Does a DNA barcoding gap exist in bioactive marine bacteria?
Evidence from analyses of 16S rRNA gene sequence data of pigmented pseudoalteromonads

[1]Arizaldo E. Castro, MSc and [2]Ian Kendrich C. Fontanilla, PhD

[1]Microbial Ecology of Terrestrial and Aquatic Systems Laboratory, Institute of Biology,
College of Science, University of the Philippines Diliman

[2]DNA Barcoding Laboratory, Institute of Biology,
College of Science, University of the Philippines Diliman

Abstract

16S rRNA gene sequences have been extensively used in assigning taxonomic identities and in elucidating phylogenetic relationships among prokaryotes. One bacterial genus that is largely identified using 16S rRNA gene sequence data is the *Pseudoalteromonas*. To date, there is limited to no published evidence supporting the certainty of using 16S rRNA gene sequence data in identifying pseudoalteromonads and describing phylogenetic relationships within the group. In this *in silico* study, the suitability of using 16S rRNA gene sequences in taxonomic and phylogenetic resolution within *Pseudoalteromonas* was examined. 195 partial 16S rRNA gene sequence data were analyzed for difference in intraspecific and interspecific genetic distances and for the presence of a DNA barcoding gap. Gene trees were also constructed using Maximum Likelihood and Maximum Parsimony approaches. Using the uncorrected and Kimura-2-parameter (K2P)-corrected pairwise distances, it was found that interspecific distances were 3.5 to 4 times greater than intraspecific distances and there is an absence of a DNA barcoding gap. In addition, the constructed ML and MP gene trees supported the monophyly of 8 out of the 13 species of pigmented pseudoalteromonads. These findings imply that 16S rRNA gene sequence divergence among the members of pigmented pseudoalteromonads are small and there is a need for caution in using 16S rRNA gene sequence data alone when identifying putative marine pigmented *Pseudoalteromonas* species. In contexts where resources are not a limiting factor, Multi-Locus Sequence Analysis (MLSA), which concatenates gene sequences from multiple conserved genes, is a more robust approach than single-gene molecular identification and phylogeny reconstruction.

Keywords: DNA barcoding gap, 16S rRNA, Pseudoalteromonas, Genetic distances, Maximum likelihood, Maximum Parsimony

Introduction

New molecular entities identified and approved by the United States Food and Drug Administration are growing smaller each year, leading to the exploration of untapped natural sources of antimicrobial substances (Spellberg et al., 2008). In response, worldwide attention has been given to identifying and purifying novel antimicrobial substances. Key organisms of interest include actinobacteria (Velho-Pereira & Kamat, 2011), plants (Coates et al., 2011), endophytic fungi (Hazalin et al., 2009), and diverse marine organisms capable of producing chemically-unique and active secondary metabolites (Mayer et al., 2013). Marine bacteria are an example of organisms that can be potential sources of novel antimicrobials.

One noteworthy marine bacterial genus that is currently studied for its anti-infective activities is the *Pseudoalteromonas.* This genus includes marine, heterotrophic, gram-negative, rod-shaped, and aerobic bacteria that are non-spore forming, non-bioluminescent, and motile with single polar flagellum that may be sheathed or not. When grown in solid agar, pseudoalteromonad colonies may be pigmented or not (Bowman, 2007), with pigmented species usually associated to producing bioactives (Vynne et al., 2011). In terms of byproducts, *Pseudoalteromonas* species produce non-protein small molecules and protein substances such as enzymes and small peptides (Vynne et al., 2011). Active compounds such as brominated metabolites (Feher et al., 2010), tambjamines (Franks et al., 2005), and polyketides (Mayer et al., 2013) are just few of the chemotherapeutic-significant products of the genus. Some bioactive compounds associated with pigmented and non-pigmented *Pseudoalteromonas* species have been systematically reviewed by Offret et al. (2016). As antagonistic marine bacteria, members of the group were found to be frequently associated with biotic and abiotic surfaces. Many reports cover in detail the isolation of pseudoalteromonads from sources such as nudibranchs (Feher et al., 2010), seawater (Oh et al.,

2011), algae (Egan et al., 2001), and corals (Shnit-Orland & Kushmaro, 2009). Although these reports detail empirical characteristics of several pseudoalteromonads, one major challenge is the identification of pigmented species from one another. Biochemical and morphology-based methods of identification offer small degrees of certainty. Currently, no standard biochemical panel test can be used to distinguish pigmented *Pseudoalteromonas* species from each other and phenotypic data like colony and cellular morphology only offer little to no substantial information. In majority of published literature, molecular data are used to identify members of the genus down to the species level.

