Phylogenetic Relationship Among Brackishwater Vibrio Species

J Ashok Kumar1, K Vinaya Kumar1, S Avunje2, V Akhil1, S Ashok1, Sujeet Kumar2, B Sivamani1, Monendra Grover3, Anil Rai3, SV Alavandi2 and KK Vijayan1

1Nutrition Genetics and Biotechnology Division, ICAR-Central Institute of Brackishwater Aquaculture, Chennai, India. 2Aquatic Animal Health and Environment Division, ICAR-Central Institute of Brackishwater Aquaculture, Chennai, India. 3Center for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India.

ABSTRACT: Vibriosis is regarded as an important disease of penaeid shrimps affecting larvae in hatcheries. Among the Vibrio species, Vibrio parahaemolyticus, Vibrio vulnificus, Vibrio furnissii, Vibrio campbellii, Vibrio harveyi, Vibrio alginolyticus, and Vibrio anguillarum are often associated with diseases in finfish and shellfish of brackishwater ecosystem. Accurate species differentiating methods for the organisms present in an ecosystem are required for precise classification of the species and to take steps for their management. Conventional methods like 16s rRNA phylogeny and multilocus sequence typing (MLST) have often failed to correctly identify Vibrio species. This has necessitated a comprehensive investigation on methodologies available to distinguish Vibrio species associated with brackishwater aquaculture system. To achieve this, 35 whole genomes belonging to 7 Vibrio species were subjected to phylogenetic analysis based on 16s rRNA gene, MLST genes, single-copy orthologous genes, and single-nucleotide polymorphisms. In addition, genome-based similarity indices like average nucleotide identity (ANI) and in silico DNA-DNA hybridization (DDH) were computed as confirmatory tests to verify the phylogenetic relations. There were some misclassifications occurred regarding phylogenetic relations based on 16s rRNA genes and MLST genes, while phylogeny with single-copy orthologous genes produced accurate species-level clustering. Study reveals that the species identification based on whole genome-based estimates or genome-wide variants are more precise than the ones done with single or subset of genes.

KEYWORDS: Vibrio, 16s rRNA, MLST single-copy orthologous genes, phylogenetics, ANI, isDDH

Evolutionary Bioinformatics
Volume 16: 1–8
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934320903288

In all, 35 genome sequences were downloaded from GenBank and made readily available for reanalysis studies. Present study aims at finding appropriate methods for differentiating and finding evolutionary distances of Vibrio spp., that are more commonly found in brackishwater ecosystem with the available complete genomes and modern bioinformatics tools.

Materials and Methods
Genomic data

In all, 35 genome sequences were downloaded from GenBank including LB102 strain of V campbellii, a brackishwater isolate sequenced in-house. Complete genomes representing 7 Vibrio

Creative Commons Non Commercial CC BY-NC. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/ham/open-access-at-sage).
spp., *V* parahaemolyticus, *Vibrio vulnificus*, *Vibrio furnissii*, *V. campbellii*, *V. harveyi*, *V. alginolyticus*, and *Vibrio anguillarum* were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/). One genome of *Vibrio cholerae* was also included to serve as an out group for phylogenetic comparisons.

**Phylogenetic analysis**

Data sets with different genomic features representing single 16s rRNA genes and different gene clusters were prepared for phylogenetic analysis. RNAmer v1.2 was used to extract 16s rRNA sequences from genome datasets.\textsuperscript{11} Fetched sequences were trimmed using Bioedit v7.0.5.3, aligned with MEGA version 7, and subjected to phylogenetic analysis using maximum likelihood (ML) method of RAxML v8.2.12.\textsuperscript{12}-\textsuperscript{14} Similarly, housekeeping genes listed at http://pubmlst.org for *Vibrio* spp. were used for phylogenetic analysis using RAxML (Randomized Accelerated ML).\textsuperscript{15}

