Genomic Epidemiology of *Vibrio* Species in the Coastal Ecosystem

By

PETER JOSEPH SEBASTIAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Christine Kreuder Johnson, Chair

_____

Bart C Weimer

_____

Barbara A Byrne

Committee in Charge

2023

TABLE OF CONTENTS

DEDICATION

I would like to dedicate this dissertation to my family who have supported my goals and aspirations at every step: my wife Dana, my mother and stepfather Pam and Ed, my father and stepmother Paul and Shari. Dana, you have been by my side throughout almost my entire PhD and have been patient and encouraging of my professional growth. I can't wait to experience the next phase of life together with you wherever that takes us. To Pam and Ed, thanks for your curiosity in the work I do and for making me feel cared for from half the country away. To Paul and Shari, thanks for reminding me to believe in myself and to not be afraid to take chances. Finally, to Kirby, who was my daily companion and a source of inspiration during both veterinary school and much of my PhD.

ACKNOWLEDGMENTS

Firstly, I would like to thank my dissertation committee, Christine K. Johnson, Barbara A. Byrne, and Bart C. Weimer, each of whom provided unique and needed expertise as well as guidance and support throughout my dissertation. Working with this committee has provided me invaluable insight for how to conduct cross-disciplinary research that is the cornerstone of the One Health approach.

Chris, as my major professor you have consistently provided opportunities for me to grow and succeed including hiring me as a graduate researcher during my first year to encouraging me to apply for highly competitive grants like the T32 award. You have exemplified so much of how to succeed as a professional researcher and how to think beyond the label of a wildlife epidemiologist as an ambassador for many aspects of global health. I am inspired by you and incredibly thankful for your mentorship.

Barb, this dissertation would not have been possible without your collaborative and supportive nature. You have been incredibly gracious with your time including helping me dig through freezer boxes for isolates over 20 years old and pouring through old files to make sense of metadata. I am in awe that a wildlife project of this size scope was possible thanks to your vision.

Bart, collaboration with your lab took this project in an extremely new direction that has enhanced the work immeasurably. You helped demystify genomics and fostered a group of like-minded students to think critically about how to use genomic data. I am incredibly grateful the

amount of time and advice you have provided that has made this dissertation considerably more impactful.

I would also like to thank Cory Schlesener, who has been the go-to expert of the Weimer lab for bioinformatics analysis and who has assisted non-genomics researchers like me with indispensable expertise. The genomic epidemiology approach in this dissertation was only possible because of the kind of collaboration that I experienced working with you.

Thanks to Carol Huang, who is the engine of the Weimer lab. Her expertise in genomic library preparation and molecular laboratory procedures has made it possible to sequence genomes on a bootstrapped budget.

Thanks to the past and present researchers that I have had spent time with at the Epicenter for Disease Dynamics that have been sounding boards and great sources of inspiration including but definitely not limited to Megan Doyle, Tierra Smiley Evans, Nicole Gardner, Megan Moriarty, Pranav Pandit, Jessica Sanchez, and Cristin Young. Thanks to the similarly helpful scientific discussions of the weekly Weimer Lab meetings that have improved my research and the many students that have attended them, especially DJ Darwin Bandoy and Katie Lee who have conducted inspirational genomic epidemiology investigations. Thanks to the additional Graduate Group in Epidemiology students that have provided much needed comradery during the graduate school journey including Sarah Abusaa, Laura Backus, Jose Pablo Gomez, Kimberly Prado, Karla Rascon-Garcia, and Monica Zuercher.

Finally, thanks to the many other collaborators and professionals that have inspired me and provided helpful insights and advice, including but not limited to: the late Jack Ames, Heejung

Bang, Francesca Batac, Brian Bird, the late Matt Blake, Kathy Burek-Huntington, Carrie Goertz, Tracey Goldstein, Heather Harris, Danielle Harvey, Danielle Carrade-Holt, Natalie Hunter, Kristin Jankowski, Melissa Miller, Mike Murray, Natalie Rouse, Brett Smith, Woutrina Smith, Michelle Staedler, Kelly Stetter, Lorraine Thompson-Barbosa, Marissa Young, Michael Ziccardi.

ABSTRACT

Genomic Epidemiology of *Vibrio* Species in the Coastal Ecosystem

Humans and marine mammals such as sea otters (*Enhydra lutris*) are exposed to pathogenic *Vibrio* species through consumption of marine invertebrates and contact with coastal waters. Despite the disease threats of vibriosis in humans and marine mammals, One Health focused *Vibrio* spp. comparative epidemiology studies are limited. This dissertation combines often-disparate approaches of comparative genomics and epidemiology to advance research for vibriosis management in three aspects: 1) antimicrobial resistance, 2) virulence detection, and 3) diagnostic resolution.

Chapter 1 characterized both phenotypic (n=165) and genomic (n=444) antimicrobial resistance patterns of *Vibrio* spp. isolated between 2000-2019 from coastal Alaska, California, and Washington and from various sources including southern sea otters (*Enhydra lutris nereis*), northern sea otters (*Enhydra lutris kenyoni*), shellfish, and environmental samples. Using genomic methods, a previously misidentified species, *Vibrio diabolicus*, was detected in northern and southern sea otters and their coastal environment. Ampicillin resistance was common in *Vibrio* spp. although there was discordance between genotype and phenotype best exemplified in *V. parahaemolyticus* isolates. Discordance was partially explained by allelic variation in the beta-lactamase gene bla_{CARB}. This study provides useful insights into antimicrobial treatment of vibriosis in sea otters and identifies genotypes associated with ampicillin resistance.

Chapter 2 characterized genetic markers of virulence in *Vibrio* spp. across isolation sources using the Chapter 1 genomic dataset (n=444) along with public domain *V. parahaemolyticus*

genomes (n=126) isolated from humans and shellfish in North America. *Vibrio parahaemolyticus* (n=287) isolated from humans and sea otters were more likely to carry virulence factors such as hemolysins and secretion system genes than isolates from environmental sources. Despite general genomic diversity of *V. parahaemolyticus*, genome clusters with high similarity were detected which included virulence factor positive genomes from both humans and sea otters. *V. parahaemolyticus* strains with co-occurrence of two virulence-related secretion systems were more likely to be isolated from dead southern sea otters with septicemia, enteritis, and moderate to severe melena than *V. parahaemolyticus* strains without either secretion system. This study provides evidence that *V. parahaemolyticus* undergoes selection pressures in both humans and otters resulting in apparent expansion of virulent strains and provides insights into *V. parahaemolyticus* pathogenicity in sea otters.

Chapter 3 assessed the diagnostic resolution of closely related species in the Harveyi clade by combining the *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* genomes from Chapter 2 with 150 additional genomes from the public domain. Within public domain genomes, 28 *V. diabolicus* were incorrectly labeled as *V. alginolyticus* resulting in 10 *V. diabolicus* genomes from human cases. A pangenome-wide association study identified species-selected gene clusters as well as *V. alginolyticus* and *V. diabolicus* virulence-related genes associated with either humans or sea otter isolates. Novel associations between putative virulence genes or alleles and either human or sea otter hosts were identified that were not present in virulence factor databases. These findings suggest *V. diabolicus* is an overlooked pathogen of humans and sea otters due to its close phylogeny with *V. alginolyticus* and

provides additional diagnostic information to better differentiate clinical strains of each species in otters and humans.

This research utilized genomic data for comparative epidemiological investigations of *Vibrio* spp. in both northern and southern sea otters and presents evidence that sea otters are at risk for antimicrobial resistant and virulent *Vibrio* strains, including the first documentation of *V. diabolicus* in sea otters. These chapters provide a baseline for further surveillance and include insights to improve the diagnosis and treatment of pathogenic *Vibrio* spp. in sea otters. Additional spatiotemporally linked genomic surveillance of *Vibrio* spp. in sea otters and humans is needed, which could further elucidate the potential for shared transmission routes of virulent and antimicrobial resistant vibriosis.

INTRODUCTION

Epidemiological considerations for genomics of marine pathogens

The first ever bacterial genome sequenced in 1995 revolutionized molecular biology and eventually ushered in a new age of microbial genomics.[1] As the number of microbial genomes sequenced have exponentially increased over the last two decades, the need for informed genomic comparisons led to the field of genomic epidemiology. Genomic epidemiology expands on molecular genetics by combining two previously disparate approaches of genomic phylogeny and epidemiologic investigation to better understand disease transmission and manage disease outbreaks.[2,3] Genomic epidemiology has developed approaches to address challenges in large data integration and analyses combining both whole genomes and metadata. Complications in applying genomic epidemiology to marine systems are compounded by major differences between terrestrial and marine systems.[4] Marine systems are often presumed to be more taxonomically diverse, have more open population recruitment of wildlife species, contain more free-living pathogens, and studies to date have emphasized the importance of different anthropogenic impacts compared to terrestrial systems including overharvesting, eutrophication, and introduction of microbial species from land-to-sea transfer, global shipping, and migration.[4]

Vibrio spp. are representative of many of the differences between terrestrial and marine systems and provide excellent examples of how genomic epidemiology can be successfully applied to marine pathogens. The *Vibrio* genus is diverse with over 100 species that inhabit freshwater, estuarine, and marine habitats globally including at least a dozen species implicated in human infections and more attributed to marine wildlife disease.[5,6] Pathogenic *Vibrio* spp.

can often survive as free-living microbes including in a viable but non-culturable state when environmental conditions are less habitable such as colder water temperatures.[7] Population distributions of *Vibrio* spp. also experienced anthropogenic impacts including increased global population mixing in the last few decades attributed to increased shipping and human migration.[8] Starting with the first whole genome sequencing of *Vibrio cholerae* in 2000[9] and then *Vibrio parahaemolyticus* in 2003,[10] both species have thousands of publicly available genomes, although other *Vibrio* species still have significantly fewer genomic comparisons readily available. Despite marine *Vibrio* spp. being generally understudied compared to common terrestrial pathogens, there have been many novel applications of whole genome sequencing in *Vibrio* spp. including utility for epidemiologic investigations.[5] The goal of this review is to highlight innovative applications of genomic epidemiology for *Vibrio* spp. investigations.

Assessing relatedness

The foremost goal of genomic epidemiology is to determine relatedness between genomes of pathogenic species and use the epidemiologic metadata to make meaningful inferences.[1] Traditional methods of genomic relatedness require alignment of genomes with a reference genome to create phylogenetic trees and visualize evolutionary relationships.[1] More recent alternative methods employ alignment-free pairwise comparisons between genomes by randomly sampling variable-length sequence fragments called *k*-mers to determine the amount of sequence similarity between genomes.[11–14] Advantages of utilizing *k*-mers include the ability to run faster and larger genomic comparisons and to make de novo comparisons with new

genomes which can facilitate real-time surveillance.[11,13] Specific tools that use *k*-mers for genomic relatedness include PopPUNK (Population Partitioning Using Nucleotide *K*-mers) which groups closely related isolates and can be beneficial for identifying outbreak-relevant clusters, [14] and sourmash, which provides quick and useful heatmaps for identifying patterns of genome similarity.[12,13]

Whole genome sequencing has also greatly improved taxonomic classifications of *Vibrio* spp. The previous standard method of 16s sequencing was often limited to genus level resolution whereas whole genome sequencing allowed for species level classification, even for closely related sister species such as *V. parahaemolyticus* and *V. alginolyticus*.[15] High intraspecies genomic diversity and closely related species have provided considerable difficulties in taxonomic classification of *Vibrio* spp., but additional whole genome taxonomic studies have further elucidated these complex relationships. For example, *Vibrio diabolicus* first detected from a deep-sea annelid,[16] was identified as a close relative of *V. alginolyticus* through phylogenomics, which also detected *V. diabolicus* isolates previously identified as *V. alginolyticus*.[17] As potentially novel species have been detected such as the proposed *Vibrio chemaguriensis*,[18] iterative improvements to whole genome taxonomy have been crucial in properly separating and combining related species; for example, the most recent *Vibrio* taxonomy shows that *V. chemaguriensis* should not be considered a separate species from *V. diabolicus*.[19]

Assessment of genomic relatedness can be applied to epidemiologic studies in novel ways. A large-scale phylogenomic analysis of *V. parahaemolyticus* outbreaks in Shenzhen, China over a 17-year period highlights innovative ways in which robust epidemiologic data can provide

novel inferences from genomic relatedness.[20] Over 3,000 isolates were sequenced with accompanying epidemiologic data from three separate surveillance systems: a) outbreak surveillance, b) outpatient diarrheal surveillance, and c) food safety surveillance. By clustering genomes by spatiotemporal metrics and multi-tiered levels of relatedness determined by the number of single nucleotide polymorphisms, this investigation determined that 71% of cases were from strain outbreaks occurring within 1-month clusters and that many outbreaks were either missed or underreported by the outbreak surveillance system due to related genomes from diarrheal outpatients not part of the outbreak surveillance. Some outbreak strains persisted 2 to 70 months later before going dormant, suggesting heretofore undiscovered reservoirs in marine environments. In fact, this study contained genomes of some pathogenic lineages that were stable and recurring for more than 10 years.[20] The stability of pathogenic *V. parahaemolyticus* genomes was similarly reported in the Americas.[21]

Sampling Considerations

Epidemiological sampling considerations are needed in genomic studies as the power of genomic comparisons is limited by the quality of informative metadata[1]. Because of the duality between high intraspecies genomic diversity and the semi-clonal population structure of some *Vibrio* spp., genome sample sets are likely to include redundant genomic information that could cause oversampling of clonal groups and under-sampling of rare or emergent strain. In a global assessment of *V. parahaemolyticus* population structure, 469 of 1,103 starting genomes were removed to create a “non-redundant” set of strains without clonally related genomes.[8] Wildlife disease surveillance studies may collect multiple isolates sub-cultured from the same sample to

ensure the correct pathogen is captured; similarly, studies may culture multiple tissue samples from a diseased host or collect isolates from multiple individuals or species in the same population.[22,23] Genomic sequencing can be useful to remove redundant isolates from an epidemiologic study. For example, two *V. parahaemolyticus* isolates sub-cultured from the same sample could be unique co-occurring strains or redundant sampling. Geographical differences in pathogen populations must also be carefully considered, especially when comparing across studies. Samples from coastal Asia exhibit a different population structure[8] dominated by clonal groups including CC3 and CC189 [20] than populations in coastal North America where sequence types ST36 and ST631 are more prominent.[24–27]

Improved traceability

Another common utilization of genomics with sufficient accompanying metadata is to exploit the finer resolution for improved subtyping and outbreak tracing.[1,28] Whole genome sequence has become the gold standard over previous molecular subtyping methods such as pulse field gel electrophoresis (PFGE) to trace or source-track pathogens and to improve diagnostic detection.[1] One of the most recognized example of outbreak tracing involves the cholera epidemic following the devastating 2010 earthquake in Haiti.[29] As Haiti had been cholera-free for more than a century, two prevailing hypotheses emerged for the outbreak of *Vibrio cholerae* in Haiti: 1) the climatic hypothesis in which a nonpathogenic strain from coastal Haiti evolved into a pathogenic strain, and 2) the human transmission hypothesis in which a foreign traveler introduced the pathogenic strain to Haiti.[29] A series of phylogeographic analyses using whole genome sequencing of *Vibrio cholerae* from Haiti and Southern Asia

strains as well as epidemiologic information helped support the human transmission hypothesis through a single-source introduction by Nepalese United Nations relief workers.[29–31]

Tracing studies can be beneficial for local as well as global scale studies. One approach for global tracing known as “molecular clock”, estimates the rates of evolution within separate phylogenomic branches; this approach was used to identify 7 distinct historical waves of global cholera transmission and identified the most recent pandemic wave originating in the 1950s from the Bay of Bengal.[32] Other cross-continental tracing studies have shown the evolution of a pathogenic *V. parahaemolyticus* clonal sequence type 36 originating in Pacific Northwest United States as it expanded its transmission to other continents.[25,26] On a country scale, the combination of outbreak data with comparative genomics identified multiple “farm-to-table” transmission events of *V. parahaemolyticus* in China by connecting closely related isolates from production, circulation, and consumption which exhibited single-source radial spread.[33] Such studies all utilize the improved resolution of whole genome sequencing to provide a more informed view on spatiotemporal patterns[5].

