Bioinformatic Genome Comparisons for Taxonomic and Phylogenetic Assignments Using *Aeromonas* as a Test Case

Sophie M. Colston,a Matthew S. Fullmer,a Lidia Beka,a Brigitte Lamy,a,b,c J. Peter Gogarten,a Joerg Grafab

Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA; Laboratoire de Bactériologie-Virologie, UMR 5119, Equipe Pathogènes et Environnements, Université Montpellier, Montpellier, France; Laboratoire de Bactériologie, Centre Hospitalier Universitaire de Montpellier, Montpellier, France

S.M.C. and M.S.F. contributed equally to this article.

ABSTRACT Prokaryotic taxonomy is the underpinning of microbiology, as it provides a framework for the proper identification and naming of organisms. The “gold standard” of bacterial species delineation is the overall genome similarity determined by DNA-DNA hybridization (DDH), a technically rigorous yet sometimes variable method that may produce inconsistent results. Improvements in next-generation sequencing have resulted in an upsurge of bacterial genome sequences and bioinformatic tools that compare genomic data, such as average nucleotide identity (ANI), correlation of tetrancleotide frequencies, and the genome-to-genome distance calculator, or in silico DDH (isDDH). Here, we evaluate ANI and isDDH in combination with phylogenetic studies using *Aeromonas*, a taxonomically challenging genus with many described species and several strains that were reassigned to different species as a test case. We generated improved, high-quality draft genome sequences for 33 *Aeromonas* strains and combined them with 23 publicly available genomes. ANI and isDDH distances were determined and compared to phylogenies from multilocus sequence analysis of housekeeping genes, ribosomal proteins, and expanded core genes. The expanded core phylogenetic analysis suggested relationships between distant *Aeromonas* clades that were inconsistent with studies using fewer genes. ANI values of $\geq 96$% and isDDH values of $\geq 70$% consistently grouped genomes originating from strains of the same species together. Our study confirmed known misidentifications, validated the recent revisions in the nomenclature, and revealed that a number of genomes deposited in GenBank are misnamed. In addition, two strains were identified that may represent novel *Aeromonas* species.

IMPORTANCE Improvements in DNA sequencing technologies have resulted in the ability to generate large numbers of high-quality draft genomes and led to a dramatic increase in the number of publicly available genomes. This has allowed researchers to characterize microorganisms using genome data. Advantages of genome sequence-based classification include data and computing programs that can be readily shared, facilitating the standardization of taxonomic methodology and resolving conflicting identifications by providing greater uniformity in an overall analysis. Using *Aeromonas* as a test case, we compared and validated different approaches. Based on our analyses, we recommend cutoff values for distance measures for identifying species. Accurate species classification is critical not only to obviate the perpetuation of errors in public databases but also to ensure the validity of inferences made on the relationships among species within a genus and proper identification in clinical and veterinary diagnostic laboratories.

Received 13 October 2014 Accepted 17 October 2014 Published 18 November 2014

Citation Colston SM, Fullmer M, Beka L, Lamy B, Gogarten JP, Graf J. 2014. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. mBio 5(6):e02136-14. doi:10.1128/mBio.02136-14.

Editor Edward G. Ruby, University of Wisconsin Madison

Copyright © 2014 Colston et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to J. Peter Gogarten, gogarten@uconn.edu, or Joerg Graf, joerg.graf@uconn.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology.

Rapid improvements in DNA sequencing technologies are providing new approaches to address prevailing questions in the field of microbiology (1–3). For example, next-generation sequencing greatly enhanced the discovery of virulence factors through comparative genomics (4), enabled epidemiological studies of recent disease outbreaks (5), led to the discovery of the rare biosphere (6), and provided insights into the physiology of uncultured microbes through metatranscriptomics (7). The increasing amounts of data also brought challenges in ensuring the accuracy of annotations in databases (8). Since many analyses are based on comparisons to known sequences, errors in a database can be easily propagated in other databases and affect subsequent studies. Microbial taxonomy is one area in which the advances in next-generation sequencing have yet to be implemented to their full potential, even though several applications have shown great promise (9, 10). Prokaryotic taxonomy has been traditionally regarded as consisting of three interrelated components: classification, nomenclature, and characterization (11). Only nomenclature is strictly regulated in the International Code of Nomenclature of Bacteria (12). It is important to reconcile nomenclature when rigorous classification and characterization methods reveal an inconsistency in the composition of a particular named species. The organizing principle of microbial taxonomy is to group...
related organisms together that are distinct from other groups. DNA-DNA hybridization (DDH) is the traditional “gold standard” of circumscribing a bacterial species, as this method provides an assessment of the overall similarity of the heritable material, with phylogenetic data providing information about neighboring organisms. The current DDH standard for strains to be considered belonging to the same species is that ≥70% of the DNA from the two strains reassociates with a ±5°C difference in melting temperatures (13). However, laboratory-based DDH measurements are not without challenges, given that DDH values can be difficult to reproduce and therefore may vary, depending on the reannealing temperature used or a laboratory’s particular method employed (14). In addition, the data cannot be archived, nor are they portable between laboratories, and as such the data cannot be readily built upon when describing a new species (15).

In contrast to DDH, DNA sequence information can be easily archived and readily transferred between laboratories. Standardized bioinformatic analyses on the same data set can be performed by different laboratories, which facilitates collaborations and, potentially, the resolution of disagreements (16). Examples of such molecular methods include multilocus sequence analysis (MLSA), which provides important information about the evolutionary relationships of bacteria and allows grouping of related strains (14). MLSA has emerged as a powerful tool for classifying bacterial strains, as it relies on the allelic differences among multiple conserved housekeeping genes (17). In MLSA, the sequences are typically concatenated to overcome the lack of resolution seen in the topology of single-gene trees, but this method may mask the different evolutionary processes underlying the individual genes (18, 19). In addition, there is no consensus as to what degree of sequence variation correlates with species boundaries, which is partly due to different genes evolving at different rates and also that a few selected genes represent only a fraction of the vast amount of information contained within an entire genome.

The field of microbiology is undergoing dramatic changes, with more genomes becoming available due to the rapidly improving technology and declining cost of sequencing. In addition to closed or finished genomes, “improved” high-quality draft genomes for which the annotations have been validated have been deemed suitable for comparative genomic studies (20). The relative ease of producing such genomes provides new opportunities for assessing taxonomic relationships, discovering new taxa, and sharing data between researchers. As a result, new tools are being developed to make use of these data, including a bioinformatic approach for calculating the DDH. One of these, the genome-to-genome distance calculator, referred to here as in silico DDH (isDDH), produces values that compare closely with experimentally derived DDH values (9, 21). Another method calculates the average nucleotide identity (ANI) among conserved and shared genes. The use of ANI has been proposed as a new standard for defining microbial species, and it is gaining wide acceptance (16, 22). The most current proposal recommends use of an ANI threshold of 95 to 96% along with support from tetranucleotide frequency correlation coefficient values (23, 24). Recently, a few studies combined either MLSA or the analysis of genes common to all members of a genus (core genome) with the overall similarity of the genome by using ANI for species identification (15, 25). We wanted to compare isDDH and ANI for species identification combined with phylogenetic approaches, using a genus with a complicated but relatively well-described phylogeny.

The genus *Aeromonas* makes for an ideal test case, because it contains a large number of species, biovars, and subspecies and its taxonomy has been the subject of much debate (26). Collectively, *Aeromonas* members are found in a number of habitats and in association with various animals, ranging from beneficial symbionts of leeches and zebrafish to pathogens of amphibians, fish, and humans (26, 27). Fourteen species of *Aeromonas* were recognized in the latest addition of *Bergey’s Manual of Systematic Bacteriology* in 2005 (28). Since then, over a dozen have been proposed, while the statuses of five species and two subspecies have been called into question. An accurate taxonomy for this genus is not only critical as a tool to differentiate benign from potentially virulent species, but it is also essential as the foundation for ecological studies.