For phylogeny of orthologous genes, at first coding sequences from the genomes were predicted using Prodigal v2.6.16 Prodigal output was subjected for gene clustering using OrthoMCL v2.0, which has resulted 2085 single-copy orthologous genes.\textsuperscript{17} Each gene present across 36 genomes was aligned using Molecular Evolutionary Genetic Analysis (MEGA) and trimmed using trimAl v1.4 with strictplus option.\textsuperscript{18} All the 2085 genes from each genome were concatenated and subjected to phylogenetic analysis using RAxML. Figgtree v1.4.2 (http://evomis.org/resources/software/molcular-evolution-software/figgtree/) was used for visualizing all the consensus trees generated by RAxML. To construct whole genome–based phylogenetic trees, variant detection and phylogenetic analysis pipeline kSNP3.0 were used. kSNP3.0 accepts genomes as input and does not require genome alignments and reference genomes. It estimates phylogenetic trees by parsimony, neighbor-joining and ML methods.\textsuperscript{19} The number of replicates for bootstrapping was set at 500 for all phylogenetic trees built in this study.

**Genome similarity indices**

Two confirmatory metrics of closeness between the genomes, namely average nucleotide identity (ANI) and Genome-to-Genome Distance, were computed using pyANI v0.20 and GGDC server (http://ggdc.dsmz.de/), respectively.\textsuperscript{20} Heatmaps for visualizing the genome similarity based on ANI and isDDH were generated using R stats library ggplot2.

**Results**

Genome statistics of all the 35 strains along with the out-group entry, that is, *V. cholerae* are given in Table 1. The average number of genes present in each genome is around 4700 genes. Smallest genome among the genomes studied is *V. anguillarum* which contains 3686 genes and the largest being 5818 genes containing *V. campbellii*. *Vibrio campbellii* BB120 has the lowest gene translated to protein with 93% and *V. parahaemolyticus* O3:K6 substr. RIMD 2210633 has highest gene translated percentage of 97%. The guanine–cytosine (GC) content of genomes ranges from 44.37% to 47.49% with an average of 45.68% except for NCTC 11218 isolate of *V. furnissii* whose GC content is 50.63%. Draft assembly of *V. campbellii* LB102 which was isolated from tiger shrimp hatchery located at south east coast of India contains 90 scaffolds and 5145 genes.

The phylogenetic tree with ML method was generated based on 16s rRNA genes from 36 Vibrio genomes (Figure 1) along with branch lengths and bootstrap values. There was no clear distinction observed between the *Vibrio* sp. in this phylogeny. Misclassifications were observed regarding *V. alginolyticus* strain ATCC33787 and *V. vulnificus* YJ016. Even though the members of the harveyi clade, namely *V. harveyi* and *V. campbellii* were found to be clustered together, they were also found in close proximity to other members, such as *V. alginolyticus* and *V. parahaemolyticus*. Lower bootstrap support for many of the nodes is a notable observation in this tree. It is has been observed that bootstrap value as low as 13% for one of the branches signifies poor clustering pattern.

The phylogenetic tree built using housekeeping genes has been depicted in Figure 2. Relatively a better clustering pattern observed in this tree compared to the one built using 16s rRNA genes. Distinct clades between the species were observed except for *harveyi* and *campbellii*. Bootstrap support for branching between species is ranged from 64% to 100%.

The ML tree constructed for 36 *Vibrio* spp. based on orthologous genes has been depicted in Figure 3. The tree indicated distinct monophyletic clades for each of the species considered in this study. Bootstrap support is 100% for between species indicate significance of this feature set. Wrong clustering patterns noticed regarding entries of harveyi clade have been corrected with this method of phylogenetic analysis. Tree generated using kSNP3 (Figure 4) pipeline is observed to be comparable with the tree generated using single-copy orthologous genes.

The ANI values between all the genomes of the study ranged from 0.8323 to 0.9998. The lowest ANI value observed is between *V. anguillarum* 90-11-286 and *V. furnissii* NCTC11218 strains, while the highest is between *V. parahaemolyticus* strains FORC 008 and FORC 018. The heat map generated with ANI values is depicted in Figure 5. Pair-wise computed DDH values for the 36 genomes including out-group have been plotted in Figure 6 to display similarities between the Vibrio species. The values are in ranges from 14.8% to 100%. DDH values of more than 70% are considered to be an important factor in classifying species.\textsuperscript{21}