Associating epidemiologic data with gene function

Whole genome sequencing allows for both known and novel gene functions to be assessed in relation to host, environmental, or pathogen phenotypic metadata with potential to address myriads of questions relating to function.[1] Common applications include studies that link specific genes or alleles with phenotypes of antimicrobial resistance, biofilm formation, or increased virulence and pathogenicity.

Due to concerns of antimicrobial resistance in marine microbes, especially through land-to-sea runoff,[34] there is a desire for quick, cost-effective, and informative detection of

antimicrobial resistance. As whole genome sequencing has become increasingly affordable, genomics has become a critical tool for predicting emergence and spread of antimicrobial resistance genes.[35] Despite the data-rich output of genomic sequencing, further genomic epidemiology studies are still needed to improve the associations between phenotypes and genotypes and make whole genome sequencing a more practical clinical tool. This is evident for *Vibrio* spp. for which antimicrobial resistant genes do not consistently match their presumed phenotypes. In a comparison of phenotypic and genotypic antimicrobial resistance from non-O1/O139 *V. cholerae* strains, there was perfect concordance between ampicillin resistance and presence of bla_{CARB} genes (12 out of 54 total genomes with bla_{CARB} genes), but extreme discordance for genes associated with other drug classes and their supposed phenotypes.[36] This was exemplified by phenotypic susceptibility to both chloramphenicol and tetracycline despite the frequent detection of genes that supposedly confer resistance (*tet34* and *catB9*).[36]

Genomic epidemiology studies that further examine function of antimicrobial resistance genes will be needed until available databases such as CARD[37] and ResFinder[38,39] can adequately predict the observed phenotypes for marine pathogens as well as they do for frequently studied terrestrial pathogens like *Salmonella enterica*.[40] Finer resolution of genotype should go beyond gene presence but include point mutations that can result in gain or loss of function.[41] For example, genes *gyrA* and *parE*, which are involved in essential bacterial function, can result in quinolone resistance after gaining a point mutation.[42] Additionally, databases of gene function must continually improve to include genes appropriate for each species, which may necessitate identifying species-specific homologs as well as new genes. The most robust virulence gene database currently relies heavily on genes from *V. cholerae*, *V.*

parahaemolyticus, and *Vibrio vulnificus*.[43,44] Virulence gene patterns in other *Vibrio* spp. such as *V. alginolyticus* and *V. diabolicus* will still be in relation to better characterized species like *V. parahaemolyticus*.[45] Lowering the percent identity and coverage of gene detection could result in detection of more genes in under-studied species, but will also result in more false positives of homologous genes that do not share the same function.

Machine learning modeling is a powerful tool to assist with functional comparisons and is likely to become more common with the availability of affordable and fast computing. Machine learning can parse genomic sequences and associated epidemiologic data to identify novel data patterns such as mutations in a virulence gene associated with disease. A machine modeling method, XGBoost, was recently used to identify single nucleotide polymorphisms in a virulence gene *porA* from extraintestinal (abortive) versus intestinal (non-abortive) *Campylobacter jejuni* strains and identified unique alleles associated with abortion causing strains.[46] While the *porA* study exemplifies the utility of machine learning for identifying alleles associated with greater disease virulence, the approach can be used to identify genetic loci associated with any phenotypic traits or epidemiologic data.[47] Future *Vibrio* genomics studies should take advantage of dramatically improving machine learning tools.

Utility of the Pangenome

Pangenome describes the combination of the core genome, comprised of genes ubiquitous to all strains within a given species, and the accessory genome, comprised of genes present in less than all genomes. The term pangenome was coined to compare the available independent genomes to understand the complete genetic diversity of a bacterial species.[48] Genomic

phylogenies are frequently based on genetic variation in the core genome alone; however, core genome-only analyses neglect a significant amount of information, especially for highly similar core genomes which may not have enough genotyping resolution.[49]

Pangenomes can be uniquely informative to the life history of a pathogen. Pangenomes can be described as "open" or "closed" depending on whether the addition of more genomes will result in the addition of new genes not previously detected. When open pangenomes include disproportionately large accessory genomes, they can be inferred to have higher rates of horizontal gene transfer.[50] *Vibrio* spp. are notoriously diverse with extremely large accessory gene pools,[27,51–53] and warrant concerns for horizontal gene transfer of genes related to colonization, biofilm formation, virulence, or antimicrobial resistance. While small pangenomes can provide beneficial descriptive information for under-sequenced emerging species such as *V. diabolicus*,[45] larger sample sizes are more likely to represent a true species pangenome and provide enough power for genome wide association studies (GWAS).

To further illustrate the impact of sample size on pangenome analyses, consider the differences between a moderately sized pangenome analysis comprising 132 *V. parahaemolyticus* from humans and oysters across North America with a large-scale global pangenome comprised of 1,103 *V. parahaemolyticus* genomes.[8,27] In the North America study, the gene finding effort was presumed complete but only detected 8,191 unique genes despite a large accessory gene pool[27]. Meanwhile, the global pangenome detected over 40,000 unique genes, suggesting that selection bias may have resulted in oversampling related strains in the North America study.[8,27] In fact, the population structure of both North American and Asian *V. parahaemolyticus* populations have predominant sequence types representing expansion of

related pandemic clonal groups, even though the predominant sequence types differ between locations.[8,20,27] Gene identity cutoffs will also impact the number of unique genes depending on whether homologs all called as separate or binned genes. The scope of investigation should also be considered when choosing the appropriate sample size for pangenome analyses. For example, analysis of a predominant sequence type may not require as large of a sample size as a study interested in assessing the population structure.

Utility of pangenomes have been proposed for a variety of applications including characterization of resistomes, or the complete collection of genes in a pangenome which encode antimicrobial resistance, mobilomes, or the complete collection of highly mobile genes involved in horizontal gene transfer, and virulence factors, as well as improved classification of species down to even more refined levels.[50] A pangenome analysis of *V. alginolyticus* genomes from the Kiel-Fjord, Germany helped classify a novel ecotype associated with *Syngnathus* spp. pipefish hosts.[54] The Kiel-Fjord ecotype had a significantly reduced accessory genome and provided evidence that *Vibrio* spp. can undergo extreme host adaptation. Another study utilized the pangenome of 1,727 *V. parahaemolyticus* to characterize 4 distinct type VI secretion systems, providing a much-needed step to improve the diagnostic relevance of the type VI secretion systems for virulence.[55] Similarly, a pangenome of *V. parahaemolyticus* isolates causing acute hepatopancreatic necrosis disease (AHPND) in cultured shrimp was used to identify three separate clades that associated by geographic locations to inform on the most plausible origins of international disease introduction.[56]

Finally, Pangenome studies can utilize the GWAS approach to identify accessory genes associated with epidemiologic metadata. In a study comparing *V. parahaemolyticus* from ready-

to-eat foods to those from humans or environmental sampling, genes putatively involved in biofilm formation were over-represented in ready-to-eat foods.[57] While an increasing number of bioinformatics tools are becoming readily available, the ready-to-eat foods study utilized the powerful bioinformatics tools Roary for pangenome assembly[58] and Scoary for statistical comparisons of trait-gene associations.[59] Roary is a free tool that drastically decreases the time needed to build large-scale pangenomes capable of processing thousands of genomes.[58] Scoary is also a free tool which can utilize the Roary output and runs a series of Fisher's exact tests comparing presence or absence of a gene with presence or absence of any binary environmental, host, or pathogen trait.[59] Metadata can also be visualized with Roary created pangenomes using the publicly available Phandango to show epidemiologic relationships between epidemiologic traits and gene presence.[60]

Advancements to *Vibrio* spp. genomic epidemiology

Through this dissertation research, a One Health approach which combined epidemiology and comparative genomics was utilized to improve the collective understanding of *Vibrio* spp. from an understudied host (sea otters) in relation to isolates from shellfish, humans, and environmental samples. This dissertation employed whole genome sequencing and sampling metadata to characterize the pangenomes of *Vibrio* spp. isolates collected between 2000-2019 from northern sea otters in Alaska, southern sea otters and shellfish from California, as well as water and sediment sampling from California and Washington. Genomic methods of species identification were implemented and contrasted with common clinical methods of MALDI-ToF and biochemical testing.

In chapter 1, antimicrobial resistance patterns of *Vibrio* spp. from various sources were characterized using both phenotypic antimicrobial susceptibility testing and whole genome sequencing. Concordance between phenotypic and genomic methods of antimicrobial resistance detection was assessed with a deeper investigation of allelic variation of antimicrobial resistance genes associated with improved concordance to resistance phenotypes. Additionally, associations between antimicrobial resistance genotypes and epidemiologic metadata were examined.

In chapter 2, genetic markers of virulence of *Vibrio* spp. were characterized and the associations between virulence factors, isolation source, and genomic relatedness were explored with the additional comparisons of human and shellfish isolation sources using public domain North American *V. parahaemolyticus* genomes. Pathology of dead sea otters that were *Vibrio* spp. positive were compared to detection of virulence-related genes to determine the clinical importance of *Vibrio* spp. infections in sea otter hosts.

Finally, chapter 3 utilized pangenomes of three closely related species in the Harveyi clade, *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* to identify genetic markers for improved resolution of species identification. Using a GWAS approach, genes useful for identifying host-adapted human or sea otter clinical strains in the understudied species *V. alginolyticus* and *V. diabolicus* were detected.

Assessment of the relatedness between genomes and the metadata that explains the patterns observed is one of the most critical goals of genomic epidemiology. The chapters presented here collectively determined the relatedness between *Vibrio* spp. strains from sea

otters, humans, and their shared marine environment and food sources using many of the timely and informative bioinformatics tools discussed here. The analyses presented addressed ways in which genomics can inform common microbiology sampling considerations that will continue to impact studies in the budding field of genomic epidemiology. Additionally, each chapter informed clinically relevant insights into gene function through associations between gene presence and epidemiological data including antimicrobial resistance, virulence, and host-adaptation to advance the current knowledge on vibriosis for the benefit of both human and marine wildlife health.

Literature Cited

1. Baker KS. Microbe hunting in the modern era: reflecting on a decade of microbial genomic epidemiology. Current Biology 2020; 30:R1124–30.

2. Tang P, Croxen MA, Hasan MR, Hsiao WWL, Hoang LM. Infection control in the new age of genomic epidemiology. American Journal of Infection Control 2017; 45:170–9.

3. Hill V, Ruis C, Bajaj S, Pybus OG, Kraemer MUG. Progress and challenges in virus genomic epidemiology. Trends in Parasitology 2021; 37:1038–49.

4. McCallum HI, Kuris A, Harvell CD, Lafferty KD, Smith GW, Porter J. Does terrestrial epidemiology apply to marine systems? Trends Ecol Evol 2004; 19:585–91.

5. Lau DYL, Aguirre Sánchez JR, Baker-Austin C, Martinez-Urtaza J. What whole genome sequencing has told us about pathogenic *Vibrios*. In: Almagro-Moreno S, Pukatzki S, editors. *Vibrio* spp. infections. Advances in experimental medicine and biology. Springer, Cham; 2023. page 337–52.

6. Baker-Austin C, Oliver JD, Alam M, Ali A, Waldor MK, Qadri F, Martinez-Urtaza J. *Vibrio* spp. infections. Nat Rev Dis Primers 2018; 4:1–19.

7. Yoon JH, Lee SY. Characteristics of viable-but-nonculturable *Vibrio parahaemolyticus* induced by nutrient-deficiency at cold temperature. Critical Reviews in Food Science and Nutrition 2020; 60:1302–20.

8. Yang C, Pei X, Wu Y, Yan L, Yan Y, Song Y, Coyle NM, Martinez-Urtaza J, Quince C, Hu Q, et al. Recent mixing of *Vibrio parahaemolyticus* populations. The ISME Journal 2019 13:10 2019; 13:2578–88.

9. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 2000; 406.

10. Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima M, Nakano M, Yamashita A, et al. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V cholerae*. Lancet 2003; 361:743–9.

11. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 2016; 17.

12. Brown CT, Irber L. sourmash: a library for MinHash sketching of DNA. The Journal of Open Source Software 2016; 1:27.

13. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. Large-scale sequence comparisons with sourmash. F1000Res 2019; 8.

14. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res 2019; 29:304–16.

15. Thompson CC, Vicente ACP, Souza RC, Vasconcelos ATR, Vesth T, Alves N, Ussery DW, Iida T, Thompson FL. Genomic taxonomy of *Vibrios*. BMC Evol Biol 2009; 9.

16. Raguénès G, Christen R, Guezennec J, Pignet P, Barbier G. *Vibrio diabolicus* sp. nov., a new polysaccharide-secreting organism isolated from a deep-sea hydrothermal vent polychaete annelid, *Alvinella pompejana*. Int J Syst Bacteriol 1997; 47:989–95.

17. Turner JW, Tallman JJ, Macias A, Pinnell LJ, Elledge NC, Azadani DN, Nilsson WB, Paranjpye RN, Armbrust E V., Strom MS. Comparative genomic analysis of *Vibrio diabolicus* and six taxonomic synonyms: a first look at the distribution and diversity of the expanded species. Front Microbiol 2018; 9:1893.

18. Ghosh A, Bhadury P. *Vibrio chemaguriensis* sp. nov., from Sundarbans, Bay of Bengal. Curr Microbiol 2019; 76:1118–27.

19. Jiang C, Tanaka M, Nishikawa S, Mino S, Romalde JL, Thompson FL, Gomez-Gil B, Sawabe T. *Vibrio* clade 3.0: new *Vibrionaceae* evolutionary units using genome-based approach. Curr Microbiol 2022; 79:1–15.

20. Yang C, Li Y, Jiang M, Wang L, Jiang Y, Hu L, Shi X, He L, Cai R, Wu S, et al. Outbreak dynamics of foodborne pathogen *Vibrio parahaemolyticus* over a seventeen year period implies hidden reservoirs. Nature Microbiology 2022 7:8 2022; 7:1221–9.

21. Guerrero A, Gomez-Gil B, Lizarraga-Partida ML. Genomic stability among O3:K6 *V. parahaemolyticus* pandemic strains isolated between 1996 to 2012 in American countries. BMC Genom Data 2021; 22.

22. Brownstein D, Miller MA, Oates SC, Byrne BA, Jang S, Murray MJ, Gill VA, Jessup DA. Antimicrobial susceptibility of bacterial isolates from sea otters (*Enhydra lutris*). J Wildl Dis 2011; 47:278–92.

23. Oates SC, Miller MA, Byrne BA, Chouicha N, Hardin D, Jessup D, Dominik C, Roug A, Schriewer A, Jang SS, et al. Epidemiology and potential land-sea transfer of enteric bacteria from terrestrial to marine species in the Monterey Bay region of California. J Wildl Dis 2012; 48:654–68.

24. Xu F, Gonzalez-Escalona N, Haendiges J, Myers RA, Ferguson J, Stiles T, Hickey E, Moore M, Hickey JM, Chillaci C, et al. Sequence type 631 *Vibrio parahaemolyticus*, an emerging foodborne pathogen in North America. J Clin Microbiol 2017; 55:645.

25. Abanto M, Gavilan RG, Baker-Austin C, Gonzalez-Escalona N, Martinez-Urtaza J. Global expansion of Pacific Northwest *Vibrio parahaemolyticus* sequence type 36. Emerg Infect Dis 2020; 26:323.

26. Martinez-Urtaza J, Van Aerle R, Abanto M, Haendiges J, Myers RA, Trinanes J, Baker-Austin C, Gonzalez-Escalona N. Genomic variation and evolution of *Vibrio parahaemolyticus* ST36 over the course of a transcontinental epidemic expansion. mBio 2017; 8.

27. Miller JJ, Weimer BC, Timme R, Lüdeke CHM, Pettengill JB, Bandoy DJD, Weis AM, Kaufman J, Huang BC, Payne J, et al. Phylogenetic and biogeographic patterns of *Vibrio parahaemolyticus* strains from North America inferred from whole-genome sequence data. Appl Environ Microbiol 2021; 87.

28. Mäklin T, Kallonen T, Alanko J, Samuelsen Ø, Hegstad K, Mäkinen V, Corander J, Heinz E, Honkela A. Bacterial genomic epidemiology with mixed samples. Microb Genom 2021; 7.

29. Orata FD, Keim PS, Boucher Y. The 2010 cholera outbreak in Haiti: how science solved a controversy. PLoS Pathog 2014; 10:e1003967.

30. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. mBio 2011; 2.