Prokaryotic taxonomy and phylogeny have been examined largely through analyses and comparisons of the 16S rRNA gene sequences. Identification of prokaryotic operational taxonomic units (OTUs) and their assignment to known taxa are based on predetermined fixed thresholds in genetic distances of the 16S rRNA gene ranging from 97% to 99%. The determination of a DNA barcoding gap would be helpful in defining thresholds to use and deciding whether or not a particular gene could really be used as basis for identification. There is limited published evidence supporting the use of 16S rRNA gene data sequence alone in delineating and determining identities of potentially bioactive pigmented pseudoalteromonads. In response, this study was conducted with the following specific objectives: (1) compare sequence divergence of pigmented *Pseudoalteromonas* species included in this *in silico* analysis; (2) construct 16S rRNA gene trees to confirm and support the clustering of species that have been previously reported; and (3) determine the presence of a DNA barcoding gap in pseudoalteromonads' 16S rRNA gene sequences.

Materials and methods

Data sources

16S rRNA gene sequences of 195 pigmented pseudoalteromonad strains and isolates were downloaded from GenBank®. Partial gene sequences were limited to 1000-1500 base pairs (bp) in length and those that were reported from 2013 up to 2017.

Data treatment and analysis

Downloaded sequences were aligned using BioEdit Sequence Alignment Editor (v. 7.2.6; Hall, 1999). Afterwards, the optimum model of DNA substitution was chosen by model testing in jModelTest 2.1.10 (v. 20160303; Darriba et al., 2012) following the Akaike Information Criterion (AIC). In order to determine if the data set is saturated or not, a test for substitution was conducted in Data Analysis in Molecular Biology and Evolution (DAMBE v. 5.2.30) following the protocol of Xia et al. (2003).

16S rRNA gene tree construction

Gene trees were produced using Maximum Likelihood (ML) and Maximum Parsimony (MP) methods. PhyML (v. 3.1; Guindon et al., 2010) was used for the construction of the ML tree following the optimal DNA substitution model. On the other hand, the MP gene tree was constructed using MEGA X (v. 10.0.5; Kumar et al., 2018). In both gene tree constructions, bootstrap test using 1000 replicates were included.

DNA barcoding gap

Sequence divergences of individuals within the same species (intraspecific) and individuals from different species (interspecific) were computed in Phylogenetic Analysis Using Parsimony program (PAUP* v. 4.0b10; Swofford, 2002). On the other hand, K2P-corrected distances were computed using MEGA X (v. 10.0.5; Kumar et al., 2018). Using Microsoft Excel©, frequency of

intraspecific and interspecific pairwise distances were plotted as superimposed histograms in order to graphically visualize the presence of a DNA barcoding gap.

Results and discussion

Alignment of 16S rRNA gene sequences, test for saturation, and optimal substitution model

The aligned 16S rRNA dataset included a mix of 195 strains and isolates of pigmented pseudoalteromonads belonging to 13 species. The length of the aligned sequences was 1342 nucleotides. Moreover, saturation test using Xia et al.'s protocol in DAMBE showed that the observed index of substitution saturation (0.041) for 32 operational taxonomic units was significantly smaller than the indices of substitution saturation for completely symmetrical (0.743) and asymmetrical (0.041) trees. This means that the dataset was not saturated and was useful for phylogenetic analyses. Prior to the construction of gene trees, model testing was done and using the AIC, TPM2uf+G was determined as the optimal DNA substitution model for the dataset.

16S rRNA gene sequence divergences

Sequence divergences using average uncorrected distances of individuals within the same species were found to be less than 1% (0.98%) compared to genetic distances of individuals from different species, which were at 3.54%. Using K2P, similar observations were also seen with distances of individuals in the same species at 1% compared to genetic distances of individuals from different species, which were at 4.08%. Intraspecific genetic distances ranged from 0 to 0.140 while interspecific genetic distances ranged from 0 to 0.168.

