and LMG 25430T were obtained from the Belgian Co-Ordinated Collection of Microorganisms (BCCM/LMG). Strain ATCC 25919 was obtained from the American type Culture Collection (ATCC). Strain MWB 21 was a part of a collection of bacteria isolated by Martinus W. Beijerinck in the 1920s and stored at the Netherland Culture Collection of Bacteria (NCCB); see Figge et al., (2011) for details of the strain recovery. Strains 090810c and 200612B were isolated from seawater samples collected off the coast of Miyazaki (Japan) in 2010 and 2012, respectively. Isolation was done using a method described by Dunlap & Urbanczyk (2013). All strains were grown on the LSW-70 broth (Dunlap & Urbanczyk, 2013) at 22–28 °C.

Whole genome sequences. Whole genome sequence data from strains in the Harveyi clade were also obtained from GenBank. Initially, whole genome sequence data were downloaded for 21 strains identified in GenBank as members of species in the Harveyi clade, and for 13 strains identified as members of the family Vibrionaceae but not assigned to any species. In order to assess if the strains could be classified as members of the Harveyi clade, sequences of housekeeping genes of these strains (including gyrB, dapA, topA and recA) were used for phylogenetic analyses in the context of type strains of 12 species in the Harveyi clade (data not shown). Furthermore, ANI between whole genome sequence data of these strains and seven type strains for species in the Harveyi clade were also calculated (data not shown). Based on results of these analyses, 22 strains were recognized as closely related to bacteria in the Harveyi clade and used in subsequent analyses (see Table S1, available in IJSEM Online, for a list of strains used and GenBank accession numbers for the genome sequence data).

It should be noted here that phylogenetic analyses of sequences of housekeeping genes of strain 16, deposited to GenBank as V. parahaemolyticus (GenBank BioProject ID: 54741), showed closer evolutionary relationship to Vibrio orientalis than to any species in the Harveyi clade. Also, the strain had less than 74.64 % ANI to any strain in the Harveyi clade, including eight strains identified as V. parahaemolyticus (data not shown). These results suggested that strain 16 was incorrectly classified as V. parahaemolyticus and that the strain is only distantly related to Harveyi clade bacteria. As a result, genome sequence data of strain 16 were not used in this work.

Whole genome sequencing and ANI. Genomic DNA was isolated using a DNeasy Blood and Tissue kit (Qiagen) following the manufacturer’s protocol for Gram-negative bacteria. Draft genome sequences were generated using the Illumina MiSeq platform. Libraries of genomic DNA were prepared using a Nextera XT DNA Sample Prep kit (Illumina) and pooled libraries were subjected to multiplexed paired-end sequencing (251 cycles × 2). Sequencing reads were assembled using Velvet version 1.2.05 (Zerbino & Birney, 2008). The draft genome sequences have been submitted to the DNA Data Bank of Japan (DDBJ) (see Table S1 for accession numbers). ANI was calculated using whole available genome sequence data for
each strain, using the JSpecies program version 1.2.1 (Richter & Rosselló-Móra, 2009) using default settings for ANIb, i.e. BLAST options: -X 150 -q -1 -F f -e 1-15 -a 2; fragments length: 1020; fragments alignment length: 70% or longer; and fragments identity: 30% or higher.