A number of taxonomic controversies exist within the *Aeromonas* genus, namely, the synonymy of the following groups: the proposed novel species *A. cuniculata* and *A. ichthiosmia* with *A. veronii* (29–31), *A. enteropelogenes* with *A. trota* (31–34), *A. allosaccharophila* with *A. veronii* (30), *A. hydrophila* subsp. anaerogenes with *A. caviae* (28, 35), and *A. hydrophila* subsp. *dhakensis* with *A. aquariorum*, which ultimately led to a proposal of a new species, *A. dhakensis* (36–38). All of these controversies are likely due, at least in part, to the limitations of past and current methods to consistently distinguish to the species level. Some of these controversies (e.g., whether the taxon *A. allosaccharophila* reaches the species level) could not even be unambiguously clarified with the most recent methods, with several MLSA schemes with partial sequences of up to seven housekeeping genes (33, 34, 39–41). A finding of some of these studies and of a study investigating discrepancies in the analysis of 16S rRNA genes (42, 43) was that recombination occur frequently between members of this genus, which renders phylogenies with single or a few genes challenging.

The use of whole genome sequences has been regarded as a promising avenue for the future of *Aeromonas* taxonomic and phylogenetic studies (41). In the present study, we generated improved, high-quality draft genome sequences from 27 type strains and 6 additional strains. These genomes were supplemented by 23 additional genomes of *Aeromonas* strains available in public databases. Our approach was to determine the phylogeny in three ways, by using (i) 16 housekeeping genes that were used in four recent MLSA classifications (HK), (ii) ribosomal protein coding gene (RG), and (iii) the expanded core (EC), which are the genes present in at least 90% of the 56 strains. In addition, we performed ANI analysis and isDDH (9, 16, 21, 22) to determine the overall similarity of the genomes. We examined our data with regard to the above-mentioned taxonomic controversies, as these provided the means to validate our approach. We also investigated the relationships of deeper phylogenetic branches in the *Aeromonas* genus. This approach led to the identification of candidate novel species and is presented as a methodology that may be applied to other genera as well.

RESULTS

Genome sequences. A total of 56 *Aeromonas* genomes were used in this study, representing type strains of 29 currently recognized or proposed species, of which 27 were sequenced in-house and 2 were available in GenBank. The additional 23 genomes were non-type strains and auxiliary strains of interest. For seven of the *Aeromonas* species, multiple strains were used in this study, and strain designations were employed to distinguish among them (*A. allosaccharophila*, *A. caviae*, *A. dhakensis*, *A. hydrophila*, *A. media*, *A. ichthiosmia*, and *A. aq
A. salmonicida, and A. veronii); for the remainder of the species, only the type strain was used, which is indicated by a superscript T.

For the 33 genomes obtained for this study, the average genome coverage ranged from 30- to 260-fold and the number of scaffolds ranged from 22 to 332 with an average of 88 (Table 1). The completeness of the genomes was assessed by screening the genomes for 16 housekeeping genes and 47 ribosomal protein-coding genes. All 63 genes were present in the 56 genomes. The genome sizes estimated from the draft genomes generated for this study ranged from 3.90 Mbp (A. fluvialis) to 5.18 Mbp (A. piscicola), with an average of 4.51 Mbp. The average G+C content of the aeromonads ranged from 58.1% (A. australiensis) to 62.8% (A. taiwanensis), with a mean of 60.2%. Based on the quality of the genomes and verification of the automated annotation, we consider these genomes to be improved, high-quality draft genomes (20).

Phylogenetic analysis. One goal of our study was to reevaluate the phylogenetic relationships of the Aeromonas species by using three phylogenies, HK, RP, and EC, derived from different sets of genes: 16 housekeeping genes, 47 ribosomal protein-coding genes, and the expanded core, which included 2,710 ortholog groups (OG), respectively. Due to the differences in the number of informative sites, the EC phylogeny had the strongest support values for all of the nodes, although both the HK and EC phylogenies provided new insights into the relationships of distant clades (Fig. 1). The RP phylogeny had the lowest support values, as these genes are more conserved (see Fig. S1 in the supplemental material). In both the HK and EC phylogenies, we found the same eight major monophyletic groups, or clades, which are defined as groups of taxa in a phylogeny that each share an ancestor, to the exclusion of all other taxa included in the analysis (Fig. 1). Interestingly, we found several differences between the HK and EC phylogenies. In the HK phylogeny, clades 6 and 7 represent shallow branches that are nested within larger groups formed by clades 2 to 7, 3 to 7, and 4 to 7; however, in the EC phylogeny, clade 6 is basal to the large clade containing clades 2 to 5, 7, and 8. Moreover, in the EC phylogeny, clades 2 and 7 form one clade, while clades 3 to 5 form another clade, which is also inconsistent with the HK phylogeny where clade 7 forms a clade with 6 that is nested within a large group containing clades 3 to 7. As the expanded core did not require each ortholog group (i.e., homologs that appear to have evolved from the same ancestral gene in the organismal most recent ancestor of the group) to be present in every genome, we repeated the analysis using the strict core with only those ortholog groups that were present in all genomes. The strict core phylogeny was consistent with the EC phylogeny (see Fig. S2 in the supplemental material), indicating that the ortholog groups present in all genomes did not represent variations in the topology observed between the strict versus expanded core.

Most of the general relationships observed in our study were consistent with those reported in the published literature. The recently proposed species, A. dhakensis, which was determined to be synonymous with A. aquariorum (44), was originally a subspecies of A. hydrophila. All three phylogenies support that these strains form one well-supported clade that is distinct from A. hydrophila. Interestingly, six A. hydrophila genomes that we obtained from GenBank clearly clustered within A. dhakensis. Our study also grouped the strain SSU with A. dhakensis, which supports its recent recategorization from A. hydrophila to A. dhakensis (45). Misnamed genomes in GenBank should be corrected and resolved with thorough classification data to prevent further misidentifications.

Our comprehensive analysis revealed an important difference compared to the previous MLSA by Murcia-Martinez et al., which was based on partial sequences of seven genes (34). In that study, the A. trota isolates (which included A. enteropelogenes) grouped with A. hydrophila and A. aquariorum, whereas in the HK and EC phylogenies of our study, A. enteropelogenes and A. trota formed a clade with a group that included the A. veronii group, or AVG (A. veronii bv. sobria, A. veronii bv. veronii, and A. allosaccharophila), and A. jandaei. This finding is in agreement with those of the study by Roger et al. (33). Examination of individual gene trees suggests that the varied placement was due to the use of different housekeeping genes in these two studies (see Fig. S3 to S6 in the supplemental material) and underscores the limitations of MLSA approaches that use shorter fragments of fewer genes, compared to studies using the expanded core or a large set of full-length housekeeping genes. Our study also confirmed the synonymity of A. trota and A. enteropelogenes (31, 32).

The AVG itself is a controversial collection of species, which includes A. culicicola and A. ichthiomycia, both initially described as new species but subsequently reclassified as A. veronii based on DNA relatedness and biochemical characterization (29–31). Our data support the synonymity of A. culicicola and A. ichthiomycia with A. veronii, as the two strains grouped together with the A. veronii strains in one well-supported clade (Fig. 1A and B). An interesting aspect of this species is that there are two reported A. veronii biovars, which differ phenotypically in that A. veronii bv. veronii is positive (100%) for esculin hydrolysis and ornithine decarboxylation while A. veronii bv. sobria is negative for both reactions (46). In our analysis, the three strains of A. veronii bv. veronii (CECT 4257, AMC35, AER397) grouped together with A. veronii B565 in a strongly supported clade within the larger A. veronii clade, which supports A. veronii bv. veronii as a bona fide biovar. Comparisons of the A. veronii genomes revealed that members of A. veronii bv. veronii encode a β-glucosidase (EC 3.2.1.21; 793 aa) and an ornithine decarboxylase (EC 4.1.1.17; 745 aa) not found among members of A. veronii bv. sobria, suggesting that these two enzymes may facilitate the reactions involving esculin and ornithine, respectively. Based on this data, A. veronii B565, whose genome contains both genes, is a presumptive member of the A. veronii bv. veronii. The two A. allosaccharophila strains (CECT 4199 and BVH88) also formed a strongly supported clade that was near but distinct from A. veronii, which suggests that A. allosaccharophila is a separate species. In our analysis, we also included the newest proposed Aeromonas species, A. australiensis, which is monophyletic with A. fluvialis and A. sobria and the AVG.