31. Eppinger M, Pearson T, Koenig SSK, Pearson O, Hicks N, Agrawal S, Sanjar F, Galens K, Daugherty S, Crabtree J, et al. Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. mBio 2014; 5.

32. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 2011; 477:462–5.

33. Yang C, Zhang X, Fan H, Li Y, Hu Q, Yang R, Cui Y. Genetic diversity, virulence factors and farm-to-table spread pattern of *Vibrio parahaemolyticus* food-associated isolates. Food Microbiol 2019; 84.

34. Singh AK, Kaur R, Verma S, Singh S. Antimicrobials and antibiotic resistance genes in water bodies: pollution, risk, and control. Front Environ Sci 2022; 10:397.

35. Waddington C, Carey ME, Boinett CJ, Higginson E, Veeraraghavan B, Baker S. Exploiting genomics to mitigate the public health impact of antimicrobial resistance. Genome Medicine 2022; 14.

36. Lepuschitz S, Baron S, Larvor E, Granier SA, Pretzer C, Mach RL, Farnleitner AH, Ruppitsch W, Pleininger S, Indra A, et al. Phenotypic and genotypic antimicrobial resistance traits of *Vibrio cholerae* non-O1/non-O139 isolated from a large Austrian lake frequently associated with cases of human infection. Front Microbiol 2019; 10:2600.

37. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 2022.

38. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy 2020; 75:3491–500.

39. Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom 2022; 8.

40. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu CH, McDermott PF, et al. Validating the AMRFINder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother 2019; 63:e00483-19.

41. Jin J, Zhou Y, Zhang Z, Wang H, Hou W, Wang H, Li R, Zhou M. Characteristics of antimicrobial-resistant *Vibrio parahaemolyticus* strains and identification of related antimicrobial resistance gene mutations. Foodborne Pathog Dis 2021; 18:873–9.

42. Rodkhum C, Maki T, Hirono I, Aoki T. *gyrA* and *parC* associated with quinolone resistance in *Vibrio anguillarum*. J Fish Dis 2008; 31:395–9.

43. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res 2016; 44:D694–7.

44. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res 2022; 50:D912–7.

45. Pérez-Duque A, Gonzalez-Muñoz A, Arboleda-Valencia J, Vivas-Aguas LJ, Córdoba-Meza T, Rodriguez-Rey GT, Díaz-Guevara P, Martinez-Urtaza J, Wiesner-Reyes M. Comparative genomics of clinical and environmental isolates of *Vibrio* spp. of Colombia: implications of traits associated with virulence and resistance. Pathogens 2021; 10:1605.

46. Bandoy DDR, Weimer BC. Biological machine learning combined with *Campylobacter* population genomics reveals virulence gene allelic variants cause disease. Microorganisms 2020; 8:549.

47. Guo Y, Wu C, Yuan Z, Wang Y, Liang Z, Wang Y, Zhang Y, Xu L. Gene-based testing of interactions using XGBoost in genome-wide association studies. Front Cell Dev Biol 2021; 9.

48. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli S V, Crabtree J, Jones AL, Scott Durkin A, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial" 'pan-genome.'" Proceedings of the National Academy of Sciences 2005; 102:13950–5.

49. Castillo-Ramírez S. Beyond microbial core genomic epidemiology: towards pan genomic epidemiology. The Lancet Microbe 2022; 3:e244–5.

50. Rouli L, Merhej V, Fournier PE, Raoult D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. New Microbes New Infect 2015; 7:72–85.

51. Janecko N, Bloomfield SJ, Palau R, Mather AE. Whole genome sequencing reveals great diversity of *Vibrio* spp in prawns at retail. Microb Genom 2021; 7:647.

52. Urmersbach S, Alter T, Koralage MSG, Sperling L, Gerdts G, Messelhäusser U, Huehn S. Population analysis of *Vibrio parahaemolyticus* originating from different geographical regions demonstrates a high genetic diversity. BMC Microbiol 2014; 14:1–14.

53. Sampaio A, Silva V, Poeta P, Aonofriesei F. *Vibrio* spp.: life strategies, ecology, and risks in a changing environment. Diversity 2022, Vol 14, Page 97 2022; 14:97.

54. Chibani CM, Roth O, Liesegang H, Wendling CC. Genomic variation among closely related *Vibrio alginolyticus* strains is located on mobile genetic elements. BMC Genomics 2020; 21:1DUMM.

55. Jana B, Keppel K, Fridman CM, Bosis E, Salomon D. Multiple T6SSs, mobile auxiliary modules, and effectors revealed in a systematic analysis of the *Vibrio parahaemolyticus* pan-genome. mSystems 2022; 7:e00723-22.

56. González-Gómez JP, Soto-Rodriguez S, López-Cuevas O, Castro-del Campo N, Chaidez C, Gomez-Gil B. Phylogenomic analysis supports two possible origins for Latin American strains of *Vibrio parahaemolyticus* associated with acute hepatopancreatic necrosis disease (AHPND). Curr Microbiol 2020; 77:3851–60.

57. Pang R, Xie T, Wu Q, Li Y, Lei T, Zhang J, Ding Y, Wang J, Xue L, Chen M, et al. Comparative genomic analysis reveals the potential risk of *Vibrio parahaemolyticus* isolated from ready-to-eat foods in China. Front Microbiol 2019; 10.

58. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015; 31:3691–3.

59. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 2016; 17:1–9.

60. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics 2018; 34:292–3.

Antimicrobial resistance of *Vibrio* spp. from coastal California: discordance between genotypic and phenotypic patterns

Peter J. Sebastian[1,2], Cory Schlesener[3], Barbara A. Byrne[4], Melissa Miller[2,5], Woutrina Smith[2], Francesca Batac[5], Caroline E.C. Goertz[6], Bart C. Weimer[3*], Christine K. Johnson[1,2*]

[1]EpiCenter for Disease Dynamics, One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA
[2]Karen C. Drayer Wildlife Health Center, One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[3]Department of Population Health and Reproduction, 100K Pathogen Genome Project, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[4]Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[5]Marine Wildlife Veterinary Care and Research Center, California Department of Fish and Wildlife, Santa Cruz, CA, USA

[6]Alaska SeaLife Center, Seward, AK, USA

***Correspondence:**
Christine K. Johnson
ckjohnson@ucdavis.edu
Bart C. Weimer
bcweimer@ucdavis.edu

Keywords: antimicrobial resistance, antibiotic, WGS, whole genome sequencing, vibriosis,

***Vibrio*, sea otter, *Enhydra lutris*.**

Formatted for submission to *Applied and Environmental Microbiology*.

Abstract

Antimicrobial resistance in *Vibrio* species is a growing concern that could affect the medical treatment of humans and marine mammals. Whole genome sequencing of *Vibrio* spp. can be utilized to screen for antimicrobial resistance genes and allelic variants to provide mechanistic insights in ways that PCR screening and phenotypic interpretation cannot. Our goals were to: 1) characterize antimicrobial resistance patterns of *Vibrio* spp. pathogens isolated from southern sea otters (*Enhydra lutris nereis*), northern sea otters (*Enhydra lutris kenyoni*), and environmental samples from the central California coast using whole genome sequencing, and 2) compare the presence of antimicrobial resistance genes with phenotypic interpretation from antibiotic susceptibility testing. Unexpectedly, genomic classification identified an understudied species, *Vibrio diabolicus*, in sea otter and environmental isolates that were previously identified as *Vibrio alginolyticus*. A total of 489 *Vibrio* spp. isolates were sequenced, and frequently detected antimicrobial resistance genes included multidrug efflux pumps and genes associated with resistance to ß-lactams and tetracyclines. Genes associated with resistance to fluoroquinolones, aminoglycosides, chloramphenicol, and sulfonamides were uncommon. Sea otter isolates were phenotypically susceptible to tetracycline despite carrying genes *tet34* and *tet35*. Similarly, bla_{CARB} genes were ubiquitous in *V. alginolyticus*, *V. diabolicus*, and *Vibrio parahaemolyticus* but ampicillin resistance was observed in only 78.4%, 76.2%, and 33.9% of tested isolates, respectively. Discordance between phenotypic and genotypic ampicillin resistance for *V. parahaemolyticus* was partially attributed to allelic variation of the bla_{CARB} genes. Tetracyclines and fluoroquinolones, but not ß-lactams, are likely to be effective treatments for vibriosis in sea otters.

Importance

Vibriosis (infection with non-cholera *Vibrio* spp.), is the most common seafood-borne illness globally, with major impacts on public health, food security, and wildlife health. Potential treatments of antimicrobial resistant *Vibrio* spp. in humans, aquaculture, and marine wildlife rehabilitation are complicated by current diagnostic challenges regarding bacterial species identification and interpretation of antimicrobial resistance patterns. Unexpected detection of previously misidentified *Vibrio diabolicus* in sea otters suggests that a broader taxonomic group of *Vibrio* infect sea otters than previously described. We also determined that presence of ß-lactamase genes alone in sea otter isolates does not necessarily correlate with an ampicillin resistant phenotype, likely due to deleterious amino acid substitutions in certain bla_{CARB} alleles. Continued monitoring of *Vibrio* spp. phenotypes and genotypes in sea otters is warranted to observe biologically relevant changes in antimicrobial resistance.

Introduction

The *Vibrio* genus is a diverse group of gram-negative, rod-shaped bacteria with two circular chromosomes that inhabit primarily warm, brackish marine, and estuarine environments. There are over 100 *Vibrio* species, including many that impact marine wildlife, and several that are implicated in human illness. Vibriosis (infection with non-*cholerae Vibrio* spp. and marine-inhabiting non-O1/O139 *Vibrio cholerae*), accounts for approximately 80,000 human illnesses in the United States each year (1). *Vibrio parahaemolyticus* is the most common bacterial cause of shellfish-related illness, resulting in gastroenteritis and sepsis in severe cases. *Vibrio alginolyticus*, one of the top four most reported *Vibrio* pathogens, is primarily associated with wound and ear infections following seawater exposure (1). *Vibrio diabolicus*, first isolated near a deep-sea vent, has since been identified in coastal habitats, is closely related to *V. alginolyticus*, and in some cases has been misclassified as *V. alginolyticus* (2, 3).

Vibrio spp. are frequently detected in benthic invertebrates consumed by marine mammals and humans (4–7). Potentially pathogenic *Vibrio* spp., primarily non-O1/O139 *V. cholerae*, *V. parahaemolyticus*, and *V. alginolyticus*, have been found in healthy (8–12), clinically ill, and dead marine mammals (13–16) along the North American Pacific coast. Sea otters have been proposed as highly sensitive ecological bioindicators; study of disease in these animals can highlight the presence of multiple pathogens and contaminants in the near-shore environment (17, 18). Unique biological attributes that make sea otters outstanding as environmental sentinels include their relatively small near-shore home range and their consumption of high volumes of shellfish and benthic invertebrates, including clams, mussels, and other species that are also consumed by humans and are important sources of human vibriosis (17, 18). Multiple

disease processes with public health importance have been successfully modeled in southern sea otters (*Enhydra lutris nereis*) including *Microcystis aeruginosa*, *Toxoplasma gondii*, and domoic acid toxicosis (19–21). *Vibrio* infections in northern sea otters (*Enhydra lutris kenyoni*) from Alaska have been proposed as indicators of climate change, especially *V. parahaemolyticus,* which has begun to be detected with Alaskan coastal water temperatures rising past 15°C (10). *Vibrio* infections in sea otters are thought to occur as secondary infections or sequelae to traumatic injury which may require effective antimicrobial therapy during rehabilitation (8, 16).

While *Vibrio* spp. are traditionally susceptible to most clinically relevant antimicrobials including tetracyclines, fluoroquinolones, and 3rd generation cephalosporins, there has been an increase in the awareness and incidence of antimicrobial resistance (AMR) of *Vibrio* spp. isolates (22, 23). Characterization of AMR in *Vibrio* spp. from southern sea otters is limited with *V. cholerae* (n=3) isolates interpreted as having clindamycin resistant, tylosin resistant, and erythromycin intermediate phenotypes, and *V. parahaemolyticus* (n=15) isolates interpreted as exhibiting variable phenotypes including either ampicillin intermediate and erythromycin, cefazolin, clindamycin, and tylosin resistant phenotypes (8). Characterization of antimicrobial resistance phenotypes of *Vibrio* spp. across diverse sources within coastal communities such as marine wildlife, marine invertebrates, and environmental samples could help protect human health, facilitate assessment of epidemiological patterns and trends, and optimize care for stranded marine mammals.

Studies of AMR in *Vibrio* spp. typically involve either antimicrobial susceptibility testing to determine phenotype, or amplification and genotypic assessment of target antimicrobial

resistance genes (5, 24–26). Study designs are often limited by the limited number of antibiotics tested, or the number and specificity of the antimicrobial resistance gene primers used. As whole genome sequencing becomes more cost-effective, it is increasingly used to detect both known and unknown antimicrobial resistance genes and determine allelic variants associated with phenotypic resistance (27, 28). However, discordance between genotypic and phenotypic AMR predictions is common for *Vibrio* species. For example, *Vibrio* spp. from Norway carrying ß-lactamase genes were still susceptible to cephalosporins (29).

Given the multiple gaps in knowledge around antimicrobial resistance patterns of *Vibrio* spp., our objectives were to: 1) improve the phenotypic and genotypic characterization of AMR in *Vibrio* spp. isolated between 2000-2019 in an understudied putative bioindicator species in the near-shore environment shared with humans and 2) examine the associations and discordance between genome-based detection of antimicrobial resistance genes with the phenotypic patterns observed from antibiotic susceptibility testing. We hypothesized that *Vibrio* spp. isolates from sea otters and their environment will carry antimicrobial resistance genes with treatment implications for vibriosis and that there will be discordance between phenotypic AMR patterns and their putative genetic determinants.

Materials and Methods

Sample Collection

Vibrio spp. isolates stored at -80°C at the University of California, Davis - Veterinary Medical Teaching Hospital from 2000-2019 were used for whole genome sequencing. Isolates originated from previous California-based projects including avian feces from wildlife hospital admissions

and wild *Larus* spp. gulls (9, 30), live-sampled and necropsied southern sea otters (8, 9, 14, 31), shellfish samples from *Mytilus* spp. mussels and pismo clams (*Tivela stultorum*) (32), water samples collected near Big Sur, California, environmental samples (water, sediment, algae, seagrass, and kelp swabs) from Elkhorn Slough, California, and 3 water samples from the Washington State coastline. Additional *Vibrio* spp. isolates were collected from live and dead northern sea otters in Alaska from 2004-2015 (10, 16).

Whole genome sequencing and genome assembly

Frozen *Vibrio* spp. isolates banked on Microbank beads (Pro-Lab Diagnostics, Fisher Scientific, Round Rock, TX) were plated on 5% sheep blood agar (Biological Media Services, University of California, Davis; Hardy Diagnostics, Santa Maria, CA) at 37°C with 5% CO_2 for 24 hours. Isolates were clinically speciated using biochemical testing (API20E, BioMerieux, Durham, NC) and a subset of isolates were confirmed by matrix-assisted laser desorption-ionization mass spectrometry (MALDI-ToF; Bruker Daltonics, Fremont, CA) (33). DNA was extracted using the spin column protocol from the QiaAMP UCP Pathogen extraction kit with the additional mechanical pre-lysis protocol using small pathogen lysis tubes (Qiagen, Hilden, Germany). Due to shipping/manufacturing delays, some extractions were performed using the Wizard genomic DNA purification kit (Promega Corporation, Madison, WI). The quality and yield of genomic DNA was verified using a 2200 TapeStation with genomic DNA ScreenTape (Agilent Technologies, Santa Clara, CA) (34).

Whole genome sequencing (WGS) was performed by the Weimer laboratory at the University of California, Davis as part of the 100K Pathogen Genome Project

(http://www.genomes4health.org/) (7, 34–38). Briefly, WGS was performed using Illumina HiSeq XTEN with PE 150 plus index read (Illumina, San Diego, CA). Libraries were constructed from fragmented DNA (1 μg) using the HTP library preparation kit (Kapa Biosystems, Wilmington, MA) (33, 39). Adapters and Illumina standards were trimmed from paired-end FASTQ files using Trimmomatic (v0.39) (40) to remove TruSeq universal adapters following previously described parameters (41) before quality control of reads using FastQC (v0.11.9) (42). Genomes were assembled from trimmed reads using Shovill (v1.0.4) (43) and genes were annotated with Prokka (v1.14.6) (44).