Gene trees of pigmented pseudoalteromonads

The Maximum Parsimony (MP) approach finds the tree topology for a set of aligned sequences that can be explained with the smallest number of character changes or evolutionary

steps, i.e., substitutions. After a reasonable number of topologies have been evaluated, the tree that requires the minimum number of changes is selected as the MP tree. In the analyses, the MP approach produced the consensus tree (Figure 1) inferred from three most parsimonious trees. Individuals from *P. luteoviolacea, P. ruthenica, P. spongiae, P. byunsanensis, P. ulvae, P. tunicata, P. phenolica* and *P. aurantia* were all supported as monophyletic groups. All of the eight groups produced 100% bootstrap supports in the MP tree. On the other hand, strains and isolates from *P. piscicida, P. citrea, P. flavipulchra, P. rubra,* and *P. maricaloris* were not supported as monophyletic groups.

A second gene tree was constructed using the Maximum Likelihood (ML) approach. The ML method of tree construction involves the computation of the likelihood of a phylogenetic tree, which is the probability of observing the data at hand under a given tree and a specified optimal model of character state changes. Using ML involves finding the tree with the highest likelihood value. In the analyses, the ML approach produced a tree (Figure 2) showing similar topology with that of the MP tree. Individuals from *P. luteoviolacea, P. ruthenica, P. spongiae, P. byunsanensis, P. ulvae, P. tunicata, P. phenolica* and *P. aurantia* were all supported as monophyletic groups and with bootstrap support values of 99%, 72%, 95%, 100%, 100%, 100%, 70%, and 78% respectively. Similarly, strains and isolates from *P. piscicida, P. citrea, P. flavipulchra, P. rubra,* and *P. maricaloris* were not supported as monophyletic groups as can be inferred from the ML tree.
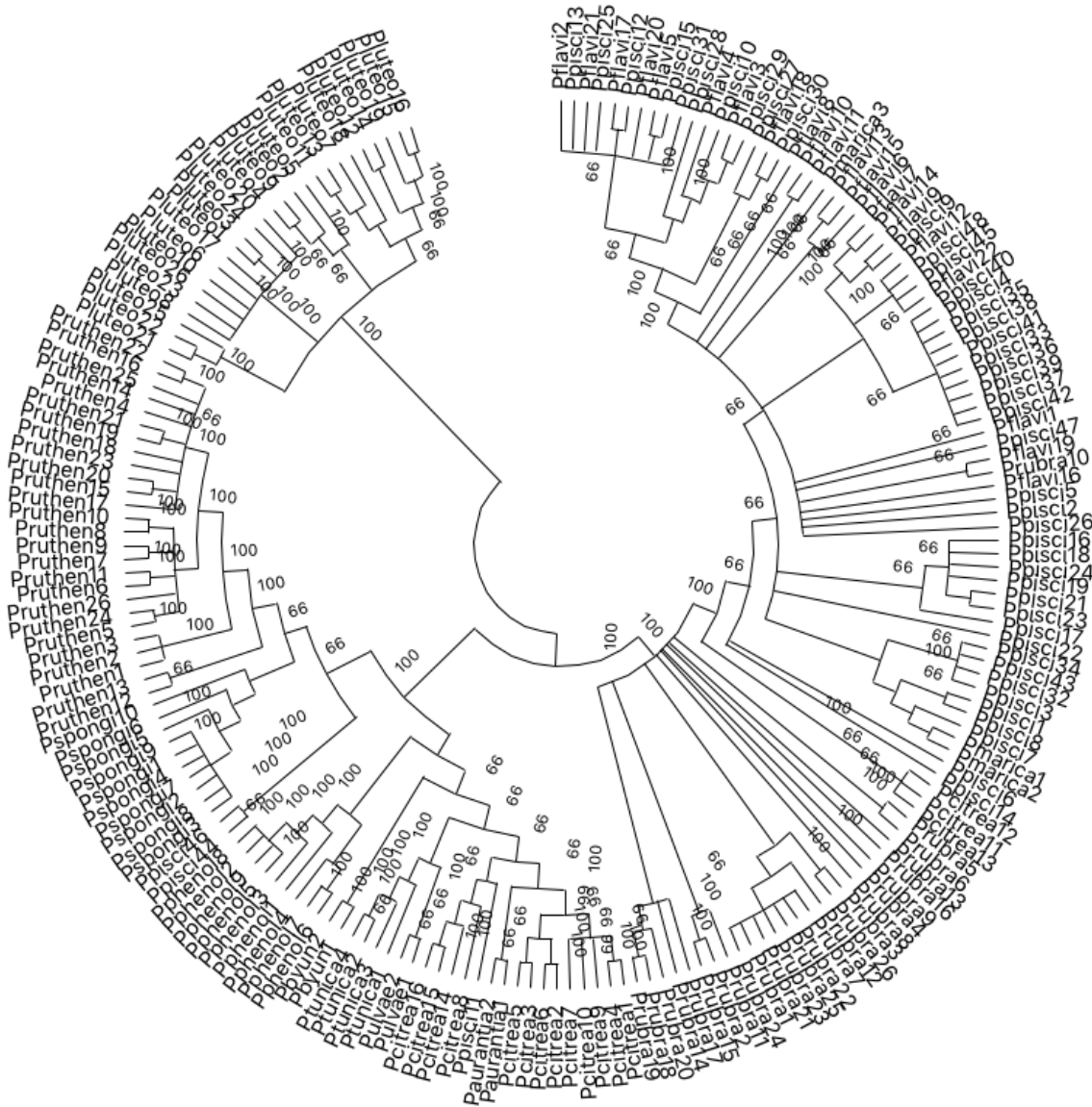
Figure 1. Majority consensus maximum parsimony tree of the pigmented pseudoalteromonads based on 1342 nucleotides of the 16S rRNA gene and 3 equally-parsimonious trees. The tree was produced using Subtree-Pruning-Regrafting (SPR) algorithm. The tree is rooted on *P. luteoviolacea*, which possesses the most number of biosynthetic pathways among the pigmented strains and isolates. Numbers on nodes represent bootstrap values out of 1000 replicates. Values less than 50% are not shown.
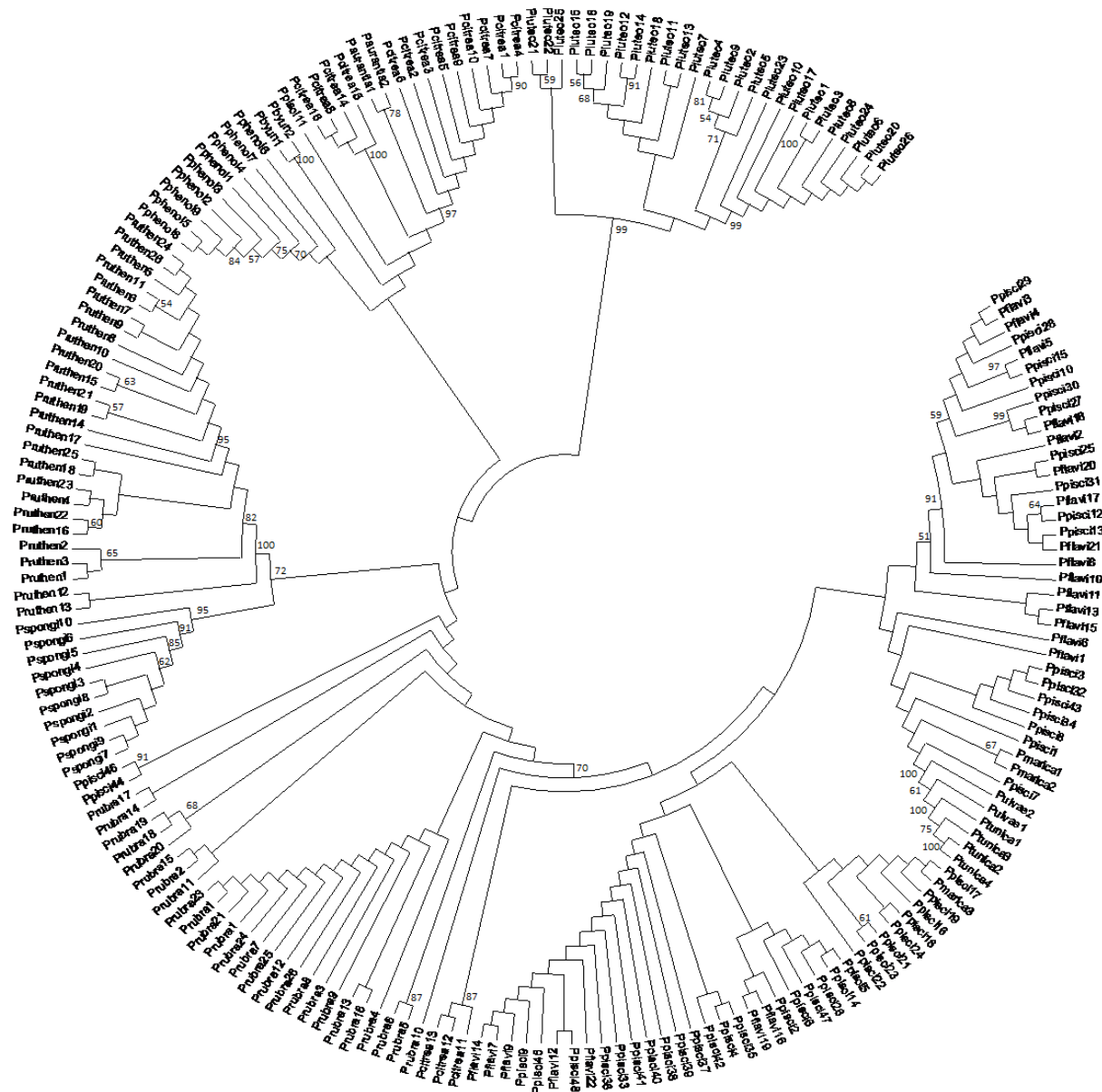
Figure 2. Maximum likelihood tree of the pigmented pseudoalteromonads based on 1342 nucleotides of the 16S rRNA gene and using the TPM2uf+G model of DNA substitution. The tree is rooted on *P. luteoviolacea*, which possesses the most number of biosynthetic pathways among the pigmented strains and isolates. Numbers on nodes represent bootstrap values out of 1000 replicates. Values less than 50% are not shown.