**Data handling and phylogenetic analyses.** In order to obtain sequences of genes shared by all Harveyi clade strains used in this study, the following procedure was used. Four strains from the clade with fully assembled and annotated genomes (only draft genome sequence data were available for the other Harveyi clade bacteria) were obtained from GenBank. These included *V. campbellii* ATCC BAA-1116, *Vibrio* sp. EJY3, *V. paraaerumolyticus* RIMD 2210633 and *V. parahaemolyticus* BB22OP. Nucleotide sequences of protein-coding genes of the four strains were clustered together using the BLASTCLUST algorithm, part of the BLAST 2.2.26 package. Sequences in each cluster had at least 80% sequence identity over 90% of each sequence length. The analysis found 472 clusters that included one sequence from each of the four strains. Sequences of *V. campbellii* ATCC BAA-1116 from the 472 clusters were used as queries in MEGABLAST searches (with 80% sequence identity cut-off) of databases containing genome sequence data of all analysed strains in the Harveyi clade. Results of these searches were excerpted using a custom PERL script and manually aligned based on inferred amino acid sequences using MacClade 4.08. From the initial 472 genes selected after BLASTCLUST clustering, we eliminated those that had no homologues in one or more of the analysed 34 Harveyi strains. During alignment, we also removed genes that had a significant portion of the gene sequence missing in one or more strains. Alignments were truncated to the shortest sequence, so that each gene sequence had the same length. Alignments shorter than 250 characters were not used in the subsequent phylogenetic analyses. In the end, complete or nearly complete sequences of 138 genes present in all analysed Harveyi clade strains were used in phylogenetic analyses (see Table S2 for a list of the gene sequences used). Alignment of the 138 gene sequences is available on request.

16S rRNA gene sequences were obtained from GenBank (see Table S3 for a list of accession numbers and loci numbers); for strains *svers*.1.1, 090810c and 200612B, the 16S rRNA genes were PCR amplified as described by Ast et al. (2007) and sequenced using an ABI Prism 3100 Genetic Analyzer (Applied Biosystems) with an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems). 16S rRNA gene sequences were aligned manually in MacClade 4.08.

TNT version 1.1 (Goloboff et al., 2008) was used for analyses using the parsimony criterion. The analyses were done using 1000 replicates of new technology searches; the most parsimonious tree was analysed with 50 000 replicates of the parsimony ratchet; jackknife resampling values are based on 10 000 replicates with 34% chance of individual character deletion per replicate. For maximum-likelihood analysis, PAUP* version 4.08b10 for UNIX was used (Swoford, 2003). Data were analysed using 319 replicates of heuristic search with the likelihood criterion; jackknife resampling values are based on 100 replicates with 34% chance of individual character deletion per replicate. Dendrograms were obtained in PAUP*, using UPGMA. Trees were visualized in FigTree version 1.3.1. Neighbour-net networks were produced using SplitsTree4 version 4.12.8 (Huson & Bryant, 2006), using default settings.

**RESULTS**

**Whole genome sequencing**

Initially, type strains of seven species in the Harveyi clade were selected for whole genome sequencing: *V. sagarniensis* NBRC 104589T, *V. azureus* NBRC 104587T, *V. harveyi* NBRC 15634T, *V. rotiferianus* LMG 21460T, *V. campbellii* NBRC 15631T, *V. jasicida* LMG 25398T and *V. owensii* LMG 25443T. These strains provided a good representation of species in the Harveyi clade.

In addition, strain LMG 25430, previously designated the type strain of *V. communis*, was included in this study to assess the proposed reclassification of *V. communis* as *V. owensii* (Cano-Gomez et al., 2011). Two strains (MWB 21 and 090810c) from the ‘beijerinckii’ lineage (Figgie et al., 2011) were also sequenced in order to test their relationship to other species in the family *Vibrionaceae*. Strain MWB 21 was classified as a member of the ‘beijerinckii’ lineage based on phylogenetic analyses of nucleotide sequences of six unlinked genes (Figgie et al., 2011). Strain 090810c was predicted to be a member of the ‘beijerinckii’ lineage based on phylogenetic analyses of nucleotide sequences of three genes (16S rRNA, *gyrB* and *luxA*) (data not shown).

In order to test the usefulness of whole genome sequence data for classification of bacteria in the Harveyi clade, two additional strains were selected for whole genome sequencing. Strain ATCC 25919 was previously classified as ‘*Benekea neptuna*’ primarily based on results of biochemical tests (Baumann et al., 1971). Recently, the strain was included in multi-locus sequence analyses using sequences of six genes (Figgie et al., 2011). Based on results of the MLSA, strain ATCC 25919 was shown to be related to bacteria in the Harveyi clade, but the strain could not be confidently assigned to any species. Strain 200612B was found to be closely related to ATCC 25919 based on phylogenetic analyses of sequences of the 16S rRNA, *gyrB* and *luxA* genes (data not shown), but also could not be confidently assigned to any species.