Genome quality was assessed using FastQC and CheckM (v1.1.2) (45); genomes with <90% genome completeness, >5% contamination, >300 contigs, or <20X estimated genome coverage were removed from the analysis. Taxonomic classification to the species level was confirmed with Kraken2 (v2.0.8) used with a RefSeq microbial genomes database (downloaded May 5, 2021) constructed with standard tools and microbe categories followed by Bayesian re-estimation of Kraken2 hits with Bracken (v2.6.1) (46, 47). An all-against-all genomic similarity comparison between pairwise genomes was visualized by MinHash sketches using Sourmash (v3.2.3) with a k-mer size of 31 and a genome sketch size of 100,000 k-mers per megabase (48, 49). Genomes with Jaccard Similarity Index scores equal to 1 were considered genetically near-identical or identical and were condensed when these isolates arose from the same sample or individual host.

Antimicrobial Resistance Gene Detection

Contigs from assembled genomes were screened for antimicrobial resistance genes using ABRicate (v1.0.1) (T. Seemann, https://github.com/tseemann/abricate) to search the following databases downloaded on March 17, 2021: Comprehensive Antibiotic Resistance Database (CARD, https://card.mcmaster/ca) (50), ResFinder (51–54), MegaRes (55), NCBI AMRFinder (56), and ARG-ANNOT (57). Gene hits were compiled from all databases and minimum cutoffs of 70% identity and 70% coverage were used to remove spurious gene hits. Allelic variants of ß-lactamase genes were called using snippy (v4.6.0, https://github.com/tseemann/snippy) (58) with the following reference isolates: BCW_12327 (*V. parahaemolyticus*, $bla_{CARB-21,}$ locus tag GGOADLPD_00464), BCW_11123 (*V. cholerae* non-O1/O139, bla_{CARB-7}, locus tag EHPCMFLH_03524), and BCW_12271 (*V. alginolyticus*, $bla_{CARB-42}$, locus tag KHCBNDJJ_00127, *ampC*, locus tag KHCBNDJJ_00781).

Antimicrobial Susceptibility Testing

A subset of 165 southern sea otter isolates (*V. alginolyticus* =37, *V. diabolicus* =20, *V. cholerae* =46, *V. parahaemolyticus* =62) were tested for antimicrobial susceptibility using a microbroth dilution method based on Clinical Laboratory Standards Institute guidelines (59) using a standardized National Antimicrobial Resistance Monitoring System (NARMS Gram negative) panel (ThermoFisher Scientific, Sensititre, Waltham, MA). Each panel tested susceptibility to the following 13 antibiotics: ampicillin, amoxicillin/clavulanic acid, ceftriaxone, ceftiofur, cefoxitin, ciprofloxacin, nalidixic acid, chloramphenicol, sulfisoxazole, sulfamethoxazole/trimethoprim, tetracycline, azithromycin, and gentamicin. Each panel included positive and negative controls and weekly quality controls were conducted with ATCC

strains of *Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa,* and *Enterococcus faecalis*.

Interpretation of the minimum inhibitory concentrations (MIC) were based on Clinical Laboratory Standards Institute guidelines for *Vibrio* spp. using the intermediate and resistant breakpoints validated from human sources (59). While no interpretations currently exist for isolates from sea otters, a designation of intermediate was used to indicate the observed MIC (μg/ml) is associated with an uncertain therapeutic effect, whereas the resistant designation was associated with a likelihood of therapeutic failure in human infections. For sulfisoxazole, chloramphenicol, and nalidixic acid, the breakpoints have only been validated for *V. cholerae* but were used here as a best estimate for non-*cholerae Vibrio* spp. (59). No clinical breakpoints were available for ceftiofur and ceftriaxone.

Statistical Analysis

Minimum inhibitory concentrations from antibiotic susceptibility tests were $\log_2$ transformed to linearize the data, and the $\log_2$ of the MIC were compared between *Vibrio* species for select antibiotics using Kruskal-Wallis tests and pairwise Wilcoxon rank sum tests with a Benjamini and Hochberg correction for significant results. The proportion of isolates categorized as intermediate or resistant for each antimicrobial were compared between species using univariate logistic regression. Univariate logistic regression was used to investigate the presence/absence of antimicrobial resistance genes with categorical variables including *Vibrio* species, and within species comparisons of sex (male, female, unknown), age (adults/aged

adults, juveniles/subadults, unknown), year, season, and sample types (environmental, southern sea otter, and northern sea otter).

Presence of antimicrobial resistance genes or specific bla_{CARB} allelic variants were compared to antimicrobial susceptibility test susceptible versus intermediate/resistant interpretations using Chi-squared tests for a subset of 158 isolates that also passed sequencing quality control metrics (35 *V. alginolyticus*, 44 *V. cholerae*, 18 *V. diabolicus*, and 61 *V. parahaemolyticus*). For antimicrobials with MIC below resistant breakpoints, Wilcoxon rank sum tests, were used to compare the $\log_2$ of the MIC with presence of antimicrobial resistance genes, or alternatively, a Chi-squared test was used for antimicrobials with only two MIC values. The above analyses were performed in R (v4.0.3). SatScan software (v10.0) was used to investigate spatial patterns of antimicrobial resistance gene detection and phenotypic patterns of antibiotic resistance based on MIC data. Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used for multiple sequence alignment of bla_{CARB} genes across isolates (60). Outputs from Clustal Omega were imported to JalView (v.2.11.2.0) for data visualization (61).

Results

Species identification and genome similarity

A total of 489 *Vibrio* spp. isolates passed quality control metrics and a large majority (410; 84%) showed consensus between the genomic species identity and the identity assigned using phenotypic methods in the clinical laboratory (Supplementary Figure 1.1). The biggest discrepancy between clinical and genomic species identification was that the clinic databases did not include *V. diabolicus*, which was clinically identified as either *V. alginolyticus* (n=54) or

V. parahaemolyticus (n=3). Additionally, 11 *V. parahaemolyticus* genomes were clinically identified as *V. alginolyticus*. Discordance also occurred with species considered nonpathogenic to humans including isolates identified as *V. ziniensis* and *V. anguillarum* by genomics but identified as *V. diazotrophicus* (n=4) or *V. aestuarianus* (n=2) respectively in the clinic.

The all-against-all genome comparative analysis indicated that all genomes appropriately clustered by genomic species identification methods, further validating genomic methods as a higher resolution method of species identification than clinical methods (Figure 1.1). Visualization of the pangenome rarefaction curves for *V. alginolyticus* (14,907 unique genes), *V. cholerae* (16,569 unique genes), *V. diabolicus* (15,152 unique genes), and *V. parahaemolyticus* (19,132 unique genes) indicated open pangenomes for all species (Supplemental Figure 1.2). Relatively low pairwise genomic similarity within each species indicated considerable within-species genomic diversity that was supported by the low contribution of the core genome to the total pangenome (e.g., 18.7% in *V. parahaemolyticus*).

Pairwise genome similarity comparisons also identified near-identical or identical genomes. Genomes were categorized as duplicates (n=45) and excluded from further analysis if another near-identical or identical genome a) originated from the same sample with multiple colonies sub-cultured, or b) was collected from a different tissue from the same host (Table 1.1). The remaining 444 genomes that were used in subsequent analyses consisted of 65 clusters of between 2 and 10 near-identical or identical genomes and 244 singleton genomes that were different from all other isolates, for a combined 309 unique genome types.

Genome clusters were rare in *V. alginolyticus* and *V. diabolicus* and occurred in genome pairs sampled the same day up to two months apart (Table 1.1). Clusters of similar genomes were common in *V. cholerae* and *V. parahaemolyticus*, and many arose from spatiotemporally localized sampling events of environmental samples, southern sea otter feces, or avian feces. Ten out of eleven clusters with five or more genomes contained at least two genomes collected from the same location within a month period, with four clusters composed of only environmental strains from the same sampling event (Table 1.2). Seven of the eleven clusters contained at least one genome from disparate sampling efforts with five clusters from mixed isolation sources. One cluster of 8 *V. parahaemolyticus* genomes ranged from Moss Landing, CA to Cayucos, CA and included samples from shellfish (2002), southern sea otters (2002-2006), and the environment (2019).

Antimicrobial resistance phenotypes in southern sea otters

The expected minimum inhibitory concentrations of 8 tested antibiotics differed between *Vibrio* spp., although resistance was only observed for ampicillin and sulfisoxazole; intermediate interpretations were also observed for cefoxitin (Figure 1.2). Ampicillin was the most frequently observed resistance; *V. alginolyticus* (78.4%, 95% CI: 62.8, 88.6) and *V. diabolicus* (75%, 95% CI: 53.1, 88.8) were 7.1 (95% CI: 2.8,18.2, p<0.001) and 5.9 (95% CI: 1.9,18.3, p=0.002) times more likely, respectively, to be resistant to ampicillin than *V. parahaemolyticus* (33.9%, 95% CI: 23.3, 46.3), while only one *V. cholerae* isolate was intermediate to ampicillin. Cefoxitin intermediate isolates were 4.7 (95% CI: 1.3-16.5, p =0.02) times more common in *V. alginolyticus* (24.3%,

95% CI: 13.4, 40.1) than *V. parahaemolyticus* (6.5%, 95% CI: 2.5, 15.4) and were also uncommon in *V. diabolicus* (5.0%, 95% CI: 0.9, 23.6) and *V. cholerae* (2.2%, 95% CI: 0.4, 11.3). Only 9 isolates were sulfisoxazole resistant (5 *V. alginolyticus*, 1 *V. cholerae*, 1 *V. diabolicus*, and 2 *V. parahaemolyticus*). *Vibrio* spp. phenotypes of antimicrobial resistance were not associated with southern sea otter age class or sex, nor were there spatial clusters observed.

No multidrug resistant isolates were detected based on a criterion of phenotypic resistance to 3 or more drug classes. However, the sulfisoxazole resistant *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* isolates were also resistant to ampicillin, with 1 *V. parahaemolyticus* isolate also interpreted as intermediate to cefoxitin. Only one northern sea otter isolate (*V. cholerae*) had antimicrobial susceptibility testing results and it was susceptible to amoxicillin-clavulanic acid, ampicillin, ceftiofur, gentamicin, tetracycline, and trimethoprim-sulfamethoxazole.

Comparison of genotypic and phenotypic resistance

Antimicrobial resistance genes detected in *Vibrio* spp. putatively confer resistance to multiple drug classes including ß-lactams, tetracyclines, sulfonamides, fluoroquinolones, and chloramphenicol (Table 1.3). Genes associated with multidrug efflux pumps and resistance to ß-lactams and tetracyclines were frequently detected while genes associated with resistance to fluoroquinolones, aminoglycosides, chloramphenicol, and sulfonamides were uncommon. However, the frequencies of gene detections were not reliably associated with phenotypic antibiotic resistance.

Tetracycline resistance genes *tet34* and *tet35* were ubiquitous in all *Vibrio* species, although southern sea otter isolates with antimicrobial susceptibility testing (n=62) were susceptible to tetracycline. All but one *tet34* gene had a percent coverage of at least 98.9%, but the percent identity was between 75.1-83.9%. For *tet35* genes, the percent coverage ranged between 70.8-100% and the percent identity ranged between 71.7-99.4%. A similar discordance was observed between the sulfisoxazole resistance phenotype and genotype; sulfonamide drug class resistance genes *sul1* and *sul2* were detected in a single *V. parahaemolyticus* isolate from shellfish and was not detected in the five sulfisoxazole resistant SSO isolates. The *sul1* and *sul2* positive shellfish isolate was also the only isolate to carry the following genes: *tetA*, bla_{CMY-2}/*act*, *aph3-lb*, *aph6-ld*, *dfrA1*, and *floR*.

Fluoroquinolone resistance *qnr* genes were detected in only 2 *V. alginolyticus* genomes, 1 *V. diabolicus* genome, and 6 *V. cholerae* genomes. The fosfomycin-resistance conferring *fosA* gene was rare in *Vibrio* spp. except in *V. diabolicus* (40.4%, 95% CI 27.3, 54.9). The chloramphenicol resistance gene *catB9* was only detected in *V. cholerae* (11%, 95% CI 6.9, 17.1) although all tested *V. cholerae* isolates were susceptible to chloramphenicol. The peptide resistance gene *almG* was nearly ubiquitous in *V. cholerae* but absent from the other three *Vibrio* species. Various multiple drug efflux pumps, including the nearly ubiquitous *vexAB*, were detected but were not associated with phenotypic resistance patterns.

At least one gene that putatively confers ß-lactam resistance was present in all species included in this study except *V. anguillarum* and *V. ziniensis*, but species differed in the specific resistance gene that was present. The genes *varG* and bla_{CARB-7} were only detected in *V. cholerae* with the bla_{CARB-7} gene detected in 11.7% of *V. cholerae* isolates (95% CI: 7.3, 17.8), but

only those from avian feces and environmental sources. The gene *varG* was detected in 44.7% (95% CI: 37.1, 52.8) of *V. cholerae* isolates, but the presence of *varG* was not associated with the MIC of amoxicillin/clavulanic acid or ampicillin. The $bla_{CARB-42}$ gene was ubiquitous in *V. alginolyticus* and *V. diabolicus* isolates, although a blast of the protein-coding sequence from *V. diabolicus* genomes most closely matched a recently discovered allelic variant $bla_{CARB-57}$ that was not yet added to the reference databases (62). The *ampC* gene, which putatively confers resistance to cephalosporins, was detected in 74.5% (95% CI: 60.7, 84.9) of *V. alginolyticus* and 78.8% (95% CI: 64.9, 88.5) of *V. diabolicus*, but presence of *ampC* was not associated with an increased MIC to ß-lactams.

Allelic variants of the bla_{CARB} gene were present in all *V. parahaemolyticus* genomes. Either $bla_{CARB-18}$, $bla_{CARB-20}$, or $bla_{CARB-21}$, which differ only at amino acid residues 46 and 88 in their protein coding region, were present in all ampicillin-resistant isolates (Figure 1.3). Presence of additional amino acid substitutions other than AA46 and AA88 in the protein coding region occurred in 32.8% (95% CI: 21.6, 46.1) of the *V. parahaemolyticus* isolates with antimicrobial susceptibility testing. Isolates with additional amino acid substitutions were 9.02 times (95% CI: 2.6, 30.8) more likely to be susceptible to ampicillin than those with $bla_{CARB-18}$, $bla_{CARB-20}$, or $bla_{CARB-21}$ alleles (x^2 =14.0 , p<0.001). Presence of additional substitutions did not differ between wet (December-May) and dry seasons (June-November), male and female otters, adults and juvenile/subadult otters, live versus dead sampling, and spatial distribution. Additional substitutions in the translated bla_{CARB} sequence did not significantly differ between environmental isolates and southern sea otters, but were 3.4 times (95% CI: 1.3, 9.1) more common in northern sea otter isolates compared to southern sea otters (p=0.009).

Discussion

Species identification and genomic similarity

Because pathogenic *Vibrio* species are often closely related to more benign species, accurate species identification is crucial to assess health threats to humans, seafood, and marine wildlife (63, 64). Clinical typing methods, including MALDI-ToF do not accurately identify *Vibrio* spp. and may mischaracterize understudied and emerging pathogens (65). By utilizing genomic-based species identification, we differentiated 52 *V. diabolicus* isolates (49 unique genome types) that were previously identified as *V. alginolyticus* or *V. parahaemolyticus*. *Vibrio diabolicus* is most closely related to *V. alginolyticus*, has been detected in diverse sources and origins, and due to its recent discovery, has limited numbers of sequenced genomes (2, 3, 66). Improved diagnostics for *V. diabolicus* identification will allow for future studies to investigate the pathogenicity and epidemiology in human and marine hosts. Biochemical testing and MALDI-ToF identification methods are not currently optimized for *V. diabolicus* detection due to limited data (67). To improve the resolution of MALDI-ToF for *V. diabolicus*, a custom set of reference mass spectra profiles informed by genomics to represent species diversity could be added; a custom database approach has been effective to improve species identification in other Harveyi clade species (65). The additional *V. diabolicus* genomes in this data set represents the largest publicly available set of *V. diabolicus* genomes and include the first confirmation of northern and southern sea otter infection.