Clustering of pigmented pseudoalteormonads as seen in MP and ML trees conform to the phylogeny inferred by Vynne et al. (2011) and Bowman (2007) using 16S rRNA data. Individuals belonging to the eight species whose monophyly were supported by the constructed trees exhibited similar topology with 16S rRNA gene trees from Vynne et al.'s study (2011) and Bowman's study (2007). Individuals belonging to the cluster of pseudoalteromonads separating *P. rubra* from *P. flavipulchra*, *P. maricaloris*, and *P. piscicida* did not form consistent monophyletic groups, something that may be attributed to the small level of genetic divergence within the cluster.

The absence of a DNA barcoding gap in bioactive pseudoalteromonads

Plotting the frequency of pairwise interspecific and intraspecific distances, no barcoding gap could be visualized as the smallest interspecific distance is flanked by intraspecific ones (See Figure 3). The absence of a barcoding gap implies uncertainty in using genetic distances as basis for assigning isolates to known species using 16S rRNA gene similarity. The average distance between species of each of the 13 *Pseudoalteromonas* species were 3.5-4 times higher than the average distance between individuals within the same species as indicated by uncorrected and K2P corrected pairwise distances. It can be inferred that pseudoalteromonads of the pigmented type fall into discrete clusters that are closely related. Such observations on genetic distances might be attributable to the fact that pigmented pseudoalteromonads exhibit niche specificity (Vynne et al. 2011) or mechanisms in their respective environments are not that competitive to initiate diversification from within clusters.
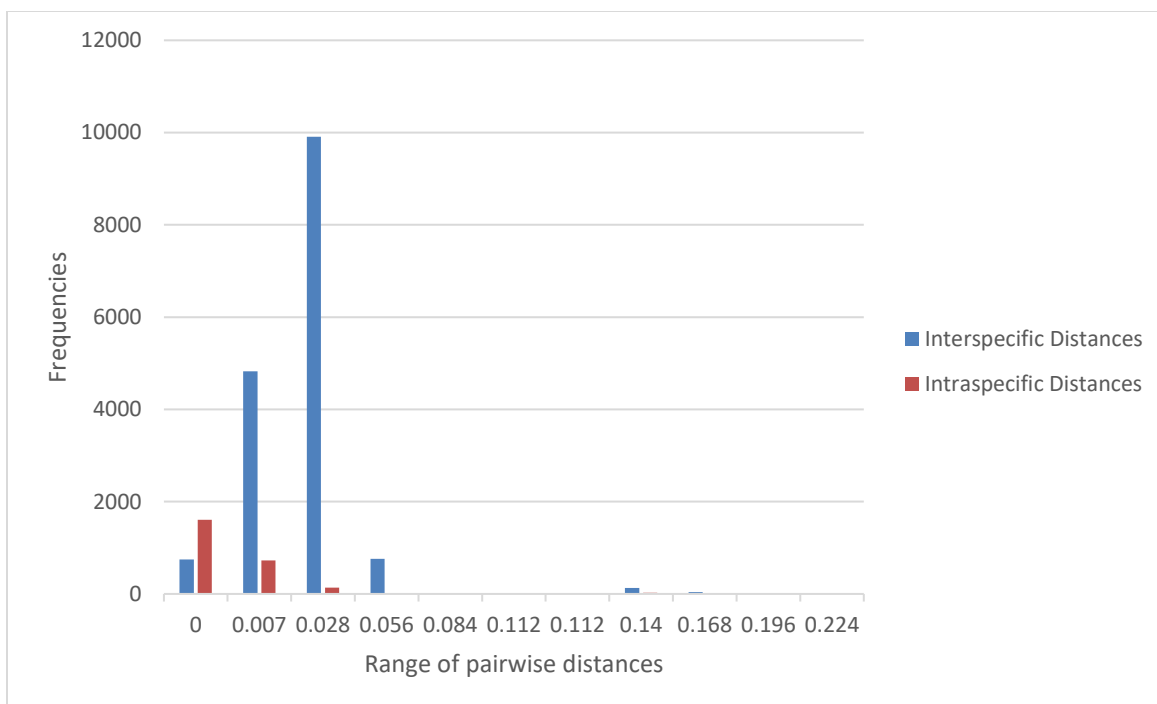
Figure 3. Distribution of intraspecific and interspecific distances within the 195 pigmented