The other phylogenetic relationships supported the relationships described in previously published reports, such as the well-supported clade formed by A. similae, A. diversa, and A. shuberiti, which is distinct from all the other Aeromonas species (Fig. 1) and observed in all three phylogenies. The close relatedness between A. piscicola and A. bestiarum (47) was also recovered in our analyses. Our results also support that strain CECT 4221, described as A. hydrophila subsp. anaerogenes, clusters within the A. caviae taxon.
**TABLE 1** General features of the *Aeromonas* genomes

| Species                        | Strain       | Genome size (Mbp) | No. of scaffolds | Avg genome coveragea | G+C content (%) | No. of predicted CDSb | Accession no. | Reference          |
|-------------------------------|--------------|-------------------|------------------|----------------------|----------------|-----------------------|---------------|-------------------|
| *A. allosaccharophila*         | CECT 4199T   | 4.66              | 120              | 87                   | 114,541        | 58.4                  | 4,173         | PRJEB7019*        |
| *A. dhakensis*                 | CECT 7289T   | 4.69              | 78               | 117                  | 163,504        | 61.7                  | 4,266         | PRJEB7020*        |
| *A. dhakensis*                 | CIP 105493T  | 4.76              | 76               | 48                   | 108,949        | 58.2                  | 3,609         | PRJEB7030*        |
| *A. dhakensis*                 | DSM 22539r   | 4.53              | 102              | 99                   | 155,151        | 60.0                  | 4,149         | PRJEB7035*        |
| *A. salmonica subsp. salmonica*| CIP 103209r  | 4.74              | 128              | 117                  | 89,543         | 58.5                  | 4,442         | PRJEB7036*        |
| *A. aquorella*                 | CECT 4227T   | 4.68              | 41               | 53                   | 208,249        | 62.0                  | 4,090         | ANPN01d           |
| *A. allosaccharophila*         | BVH88        | 4.71              | 131              | 204                  | 74,486         | 58.6                  | 4,295         | PRJEB7045*        |
| *A. cavie*                    | Ac398        | 4.44              | 149              | UNK                  | 76,364         | 61.4                  | 3,866         | CACP01d           |
| *A. cavie* [A. hydrophila subsp. anaerogenes] | CECT 4221 | 4.58              | 332              | 66                   | 31,465         | 61.0                  | 4,207         | PRJEB7046*        |
| *A. dhakensis* [A. aquorella]  | CIP 107500   | 4.71              | 73               | 84                   | 165,885        | 61.8                  | 4,284         | PRJEB7048*        |
| *A. dhakensis*                 | 173          | 4.79              | 74               | 46                   | 119,625        | 61.6                  | 4,134         | AOBN01d           |
| *A. dhakensis*                 | 277          | 4.79              | 41               | 76                   | 282,384        | 61.6                  | 4,213         | AOBQ01d           |
| *A. dhakensis*                 | 14           | 4.67              | 75               | 45                   | 130,840        | 62                    | UNK           | AOBM01d           |
| *A. dhakensis*                 | 116          | 4.61              | 45               | 66                   | 208,249        | 62                    | 4,090         | ANPN01d           |
| *A. dhakensis*                 | 259          | 4.70              | 80               | 39                   | 117,245        | 61.7                  | 4,098         | AOBP01d           |
| *A. dhakensis*                 | 187          | 4.78              | 59               | 111                  | 197,352        | 61.6                  | 4,205         | AOB01d            |
| *A. dhakensis*                 | SSU          | 4.94              | 2                | 285                  | 4,791,870      | 61.5                  | 4,449         | AGWRO1d           |
| *A. hydrophila*                | ML09.119     | 5.02              | UNK               | UNK                  | UNK            | 60.8                  | 4,433         | CP005966.1d       |
| *A. hydrophila*                | SNPUPC.8     | 4.97              | 41               | 37                   | 234,812        | 60.8                  | 4,352         | AMjQ01d           |
| *A. hydrophila* subsp. ranae  | CIP 107985   | 4.68              | 107              | 140                  | 90,304         | 61.6                  | 4,268         | PRJEB7049*        |
| *A. media*                    | WS           | 4.78              | 1                | UNK                  | 4,788,430      | 60.7                  | 4,385         | CP007567.1d       |

(Continued on following page)
Assessment of genome similarity using \(i\)DDH and ANI. The information gained from the phylogenetic analyses provides an important depiction of the evolutionary relationships of different strains but does not translate directly into the overall similarity of the genomes, which was determined through DDH. We used two different \textit{in silico} or bioinformatics approaches, \(i\)DDH and ANI, that have been proposed to overcome the challenges of conventional laboratory-based DDH to evaluate the genomic similarity of bacteria, and we evaluated the congruence of these methods (DDH and ANI).

Two excellent examples for validating this approach are \textit{A. cullicicola} and \textit{A. ichthiosmiasis}, which were initially proposed as novel species and later reclassified as \textit{A. veronii} based in part on DDH values that exceeded 70%. The predicted point estimates of the \(i\)DDH values we obtained for these two strains were all slightly below 70% (69.1 to 69.6% and 67.4 to 68.2%, respectively) compared to all other named \textit{A. veronii} strains. Additionally, the ANI values are all ~94%. These data support the status of \textit{A. allosaccharophila} as a bona fide species that is closely related to \textit{A. veronii}. Interestingly, while the HK, RP, and EC phylogenies all grouped the two \textit{A. allosaccharophila} genomes (CECT 4199\(^T\) and BVH88) together and separate from \textit{A. veronii}, the ANI and the upper 95% CI \(i\)DDH values between the two \textit{A. allosaccharophila} genomes were both just under the species cutoff boundary, at 95.8% and 68.7%, respectively. These data suggest that BVH88 may not be a member of the \textit{A. allosaccharophila} species, but a greater number of strains in this clade will need to be evaluated to clarify their relationships. Two other species, \textit{A. fluvalis} (ANI, ~92%) and \textit{A. australiensis} (ANI, ~93%), also group near \textit{A. veronii}. Their \(i\)DDH estimates register ~52% compared to \textit{A. veronii}.

Another group of species that has recently attracted attention is \textit{A. aquariorum}, \textit{A. hydrophila} subspp. \textit{dhakensis}, and \textit{A. hydrophila}. The partition of the group comprised of \textit{A. aquariorum}/\textit{A. hydrophila} subspp. \textit{dhakensis} strains from the \textit{A. hydrophila} group, which includes the type strain (CECT 839), was recovered conclusively by every method we used in our study. The branch lengths of the HK phylogeny between \textit{A. dhakensis} and \textit{A. hydrophila} (~0.075 substitutions/site) were similar to those separating many named species in the HK reconstruction, such as those between \textit{A. eucnepholia} and \textit{A. tecta} (~1.0 substitutions/site), \textit{A. schuberti} and \textit{A. diversa} (~0.09 substitutions/site), \textit{A. rivalii} and \textit{A. molluscorum} (~0.06 substitutions/site), and \textit{A. piscicola} and \textit{A. besiiatrum} (~0.04 substitutions/site). Similar relationships were observed in the RP and EC phylogenies. Further evidence comes from the ANI data, which showed only 93% similarity between the two different clades. This is well below the 96% species cutoff recommended by Richter (23). This conclusion was further sup-