Genome sequencing using the Illumina technology resulted in nearly complete, high quality draft genome sequences (see Table 1 for details). Predicted genome sizes were between 4.56 Mb and 6.37 Mb and DNA G+C contents of the genomes were between 42.88 and 46.44%.

**Average Nucleotide Identity**

Calculation of the ANI was performed between draft genome sequences of 12 strains from the Harveyi clade obtained in this study, 22 whole genome sequences available in GenBank for strains in the Harveyi clade, and four whole genome sequences of representative members of the family *Vibrionaceae* available in public databases. These included *Aliivibrio fischeri* ES114 (representative strain from the type species of the genus *Aliivibrio*; Ruby et al., 2005), *Vibrio cholerae* N16961 (representative strain from the type species of the genus *Vibrio*; Heidelberg et al., 2000) and ‘*Photobacterium mandapamensis*’ *svers*.1.1 (representative strain from the genus *Photobacterium*) (Urbanczyk et al., 2011b). Strain *svers*.1.1 was chosen for the analysis because among species of the genus *Photobacterium* with a
Table 1. Genome sequencing statistics.

<table>
<thead>
<tr>
<th>Strain</th>
<th>GenBank accession number</th>
<th>Coverage</th>
<th>Max. contig length</th>
<th>No. contigs</th>
<th>DNA G+C content (mol%)</th>
<th>GenBank accession number</th>
<th>Coverage</th>
<th>N50 statistic Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAOB01000001–BAOB01000783</td>
<td>BA01000001–BA0100000138</td>
<td>2637</td>
<td>25321</td>
<td>48</td>
<td>43.5</td>
<td>NBRC 104587T</td>
<td>91.32</td>
<td>25321</td>
</tr>
<tr>
<td>BAOF01000001–BAOF01000387</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>44.94</td>
<td>V. campbellii</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>200612B*</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.31</td>
<td>V. campbellii</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>BANY01000001–BANY01000229</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.55</td>
<td>V. harveyi</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>BAOD01000001–BAOD01000245</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.55</td>
<td>V. harveyi</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>V. jasicida</td>
<td>BAOG01000001–BAOG01000278</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.55</td>
<td>V. jasicida</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>MV 21</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.55</td>
<td>V. Jasicida</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>V. jasicida</td>
<td>BAOG01000001–BAOG01000278</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.46</td>
<td>V. jasicida</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>V. jasicida</td>
<td>BAOG01000001–BAOG01000278</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.46</td>
<td>V. jasicida</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>090810c*</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.83</td>
<td>V. jasicida</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>BALZ01000001–BANZ01000174</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.83</td>
<td>V. owensii</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>BAOE01000001–BAOE01000341</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.83</td>
<td>V. parahaemolyticus</td>
<td>91.32</td>
<td>17534</td>
</tr>
<tr>
<td>BAOI01000001–BAOI01000318</td>
<td>BA01000001–BA0100000138</td>
<td>324</td>
<td>17534</td>
<td>30</td>
<td>45.83</td>
<td>V. parahaemolyticus</td>
<td>91.32</td>
<td>17534</td>
</tr>
</tbody>
</table>

*Strain isolated from seawater samples collected off the coast of Miyazaki, Japan.
§Strain previously identified as ‘Beneckea neptuna’ (Baumann et al., 2011).
\[Strain previously classified as Vibrio campbellii.\]
\[Strain previously identified as ‘beijerinckii’ lineage, MWB 21 and 090810c, had ANI of 97.63% or higher with V. jasicida LMG 25398T, indicating that the two strains should be reclassified as V. jasicida.\]
\[The whole genome sequence available, ‘P. mandapamensis’ is most closely related to the type species of the genus, Photobacterium phosphoreum (Urbanczyk et al., 2011a). An additional representative strain from the genus Vibrio, Vibrio vulnificus CMCP6, was also included in the analysis (Kim et al., 2011). All results of the ANI calculations are presented in Table S4.\]