**Discussion**

The most commonly used bacterial signature sequences are 16s rRNA genes due to their presence in all the bacteria, limited evolutionary changes over time, and sufficiency of their length for analytical purposes.\textsuperscript{22} These gene sequences are widely applied in the studies related to species identification and to calculate evolutionary distances between or within the species. Misclassifications or ambiguities in identification of the species...
Table 1. The genomes of *Vibrio* species used in the study.

| GENOME NUMBER | ORGANISM NAME     | STRAIN         | BIOPROJECT/ACCESSION NUMBER | ASSEMBLY     | GC%  | GENES | PROTEINS | RELEASE DATE |
|---------------|-------------------|----------------|-----------------------------|--------------|------|-------|----------|--------------|
| 1             | *Vibrio alginolyticus* | ATCC 17749    | PRJNA184046                  | GCA_000354175.2 | 44.7 | 4638  | 4457     | 10-09-13     |
| 2             | *Vibrio alginolyticus* | ATCC 33787    | PRJNA305220                  | GCA_001469735.1 | 44.48| 5276  | 5028     | 09-12-15     |
| 3             | *Vibrio alginolyticus* | ZJ-T          | PRJNA326213                  | GCA_001679745.1 | 44.67| 4866  | 4664     | 05-07-16     |
| 4             | *Vibrio anguillarum* | 775           | PRJNA51883                   | GCA_000217675.1 | 44.48| 3656  | 3406     | 30-07-15     |
| 5             | *Vibrio anguillarum* | M3            | PRJNA211964                  | GCA_000462975.1 | 44.45| 3705  | 3479     | 30-08-13     |
| 6             | *Vibrio anguillarum* | NB10          | PRJEB5701                    | GCA_000786425.1 | 44.37| 3985  | 3772     | 02-06-14     |
| 7             | *Vibrio anguillarum* | 90-11-266     | PRJNA266919                  | GCA_001660505.1 | 44.43| 3883  | 3689     | 09-06-16     |
| 8             | *Vibrio campbellii* | LMB29         | PRJNA361283                  | GCA_001969325.1 | 45.02| 5818  | 5537     | 23-01-17     |
| 9             | *Vibrio campbellii* | ATCC BAA-1116; BB120 | PRJNA19857                  | GCA_00017705.1 | 45.4 | 5596  | 5195     | 28-08-07     |
| 10            | *Vibrio campbellii* | LB102         | PRJNA377806                  | GCA_002027615.1 | 45.5 | 5145  | 4920     | 15-03-17     |
| 11            | *Vibrio furnissii* | NCTC 11218    | PRJNA53247                   | GCA_000184325.1 | 50.63| 4541  | 4351     | 17-12-10     |
| 12            | *Vibrio harveyi*    | ATCC 33843 (392 [MAV]) | PRJNA260398                  | GCA_000770115.2 | 44.96| 5332  | 5112     | 31-10-14     |
| 13            | *Vibrio harveyi*    | ATCC 43516    | PRJNA231221                  | GCA_001558435.1 | 44.9 | 5478  | 5248     | 11-02-16     |
| 14            | *Vibrio harveyi*    | —             | PRJNA340970                  | GCA_001908435.2 | 45.05| 5731  | 5427     | 19-12-16     |
| 15            | *Vibrio parahaemolyticus* | O3: K6 substr. RIMD 2210633 | PRJNA360                  | GCA_000196905.1 | 45.4 | 4991  | 4831     | 05-03-03     |
| 16            | *Vibrio parahaemolyticus* | BB22OP        | PRJNA170885                  | GCA_000328405.1 | 45.33| 4636  | 4442     | 21-12-12     |
| 17            | *Vibrio parahaemolyticus* | FDA_R31      | PRJNA203445                  | GCA_000430405.1 | 45.33| 4795  | 4563     | 16-07-13     |
| 18            | *Vibrio parahaemolyticus* | CDC_K4557   | PRJNA203445                  | GCA_000430425.1 | 45.34| 4658  | 4461     | 16-07-13     |
| 19            | *Vibrio parahaemolyticus* | UCM-V493    | PRJNA229758                  | GCA_000568495.1 | 45.32| 4821  | 4612     | 12-02-14     |