### TABLE 1 (Continued)

| Species                  | Strain | Genome size (Mbp) | No. of scaffolds | Avg genome coverage\(^c\) | \(N_{50}\) (nt) | G+C content (%) | No. of predicted CDS\(^d\) | Accession no. | Reference                   |
|--------------------------|--------|-------------------|-----------------|---------------------------|----------------|-----------------|--------------------------|---------------|------------------------------|
| \textit{A. salmonicida} \textit{subsp. achronogenes} | AS03   | 4.96              | 69              | 21                        | 124,543        | 58.3            | UNK                      | AMQG02\(^d\) | 82                           |
| \textit{A. salmonicida} | A449   | 5.04              | 1               | UNK                       | 5,040,536      | 58.2            | 4,436                    | CP000644.1\(^d\) | 83                           |
| \textit{A. salmonicida} \textit{subsp. salmonicida} | 01-B526 | 4.92              | 604             | 40                        | 83,743         | 58.4            | 4,529                    | AGVO01\(^d\) | 84                           |
| \textit{Aeromonas sp.} \{\textit{A. hydrophila}\} | AH4    | 4.87              | 41              | 90                        | 258,555        | 59.6            | 4,453                    | PRJEB6940\(^e\) | This study                    |
| \textit{Aeromonas sp.} \{\textit{A. veronii}\} | AMC 34 | 4.58              | 1               | 288                       | 4,578,728      | 58.5            | 4,117                    | AGWU01\(^d\) | The Broad Institute           |
| \textit{A. veronii}      | B565   | 4.55              | 1               | UNK                       | 4,551,783      | 58.7            | 4,073                    | CP002607\(^d\) | 85                           |
| \textit{A. veronii} \textit{bv. sobria} | AER 39 | 4.42              | 4               | 283                       | 1,516,045      | 58.9            | 3,948                    | AGWT01\(^d\) | The Broad Institute           |
| \textit{A. veronii} \textit{bv. sobria} | Hm21   | 4.68              | 50              | 200                       | 179,631        | 58.7            | 4,245                    | ATFBO1\(^d\) | 62                           |
| \textit{A. veronii} \textit{bv. sobria} | LMG 13067 | 4.74          | 72              | 46                        | 147,470        | 58.3            | 4,171                    | PRJEB7051\(^e\) | This study                    |
| \textit{A. veronii} \textit{bv. veroni} | AER 397 | 4.50             | 5               | 378                       | 3,260,625      | 58.9            | 3,986                    | AGWV01\(^d\) | The Broad Institute           |
| \textit{A. veronii} \textit{bv. veroni} | AMC 35 | 4.57              | 2               | 285                       | 4,172,420      | 58.6            | 4,036                    | AGWV01\(^d\) | The Broad Institute           |

\(^a\) Obtained from the EMBL Nucleotide Sequence Database.
\(^b\) Previously published names are indicated inside braces.
\(^c\) UNK, unknown.
\(^d\) Obtained from GenBank, National Center for Biotechnology Information.
\(^e\) The average genome coverage is expressed in bp sequenced divided by genome size.
\(^f\) The \(N_{50}\) (reported in nucleotides) represents the smallest of the largest contigs covering 50% of the total size of all contigs.

\(^g\) CDS, coding sequence.
FIG 1. (A) Maximum likelihood reconstruction of 16 single-copy housekeeping genes. Support values are represented by dots: red (90%+), yellow (70%+), green (60%+), and gray (<60%) (90%+). (B) Approximate maximum likelihood reconstruction of 2,710 orthologous groups found in 90% or more of the taxa. A. veronii AM14 and A. hydrophila HA1 are shown in red and dark gray, respectively. Eight well-supported clades were shared between the two reconstructions.

**Branch Supports:**
- 90%+ Bootstraps
- 80%+ Bootstraps
- 70%+ Bootstraps
- 60%+ Bootstraps
FIG 2 ANI and isDDH values. The lower triangle displays ANI values, and the upper triangle shows the isDDH values. ANI values are colored according to three historical species cutoff values: 94% (yellow), 95% (orange), and 96%+ (red). The isDDH values displayed are the upper limits of the 95% confidence intervals and are colored red if the met the laboratory DDH species cutoff of 70% hybridization. ANI of 96% correlates well with 70% isDDH values, with only the *A. allosaccharophila* isolates failing to match (68.7%).
ported by isDDH data, in which *A. dhakensis* and *A. hydrophila* strains all scored below 60% between species when using the upper border of the 95% CI, while within each partition all values were well above 70%. These data confirm that these two clades represent two discrete species rather than constituents of one, as was originally proposed (48).

*A. piscicola*\(^T\) and *A. bestiarum*\(^T\) grouped together and formed one clade with *A. popoffii*\(^T\). The ANI between *A. piscicola*\(^T\) and *A. bestiarum*\(^T\) was 95.2%, which is near the 96% suggested species cutoff (23). However, while their *isDDH* values were higher than most between-species comparisons (61.1% point estimate, 64.4% at the upper 95% CI), they still fell short of what one would expect for members of the same species. It will be important to add more strains of these two groups in future analyses to gain better insight into the relationships between these taxa. Based on the current data, a 96% cutoff for the ANI value seems appropriate for *Aeromonas* species delineations.

**Discovery of novel species.** We also included two strains in our analysis that seemed unusual based either on previous studies or preliminary data. AMC 34, a clinical isolate described as *A. veronii* bv. veronii, had a long branch length and clustered away from other *A. veronii* bv. veronii strains in a previous study (41). Strain AH4 was published as *A. hydrophila* by investigators that had obtained this isolate from the water of a storage container for medicinal leeches (49). In the HK phylogeny, AMC 34 clustered well outside the *A. veronii* clade, near *A. jandaei*\(^a\) and *A. fluvalia*\(^a\), with bootstrap support values in excess of 90% (Fig. 1A). Similarly, the EC phylogeny placed AMC 34 outside of *A. veronii* with high support (Fig. 1B). The ANI between AMC 34 and the AVG was ~94%, while the *isDDH* was only ~58% compared to the same taxa (Fig. 2). Taken together, the data strongly support AMC 34 as a new species.

The other strain, AH4, was identified by a clinical diagnostic laboratory as *A. hydrophila* (49). In all of our phylogenetic analyses, AH4 grouped with *A. piscicola*\(^T\) and *A. bestiarum*\(^T\) with high support. This placement and its distance from *A. hydrophila* were strongly supported by the ANI and *isDDH* data (Fig. 2). AH4 registered only ~89% to both the *A. hydrophila* and *A. dhakensis* groups but much higher values to *A. bestiarum*\(^T\) (~94%) and *A. piscicola*\(^T\) (~93%). *isDDH* also supported the conclusion that AH4 is not likely a member of *A. bestiarum* (~55%) or *A. piscicola* (~52%) and is distinct from the *A. hydrophila* (~38%) and *A. dhakensis* (37%) groups.

All of our bioinformatic analyses indicated that the strains AMC 34 and AH4 represent two new species; however, we were restricted to a single isolate of each, which precluded the assessment of the variabilities of biochemical tests (see Table S1 in the supplemental material). In addition, we were unable to include one recently published type strain, *A. cavernicola* CCM7641\(^T\) (50) or one proposed new species, *A. lusitana* (34), which has not yet been officially described. Using the available MLST data, we were able to show that AMC 34 and AH4 did not cluster near these two species and are thus not likely members of either *A. cavernicola* or *A. lusitana* (see Fig. S8 in the supplemental material). The accessibility of the genomes published for this study will provide other researchers with the opportunity to determine the probable taxonomic position of candidate novel species, an important capability in light of the number of taxonomic problems described for *Aeromonas*.