Genome distance comparisons confirmed that the *Vibrio* spp. genomes were appropriately binned by species and indicated both high inter- and intra-species genome diversity, as expected for *Vibrio* species (7, 66, 68, 69). The pairwise genomic comparisons also identified clusters of near-identical or identical genomes, which can be used to identify potential selection bias. Samples were not obtained completely independently, as some isolates were sub-cultured from the same sample or multiple tissues from a single host. Selection bias can result in a considerably smaller number of independent isolates than the total number of isolates sampled due to oversampling of specific strains (70). Interdependent sampling can increase the capture likelihood for desired bacterial species. However, statistical adjustments for interdependent sampling not informed by genomics could incorrectly treat two genetically distinct strains the same as two identical strains. Genomic similarity should be utilized to assess the interdependence between samples and choose the appropriate genome set for further analyses (70–72).

Comparisons of genome similarity can also generate hypotheses around sharing of *Vibrio* spp. strains within localized populations, as well as more widespread transmission patterns. Intensive sampling events included feces from multiple sea otters or multiple gulls (*Larus* spp.) from the same beach, or collection of environmental samples from multiple locations within the same watershed (8, 9). These events were characterized by multiple nearly-identical or identical genomes, which may suggest sharing and localized dissemination of specific strains. Certain strains of *V. parahaemolyticus* are primarily detected within specific isolation sources and can be localized in outbreak scenarios (7, 73).

Other genome clusters were composed of isolates from disparate space-time points indicative of long-distance transmission and environmental persistence. Natural processes such as ocean currents or human activities such as shipping provide opportunities to disseminate highly related strains to geographically distant locations (71, 74). Genomic stability of phylogenomic *V. parahaemolyticus* clusters detected over 10 plus years suggest that a) at least some *Vibrio* lineages can be extremely stable over time and b) related genomes may arise from persistent but undetected reservoirs that can appear sporadically in different times and locations (73, 75).

Some genome clusters were comprised of isolates from mixed sources and indicate potential for cross-species transmission, with marine wildlife potentially acting as reservoirs. Aquatic birds have been proposed as potential reservoirs and vectors for pathogenic *Vibrio* spp. (76–78); sea otters and other marine mammals could represent additional reservoirs that should be further investigated. Screening for *Vibrio* spp. among marine mammals has been limited, and very few of these isolates have been sequenced.

Antimicrobial susceptibility in sea otters

The SSO isolates tested for antimicrobial susceptibility were susceptible to most clinically appropriate drugs tested. Ampicillin resistance was the most common drug resistance observed even though 100% of the *V. cholerae* isolates were susceptible to ampicillin. Few isolates were characterized as cefoxitin intermediate. Previous *V. parahaemolyticus* studies in the United States reported approximately 50% prevalence of ß-lactam drug resistance or higher, while

globally some studies reported as high as 100% prevalence (5, 22, 24, 25). Ampicillin MICs were never higher than 16µg/ml (intermediate) in the only previous SSO antimicrobial susceptibility study (8). Few studies have reported antimicrobial susceptibility testing for whole genome sequenced *V. alginolyticus* and *V. diabolicus* (listed as formerly separate species *V. antiquarius*), but both species commonly exhibited ß-lactam drug resistance in a Norwegian environmental study (29). Prevalence of ß-lactam resistant *V. cholerae* appears to be less frequent than the other three species, which corresponds with the lack of bla_{CARB-7} positive and ß-lactam resistant *V. cholerae* SSO isolates in our study (79–81).

Presumptive sulfisoxazole resistance was observed infrequently in all 4 *Vibrio* species, but the breakpoint for sulfisoxazole (512 µg/ml) is currently only valid for *V. cholerae* (59). A slightly more lenient breakpoint for sulfisoxazole of 256 µg/ml could be useful given the distribution of MIC values and the lack of breakpoints for all *Vibrio* species, especially those intended to inform treatment of stranded marine wildlife. There are inherent reproducibility issues with broth microdilution as results are measured in 2-fold dilutions, and small variations in test conditions could theoretically impact results despite rigorous quality control (82). While we did not find evidence of resistance to additional drug classes beyond ß-lactams and sulfonamides, other studies detected resistance to higher class cephalosporins, chloramphenicol, erythromycin, fluoroquinolones, tetracyclines, or trimethoprim (25, 83, 84).

Multiple drug resistant *V. parahaemolyticus* has previously been reported (85), but was not detected in the current study. Multiple drug resistance should be interpreted both in the context of clinical appropriateness of each antimicrobial and the number of antimicrobials tested. For example, resistance to erythromycin and streptomycin were not assessed in the

current study, but a high prevalence of resistance to both has been reported for *V. parahaemolyticus* (84, 86). Human clinical cases of vibriosis in the United States are treated with quinolones (56.1%), cephalosporins (24.1%), tetracyclines (23.5%), and penicillins (15.4%), although penicillins alone are not typically effective (87). Our data suggest the antibiotics used in human infections (quinolones, cephalosporins, and tetracycline) are also good choices to treat vibriosis in stranded sea otters.

Host and sample type factors

Some of the SSO isolates used in this study originated from prior research that found the prevalence of *V. parahaemolyticus* and *V. cholerae* detection was higher in males than females, and in areas with higher density human populations, or higher coastal freshwater outflows (14). We did not observe sea otter host factor differences in the presence of antimicrobial resistance genes or phenotypic patterns of antimicrobial resistance, nor did we observe spatial clusters of genes. While the prevalence of *Vibrio* spp. may be impacted by sea otter host and environmental factors, these factors might not impact the likelihood of antimicrobial resistance gene detection as many were ubiquitous across isolates. For *V. cholerae*, the bla_{CARB-7} gene was detected in avian and environmental *V. cholerae* isolates only, suggesting potential niche adaptation.

In their potential role as reservoirs for *Vibrio* spp., wild aquatic birds may transfer antimicrobial resistance genes between marine and terrestrial environments (9, 30, 77, 88–90). Substitutions to the bla_{CARB} genes of *V. parahaemolyticus* were more frequently observed in

NSO than SSO, and from California environmental isolates. This may indicate that NSO isolates are more likely to be susceptible to ampicillin than their Southern counterparts; however, additional phenotypic testing is needed to further investigate differences in antimicrobial resistance phenotypes between SSO and NSO.

Discordance between AMR genotype and phenotype

Previous studies that were able to accurately predict antimicrobial resistance phenotypes using whole genome sequencing with gene databases utilized bacterial species with known mechanisms of resistance and well-validated antimicrobial susceptibility test interpretation (56, 82, 91). Genotype-phenotype comparison is complicated in *Vibrio* spp. by extrapolated phenotypic interpretations, incomplete databases for *Vibrio*-specific genes, and gaps in knowledge about the genetic mechanisms of resistance, including factors impacting variable gene expression (29). Comparison of the bla_{CARB-7} genotype and ampicillin resistance phenotype for *V. cholerae* could not be performed because no sea otter isolates carried bla_{CARB-7}. Previously, 57% of non-O1/O139 *V. cholerae* isolates characterized as ampicillin intermediate or resistant harbored bla_{CARB-7}, suggestive of a partial association between genotype and phenotype (79). All *V. parahaemolyticus*, *V. alginolyticus*, and *V. diabolicus* genomes in the current study possessed a copy of a bla_{CARB} family gene, but despite their ubiquity, only 78.4%, 76.2%, and 33.9% of, *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* isolates, respectively, were ampicillin resistant. In-depth examination of the bla_{CARB} protein sequence

from *V. parahaemolyticus* genomes further elucidates the discordance between bla_{CARB} presence and ampicillin susceptibility.

The $bla_{CARB-17}$ gene family was only recently discovered and was thought to be intrinsic to the *V. parahaemolyticus* genome due to its position in the genome and ubiquitous detection in Genbank accessions (92). Nomenclature for AMR genes are inconsistent in literature and across AMR databases and include allelic variants that can vary by as few as one or two amino acids. In this study, 17 different bla_{CARB} genes were detected for *V. parahaemolyticus*, each with different amino acid substitutions in the protein-coding region that may alter gene function or antibiotic specificity. Alignment of the bla_{CARB} proteins revealed that rare amino acid substitutions partially explain phenotypic ampicillin resistance. Sequences that had a 100% match to the protein sequences of $bla_{CARB-18}$, $bla_{CARB-20}$, or $bla_{CARB-21}$ were more likely to be ampicillin resistant. Structural analysis of $bla_{CARB-20}$ confirmed a preference for non-cephalosporin ß-lactams like ampicillin; this plasticity in amino acid substitutions should continue to be monitored for potential functional alterations such as reduced or increased drug binding (93). Discordance between antimicrobial-resistant phenotypes and genotypes arise from multiple factors, including imprecise phenotypic methodology and differential gene expression. The two-component system *vbrK*/*vbrR* regulates expression of bla_{CARB} genes and type III secretion system genes, which results in greater bla_{CARB} gene expression in the presence of ß-lactams (94–96). Future investigations should consider these additional physiological and genetic complexities when interpreting the presence or absence of various antimicrobial resistance genes.

Unexplained discord was noted in the genotype and phenotype for other antimicrobial drug classes including sulfonamides, tetracyclines and multi-drug efflux pumps. The discordance between the rarity of *sul1* and *sul2* sulfonamide resistance genes and the variation in MICs suggests the presence of other genetic mechanisms that *Vibrio* spp. can harbor to deal with sulfonamide class drugs, including mechanisms to transform sulfonamides through metabolic pathways (97). Conversely, tetracycline resistance genes *tet34* and *tet35* were ubiquitous while isolates were phenotypically susceptible. The *tet34* gene is dependent on $MgCl_2$, so variable $MgCl_2$ concentrations could impact antibiotic susceptibility test results (98). While *tet35* works as an energy-dependent efflux pump, it may be less effective than other tetracycline efflux pumps such as *tetA* (99). Additionally, the percent identities of *tet34* and *tet35* in the study genomes were relatively low compared to the reference sequence (70-90%), which may suggest a less functional or non-functional allelic variant.

Antimicrobial resistance can also be mediated through various multidrug resistance efflux pumps such as *vcmA*, which was ubiquitously detected, and is part of the multidrug and toxin compound extrusion (MATE) family that is experimentally shown to increase resistance to multiple drug classes (100, 101). The major facilitatory superfamily (MFS) includes *emrD*, which can confer resistance to chloramphenicol (101). Resistance-nodulation-division (RND) family efflux systems include the *vex* operons *vexRAB, vexCD*, *vexEF*, *vexGH*, and *vexIJK*, although the importance of each operon is poorly understood (101, 102). In particular, *vexB and vexD* are thought to be most important for multidrug efflux; *vexB* was nearly ubiquitous in our isolates, while *vexD* was only detected in *V. cholerae* isolates and a few sea otter *V. parahaemolyticus* isolates. Other efflux pumps such as the *vceCAB* operon provide multidrug efflux of compounds

that are part of the host response, such as the bile acid compound deoxycholate (103, 104). While the efflux pumps detected were not associated with increased antibiotic resistance, many efflux pumps are not utilized exclusively for antimicrobial resistance and may also be important for efflux of virulence factors (101).

Conclusions

Using whole genome sequencing identification methods with analysis of the entire genome, we identified *V. diabolicus* genomes that were previously misclassified and share many of the same antimicrobial resistance genes as known pathogenic species. We characterized the genomic antimicrobial resistance patterns of a large sample set of *V. alginolyticus*, *V. cholerae*, *V. diabolicus*, and *V. parahaemolyticus* isolated from southern sea otters in coastal California. Antimicrobial resistance of *Vibrio* spp. has been previously understudied in California, including in sea otters, which may act as a reservoir and bioindicator of vibriosis. This study suggests that currently used antimicrobial classes of tetracyclines and fluoroquinolones are likely to treat vibriosis in sea otters, although ß-lactams may be ineffective. We have also highlighted the discordance between phenotypic AMR and putative underlying genetic determinants in *Vibrio* species. Continued surveillance of AMR phenotypes and genotypes are warranted to detect biologically relevant changes in antimicrobial resistance of *Vibrio* species over time. Through both phenotypic testing and alignment of bla_{CARB} genes in *V. parahaemolyticus*, we determined that specific allelic variants are more likely to result in ampicillin resistance, and that gene detection alone is not sufficient to predict antimicrobial resistance in *Vibrio* species.

Literature Cited

1. Baker-Austin C, Oliver JD, Alam M, Ali A, Waldor MK, Qadri F, Martinez-Urtaza J. 2018. *Vibrio* spp. infections. Nat Rev Dis Primers 4:1–19.

2. Turner JW, Tallman JJ, Macias A, Pinnell LJ, Elledge NC, Azadani DN, Nilsson WB, Paranjpye RN, Armbrust E V., Strom MS. 2018. Comparative genomic analysis of *Vibrio diabolicus* and six taxonomic synonyms: a first look at the distribution and diversity of the expanded species. Front Microbiol 9:1893.

3. Song J, Liu X, Wu C, Zhang Y, Fan K, Zhang X, Wei Y. 2021. Isolation, identification and pathogenesis study of *Vibrio diabolicus*. Aquaculture 533:736043.

4. DePaola A, Hopkins LH, Peeler JT, Wentz B, McPhearson RM. 1990. Incidence of *Vibrio parahaemolyticus* in U.S. coastal waters and oysters. Appl Environ Microbiol 56:2299–2302.

5. Baker-Austin C, Mcarthur J V., Tuckfield RC, Najarro M, Lindell AH, Gooch J, Stepanauskas R. 2016. Antibiotic resistance in the shellfish pathogen *Vibrio parahaemolyticus* isolated from the coastal water and sediment of Georgia and South Carolina, USA. J Food Prot 71:2552–2558.

6. Silva IP, de Souza Carneiro C, Saraiva MAF, de Oliveira TAS, de Sousa OV, Evangelista-Barreto NS. 2018. Antimicrobial resistance and potential virulence of *Vibrio parahaemolyticus* isolated from water and bivalve mollusks from Bahia, Brazil. Mar Pollut Bull 131:757–762.

7. Miller JJ, Weimer BC, Timme R, Lüdeke CHM, Pettengill JB, Bandoy DJD, Weis AM, Kaufman J, Huang BC, Payne J, Strain E, Jones JL. 2021. Phylogenetic and biogeographic patterns of *Vibrio parahaemolyticus* strains from North America inferred from whole-genome sequence data. Appl Environ Microbiol 87.

8. Brownstein D, Miller MA, Oates SC, Byrne BA, Jang S, Murray MJ, Gill VA, Jessup DA. 2011. Antimicrobial susceptibility of bacterial isolates from sea otters (*Enhydra lutris*). J Wildl Dis 47:278–292.

9. Oates SC, Miller MA, Byrne BA, Chouicha N, Hardin D, Jessup D, Dominik C, Roug A, Schriewer A, Jang SS, Miller WA. 2012. Epidemiology and potential land-sea transfer of enteric bacteria from terrestrial to marine species in the Monterey Bay region of California. J Wildl Dis 48:654–668.

10. Goertz CEC, Walton R, Rouse N, Belovarac J, Burek-Huntington K, Gill V, Hobbs R, Xavier C, Garrett N, Tuomi P. 2013. *Vibrio parahaemolyticus*, a climate change indicator in Alaska marine mammals, p. 41–52. *In* Responses of Arctic marine ecosystems to climate change. Alaska Sea Grant, University of Alaska Fairbanks.

11. Hughes SN, Greig DJ, Miller WA, Byrne BA, Gulland FMD, Harvey JT. 2013. Dynamics of *Vibrio* with virulence genes detected in Pacific harbor seals (*Phoca vitulina richardii*) off California: implications for marine mammal health. Microb Ecol 65:982–994.

12. Greig DJ, Gulland FMD, Smith WA, Conrad PA, Field CL, Fleetwood M, Harvey JT, Ip HS, Jang S, Packham A, Wheeler E, Hall AJ. 2014. Surveillance for zoonotic and selected pathogens in harbor seals *Phoca vitulina* from central California. Dis Aquat Organ 111:93–106.

13. Thornton SM, Nolan S, Gulland FMD. 1998. Bacterial isolates from California sea lions (*Zalophus californianus*), harbor seals (*Phoca vitulina*), and northern elephant seals (*Mirounga angustirostris*) admitted to a rehabilitation center along the central California coast, 1994-1995. Journal of Zoo and Wildlife Medicine 29:171–176.