Pseudoalteromonads based on 1342 nucleotides of the 16S rRNA gene

The overlap of intraspecific and interspecific distances in the 16S rRNA gene sequence data of the pigmented pseudoalteromonads may be attributed to incorrect identification of individual strains and isolates (Collins & Cruickshank, 2012). Misidentification may happen when there are multiple investigators on the same taxa that may lead to conflicting identifications of the same species. For the case of pigmented pseudoalteromonads, clustering of different species as monophyletic groups supports conflicting identification of some species. In both gene trees constructed, isolates from *P. piscicida, P. citrea, P. flavipulchra, P. rubra,* and *P. maricaloris* were not supported as monophyletic groups.

The absence of a barcoding gap serves as a reminder that it is not enough to use 16S rRNA gene data alone in identifying potentially bioactive marine prokaryotes.

Conclusion and recommendation

This *in silico* study reports an analysis of 16S rRNA gene sequence data of 195 individuals from 13 pigmented *Pseudoalteromonas* species. From the results, the interspecies divergence of individuals was found to be 3.5 to 4 times higher compared to intraspecific genetic distances. Moreover, based on the plotted frequency distribution of intraspecific and interspecific distances, there is an absence of DNA barcode gap in pigmented pseudoalteromonads. Phylogenetic analyses using the 16S rRNA gene and employing the ML and MP approaches produced tree topologies that conform to previously reported analyses of the group (Vynne et al., 2011 and Bowman, 2007). Since pseudoalteromonads are still underexplored sources of antimicrobials (Offret et al., 2016), identifying isolated individuals in a fast and accurate manner is a very important step in biosourcing. Using non-whole genome data, particularly 16S rRNA gene sequences, in identifying bioactive strains of marine bacteria is found to be functional as supported by the findings of this study. However, relying on 16S rRNA gene sequence alone must be avoided as much as possible. Instead a systematic polyphasic approach should be considered where other data types such as cellular, biochemical, colonial, and chemotaxonomic characteristics are taken into account when identifying unknown bacterial isolates (Gevers et al., 2005). Multi-Locus Sequence Analysis (MLSA), an alternative to single-gene molecular identification and phylogeny reconstruction, could also be employed. In MLSA, where multiple genes are concatenated and used, there is buffer against distortion brought by recombination in a single locus.

Our current knowledge base on population genetics, microbial ecology, microbial genomics, and the advances in nucleotide sequencing technology have tremendous impacts on our understanding of what constitutes a prokaryotic species. Indeed, gene sequence data revolutionized the contemporary scientific workflow of identifying bacterial isolates from natural and built

environments. Therefore, recognizing the limitations of using specific genes as markers for taxonomic identity is a must.

References

Bowman J.P., 2007. Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *Pseudoalteromonas*. Mar. Drugs 5(4), 220-241.

Coates A.R.M., G. Halls, & Y. Hu, 2011. Novel classes of antibiotics or more of the same? Br J Pharmacol 163(1): 184–194.