**Comparison of phylogenetic and genetic distance measures.** The delineation of organisms into taxonomic groups is based on their evolutionary histories and genetic distances. In this study, we utilized five different approaches, of which two were phylogeny independent (*isDDH* and ANI) and three had a phylogenetic component (HK, RP, and EC phylogenies). To guide subsequent studies, we wanted to evaluate whether these approaches were in agreement with one another and whether some were more informative than others. Even though *isDDH* and ANI use different algorithms for the calculations, e.g., ANI evaluates the similarity of shared elements between two genomes, while *isDDH* estimates the overall similarity of two genomes, the results were very consistent (Fig. 3). The \(r^2\) value was 0.957 for the entire data set, and when restricted to comparisons of more closely related strains (isDDH of ≥55%), the \(r^2\) was 0.966. These values demonstrated that at least for this data set, either method can be used for determining overall genome similarities. When isDDH (upper 95% CI) and ANI results were compared to the P-distance of the entire EC data set, the \(r^2\) values were low for both approaches, 0.599 and 0.713, respectively. When the data set was restricted to comparisons of genomes that had at least a similarity of ≥50% based on isDDH, the correlation coefficients were 0.943 and 0.965 for isDDH and ANI, respectively (see Fig. S9 in the supplemental material). This indicated that either approach works well at separating closely related genomes but not for determining more distant relationships.

Most researchers characterize strains by analyzing the sequences of only one or two genes. We wanted to ascertain whether there are particular genes that are better suited than others for an initial analysis. One important concern is that horizontal gene transfer of gene fragments and not just entire genes can occur among aeromonads and result in conflicting phylogenies (41). Thus, relying on any one gene can produce erroneous results. On the other hand, including a preponderance of genes that represent a highway of gene sharing in a concatenation may result in phylogenies that reflect neither organismal evolution nor any individual gene history (51). The individual gene trees (see Fig. S3 to 6 in the supplemental material) for the 16 housekeeping genes were compared to the phylogeny derived from the consensus tree using...
the approximately unbiased (AU) test (52). The set of maximum likelihood (ML) trees generated from bootstrap samples of the MLSA data were significantly different from the best gene tree for each gene. When maximum likelihood trees from bootstrap samples of the 16 housekeeping genes were compared to the MLSA tree, 15 of the gene tree sets were significantly different from the MLSA best tree. Only one of the bootstrap samples for recA had a P value of ≈0.05 (P = 0.93). These results reveal that no individual gene tree properly reflects or is even compatible with the phylogeny of the MLSA tree.

**DISCUSSION**

Our polyphasic genome comparison utilizing both phylogenetic and genetic distance metrics was by and large consistent with the current understanding of the phylogenetic relationships of the species contained within the genus *Aeromonas*, which had been hitherto based on laboratory-determined DDH values, biochemical tests, and multilocus sequence typing. Importantly, we were able to gain new insights into the overall relationships of the *Aeromonas* species with the phylogeny generated from the expanded core and the HK genes. There were eight major clades from the EC that were largely consistent with the HK phylogeny (Fig. 1). One major difference between the two phylogenies was the placement of *A. salmonicida* (clade 7) and *A. hydrophila* and *A. dhakensis* (clade 2). In the EC phylogeny, they form one strongly supported clade, but in the HK phylogeny they are separated by two well-supported nodes (Fig. 1). This suggests that other components of the genome are forcing *A. hydrophila* and *A. salmonicida* together in the expanded core phylogeny. Due to the limited resolution, the RP phylogeny did not provide additional support. A strict core phylogeny using only ortholog groups present in all 56 taxa shared the topology of the EC tree, suggesting that the conflict with the HK method was due to genes present in 100% of the genomes (see Fig. S2 in the supplemental material). One should consider, however, that the EC phylogeny may have inherent biases which might lead to an inaccurate depiction of organismal phylogeny. At this point, we cannot establish which topology is correct, since gene transfer between divergent groups has the potential to lead to trees from concatenated data sets that do not reflect the vertical inheritance (19). Gene transfer frequency is usually biased toward close relatives, thus reinforcing the signal due to shared ancestry (53, 54). In contrast, highways of gene sharing between more distant species can obscure the vertical phylogenetic signal due to shared ancestry (51, 55). For phylogenetic relationships within each of the clades 1 through 7, the HK and EC phylogenies appear to approximate organismal phylogeny (Fig. 1). On the other hand, relationships between these clades remain ambiguous. Differences in substitution rates and saturation with substitutions make it difficult to apply ANI and isDDH to higher taxonomic levels. Future work will need to include the evaluation of the 2,710 individual trees from the EC analysis in a combined analysis, such as the one described by Bansal, Alm, and Kellis (56), to determine the major conflicting phylogenetic signals retained in these genomes. Even so, both the HK and EC phylogenies provided more information regarding the relationships of different *Aeromonas* species than previous MLSA studies.

The psychrophilic aeromonads have been differentiated from the mesophilic strains based on growth physiology, biochemical properties, and virulence characteristics. Although there certainly are important differences among these characteristics, whole-genome information groups them clearly among the mesophilic species, near *A. hydrophila* and *A. dhakensis*. One interesting distinction of the *A. salmonicida* clade is that there is much less genetic diversity, indicated by the isDDH values for strains of the same species. The four *A. salmonicida* genomes had isDDH values ≥98.5%, in comparison to *A. hydrophila* (≥75.7%), *A. dhakensis* (≥78.3%), and *A. veronii* (≥70.4%). This was consistent with a study that suggested a clonal distribution of *A. salmonicida* subsp. *salmonicida* based on identical pulse electrophoresis DNA fingerprints, which showed identical banding patterns from strains isolated from different geographical regions (57). This difference in genetic diversity could reflect different evolutionary driving forces for *A. salmonicida* strains. One conjecture is that perhaps they are adapted for a virulent lifestyle in fish, where clonal outbreaks are more likely to occur. It is also possible that there is a sampling bias, which future studies employing more strains should help to resolve.

One of our goals was to assess the utility of bioinformatics approaches to replace traditional taxonomic approaches for species identification. Despite the shortcomings and challenges of laboratory DDH, whole-genome content comparisons collectively represent the most valuable criterion for demarcation of bacterial species. As more bacterial genomes are sequenced and the information is made accessible, the use of whole genome sequences in the characterization of bacterial species provides opportunities that should not be ignored. This approach has been used in clarifying the taxonomic positions in some cases, e.g., for *Acinetobacter* using ANI and core gene phylogeny (15) and for *Vibrio* using MLSA based on genome information (58). To our knowledge, however, an approach utilizing isDDH and ANI combined with HK, RG, and EC phylogenies has not yet been done for a genus characterized by a complicated taxonomy and using a plurality of its members.

*Aeromonas* is an interesting test case for a number of reasons. This genus is comprised of a large number of species capable of diverse associations depending on the species. The spectrum encompasses benign and virulent species, a range that can also exist within a single species. *A. hydrophila*, *A. caviae*, and *A. veronii* have long been associated with human disease (26). Recently, *A. dhakensis* was recognized as a new virulent species (59), a distinction obfuscated in part due to *A. dhakensis* strains initially regarded as *A. hydrophila*. Of the numerous *Aeromonas* species that have been proposed and characterized, many of those species have been redefined and renamed as new information has been presented. This shifting nomenclature is a manifestation of the inefficiencies inherent in current taxonomic methods for *Aeromonas*. While the number of publically available *Aeromonas* genomes has increased dramatically in the last few years, most of the type strains are yet to be fully sequenced. We produced improved, high-quality draft genomes for these type strains and for some non-type strains of interest. Our results recapitulated known phylogenetic relationships and provided further insights into several others. This study also identified the breakpoints between species, indicating that this approach can be used to identify new species. For demarcating species boundaries, isDDH and ANI produced similar results, as reflected in the correlation of the values observed when using the upper 95% CI bound to the isDDH estimates (Fig. 3). The current version of isDDH is only available in a Web-based interface that requires manually uploading the sequence information, while ANI can be easily run on local servers. Consequently, we found...
ANI to be more time-effective when dealing with a large number of strains. For smaller studies, isDDH would be equally fast for computing and would also have the benefit of confidence intervals and probability statistics.