Three strains, 40B, 12G01 and 712i, showed ANI values of 98.28% or higher, indicating that the three strains are members of the same species. Strains 40B and 12G01 are classified in GenBank as V. alginolyticus, while strain 712i was not assigned to any species of the family Vibrionaceae. For the purpose of this work, the three strains were provisionally classified as V. alginolyticus. Four strains showed ANI higher than 95% when compared with the draft genome sequence of V. campbellii NBRC 15637T. Two of the strains, ATCC BAA-1116 and HY01, were submitted to GenBank as V. harveyi, but both strains have ANI of 87.85% or lower with the type strain of V. harveyi. Based on these results, strains ATCC BAA-1116 and HY01 were reclassified for the purpose of this work as members of V. campbellii. This reclassification is in agreement with an earlier proposal to reclassify both strains as V. campbellii (Lin et al., 2010). ANI values of 96.04% or higher were also found between the whole genome sequences of V. campbellii NBRC 15637T and strain 200612B, which suggested that strain 200612B should also be classified as V. campbellii. Both analysed strains from the ‘beijerinckii’ lineage, MWB 21 and 090810c, had ANI of 97.63% or higher with V. jasicida LMG 25398T, indicating that the two strains should be reclassified as V. jasicida. Analysis of the whole genome sequence of strain EJY3 revealed ANI below the 95–96% cut-off to any of the analysed strains. However, BLASTN analyses of sequences of housekeeping genes of this strain matched those of the type strain of V. natriegens NBRC 15636T (data not shown). Pending comparison of EJY3 to the type strain of V. natriegens, for the purpose of this work EJY3 was classified as V. natriegens. The whole genome sequences of V. owensii LMG 25443T, LMG 25430 and ATCC 25919 showed ANI of 96.3% or higher, which indicates that the three strains should be classified as members of the same species. Eight analysed V. parahaemolyticus strains showed high average nucleotide identities, having ANI of 98.21% or higher. Two strains, AND4 and Ex25, could not be assigned to any species based on ANI.

In order to assess the correlation between the results of ANI and DDH, DNA–DNA reassociation values between the type strains of V. jasicida, V. campbellii, V. harveyi, V. rotiferianus, V. azureus, V. sagamiensis and V. owensii, and strain V. owensii LMG 25430 available in the literature (Gomez-Gil et al., 2003; Yoshizawa et al., 2009; Cano-Gómez et al., 2010; Yoshizawa et al., 2010; Yoshizawa et al., 2012) were compared to ANI calculated in this study. DNA–DNA reassociation values between each pair of type strains (see Table S5 for the reassociation values) were plotted against corresponding ANI values. The resulting

http://jfs.sgmjournals.org
processes neighbour-net analysis was used to evaluate phylogenetic signals resulting from reticulate evolutionary relationships among the 34 Harveyi clade strains (Fig. 4). The resulting network showed similar relationships between the analysed taxa as the phylogenetic analysis shown in Fig. 2, indicating that recombination did not obscure phylogenetic signals in the analysed sequence data. Importantly, all Harveyi strains identified as members of the same species using ANI were found to be closely related based on the results of neighbour-net analysis.

It should be noted that none of the draft genome sequences obtained in this study contained a complete sequence of the 16S rRNA gene. Bacteria in the family Vibrionaceae usually have a high number of rRNA operons in their genomes, up to 15 in the Photobacterium profundum SS9 genome (Vezzi et al., 2005). The rRNA operon sequences are nearly identical at all the loci, and assembly of multiple rRNA operons was not possible using the data obtained in this study. In order to conduct a phylogenetic analysis based on 16S rRNA gene data, nearly complete sequences of the 16S rRNA genes of strains 090810c, 200612B and ‘P. mandapamensis’ svers.1.1 were PCR amplified and sequenced. Other 16S rRNA gene sequences for most of the analysed strains were obtained from GenBank for a total of 31 taxa (see Table S3 for a list of accession numbers). Hypothesis of evolutionary relationship resulting from the analysis of 16S rRNA gene sequences could not resolve all species in the Harveyi clade and had very poor resampling support (Fig. S3). Differences in the results of 16S rRNA gene analysis and analyses based on concatenated alignment of 138 gene sequences could be attributed to insufficient phylogenetic signal in the 16S rRNA gene data to resolve relationships between analysed taxa. The alignment was 1025 characters long, but only 56 characters were informative, which is not sufficient to resolve the relationship between all analysed taxa.