14. Miller MA, Byrne BA, Jang SS, Dodd EM, Dorfmeier E, Harris MD, Ames J, Paradies D, Worcester K, Jessup DA, Miller WA. 2010. Enteric bacterial pathogen detection in southern sea otters (*Enhydra lutris nereis*) is associated with coastal urbanization and freshwater runoff. Vet Res 41.

15. Miller MA, Moriarty ME, Henkel L, Tinker MT, Burgess TL, Batac FI, Dodd E, Young C, Harris MD, Jessup DA, Ames J, Conrad PA, Packham AE, Johnson CK. 2020. Predators, disease, and environmental change in the nearshore ecosystem: mortality in southern sea otters (*Enhydra lutris nereis*) from 1998–2012. Front Mar Sci 7:582.

16. Burek Huntington KA, Gill VA, Berrian AM, Goldstein T, Tuomi P, Byrne BA, Worman K, Mazet J. 2021. Causes of mortality of northern sea otters (*Enhydra lutris kenyoni*) in Alaska from 2002 to 2012. Front Mar Sci 8:630582.

17. Jessup DA, Miller M, Ames J, Harris M, Kreuder C, Conrad PA, Mazet JAK. 2004. Southern sea otter as a sentinel of marine ecosystem health. Ecohealth 1:239–245.

18. Johnson CK, Tinker MT, Estes JA, Conrad PA, Staedler M, Miller MA, Jessup DA, Mazet JAK. 2009. Prey choice and habitat use drive sea otter pathogen exposure in a resource-limited coastal system. Proceedings of the National Academy of Sciences 106:2242–2247.

19. Burgess TL, Tim Tinker M, Miller MA, Bodkin JL, Murray MJ, Saarinen JA, Nichol LM, Larson S, Conrad PA, Johnson CK. 2018. Defining the risk landscape in the context of pathogen pollution: *Toxoplasma gondii* in sea otters along the Pacific Rim. R Soc Open Sci 5:171178.

20. Moriarty ME, Tinker MT, Miller MA, Tomoleoni JA, Staedler MM, Fujii JA, Batac FI, Dodd EM, Kudela RM, Zubkousky-White V, Johnson CK. 2021. Exposure to domoic acid is an ecological driver of cardiac disease in southern sea otters☆. Harmful Algae 101:101973.

21. Miller MA, Kudela RM, Mekebri A, Crane D, Oates SC, Tinker MT, Staedler M, Miller WA, Toy-Choutka S, Dominik C, Hardin D, Langlois G, Murray M, Ward K, Jessup DA. 2010. Evidence for a novel marine harmful algal bloom: cyanotoxin (microcystin) transfer from land to sea otters. PLoS One 5:1–11.

22. Elmahdi S, DaSilva L V., Parveen S. 2016. Antibiotic resistance of *Vibrio parahaemolyticus* and *Vibrio vulnificus* in various countries: a review. Food Microbiol 57:128–134.

23. Loo KY, Letchumanan V, Law JWF, Pusparajah P, Goh BH, Ab Mutalib NS, He YW, Lee LH. 2020. Incidence of antibiotic resistance in *Vibrio* spp. Rev Aquac 12:2590–2608.

24. Han F, Walker RD, Janes ME, Prinyawiwatkul W, Ge B. 2007. Antimicrobial susceptibilities of *Vibrio parahaemolyticus* and *Vibrio vulnificus* isolates from Louisiana Gulf and retail raw oysters. Appl Environ Microbiol 73:7096–7098.

25. Shaw KS, Rosenberg Goldstein RE, He X, Jacobs JM, Crump BC, Sapkota AR. 2014. Antimicrobial susceptibility of *Vibrio vulnificus* and *Vibrio parahaemolyticus* recovered from recreational and commercial areas of Chesapeake Bay and Maryland Coastal Bays. PLoS One 9:e89616.

26. Rowe W, Verner-Jeffreys DW, Baker-Austin C, Ryan JJ, Maskell DJ, Pearce GP. 2016. Comparative metagenomics reveals a diverse range of antimicrobial resistance genes in effluents entering a river catchment. Water Science and Technology 73:1541–1549.

27. Schürch AC, van Schaik W. 2017. Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance. Ann N Y Acad Sci 1388:108–120.

28. Monk JM. 2019. Predicting antimicrobial resistance and associated genomic features from whole-genome sequencing. J Clin Microbiol 57:e01610-18.

29. Håkonsholm F, Lunestad BT, Aguirre Sánchez JR, Martinez-Urtaza J, Marathe NP, Svanevik CS. 2020. *Vibrios* from the Norwegian marine environment: characterization of associated antibiotic resistance and virulence genes. Microbiologyopen 9:e1093.

30. Siembieda JL, Miller WA, Byrne BA, Ziccardi MH, Anderson N, Chouicha N, Sandrock CE, Johnson CK. 2011. Zoonotic pathogens isolated from wild animals and environmental samples at two California wildlife hospitals. J Am Vet Med Assoc 238:773–783.

31. Kreuder C, Miller MA, Jessup DA, Lowenstine LJ, Harris MD, Ames JA, Carpenter TE, Conrad PA, Mazet JAK. 2013. Patterns of mortality in southern sea otters (*Enhydra lutris nereis*) from 1998–2001. J Wildl Dis 39:495–509.

32. Miller WA, Miller MA, Gardner IA, Atwill ER, Byrne BA, Jang S, Harris M, Ames J, Jessup D, Paradies D, Worcester K, Melli A, Conrad PA. 2006. *Salmonella* spp., *Vibrio* spp., *Clostridium perfringens*, and *Plesiomonas shigelloides* in marine and freshwater invertebrates from coastal California ecosystems. Microb Ecol 52:198–206.

33. Lawton SJ, Weis AM, Byrne BA, Fritz H, Taff CC, Townsend AK, Weimer BC, Mete A, Wheeler S, Boyce WM. 2018. Comparative analysis of Campylobacter isolates from wild birds and chickens using MALDI-TOF MS, biochemical testing, and DNA sequencing. Journal of Veterinary Diagnostic Investigation 30:354–361.

34. Kong N, Ng W, Lee V, Weimer BC, Kelly L. 2013. Production and analysis of high molecular weight genomic DNA for NGS pipelines using Agilent DNA extraction kit (p/n 200600).

35. Lüdeke CHM, Kong N, Weimer BC, Fischer M, Jones JL. 2015. Complete genome sequences of a clinical isolate and an environmental isolate of *Vibrio parahaemolyticus*. Genome Announc 3:e00216-15.

36. Chen P, Den Bakker HC, Korlach J, Kong N, Storey DB, Paxinos EE, Ashby M, Clark T, Luong K, Wiedmann M, Weimer BC. 2017. Comparative genomics reveals the diversity of restriction-modification systems and DNA methylation sites in *Listeria monocytogenes*. Appl Environ Microbiol 83:e02091-16.

37. Kong N, Davis M, Arabyan N, Huang BC, Weis AM, Chen P, Thao K, Ng W, Chin N, Foutouhi S, Foutouhi A, Kaufman J, Xie Y, Storey DB, Weimer BC. 2017. Draft genome sequences of 1,183 *Salmonella* strains from the 100K Pathogen Genome Project. Genome Announc 5:e00518-17.

38. Weimer BC. 2017. 100K Pathogen Genome Project. Genome Announc 5.

39. Miller B, van Kets V, van Rooyen B, Whitehorn H, Jones P, Ranik M, Geldart A, van der Walt E, Appel M, Kong N, Huang C, Storey D, Weimer BC. 2015. A novel, single-tube enzymatic fragmentation and library construction method enables fast turnaround times and improved data quality for microbial whole-genome sequencingKapa Biosystems Application Note.

40. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

41. Higdon SM, Huang BC, Bennett AB, Weimer BC. 2020. Identification of nitrogen fixation genes in *Lactococcus* isolated from maize using population genomics and machine learning. Microorganisms 8:1–25.

42. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

43. Seemann T. 2020. Shovill—Assemble bacterial isolate genomes from Illumina paired-end readsGitHub: San Francisco, CA, USA.

44. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

45. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

46. Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci 3:e104.

47. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:1–13.

48. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. The Journal of Open Source Software 1:27.

49. Bandoy DDR, Weimer BC. 2020. Biological machine learning combined with *Campylobacter* population genomics reveals virulence gene allelic variants cause disease. Microorganisms 8:549.

50. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573.

51. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644.

52. Clausen PTLC, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics 19:1–8.

53. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek

K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy 75:3491–3500.

54. Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. 2022. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom 8.

55. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. 2020. MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res 48:D561–D569.

56. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu CH, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP, Klimke W. 2019. Validating the AMRFINder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother 63:e00483-19.

57. Gupta SK, Padmanabhan BR, Diene SM, Lopez-rojas R, Kempf M, Landraud L, Rolain J-M. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220.

58. Seemann T. 2015. Snippy: fast bacterial variant calling from NGS reads.

59. CLSI. 2016. M45: Methods for antimicrobial dilution and disk susceptibility testing of infrequently isolated or fastidious bacteria, 3rd ed.3rd ed. Clinical and Laboratory Standards Institute, Wayne, PA.

60. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 50:W276–W279.

61. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.

62. Liu M, Zhang W, Peng K, Wang Z, Li R. 2021. Identification of a Novel Plasmid-Mediated Carbapenemase-Encoding Gene, bla VMB-2 , in *Vibrio diabolicus*.

63. Hoffmann M, Monday SR, Fischer M, Brown EW. 2012. Genetic and phylogenetic evidence for misidentification of *Vibrio* species within the Harveyi clade. Lett Appl Microbiol 54:160–165.

64. Gabriel MW, Matsui GY, Friedman R, Lovell CR. 2014. Optimization of multilocus sequence analysis for identification of species in the genus *Vibrio*. Appl Environ Microbiol 80:5359–5365.

65. Mougin J, Flahaut C, Roquigny R, Bonnin-Jusserand M, Grard T, Le Bris C. 2020. Rapid identification of *Vibrio* species of the Harveyi clade using MALDI-TOF MS profiling with main spectral profile database implemented with an in-house database: Luvibase. Front Microbiol 11:2494.

66. Pérez-Duque A, Gonzalez-Muñoz A, Arboleda-Valencia J, Vivas-Aguas LJ, Córdoba-Meza T, Rodriguez-Rey GT, Díaz-Guevara P, Martinez-Urtaza J, Wiesner-Reyes M. 2021. Comparative genomics of clinical and environmental isolates of *Vibrio* spp. of Colombia: implications of traits associated with virulence and resistance. Pathogens 10:1605.

67. Culot A, Grosset N, Bruey Q, Auzou M, Giard JC, Favard B, Wakatsuki A, Baron S, Frouel S, Techer C, Gautier M. 2021. Isolation of Harveyi clade *Vibrio* spp. collected in aquaculture farms: How can the identification issue be addressed? J Microbiol Methods 180:106106.

68. Koslicki D, Falush D. 2016. MetaPalette: a k-mer painting approach for metagenomic taxonomic profiling and quantification of novel strain variation. mSystems 1:e00020-16.

69. Nathamuni S, Jangam AK, Katneni VK, Selvaraj A, Krishnan K, Kumar S, Avunje S, Balasubramaniam S, Grover M, Alavandi S V., Koyadan VK. 2019. Insights on genomic diversity of *Vibrio* spp. through Pan-genome analysis. Ann Microbiol 69:1547–1555.

70. Cui Y, Yang X, Didelot X, Guo C, Li D, Yan Y, Zhang Y, Yuan Y, Yang H, Wang J, Wang J, Song Y, Zhou D, Falush D, Yang R. 2015. Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. Mol Biol Evol 32:1396–1410.

71. Yang C, Pei X, Wu Y, Yan L, Yan Y, Song Y, Coyle NM, Martinez-Urtaza J, Quince C, Hu Q, Jiang M, Feil E, Yang D, Song Y, Zhou D, Yang R, Falush D, Cui Y. 2019. Recent mixing of *Vibrio parahaemolyticus* populations. The ISME Journal 2019 13:10 13:2578–2588.

72. Cui Y, Yang C, Qiu H, Wang H, Yang R, Falush D. 2020. The landscape of coadaptation in *Vibrio parahaemolyticus*. Elife 9.

73. Yang C, Li Y, Jiang M, Wang L, Jiang Y, Hu L, Shi X, He L, Cai R, Wu S, Qiu Y, Lu L, Zuo L, Chen Q, Wu Y, Martinez-Urtaza J, Wan C, Yang R, Cui Y, Hu Q. 2022. Outbreak dynamics of foodborne pathogen *Vibrio parahaemolyticus* over a seventeen year period implies hidden reservoirs. Nature Microbiology 2022 7:8 7:1221–1229.

74. González-Escalona N, Gavilan RG, Brown EW, Martinez-Urtaza J. 2015. Transoceanic spreading of pathogenic strains of *Vibrio parahaemolyticus* with distinctive genetic signatures in the recA gene. PLoS One 10.

75. Guerrero A, Gomez-Gil B, Lizarraga-Partida ML. 2021. Genomic stability among O3:K6 *V. parahaemolyticus* pandemic strains isolated between 1996 to 2012 in American countries. BMC Genom Data 22.

76. Laviad-Shitrit S, Izhaki I, Arakawa E, Halpern M. 2018. Wild waterfowl as potential vectors of *Vibrio cholerae* and *Aeromonas* species. Tropical Medicine and International Health 23:758–764.

77. Laviad-Shitrit S, Izhaki I, Halpern M. 2019. Accumulating evidence suggests that some waterbird species are potential vectors of *Vibrio cholerae*. PLoS Pathog 15:e1007814.

78. Muangnapoh C, Tamboon E, Supha N, Toyting J, Chitrak A, Kitkumthorn N, Ekchariyawat P, Iida T, Suthienkul O. 2022. Multilocus sequence typing and virulence potential of *Vibrio parahaemolyticus* strains isolated from aquatic bird feces. Microbiol Spectr 10.

79. Melano R, Petroni A, Garutti A, Saka HA, Mange L, Pasterán F, Rapoport M, Rossi A, Galas M. 2002. New carbenicillin-hydrolyzing beta-lactamase (CARB-7) from *Vibrio cholerae* non-O1, non-O139 strains encoded by the VCR region of the *V. cholerae* genome. Antimicrob Agents Chemother 46:2162–2168.

80. Laviad-Shitrit S, Sharaby Y, Izhaki I, Peretz A, Halpern M. 2018. Antimicrobial susceptibility of environmental non-O1/non-O139 *Vibrio cholerae* isolates. Front Microbiol 9:1726.

81. Verma J, Bag S, Saha B, Kumar P, Ghosh TS, Dayal M, Senapati T, Mehra S, Dey P, Desigamani A, Kumar D, Rana P, Kumar B, Maiti TK, Sharma NC, Bhadra RK, Mutreja A, Nair GB, Ramamurthy T, Das B. 2019. Genomic plasticity associated with antimicrobial resistance in *Vibrio cholerae*. Proc Natl Acad Sci U S A 116:6226–6231.

82. Zhao S, Tyson GH, Chen Y, Li C, Mukherjee S, Young S, Lam C, Folster JP, Whichard JM, McDermott PF. 2016. Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in *Campylobacter* spp. Appl Environ Microbiol 82:459–466.

83. Igbinosa EO. 2015. Detection and antimicrobial resistance of *Vibrio* isolates in aquaculture environments: implications for public health. Microbial Drug Resistance 22:238–245.

84. Silvester R, Alexander D, Ammanamveetil MHA. 2015. Prevalence, antibiotic resistance, virulence and plasmid profiles of *Vibrio parahaemolyticus* from a tropical estuary and adjoining traditional prawn farm along the southwest coast of India. Ann Microbiol 65:2141–2149.

85. Park K, Mok JS, Kwon JY, Ryu AR, Kim SH, Lee HJ. 2018. Food-borne outbreaks, distributions, virulence, and antibiotic resistance profiles of *Vibrio parahaemolyticus* in Korea from 2003 to 2016: a review. Fish Aquatic Sci. BioMed Central Ltd. https://doi.org/10.1186/s41240-018-0081-4.

86. Mok JS, Ryu A, Kwon JY, Kim B, Park K. 2019. Distribution of *Vibrio* species isolated from bivalves and bivalve culture environments along the Gyeongnam coast in Korea: virulence and antimicrobial resistance of *Vibrio parahaemolyticus* isolates. Food Control 106:106697.