Apart from the fact that our approach could confidently and consistently resolve recent taxonomic controversies, our analysis also revealed that two strains, AMC 34 and AH4, represent new Aeromonas species. This conclusion is based on the distance in the genome content according to ANI and isDDH values, as well as the phylogenetic distances of the strains. These finding highlights two important advantages of bioinformatic assessment of genome similarity: (i) the expensive generation of the raw data does not have to be repeated by other research groups, and (ii) interlaboratory variations in DDH determinations can be overcome by agreeing to a cutoff value with standardized parameters in bioinformatic analyses. To facilitate the progress of other research groups in the Aeromonas field, we have set up a website (http://aeromonasgenomes.uconn.edu) that allows users to query and download all of the available Aeromonas genomes, contains the scripts we used in our analysis, and provides a summary of our current distance measures.

Another important finding from our analysis was that, out of the 23 publicly available Aeromonas genomes that we analyzed, 8 (34.8%) are inconsistently named. In large part this was due to the recent reclassification of A. hydrophila subsp. dhakensis as A. dhakensis and the reclassification of A. aquariorum as A. dhakensis. While the initial misclassifications are understandable, efforts should be taken to correct and update the nomenclature to curtail the promulgation of inaccurate information. NCBI currently allows only the original submitter to request the name change (http://www.ncbi.nlm.nih.gov/books/NBK51157/). One possibility would be to involve the community at large to provide input on such discrepancies.

The ability to generate improved, high-quality draft genome sequences rapidly and inexpensively, and of a sufficient quality for robust phylogenetic analyses (20), is changing the landscape of how one can investigate microbial taxonomy and should lead to a robust phylogenetic analyses (20), is changing the landscape of how one can investigate microbial taxonomy and demonstrated the accuracy of the bioinformatics approach to identify new species and to correct erroneous identifications from previous studies. Utilizing the same software, code, and parameters for the data analysis, one can readily compare findings of other groups, thus supplanting arguments concerning laboratory methodologies with practical discussions on appropriate cutoff levels. For this test case study with Aeromonas, an isDDH of $\geq 70\%$ at the upper 95% confidence interval or an ANI value of $\geq 96\%$ was consistent for genomes belonging to the same species.

Distance in the EC phylogeny is another metric that can be useful in species identification; in our study, a distance of $\leq 0.026$ indicated that the genomes belong to the same species. It is likely that these types of values will also be applicable to other genera.

**MATERIALS AND METHODS**

**Strains, growth conditions, and biochemical tests.** For the genome data set, we included all of the type strains for Aeromonas with the exception A. cavernicola (50), as well as all other Aeromonas genomes deposited into public databases as of 17 July 2013. For the type strains, 2 were publicly available and 27 were sequenced in-house. For additional strains, 21 were available publically and 6 were sequenced in-house. The bacteria were grown at the optimal growth temperature for the strain in LB broth or on LB agar (1.5%) plates for 16 to 18 h (60). For biochemical tests, API 20NE strips (bioMérieux, Marcy l’Etoile, France) were used in accordance with the manufacturer’s instructions. Separate tests for ornithine decarboxylase (ODC) activity and esculin hydrolysis were assessed using ODC broth and bile esculin agar (Sigma-Aldrich, St. Louis, MO). Tests were performed in triplicate.

**Library preparation and genome sequencing.** Genomic DNA was extracted using the MasterPure DNA purification kit (Epicenter, Madison, WI) and quantified using a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA). DNA was also checked for quality by using a NanoDrop spectrophotometer and bioanalyzer (Agilent Technologies, Santa Clara, CA) prior to sequencing on a MiSeq benchtop sequencer (Illumina, Inc., San Diego, CA). Library concentrations were determined by using the Qubit fluorometer and bioanalyzer (Agilent Technologies, Santa Clara, CA) prior to sequencing on a MiSeq benchtop sequencer (Illumina, Inc.) at the Microbial Analysis Resources and Services facility at the University of Connecticut (Storrs, CT).

**Assembly and annotation.** Paired Illumina reads were trimmed and assembled into scaffolded contigs by using the de novo assembler of CLC Genomics Workbench versions 6.0.04 to 7.0.04 (CLC-bio, Aarhus, Denmark). Annotation of the contigs was accomplished using the Rapid Automated Annotation using Subsystem Technology (RAST) server (61). All Aeromonas completed and draft annotated assemblies from the NCBI ftp repository that were used in this study were downloaded, back-engineered into contigs, and submitted to RAST for reannotation to mitigate any biases in the RAST annotation algorithms by applying them equally to each genome. The completeness of the genomes was initially assessed by screening for 17 housekeeping genes and 47 ribosomal proteins. We failed to detected $ppkA$ (phosphoenolpyruvate synthase) in A. flavians. A thorough investigation employing mapping of reads to reference sequences and examining the region containing $ppkA$ in the other strains suggested that this gene may not be present in this organism, and thus we excluded $ppkA$ from the analysis.

**MLSA reference tree and individual gene tree generation.** Sixteen housekeeping genes ($atpD$, $dnaj$, $dnaK$, $dnaN$, $glnA$, $gol$, $gyrA$, $gyrB$, metG, mfd, radA, roA, $rpoC$, $rpoD$, $tsf$, and $zipA$) were used for MLSA (33, 34, 39). The DNA-directed RNA polymerase subunit beta’ ($rpoC$) was used in the MLSA dataset. Adding $rpoB$ to the dataset or switching it for $rpoC$ did not change the phylogeny resulting from the MLSA analysis depicted in Fig. 1. These genes were initially chosen in three separate MLSA studies for their conservation among all aeromonads, ease of PCR primer design, broad distribution, and single copy number in the chromosome. The full-length sequence of each gene was initially derived from the previously published genome of A. veronii Hm21 (62), and these sequences served as queries for BLAST searches against the annotated proteins of all 56 genomes. Multiple sequence alignments (MSAs) were generated by translating the genes to protein sequences in SeaView (63), aligning the proteins using MUSCLE (v.3.8.31) (64) and then back-translating to the nucleotide sequences prior to the phylogenetic analysis. Each MSA was manually evaluated, and any sequences showing poor alignment were examined further, including comparison against the nonredundant data-
base using BLAST and excluded if not found to be the correct protein. In-house scripts created a concatenated alignment of all 16 genes. A model of evolution was determined by using the Akaike information criterion with correction for small sample size (AICc), as implemented in jModelTest 2.1.4. An ML phylogeny was generated from the concatenated MSA, and individual gene phylogenies from the individual gene MSAs were determined by using PhyML (v 3.0, 300-500M) (65). PhyML parameters consisted of a GTR model, estimated p-invar, 4 substitution rate categories, estimated gamma distribution, and subtree pruning and regrafting enabled with 100 bootstrap replicates. Using the same approach, phylogenies were determined for each of the 16 housekeeping genes.

**Ribosomal reference tree generation.** Forty-seven ribosomal proteins were obtained from the BioCyc website (66). These served as queries for BLAST searches against the annotated proteins of all 56 genomes. Multiple sequence alignments were generated as described above for the MLSA tree. The AICc reported the best-fitting model to be GTR plus invariable site estimation.

MSA tree. The AICc reported the best-fitting model to be GTR plus invariable site estimation. MLSA tree. The AICc reported the best-fitting model to be GTR plus invariable site estimation. MLSA tree. The AICc reported the best-fitting model to be GTR plus invariable site estimation.

**Core genome comparison.** To define a core genome, the annotated protein open reading frames (ORFs) from each genome were used as BLAST queries against the protein ORFs of each other genome in the study, using in-house Perl scripting. The BLAST outputs were processed into OGs with MCL-edge v14-137 (67, 68) (http://micans.org/mcl/). The inflation value was set to 10 in order to break the OGs down into smaller clusters that more closely resembled individual genes rather than families. A relaxed core was defined by extracting OGs present in at least 90% of the taxa used in this study. Where a taxon had multiple entries in a single OG, the first entry reported by MCL was arbitrarily included and the others were excluded. Each OG was aligned using MUSCLE v 3.8.31 (64). In-house Perl scripting concatenated the OGs into a single alignment. Aligning to the scale of the concatenated alignment, FastTreeMP (69) was used to perform the phylogenetic reconstruction. The substitution model used was WAG.