When the concatenated alignment of 138 gene sequences was analysed using maximum-likelihood and neighbour-joining algorithms, the same strains clustered together as in the analysis shown in Fig. 2 (Figs S1 and S2). Importantly, in all phylogenetic analyses, strains identified based on ANI as V. campbellii, showed a close evolutionary relationship to NBRC 15637T and other strains classified as V. campbellii based on ANI. Furthermore, both strains from the ‘beijerinckii’ lineage formed a clade with V. jasicida LMG 25398T, providing further support for reclassification of these strains as V. jasicida. Results of the phylogenetic analysis showed strains AND4 and Ex25 to be closely related to Harveyi clade bacteria, but they could not be confidently assigned to any species. Potentially, strains AND4 and Ex25 represent lineages in the Harveyi clade which were not described previously.

A potential criticism of the phylogenetic analyses presented here would be that they did not take into consideration the possibility of recombination disrupting phylogenetic signals in the sequence data used. Previous studies found that members of the family Vibrionaceae frequently undergo recombination and exchange genes horizontally (Reen et al., 2006; Urbanczyk et al., 2008). In order to predict phylogenetic signals resulting from reticulate evolutionary processes neighbour-net analysis was used to evaluate relationships among the 34 Harveyi clade strains (Fig. 4). The resulting network showed similar relationships between the analysed taxa as the phylogenetic analysis shown in Fig. 2, indicating that recombination did not obscure phylogenetic signals in the analysed sequence data. Importantly, all Harveyi strains identified as members of the same species using ANI were found to be closely related based on the results of neighbour-net analysis.

It should be noted that none of the draft genome sequences obtained in this study contained a complete sequence of the 16S rRNA gene. Bacteria in the family Vibrionaceae usually have a high number of rRNA operons in their genomes, up to 15 in the Photobacterium profundum SS9 genome (Vezzi et al., 2005). The rRNA operon sequences are nearly identical at all the loci, and assembly of multiple rRNA operons was not possible using the data obtained in this study. In order to conduct a phylogenetic analysis based on 16S rRNA gene data, nearly complete sequences of the 16S rRNA genes of strains 090810c, 200612B and ‘P. mandapamensis’ svers.1.1 were PCR amplified and sequenced. Other 16S rRNA gene sequences for most of the analysed strains were obtained from GenBank for a total of 31 taxa (see Table S3 for a list of accession numbers). Hypothesis of evolutionary relationship resulting from the analysis of 16S rRNA gene sequences could not resolve all species in the Harveyi clade and had very poor resampling support (Fig. S3). Differences in the results of 16S rRNA gene analysis and analyses based on concatenated alignment of 138 gene sequences could be attributed to insufficient phylogenetic signal in the 16S rRNA gene data to resolve relationships between analysed taxa. The alignment was 1025 characters long, but only 56 characters were informative, which is not sufficient to resolve the relationship between all analysed taxa.
Phylogenetic analysis and ANI data supported classification of strains ATCC 25919 and 200612B as *V. owensii* and *V. campbellii*, respectively, even though preliminary analysis using sequences of three genes (16S rRNA, *gyrB*, and *recA*) predicted a close evolutionary relationship between the two strains. Inspection of the sequence data revealed that analysis based on sequences of the *gyrB* gene resulted in strains of *V. owensii*, *V. campbellii* and *V. harveyi* forming two clades where strains from the three species were grouped together (Fig. S4). The predicted relationships between strains from other Harveyi clade species were similar to those predicted by the phylogenetic analysis based on concatenated alignment of 138 genes, or based on ANI. Incongruence between the results of the analysis based on the *gyrB* gene sequences and other phylogenetic analyses could not be attributed to insufficient phylogenetic signal, since the *gyrB* gene sequence alignment included 561 informative characters and the resulting phylogenetic hypotheses were well supported. Analysis of the *gyrB* sequences using neighbour-net revealed that uncertainties regarding the relationships between the analysed taxa could be attributed to recombination between the *gyrB* sequences of some Harveyi clade strains (Fig. S5).