Collins R.A. & R.H. Cruickshank, 2013. The seven deadly sins of DNA barcoding. Mol Ecol Resour 13(6):969-75.

Darriba D., G.L. Taboada, R. Doallo, & D. Posada, 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9(8), 772.

Egan S., C. Holmstrom, & S. Kjelleberg, 2001. *Pseudoalteromonas ulvae* sp. nov., a bacterium with antifouling activities isolated from the surface of a marine alga. Int J Syst Evol Microbiol; 51(Pt 4):1499-504.

Feher D., R. Barlow, J. McAtee, & T.K. Hemscheidt, 2010. Highly brominated antimicrobial metabolites from a marine *Pseudoalteromonas* sp. J Nat Prod. 73(11): 1963-1966.

Franks A., P. Haywood, C. Holmstrom, S. Egan, S. Kjelleberg, & N. Kumar, 2005. Isolation and structure elucidation of a novel yellow pigment from the marine bacterium *Pseudoalteromonas tunicata.* Molecules 10, 1286–1291.

Franks A., S. Egan, C. Holmstrom, S. James, H. Lappin-Scott, & S. Kjelleberg, 2006. Inhibition of fungal colonization by *Pseudoalteromonas tunicata* provides a competitive advantage during surface colonization. Appl Environ Microbiol. 72(9): 6079–6087.

Gevers D, F.M. Cohan, J.G. Lawrence, B.G. Spratt, T. Coenye, E.J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F.L. Thompson, & J. Swings. 2005. Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol. 3(9):733-9.

Guindon S., J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, & O. Gascuel, 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 59(3): 307-21.

Hall T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Hazalin N.A.M.N., K. Ramasamy, S.M. Lim, I.A. Wahab, A.L.J. Cole, & A.B.A. Majeed, 2009. Cytotoxic and antibacterial activities of endophytic fungi isolated from plants at the National Park, Pahang, Malaysia. BMC Complement Altern Med. 9:46.

Kumar S., G. Stecher, M. Li, C. Knyaz, & K. Tamura, 2018. MEGA X: Molecular Evolutionary Genetic Analysis across Computing Platforms. Mol Biol Evol 35(6): 1547-1549.

Mayer A.M.S., A.D. Rodriguez, O. Taglialatela-Scafati, & N. Fusetani, 2013. Marine pharmacology in 2009-2011: marine compounds with antibacterial, antidiabetic, antifungal, anti-inflammatory, antiprotozoal, antituberculosis, and antiviral activities; affecting the immune and nervous systems, and other miscellaneous mechanisms of action. Mar. Drugs 11, 2510-2573.

Offret C., F. Desriac, P. Le Chevalier, J. Mounier, C. Jégou, & Y. Fleury, 2016. Spotlight on Antimicrobial Metabolites from the Marine Bacteria *Pseudoalteromonas*: Chemodiversity and Ecological Significance. Mar Drugs 14(7). pii: E129.

Oh Y.S., A.R. Park, J.K. Lee, C.S. Lim, J.S. Yoo, & D.H. Roh, 2011. *Pseudoalteromonas donghaensis* sp. nov. isolated from seawater. Int J Syst Evol Microbiol 61, 351-355.

Shnit-Orland M. &A. Kushmaro, 2009. Coral mucus-associated bacteria: a possible first line of defense. FEMS Microbiol Ecol. 67(3):371-80.

Swofford D.L., 2002. Phylogenetic Analysis Using Parsimony (and other methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.

Spellberg B., R. Guido, D. Gilbert, J. Bradley, H.W. Boucher, W.M. Scheld, J.G. Bartlett & et al., 2008. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin Infect Dis. 46 (2):155-164.

Velho-Pereira S. & N.M. Kamat, 2011. Antimicrobial Screening of Actinobacteria using a Modified Cross-Streak Method. Indian J Pharm Sci 73(2): 223–228.

Vynne N.G., M. Mansson, K.F. Nielsen, & L. Gram, 2011. Bioactivity, chemical profiling, and 16S rRNA-based phylogeny of *Pseudoalteromonas* strains collected on a global research cruise. Mar. Biotechnol 13:1062- 1073.

Xia X., Z. Xie, M. Salemi, L. Chen, & Y. Wang, 2003. An index of substitution saturation and its application. Mol Phylogenet Evol 26(1):1-7.