**Pairwise sequence distance calculations and identity calculations.** Sequence distances were calculated using the SaveDist function in PAUP* v4.0b10 (70). The distance type calculated was the P-distance. Average nucleotide identity/tetramer analysis. Assembled contigs were reconstituted from the RAST-generated GenBank files for all genomes by using the secret function of the EMBoss package (71). All genomes were treated in the same manner to ensure that any biases were consistent across the entire data set. iSpecies2.1.2 (23) was used to analyze these contig sets for the ANI and tetramer usage patterns, using default parameters. We report here the averages of the reciprocal comparisons.

**Tree comparisons using the approximately unbiased test.** Per site log likelihoods were generated in RAXML v.7.3.5 (72). The AU tests (52) were carried out in the CONSEL v 1.20 package (73). Comparisons were made with HK tree against the 100 bootstrap replicates from each individual gene. Likewise, each best individual gene tree was compared against 100 bootstrap replicates of the HK tree.

**In silico DNA-DNA hybridization.** Estimates of idDHH were made using the Genome-to-Genome Distance Calculator (GGDC) (9, 21). The contig files were uploaded to the GGDC 2.0 Web server (http://ggdc.dsmz.de/distcalc2.php), where idDHH calculations were performed. Formula 2 alone was used for analysis, since it calculates idDHH estimates independent of genome lengths and is recommended by the authors of GGDC for use with any incomplete genomes (9, 21). The point estimate plus the 95% model-based confidence intervals were used for analysis.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl doi:10.1128/mBio.02136-14/-/DCSupplemental. Figure S1, EPS file, 0.6 MB. Figure S2, EPS file, 0.6 MB. Figure S3, EPS file, 2.1 MB. Figure S4, EPS file, 2 MB. Figure S5, EPS file, 1.5 MB. Figure S6, PDF file, 2 MB.

**REFERENCES**

1. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat. Rev. Genet. 13:601–612. http://dx.doi.org/10.1038/nrg3437.

2. Paklen MJ, Loman NJ, Penn CW. 2010. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr. Opin. Microbiol. 13:625–631. http://dx.doi.org/10.1016/j.mib.2010.08.003.

3. Ribeca P, Valiente G. 2011. Computational challenges of sequence classification in microbial data. Brief. Bioinform. 12:614–625. http://dx.doi.org/10.1093/bib/bbr019.

4. Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. 40:D641–D645. http://dx.doi.org/10.1093/nar/gkr989.

5. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, Fitzgerald M, Godfrey F, Haas BJ, Murphy CJ, Russ C, Sykes S, Walker BJ, Wortman JR, Young S, Zeng Q, Abouelliel A, Boichichio J, Chauvin S, Desmet T, Gujia S, McGowan C, Montmayeur A, Steelman S, Frimodt-Møller J, Petersen AM, Struve K, Krogfelt KA, Bingen E, Weill FX, Lander ES, Nusbaum C, Birren BW, Hugt DT, Hanage WP. 2012. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc. Natl. Acad. Sci. U. S. A. 109:3065–3070. http://dx.doi.org/10.1073/pnas.1121491109.

6. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” Proc. Natl. Acad. Sci. U. S. A. 103:12115–12120. http://dx.doi.org/10.1073/pnas.0605127103.

7. Bomar L, Maltz M, Colston S, Graf J. 2011. Directed culturing of microorganisms using metatranscriptomics. mBio 2(2):e00012-11. http://dx.doi.org/10.1128/mBio.00012-11.

8. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLOS Comput. Biol. 5:e1000605. http://dx.doi.org/10.1371/journal.pcbi.1000605.

9. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome-sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. http://dx.doi.org/10.1186/1471-2105-14-60.

10. Chau N, Rainey FA. 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int. J. Syst. Evol. Microbiol. 64:316–324. http://dx.doi.org/10.1099/ijsem.0.054171-0.

11. Brenner DJ, Staley JT, Krieg NR. 2005. Classification of procaryotic organisms and the concept of bacterial speciation, p 27316–32. In Brenner DJ, Staley JT, Krieg NR, Garrity GM (ed), Bergey’s manual of systematic bacteriology, vol 2. The proteobacteria. Springer Verlag, New York, NY.

12. Lapage SP, Sneath PHA, Lessel EF, Skerker VBD, Seeliger HPB, Clark WA. 1992. International code of nomenclature of Bacteria: bacteriological code, 1990 revision. ASM Press, Washington, DC.

13. Stackebrandt E, Frederiksen W, Garrity GM, Grömont PA, Kämpfer P, Maiden MG, Nesme X, Rosselló-Mora R, Swings J, Trüper HG, Vauterin L, Ward AC, Whitman WB. 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. http://dx.doi.org/10.1099/ijsem.0.02360-0.

14. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J.
2005. Opinion: re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3:733–739. http://dx.doi.org/10.1038/nrmicro1236.

15. Chan JZ, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ. 2012. Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol. 12:302. http://dx.doi.org/10.1186/1471-2179-12-302.

16. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102: 2567–2572. http://dx.doi.org/10.1073/pnas.0409727102.

17. Clarke SC, Diggle MA, Edwards GF. 2002. Multilocus sequence typing and porA gene sequencing differentiates strains of Neisseria meningitidis during case clusters. Br. J. Biomed. Sci. 59:160–162.

18. Kämpfer P, Glaeser SP. 2012. Prokaryotic taxonomy in the sequencing era—the polyphasic approach revisited. Environ. Microbiol. 14:291–317. http://dx.doi.org/10.1111/j.1462-2920.2011.02615.x.

19. Lapierre P, Lasek-Nesselquist E, Gogarten JP. 2014. The impact of HGT on phylogenic reconstruction methods. Brief. Bioinform. 15:79–90. http://dx.doi.org/10.1093/bib/bbt050.

20. Chain PSG, Graham DV, Fulford RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buahy C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Consortium GSCHMPJ, Detter JC. 2005. Genome prokaryotic standards in a new era of sequencing. Science (New York, NY) 302:236–237. http://dx.doi.org/10.1126/science.1086144.

21. Auch AF, von Jan M, Klenk HP, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. 5:239–244. http://dx.doi.org/10.1111/j.1758-2146.2010.00216.x.

22. Beaz-Hidalgo R, Martínez-Murcia A, Figueras MJ. 2014. Corrigendum to “Reclassification of Aeromonas hydrophila subsp. dhakensis Huys et al. 2002 and Aeromonas aurariarum Martinez-Murcia et al. 2008 as Aeromonas dhakensis sp. nov. comb. nov. and emendation of the species Aeromonas hydrophila.” [Syst. Appl. Microbiol. 36:171–176, 2013.] Syst. Appl. Microbiol. 37:543. http://dx.doi.org/10.1016/j.syapm.2012.12.007.

23. Martino ME, Fasold L, Montemurro F, Rosteghin M, Manfrin A, Patarnello T, Novelli E, Cardazzo B. 2011. Definition of microbial diversity in Aeromonas strains based on multilocus sequence typing, phenotype and presence of putative genes of virulence. Appl. Environ. Microbiol. 77:4986–5000. http://dx.doi.org/10.1128/AEM.00708-11.

24. Greer JG, Falsetti M, Fuset MC. 2014. Molecular phylogenetics and temporal diversification in the genus Aeromonas based on the sequences of five housekeeping genes. PLoS One 9:e88805. http://dx.doi.org/10.1371/journal.pone.0088805.

25. Silver AC, Williams D, Faucher J, Horneman AJ, Gogarten JP, Graf J. 2011. Complex evolutionary history of the Aeromonas veronii group revealed by host interaction and DNA sequence data. PLoS One 6:e16751. http://dx.doi.org/10.1371/journal.pone.0016751.

26. Morandi A, Zhaxybayeva O, Gogarten JP, Graf J. 2005. Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in Aeromonas hydrophila. J. Bacteriol. 187:6561–6564. http://dx.doi.org/10.1128/JB.187.16.6561-6564.2005.