87. Wong KC, Brown AM, Luscombe GM, Wong SJ, Mendis K. 2015. Antibiotic use for *Vibrio* infections: important insights from surveillance data. BMC Infect Dis 15:1–9.

88. Miyasaka J, Yahiro S, Arahira Y, Tokunaga H, Katsuki K, Hara-Kudo Y. 2006. Isolation of *Vibrio parahaemolyticus* and *Vibrio vulnificus* from wild aquatic birds in Japan. Epidemiol Infect 134:780–785.

89. Halpern M, Senderovich Y, Izhaki I. 2008. Waterfowl—the missing link in epidemic and pandemic cholera dissemination? PLoS Pathog 4:e1000173.

90. Wang J, Ma Z-B, Zeng Z-L, Yang X-W, Huang Y, Liu J-H. 2017. The role of wildlife (wild birds) in the global transmission of antimicrobial resistance genes. Zool Res 38:55–80.

91. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davisa JJ. 2019. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal *Salmonella*. J Clin Microbiol 57:e01260-18.

92. Chiou J, Li R, Chen S. 2015. CARB-17 family of β-lactamases mediates intrinsic resistance to penicillins in *Vibrio parahaemolyticus*. Antimicrob Agents Chemother 59:3593–3595.

93. Li P, Liu C, Li B, Ma Q. 2020. Structural analysis of the CARB β-lactamase from *Vibrio parahaemolyticus* facilitates application of the β-lactam/β-lactamase inhibitor therapy. Biochimie 171:213–222.

94. Hofer U. 2016. β-lactam sensor discovered. Nature Reviews Microbiology 2016 14:4 14:195–195.

95. Cho SY, Yoon S il. 2020. Structural analysis of the sensor domain of the β-lactam antibiotic receptor VbrK from *Vibrio parahaemolyticus*. Biochem Biophys Res Commun 533:155–161.

96. Gu D, Zhang Y, Wang Q, Zhou X. 2020. S-nitrosylation-mediated activation of a histidine kinase represses the type 3 secretion system and promotes virulence of an enteric pathogen. Nature Communications 2020 11:1 11:1–19.

97. Wang Q, Wang H, Jiang Y, Lv M, Wang X, Chen L. 2021. Biotransformation mechanism of *Vibrio diabolicus* to sulfamethoxazole at transcriptional level. J Hazard Mater 411:125023.

98. Nonaka L, Suzuki S. 2002. New Mg2+-dependent oxytetracycline resistance determinant Tet 34 in *Vibrio* isolates from marine fish intestinal contents. Antimicrob Agents Chemother 46:1550–1552.

99. Teo JWP, Suwanto A, Poh CL. 2000. Novel β-Lactamase genes from two environmental isolates of *Vibrio harveyi*. Antimicrob Agents Chemother 44:1309–1314.

100. Huda MN, Morita Y, Kuroda T, Mizushima T, Tsuchiya T. 2001. Na+-driven multidrug efflux pump VcmA from *Vibrio cholerae* non-O1, a non-halophilic bacterium. FEMS Microbiol Lett 203:235–239.

101. Kitaoka M, Miyata ST, Unterweger D, Pukatzki S. 2011. Antibiotic resistance mechanisms of *Vibrio cholerae*. J Med Microbiol 60:397–407.

102. Bina XR, Provenzano D, Nguyen N, Bina JE. 2008. *Vibrio cholerae* RND family efflux systems are required for antimicrobial resistance, optimal virulence factor production, and colonization of the infant mouse small intestine. Infect Immun 76:3595–3605.

103. Borges-Walmsley MI, Du D, McKeegan KS, Sharples GJ, Walmsley AR. 2005. VceR regulates the vceCAB drug efflux pump operon of Vibrio cholerae by alternating between mutually exclusive conformations that bind either drugs or promoter DNA. J Mol Biol 349:387–400.

104. Alatoom AA, Aburto R, Hamood AN, Colmer-Hamood JA. 2007. VceR negatively regulates the vceCAB MDR efflux operon and positively regulates its own synthesis in *Vibrio cholerae* 569B. Can J Microbiol 53:888–900.

Figures

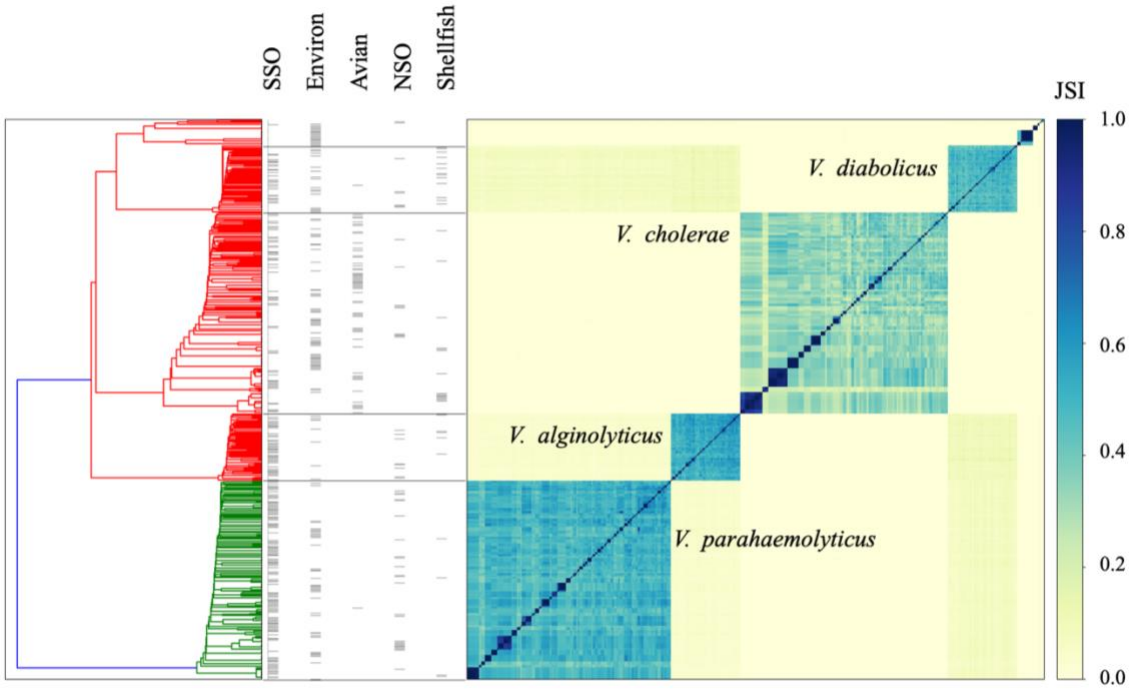


Figure 1.1. All-against-all comparative analysis of the similarity between 489 *Vibrio* species genomes collected between 2000-2019 from various sources in coastal Alaska, California, and Washington confirms genomes clustered by their genomic species identification. Within the four most frequently isolated species (labeled), there was both large genomic diversity and small clusters of near-identical or identical genomes. Darker blue colors on the heatmap indicate higher Jaccard Similarity Index (JSI) values and higher similarity between genomes while lighter yellow colors indicate lower similarity between genomes. Genomes are annotated by source from most frequent to least frequent: southern sea otters (SSO), environment, avian, northern sea otters (NSO), and shellfish.

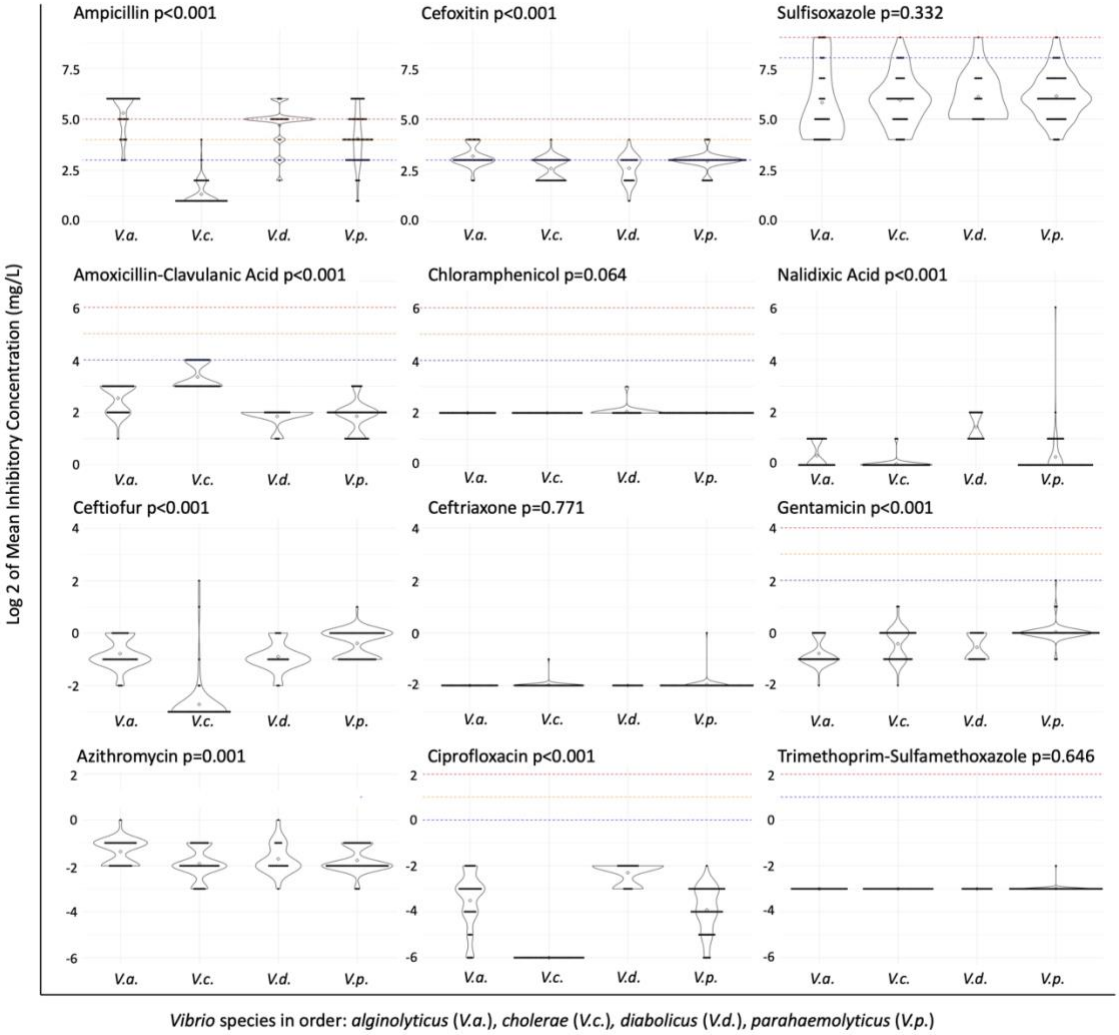


Figure 1.2. Violin plot of antimicrobial susceptibility testing to 13 antibiotics by microbroth dilution of a subset of n=37 *V. alginolyticus*, n=46 *V. cholerae*, n= 20 *V. diabolicus*, and n=62 *V. parahaemolyticus* isolated from southern sea otters off the California coast between 2000-2013. *Vibrio* species differed in their minimum inhibitory concentration to eight different antibiotics based on Kruskal-Wallis p-values less than 0.05. All isolates exhibited a minimum inhibitory concentration to tetracycline of <= 4mg/L (not pictured). Dashed lines represent the clinical laboratory standards institute's breakpoints for susceptible (blue), intermediate (orange), and resistant (red) interpretations for each antibiotic where available. Sulfisoxazole breakpoints are validated for *V. cholerae* only.

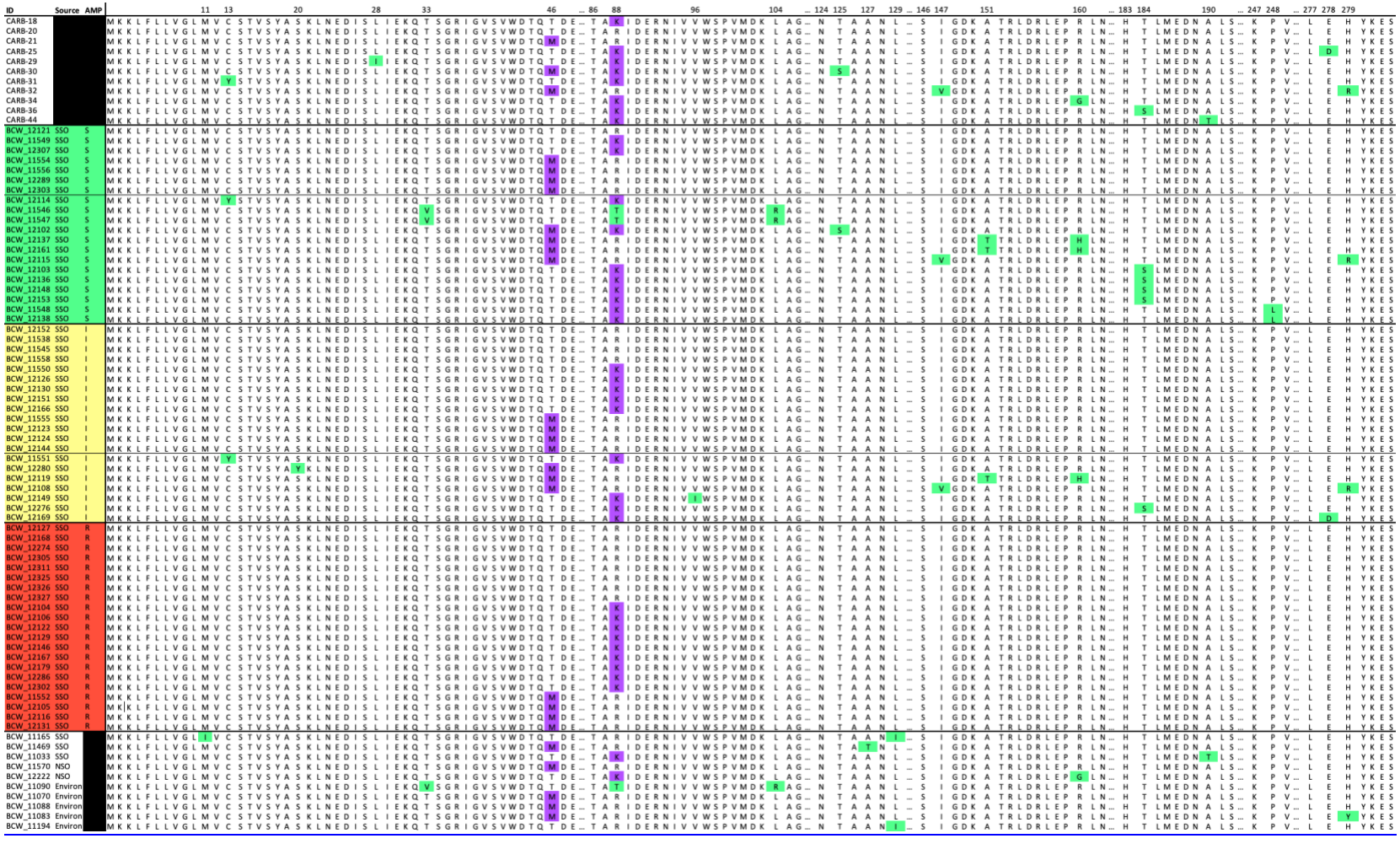


Figure 1.3. Amino acid sequence alignment of translated *bla*$_{CARB}$ genes from 61 *Vibrio parahaemolyticus* isolates from southern sea otters in California (2000-2013) separated by ampicillin susceptibility testing interpretation and compared to 11 reference sequences (top rows) from the Comprehensive Antibiotic Resistance Database. All sequences from ampicillin resistant isolates aligned exactly with proteins Bla$_{CARB-18}$, Bla$_{CARB-20}$, or Bla$_{CARB-21}$.