**DISCUSSION**

Sequencing of whole genomes of 12 strains, including seven type strains, and analysis of the genomic data resolved some ambiguities regarding the taxonomy of the Harveyi clade bacteria, and showed the usefulness of whole genome sequence analysis for the taxonomic classification of members of the family *Vibrionaceae*. Based on the values of ANI between *Vibrio owensii* LMG 25443<sup>T</sup> and LMG 25430, which are higher than the 95–96% cut-off for bacteria belonging to the same species (Goris *et al.*, 2007; Richter & Rosselló-Móra, 2009), and based on the analyses
using sequences of 138 genes that are conserved in all 34 strains analysed (Figs 2 and 4), it was shown that the two strains should be classified as members of the same species. These results are in agreement with previous findings based on MLSA of Harveyi clade bacteria using six housekeeping genes (Cano-Gomez et al., 2011) and support the previous recommendation to reclassify strains identified as \[V.\] communis (including strain LMG 25430) as \[V.\] owensii.

This work also helped to resolve ambiguities regarding the taxonomy of bacteria from the ‘beijerincki’ lineage (Figs 2, 3 and 4), which showed close evolutionary relationship of the two ‘beijerincki’ lineage strains to \[V.\] jasicida LMG 25398\(^\top\). Other strains identified by Figge et al., (2011) as members of the ‘beijerincki’ lineage, although not analysed in this work, likely should also be reclassified as \[V.\] jasicida.

Furthermore, results of this work allowed us to assign two strains from the Harveyi clade to recognized species of the genus Vibrio. Based on ANI values and the results of phylogenetic analysis (Fig. 2), strain ATCC 25919 was classified as \[V.\] owensii, while strain 200612B was classified as \[V.\] campbellii. It should be noted that the two strains were assigned to two distinct species, even though initial analysis based on sequences of three genes (16S rRNA, \(gyrB\) and \(recA\)) showed the strains to be closely related. The exact cause for differences between the results of the analysis based on whole genome sequences and the analysis based on three genes could not be established, but we predict that the \(gyrB\) gene of strain 200612B underwent one or more recombination events with homologous sequences from other bacteria (Figs S4 and S5), which may have altered the results of analysis based on the sequences of three genes. Recombination within the \(gyrB\) gene sequence was also possible in other analysed \[V.\] owensii, \[V.\] campbellii and \[V.\] harveyi strains. However, the exact nature of the recombination events, as well as the sources of homologous

\[Fig. 3.\] Genome relatedness between Harveyi clade bacteria. Dendrogram reconstructed using ANI between the whole genome sequences of Harveyi clade bacteria and related members of the family \textit{Vibrionaceae}. ANI data used to reconstruct the dendrogram are included in Table S4.
sequences could not be determined. The predicted recombination events apparently did not introduce nonsense mutations and the genes likely remained functional. Despite the possibility of a recombination in the \( \text{gyrB} \) gene sequence, with respect to the taxonomic classification of strain 200612B, we classify the strain as \( \text{V. campbellii} \) based on the results of ANI data and phylogenetic analyses using concatenated sequences of the 138 genes.

Analysis of the whole genome data of 34 strains from the Harveyi clade found two strains, AND4 and Ex25, which could not be confidently assigned to any species of the family \( \text{Vibrionaceae} \). The two strains show a close relationship to Harveyi clade bacteria (Figs 2, 3 and 4), but could not be assigned to previously described species in the clade. These results suggest that strains AND4 and Ex25 represent lineages that could constitute novel species of the family \( \text{Vibrionaceae} \). However, taxonomic characterization of these two lineages will require isolation and analysis of additional strains in the two lineages.