(Continued)
| GENOME NUMBER | ORGANISM NAME     | STRAIN       | BIOPROJECT/ACCESSION NUMBER | ASSEMBLY           | GC%   | GENES | PROTEINS | RELEASE DATE |
|---------------|------------------|--------------|----------------------------|--------------------|-------|-------|----------|--------------|
| 20            | Vibrio parahaemolyticus | FORC_008     | PRJNA266097                | GCA_001244315.1    | 45.44 | 4611  | 4408     | 05-08-15     |
| 21            | Vibrio parahaemolyticus | FORC_006     | PRJNA261558                | GCA_001304775.1    | 45.33 | 4691  | 4465     | 01-10-15     |
| 22            | Vibrio parahaemolyticus | FORC_004     | PRJNA259940                | GCA_001433415.1    | 45.49 | 4721  | 4506     | 05-11-15     |
| 23            | Vibrio parahaemolyticus | ATCC 17802   | PRJNA231221                | GCA_001558495.1    | 45.33 | 4659  | 4417     | 11-02-16     |
| 24            | Vibrio parahaemolyticus | FORC_014     | PRJNA280138                | GCA_001636035.1    | 45.35 | 4853  | 4641     | 02-05-16     |
| 25            | Vibrio parahaemolyticus | CHN25        | PRJNA274308                | GCA_001700835.1    | 45.19 | 4974  | 4781     | 09-08-16     |
| 26            | Vibrio parahaemolyticus | FORC_023     | PRJNA284329                | GCA_001758605.1    | 45.44 | 4562  | 4371     | 11-10-16     |
| 27            | Vibrio parahaemolyticus | FORC_018     | PRJNA303095                | GCA_001887055.1    | 45.44 | 4765  | 4510     | 28-11-16     |
| 28            | Vibrio vulnificus    | YJ016        | PRJNA1430                  | GCA_000009745.1    | 46.67 | 4703  | 4522     | 01-10-03     |
| 29            | Vibrio vulnificus    | CMCP6        | PRJNA349                   | GCA_000039765.1    | 46.72 | 4578  | 4374     | 22-12-02     |
| 30            | Vibrio vulnificus    | MO6-24/O     | PRJNA59881                 | GCA_000186585.1    | 46.95 | 4494  | 4324     | 24-01-11     |
| 31            | Vibrio vulnificus    | 93U204       | PRJNA256021                | GCA_000746665.1    | 46.7  | 4600  | 4387     | 29-08-14     |
| 32            | Vibrio vulnificus    | FORC_009     | PRJNA266859                | GCA_001433435.1    | 46.75 | 4513  | 4315     | 05-11-15     |
| 33            | Vibrio vulnificus    | ATL 6-1306   | PRJNA231221                | GCA_001558515.1    | 46.84 | 4436  | 4236     | 11-02-16     |
| 34            | Vibrio vulnificus    | FORC_016     | PRJNA286054                | GCA_001653775.1    | 46.74 | 4514  | 4314     | 01-06-16     |
| 35            | Vibrio vulnificus    | FORC_017     | PRJNA291949                | GCA_001675245.1    | 46.61 | 4689  | 4493     | 22-06-16     |
| 36            | Vibrio cholerae      | N16961       | PRJNA36                    | GCA_000006745.1    | 47.49 | 3693  | 3504     | 03-08-16     |

Abbreviations: GC, guanine-cytosine.
can arise due to poor quality of the sequences available at public
domain databases. But in recent times, due to advent of accurate
and high throughput sequencing technologies, more accurate
and complete prokaryotic genomes are available at public
domain databases for mining and identifying true signatures of
a species required for identification and clustering of it with
related ones. Here in, we used 35 genomes of Vibrios having
relevance with brackishwater aquaculture for finding the phylo-
genetic relations between the species and precise methods for
clustering of the genomes. Correct species-level assignment,
formation of monophyletic clades for each species, and high
degree of bootstrap support are the metrics chosen for deciding
the accuracy of species classification based on phylogenetic
analysis. Misclassifications observed in the phylogenetic tree
with 16s rRNA genes regarding *alginolyticus* ATCC33787 and
*vulnificus* YJ016 strains imply that the variation present in these
genes is not sufficient for proper delineation. With no species-
level monophyletic clades and lesser bootstrap support necessi-
tated other criteria which can take more variation present
between the genomes for accurate classification.