27. Roger F, Lamy B, Jumas-Bilak E, Kodjo A, colBVH Study Group, Marchand H. 2012. Ribosomal multi-operand diversity: an original perspective on the genus Aeromonas. PLoS One 7:e46268. http://dx.doi.org/10.1371/journal.pone.0046268.

28. Beaz-Hidalgo R, Martínez-Murcia A, Figueras MJ. 2013. Reclassification of Aeromonas hydrophila subsp. dhakensis Huys et al. 2002 and Aeromonas aurariarum Martinez-Murcia et al. 2008 as Aeromonas dhakensis sp. nov. comb. nov. and emendation of the species Aeromonas hydrophila. Syst. Appl. Microbiol. 36:171–176. http://dx.doi.org/10.1016/j.syapm.2012.12.007.

29. Grim CJ, Kozlova EV, Ponnusamy D, Fitts EC, Sha J, Kirtley MI, van Lier CJ, Tiner BL, Erova TV, Joseph SJ, Read TD, Shah JR, Joseph SW, Singletary E, Felland T, Baze WB, Horneman AJ, Chopra AK. 2014. Functional genomic characterization of virulence factors from necrotizing fasciitis-causing strains of Aeromonas hydrophila. Environ. Microbiol. 16:4162–4183. http://dx.doi.org/10.1111/1462-2920.12398.

30. Carnahan AM, Behram S, Joseph SW. 1991. Aerokoy Il: a flexible key for identifying clinical Aeromonas species. J. Clin. Microbiol. 29:2843–2849.

31. Beaz-Hidalgo R, Alper A, Bulaco N, Romaldé JL, Figueras MJ. 2010. Comparison of phenotypical and genetic identification of Aeromonas strains isolated from diseased fish. Syst. Appl. Microbiol. 33:149–153. http://dx.doi.org/10.1016/j.synap.2010.02.002.

32. Martínez-Murcia A, Monera A, Alperi A, Figueras MJ, Saavedra MJ. 2009. Phylogenetic evidence suggests that strains of Aeromonas hydrophila subsp. dhakensis belong to the species Aeromonas aurariarum sp. nov.
Curr. Microbiol. 58:76–80. http://dx.doi.org/10.1007/s00284-008-9278-6.
49. Giltner CL, Bobenchik AM, Uslan DZ, Deville JL, Humphries RM. 2013. Ciprofloxacin-resistant Aeromonas hydrophila cellulitis following leech therapy. J. Clin. Microbiol. 51:1324–1326. http://dx.doi.org/10.1128/JCM.03217-13.
50. Martinez-Muria A, Beaz-Hidalgo R, Svec P, Saavedra MJ, Figueras MJ, Sedlacek I. 2013. Aeromonas caviae caviae sp. nov., isolated from fresh water of a brook in a cavern. Curr. Microbiol. 66:197–204. http://dx.doi.org/10.1007/s00284-012-0253-x.
51. Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102:14332–14337. http://dx.doi.org/10.1073/pnas.0504681102.
52. Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492–508. http://dx.doi.org/10.1080/10635150290069913.
53. Andam CP, David W, Gogarten JP. 2010. Biased gene transfer mimics patterns created through shared ancestry. Proc. Natl. Acad. Sci. U. S. A. 107:10679–10684. http://dx.doi.org/10.1073/pnas.0904263107.
54. Pace NR, Sapp J, Goldenfeld N. 2012. Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. Proc. Natl. Acad. Sci. U. S. A. 109:1011–1018. http://dx.doi.org/10.1073/pnas.1007619109.
55. Williams D, Fournier GP, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.
56. Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247. http://dx.doi.org/10.1093/bioinformatics/17.12.1246.
57. Seshadrini R, Joseph SW, Chopra AK, Sha J, Shaw J, Graf J, Haft D, Wu M, Ren Q, Rosovitch MJ, Madupu R, Tallon L, Kim M, Jin S, Vuong H, Stine OC, Ali A, Horneman AJ, Heidelberg JF. 2005. Genome sequence of Aeromonas hydrophila ATCC 7966™: a back of all trades. J. Bacteriol. 188:8272–8282. http://dx.doi.org/10.1128/JB.00621-06.
58. Spathar O, Farfán M, Allbarral V, Sanglas A, Loren JP, Fusté MC, Bosch A. 2011. Draft genome sequence of Aeromonas salmonicida strain 848TT purified from bivalve molluscs. Genome Announc. 1(3):e00382-13. http://dx.doi.org/10.1128/genomea.00382-13. retrieve your copy from http://mbio.asm.org/Downloaded from on March 7, 2021 by guest
59. Beaton SA, the Graças de Luna M, Bachmann NL, Alkhani NF, Hanks KR, Sullivan MJ, Wee BA, Freitas-Almeida AC, Dos Santos PA, de Melo JT, Squire DJ, Cunningham AF, Fitzgerald JR, Henderson IR. 2011. Genome sequence of the emerging pathogen Aeromonas cararius. J. Bacteriol. 193:1186–1187. http://dx.doi.org/10.1128/JB.01337-10.
60. Wu CJ, Wang HC, Chen CS, Shu HY, Kao AW, Chen PL, Ko WC. 2012. Genome sequence of a novel human pathogen, Aeromonas aquariorum. J. Bacteriol. 194:4114–4115. http://dx.doi.org/10.1128/JB.00621-12.
61. Chan KG, Puthucheary SD, Chan XY, Yin WF, Wong CS, Too WS, Chua KH. 2011. Quorum sensing in Aeromonas species isolated from patients in Malaysia. Curr. Microbiol. 62:167–172. http://dx.doi.org/10.1007/s00284-010-9689-z.
62. Tekedar HC, Waldebeiser GC, Karsi A, Liles MR, Griffin MJ, Vamenta S, Sonstegard T, Hossain M, Schroeder SG, Khoo L, Lawrence ML. 2013. Complete genome sequence of a channel catfish epidemic isolate, Aeromonas hydrophila strain ML09-119. Genome Announc. 1(5):e00755-13. http://dx.doi.org/10.1128/genomea.00755-13.
63. Han JE, Kim JH, Chae YJ, Shin SP, Jun JW, Park SC. 2013. Draft genome sequence of a clinical isolate, Aeromonas hydrophila SNUFAC-88, from a moribund cherry salmon (Oncorhynchus masou masou). Genome Announc. 1(1):e00133-12. http://dx.doi.org/10.1128/genomea.00133-12.
64. Chai B, Wang H, Chen X. 2012. Draft genome sequence of high-melanin-yielding Aeromonas media strain WS. J. Bacteriol. 194:6693–6694. http://dx.doi.org/10.1128/JB.01807-12.
65. Han JE, Kim JH, Shin SP, Jun JW, Chai YJ, Park SC. 2013. Draft genome sequence of Aeromonas salmonicida subs. achronegenes AS503, an atypical strain isolated from crucian carp (Carassius carassius) in the Republic of Korea. Genome Announc. 1(5):e00791-13. http://dx.doi.org/10.1128/genomea.00791-13.
66. Reitl ME, Simko RK, Curtis B, Boyd JM, Bouevitch A, Kimball J, Munholland J, Murphy C, Sarty D, Williams J, Nash JH, Johnson SC, Brown LL. 2008. The genome of Aeromonas salmonicida subs. salmonicina A449: insights into the evolution of a fish pathogen. BMC Genomics 9:427. http://dx.doi.org/10.1186/1471-2164-9-427.
67. Charette CJ, Brochu F, Boyle B, Filion G, Tanaka KH, Derome N. 2012. Draft genome sequence of the virulent strain 01-B526 of the fish pathogen Aeromonas salmonicida. J. Bacteriol. 194:722–723. http://dx.doi.org/10.1128/JB.00276-11.
68. Li Y, Liu Y, Zhou Z, Huang H, Ren Y, Zhang Y, Li G, Zhou Z, Wang L. 2011. Complete genome sequence of Aeromonas veronii strain BS653. J. Bacteriol. 193:3389–3390. http://dx.doi.org/10.1128/JB.00347-11.