The results of this study highlight the usefulness of whole genome sequence data and ANI for taxonomic classification of Harveyi clade bacteria and other members of the family \( \text{Vibrionaceae} \). ANI between strains analysed in this study show a good correlation with results of previous estimates of relatedness using DDH (Fig. 1), similar to a linear correlation between ANI and DDH found by Goris.

Fig. 4. Neighbour-Net analysis of multi-locus sequence data. The neighbour-net dendrogram was calculated using concatenated alignment of 138 gene sequences, for a total of 142,946 characters. Boxes in the network indicate uncertainties regarding relationships between the analysed taxa, which potentially arose due to recombination and horizontal gene transfer. Bar, 1% sequence divergence.
et al. (2007). Importantly, all strains classified as members of the same species based on ANI were found to be closely related in all phylogenetic analyses conducted in this study (Fig. 2). Based on the results of this work and previous findings regarding the use of ANI in bacterial taxonomy (Konstantinidis & Tiedje, 2005; Goris et al., 2007; Richter & Rosselló-Móra, 2009), we find ANI to be a very useful tool for taxonomic classification of Harveyi clade bacteria and other members of the family Vibrionaceae. We recommend that, when applicable, future taxonomic studies of the family attempt to determine whole genome sequences of all relevant isolates, especially the type strains of all analysed species, and calculate ANI between relevant strains. We understand that not all taxonomic studies of bacteria in the family require the use of whole genome sequence data, but accumulation of whole genome sequences for all Vibrionaceae type strains and other relevant isolates in public databases will allow the inclusion of ANI and other analyses based on genomic data in increasing numbers of taxonomic studies of the family Vibrionaceae.

ANI has numerous advantages when used for bacterial taxonomy (Richter & Rosselló-Móra, 2009), including the ease of simultaneous examination of large number of strains. Unlike the DDH method, which is impractical and prohibitively expensive for more than a few strain combinations, ANI can be calculated for practically any number of whole genome sequences. The family Vibrionaceae consists of over 120 recognized species and exhaustive taxonomic studies of the whole family requires inclusion of representative strains from a large number of species, which currently can only be done using whole genome sequence data. ANI can also be used for classification of uncultured members of the family Vibrionaceae, such as obligate light organ symbionts of some deep-sea fishes (Haygood & Distel, 1993; Hendry & Dunlap, 2011). These bacteria cannot currently be cultured under laboratory conditions and their taxonomic characterization in the context of other members of the family Vibrionaceae will require the use of whole genome sequence data. Furthermore, DDH experiments require specialist skills and equipment to produce high quality data, which can be an obstacle for scientists who are not primarily focused on bacterial taxonomy research. In contrast, access to genome sequencing technologies and genomics knowhow is more widespread than to DDH, and adding ANI to the repertoire of methods used in studies of Vibrionaceae taxonomy will allow more researchers to contribute to the taxonomy of this family.

The phylogenetic analysis reported here also highlights the importance of using numerous independent loci during MLSA of members of the family Vibrionaceae. Initially, we incorrectly predicted that strain 200612B was closely related to V. owensii ATCC 25919 based on limited analysis of the sequences of three genes, one of which apparently underwent recombination. However, a more thorough examination, including a genome-wide MLSA, revealed strain 200612B as a member of V. campbellii. We also found that analysis of 16S rRNA gene sequences was insufficient to fully resolve the evolutionary relationships between species in the Harveyi clade. We recommend that further studies of the clade attempt to use numerous independent loci for MLSA, ideally after extracting the sequence information from whole genome sequence data. A list of the 138 genes conserved between the analysed 34 Harveyi clade strains identified in this study is available in Table S2. Sequences of these genes will be useful for taxonomic classification of Harveyi clade bacteria, as well as for studies of the bacterial evolutionary relationship.

ACKNOWLEDGEMENTS

We thank P. V. Dunlap for the gift of strain MWB 21. This work was supported by the Program to Disseminate Tenure Tracking System from the Japanese Ministry of Education, Culture, Sports, Science and Technology; as well as by a grant for Scientific Research on Priority Areas from the University of Miyazaki.

REFERENCES


http://ijs.sgmjournals.org