Multilocus sequence typing is a sequence-based approach
to unambiguous characterization of bacterial strains with
sequences of internal fragments of housekeeping genes.23 The
multilocus sequence typing (MLST) databases like the one at
Figure 5. Heatmap generated based on average nucleotide identities for selected Vibrio species.

Figure 6. Similarity matrix based on DDH values for selected Vibrio species. DDH indicates DNA-DNA hybridization.
www.pubmlst.org houses MLST allelic profiles and sequences for different bacterial species. Figure 2, plotted based on MLST sequences, has overcome misclassification errors unlike the one made with 16s rRNA genes. But issues of distinct monophyletic clades and bootstrap support were not found to be addressed with MLST also.

Orthologs are genes that diverged through a speciation event unlike paralogous genes, which diverged after a duplication event. Size of these genes over a set of genomes depends on their evolutionary relationship and the quality of genome assemblies. With 4700 average number of genes among the selected Vibrio spp., 2085 single-copy orthologs were found to be present, which accounts nearly 44% of genes. Phylogenetic analysis based on these single-copy orthologs met all the criteria set for best classification (Figure 3). Monophyletic clades with 100% bootstrap support for between species indicate superiority of this feature set compared to previous ones. The utility of orthologous genes to address species ambiguities was demonstrated by Ke et al in an attempt to correct the misclassification of V campbellii with the strains present in harveyi clade. But the clustering, extraction, and curation of the sequences and tree building with single-copy orthologous genes were observed to be computationally intensive and time consuming.

kSNP 3.0 pipeline identifies single-nucleotide polymorphisms (SNPs) from the input genomes and does phylogeny based on the core SNP data matrices, which include only SNPs detected at loci that were present in all genomes. The program runs faster and requires less memory compared to previous approach. Phylogenetic tree generated from kSNP 3.0 is comparable to the one built with single-copy orthologous genes (Figure 4) in terms of accuracy in classification.

The complete genome is used as reference standard to determine phylogeny which in turn determines taxonomy of the species in genome-based similarity indices. Average nucleotide identity is one such method which depends on large number of genes unlike the ones depends on 16s rRNA, and it was considered to be better measure of relatedness. When the large number of genes considered for estimating relatedness of genomes, the indices are unaffected by varied evolutionary rates of the genes as fast-evolving genes are compensated by slow-evolving genes. Dendrogram based on ANI values in Figure 5 accurately classified each species and made monophyletic clades. Genome-genome distance measure isDDH is another metric of relatedness computed using whole genomes. Average nucleotide identity values of 95% is equal to DDH values of 70%, which signifies highly related species. In Figure 6, clusters having more than 70% similarity belonged to the same species. Both the measures have confirmed the clustering patterns established through single-copy orthologous genes. These similarity metrics have become gold standard for in silico species identification in recent times.

Conclusions

Information on bacterial species present in an ecosystem along with their phylogenetic distances and right methods for identification or classification has got significance in evolutionary biology. Here, we used different subsets of whole genome data, namely, 16s rRNA gene, MLST genes, and single-copy orthologous genes pertaining to 35 Vibrio species which are native to brackishwater ecosystem. Phylogenetic trees based on 16s rRNA and MLST sequences resulted the wrong classification patterns. To clear the ambiguities in classification, it was further tested with single-copy orthologous genes dataset as well as kSNP 3.0 pipeline. Clearly distinguished clades were observed in both of these methods among which kSNP3.0 requires less computational resources. Genome similarity indices like ANI and in silico DDH methods supported the validity of trees built with single-copy orthologous genes and kSNP3.0. The work needs to be continued from time to time by including newly sequenced genomes with proposed methods to clear ambiguities in classification of a new species.

Author Contributions

JAK, KVK conceived the study, JAK, VA, and SAs conducted the analysis. JAK drafted the manuscript with inputs from SAV, KVK, SK, and BS, MG, AR, SVA, and KKV reviewed the manuscript and gave critical comments.

ORCID iD

J Ashok Kumar https://orcid.org/0000-0002-2213-2931

REFERENCES

1. Austin B, Zhang XL. Vibrio harveyi: a significant pathogen of marine vertebrates and invertebrates. Lett Appl Microbiol. 2006;43:119-124.
2. Ruwandeepika HAD, Sanjeewa Prasad Jayaweera T, Pahan Bhowmick P, Karunasagar I, Bossier P, Defioriti T. Pathogenesis, virulence factors and virulence regulation of Vibrio belonging to the Harveyi clade. Rev Aquac. 2012;4:59-74.
3. Harris L. The involvement of toxins in the virulence of Vibrio harveyi strains pathogenic to the black tiger shrimp Penaeus monodon and the use of commercial probiotics to reduce shrimp hatchery disease outbreaks caused by V. harveyi strains. Paper presented at CRC for Aquaculture, Scientific Conference Abstract; 27-29 September 1995; Bribie Island, QLD, Australia.
4. Tran L, Nusan L, Redman RM, et al. Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. Dis Aquat Organ. 2013;105:45-55.
5. Zoritchazhra MJ, Banauderakhshan R. Early mortality syndrome (EMS) as new emerging threat in shrimp industry. Adv Anim Vet Sci. 2015;3:64-72.
6. Ke H-M, Prachunuwat A, Yu C-P, et al. Comparative genomics of Vibrio campbellii strains and core species of the Vibrio harveyi clade. Sci Rep. 2017;7:41394. doi:10.1038/srep41394.
7. Grimes DJ, Singleton FL, Colwell RR. Allergic succession of marine bacterial communities in response to pharmaceutical waste. J Appl Bacterial. 1984;57:247-261.
8. Dryselius R, Itunou K, Honda T, Iida T. Differential replication dynamics for large and small Vibrio chromosomes affect gene dosage, expression and location. BMC Genomics. 2008;9:559. doi:10.1186/1471-2164-9-559.
9. Grimes DJ, Johnson CN, Dillon KS, Flowers AR, Noria NF, Berutti T. What genomic sequence information has revealed about vibrio ecology in the ocean—a review. Microb Ecol. 2009;58:447-460. doi:10.1007/s00248-009-9578-9.
10. Kumar S, Ashok Kumar J, Akhil V, et al. Draft genome sequence of the luminescent strain Vibrio campbellii LB102, isolated from a black tiger shrimp (Penaeus monodon) broodstock rearing system. Genome Announc. 2017;5:e00342-17.
11. Lagesen K, Hallin P, Rodland EA, Starefeldt HH, Rogne T, Ussery DW. RNAHammer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100-3108.
12. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95-98.

13. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870-1874. doi:10.1093/molbev/msw305.

14. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312-1313. doi:10.1093/bioinformatics/btu033.

15. Jolley KA. PubMLST website—publicly-accessible MLST databases and software (WWW Document). http://pubmlst.org

16. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119. doi:10.1186/1471-2105-11-119.

17. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178-2189. doi:10.1101/gr.1224503.

18. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972-1973.

19. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics.* 2015;31:2877-2878. doi:10.1093/bioinformatics/btv271.

20. Pirianch L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Annu Methods.* 2016;8:12-24. doi:10.1039/C5AY02550H.

21. Johnson JL, Whitman WB. Similarity analysis of DNA. In: Reddy CA, Beveridge TJ, Breznak JA, Marzluf GA, Schmidt TM, Snyder LR, eds. *Methods for General and Molecular Microbiology.* 3rd ed. Washington, DC: American Society of Microbiology; 2007:624-625.

22. Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mil Diagn.* 2001;6:313-321.

23. Maiden MCJ, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA.* 1998;95:3140-3145.

24. Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool.* 1970;19:99-113.

25. Wayne LG. International Committee on Systematic Bacteriology: announcement of the report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Zentralbl Bakteriol Mikrobiol Hyg A.* 1988;268:433-434.

26. Arzahal DR. Whole-genome analyses: average nucleotide identity. In: Goodfellow M, Sutcliffe I, Chun J, eds. *Methods in Microbiology,* vol. 41. Cambridge, MA: Academic Press; 2014:103-122.

27. Goris J, Konstantinidis KT, Klappebach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol.* 2007;57:81-91.