Figure 1.3 continued. Sequences with at least one additional amino acid substitution were more likely to be susceptible to ampicillin than sequences that exactly matched $Bla_{CARB-18}$, $Bla_{CARB-20}$, or $Bla_{CARB-21}$ proteins (OR=9.0 (2.6, 30.8), p<0.001). Isolate IDs on the left column are color coded based on antibiotic susceptibility testing as ampicillin susceptible (green, n=20), intermediate (yellow, n=20), or resistant (red, n=21). Ten additional isolates from environmental samples, northern sea otters, and southern sea otters collected between 2008-2019 are included (white) to represent the additional variability in bla_{CARB} sequences observed across all *V. parahaemolyticus* isolates sequenced. Amino acid substitutions T46M and R88K which differentiate $Bla_{CARB-18}$, $Bla_{CARB-20}$, and $Bla_{CARB-21}$ are highlighted purple while all other amino acid substitutions are highlighted in green.

Tables

Table 1.1. Comparison of the number of genomes in each *Vibrio* species sampled from various sources in coastal Alaska, California, and Washington between 2000 and 2019 before and after removal of duplicates. Pairwise analysis of genome similarity was used to identity duplicate genomes when two or more sequenced isolates arose from the same sample or host. After removal of duplicate genomes, clusters of near-identical or identical genomes from different samples and hosts were still common, especially in *V. cholerae* and *V. parahaemolyticus*. Genomes that were not closely related to any others were considered singletons. The total unique genomes were composed of the sum of singletons and the number of genome clusters.

Vibrio species	Starting Genomes	Duplicates Removed	Remaining Genomes	Clusters (size range)	Singletons	Unique Genomes
V. alginolyticus	57	2	55	1 (2)	53	54
V. anguillarum	4	1	3	0	3	3
V. cholerae	177	14	163	29 (2-10)	68	97
V. diabolicus	57	5	52	3 (2)	46	49
V. harveyi	1	0	1	0	1	1
V. metschnikovii	10	4	6	1 (6)	0	1
V. parahaemolyticus	179	18	161	30 (2-8)	73	103
V. ziniensis	4	1	3	1 (3)	0	1
Total	489	45	444	65 (2-10)	244	309

Table 1.2. Table Collection details on genome clusters comprised of five or more genomes from the broader *Vibrio* species sampling efforts in Alaska, California, and Washington from 2000-2019 including the isolation sources and notes on localized genome sharing. Sources labeled SSO are isolated from southern sea otters.

Cluster Size	Species	BCW Sample IDs	Sources	Collection Notes
10	*V. cholerae*	11447, 11449, 11451, 11457, 11463, 11464, 11424, 11427, 11429, 11440	10 Environment	Localized collection May-Jun. 2010 near Andrew Molera SP, CA.
8	*V. cholerae*	11175, 12267, 11448, 11459, 11425, 11432, 11433, 11435	1 Avian, 7 Environment	Environment localized collection May-Jun. 2010 near Andrew Molera SP, CA. Avian feces collected 2007 from a San Francisco Bay wildlife hospital.
8	*V. parahaemolyticus*	12103, 12176, 12180, 12209, 12276, 12148, 11036, 11076	2 Environment, 2 Shellfish, 4 SSO	Environment localized collection Jun. 2019 in Elkhorn Slough, near Moss Landing, CA. Shellfish collected 2002 from Monterey Bay. SSO collected 2003-2006 from Monterey, CA to Cayucos, CA.
6	*V. metschnikovii*	11063, 11066, 11080, 11082, 11085, 11087	6 Environment	Localized collection Jun. 2019 in Elkhorn Slough, near Moss Landing, CA.
6	*V. parahaemolyticus*	11034, 11074, 11166, 11543, 12166, 12167	1 Environment, 5 SSO	Environment collected Jun. 2019 in Elkhorn Slough, near Moss Landing, CA. SSO sampled in 2008 (4) and 2012 between Moss Landing, CA and Lucia, CA.
5	*V. cholerae*	11098, 11100, 11188, 11189, 12147	5 SSO	Localized collection of 4 otters Feb. 2008 near Moss Landing, CA. One additional collected 2006 near Piedras Blancas, CA.
5	*V. cholerae*	11117, 11485, 12246, 12256, 12258	4 Avian, 1 SSO	Localized collection of 3 avian feces Nov. 2007 from a San Francisco bay wildlife hospital. 1 avian feces sampled near Carmel, CA in Aug. 2007. SSO sampled Oct. 2007 near Moss Landing, CA. All isolates from 2007.
5	*V. cholerae*	11141, 11157, 11143, 11483, 11484	3 Avian, 2 SSO	Localized collection of avian feces Aug.-Sep. 2007 near Carmel, CA. SSO collected 2008 near Moss Landing, CA
5	*V. cholerae*	11178, 11442, 11444, 11426, 11434	5 Environment	Localized collection May-Jun. 2010 near Andrew Molera SP, CA.

Table 1.2 continued.

Cluster Size	Species	BCW Sample IDs	Sources	Collection Notes
5	*V. cholerae*	11443, 11445, 11454, 11456, 11458	5 Environment	Localized collection May-Jun. 2010 near Andrew Molera SP, CA.
5	*V. parahaemolyticus*	12127, 12305, 12325, 12326, 12327	5 SSO	Localized collection of two SSO Nov.-Dec. 2005 near Morro Bay, CA. Remaining three collected from 2004-2006 ranging from near Pajaro Dunes, CA to near Vandenberg Air Force Base, CA.

Table 1.3. Prevalence of antimicrobial resistance genes in genomes from four different *Vibrio* spp. collected in Alaska, California, and Washington from 2000-2019 separated by isolation sources. Not described in the table above are *V. anguillarum* (n=3), *V. harveyi* (n=1), *V. metschnikovii* (n=6), and *V. ziniensis* (n=3) which all carried *tet34*, *tet35*, *CRP, vexA*, *vexB*, and *vexH.* In addition, the *V. anguillarum* genomes carried *vcmA*, *vexF*, and *vexK* (33%); V. harveyi carried bla_{VHH-1}, *vcmA*, *vexF*, and *vexK*; *V. metschnikovii* carried *vcmA*, bla_{CARB-4}, and *ugd*; and *V. ziniensis* carried *catB9*, *vceA*, *vceB*, *vceR*, and *vexF*. Isolation sources include avian feces (Avian), environmental sampling (Environ), northern sea otters (NSO), shellfish (Shell), and southern sea otters (SSO).

		alginolyticus count (%)						*diabolicus* count (%)					
Drug Class	**Gene**	**Avian (n=0)**	**Environ (n=7)**	**NSO (n=6)**	**Shell (n=3)**	**SSO (n=39)**	**Total (n=55)**	**Avian (n=1)**	**Environ (n=14)**	**NSO (n=5)**	**Shell (n=12)**	**SSO (n=20)**	**Total (n=52)**
aminoglycoside	*aph(3")-Ib*												
	aph(6)-Id												
beta-lactam	*varG*												
	blaCARB-7												
	blaCARB-42		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	blaCARB-18 to blaCARB-46												
	ampC		5 (71.4)	5 (83.3)	3 (100)	28 (71.8)	41 (74.5)	1 (100)	9 (64.3)	5 (100)	7 (58.3)	19 (95)	41 (78.8)
	ACT/BlaCMY-2												
diaminopyrimidine	*dfrA1*												
	dfrA6												
fluoroquinolone	*qnrS5*					1 (2.6)	1 (1.8)						
	qnrVC4												
	qnrVC6								1 (7.1)				1 (1.9)
	qnrVC10					1 (2.6)	1 (1.8)						
fosfomycin	*fosA*					2 (5.1)	2 (3.6)		5 (35.7)	1 (20)	5 (41.7)	10 (41.7)	21 (37.5)
multidrug efflux	*CRP*		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	EMRD												
	vceCAB, vceR												
	VcmA		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	VexAB		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	VexC												
	VexD												
	VexE												
	VexF		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	23 (95.8)	55 (98.2)
	VexH		v	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	VexK		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
peptide (polymyxin)	*almG*												
	UGD		5 (71.4)	5 (83.3)	3 (100)	28 (71.8)	41 (74.5)		8 (57.1)	1 (20)	5 (41.7)	14 (58.3)	28 (50)
phenicol	*catB9*												
	floR												
sulfonamide	*sul1*												
	sul2												
tetracycline	*tet34*		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	tet35		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12(100)	20 (100)	52 (100)
	tetA												

Table 1.3 (Continued).

		cholerae count (%)						*parahaemolyticus* count (%)					
		Avian (n=50)	Environ (n=52)	NSO (n=5)	Shell (n=12)	SSO (n=44)	Total (n=163)	Avian (n=1)	Environ (n=29)	NSO (n=21)	Shell (n=4)	SSO (n=106)	Total (n=161)
Drug Class	**Gene**												
aminoglycoside	*aph(3")-Ib*										1 (25)		1 (0.6)
	aph(6)-Id										1 (25)		1 (0.6)
beta-lactam	*varG*	24 (48)	14 (26.9)	2 (40)	10 (83.3)	23 (52.3)	73 (44.8)						
	blaCARB-7	5 (10)	14 (26.9)				19 (11.7)						
	blaCARB-42												
	blaCARB-18 to blaCARB-46							1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	ampC												
	ACT/ BlaCMY-2										1 (25)		1 (0.6)
diaminopyrimidine	*dfrA1*										1 (25)		1 (0.6)
	dfrA6									1 (4.8)			1 (0.6)
fluoroquinolone	*qnrS5*												
	qnrVC4	5 (10)				1 (2.3)	6 (3.7)						
	qnrVC6												
	qnrVC10												
fosfomycin	*fosA*				1 (8.3)	2 (4.5)	3 (1.8)					2 (1.9)	2 (1.2)
multidrug efflux	*CRP*	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	emrD	3 (6)	7 (13.5)	2 (40)	4 (33.3)	6 (13.6)	22 (13.5)						
	vceCAB, vceR	23 (46)	34 (65.4)	5 (100)	5 (41.7)	26 (59.1)	93 (57.1)						
	VcmA	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	VexAB	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	VexC	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)						
	VexD	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)			1 (4.8)		12 (11.3)	13 (8.1)
	VexE	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)						
	VexF	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	VexH	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	VexK	50 (100)	50 (96.2)	5 (100)	12 (100)	44 (100)	161 (98.8)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
peptide (polymyxin)	*almG*	49 (98)	49 (94.2)	5 (100)	11 (91.7)	44 (100)	158 (96.9)						
	UGD	8 (16)			2 (16.7)	3 (6.8)	13 (8)	1 (100)	8 (27.6)	3 (14.3)	3 (75)	39 (36.8)	54 (33.5)
phenicol	*catB9*	3 (6)	10 (19.2)	2 (40)		3 (6.8)	18 (11)						
	floR										1 (25)		1 (0.6)
sulfonamide	*sul1*										1 (25)		1 (0.6)
	sul2										1 (25)		1 (0.6)
tetracycline	*tet34*	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	tet35	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)	1 (100)	29(100)	21 (100)	4 (100)	106 (100)	161 (100)
	tetA										1 (25)		1 (0.6)

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Peter J. Sebastian- Contributed to the conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, data visualization, and the writing and editing of the manuscript.

Cory Schlesener- Contributed to the conceptualization, data curation and visualization, formal analysis, methodology, and manuscript review.

Barbara A. Byrne- Contributed to the conceptualization, supervision of data generation and curation, investigation, resources, and manuscript review.

Melissa Miller- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Woutrina Smith- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Francesca Batac- Contributed to the data curation and manuscript review.

Caroline E.C. Goertz- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Bart C. Weimer- Contributed to the conceptualization, methodology, supervision of data generation and curation, project administration, resources, and manuscript review.

Christine K. Johnson- Contributed to the conceptualization, funding acquisition, supervision of data generation and curation, project administration, resources, and manuscript review.

Funding

P. Sebastian was supported by the National Institutes of Health under grant number T32OD011147. Additional funding was provided by an Oiled Wildlife Care Network grant.

Acknowledgments

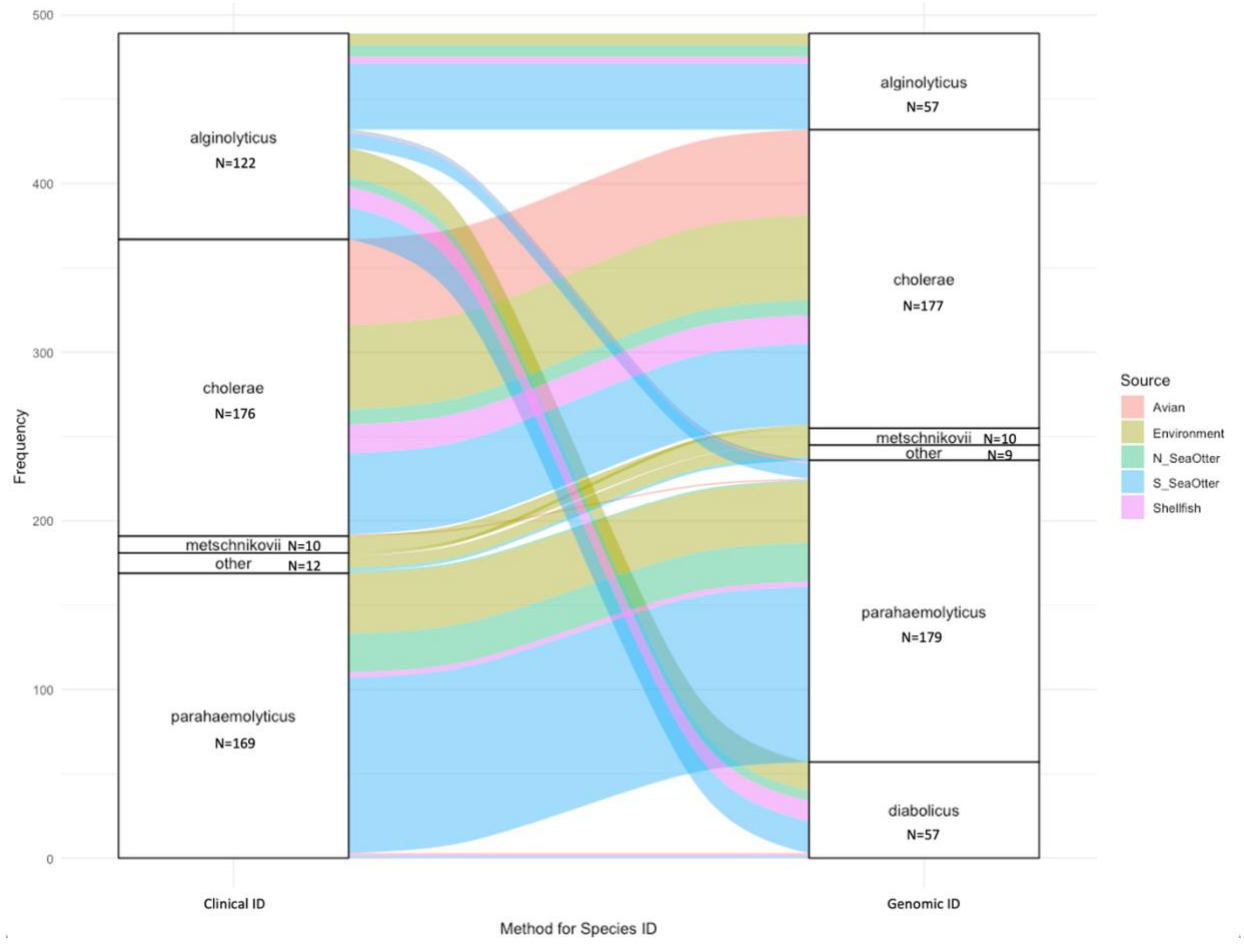
We are grateful to the large number of current and former sea otter investigators, volunteers, and researchers without whom this project would not have been possible including (by organization): University of California, Davis: Brian Bird, Danielle Carrade Holt, Patricia Conrad, Tracey Goldstein, Stori Oates, Jennifer Siembieda, Brett Smith, Robyn Stoddard; California Department of Fish and Wildlife; U.S. Geological Survey: Tim Tinker, Joe Tomoleoni; U.S. Fish and Wildlife Service: Joel Garlich-Miller, Verena Gill, Kristin Worman; Monterey Bay Aquarium: Mike Murray, Marissa Young; Alaska Veterinary Pathology Services: Kathy Burek-Huntington, Natalie Rouse; Alaska SeaLife Center: Natalie Hunter, Pam Tuomi; Sea Otter Savvy: Michelle Staedler; Washington Department of Fish and Wildlife: Katherine Haman. Thanks to the team at 100K Pathogen Project and the Weimer Laboratory especially Carol Huang and DJ Darwin

Bandoy, and to all past and present members of the Epicenter for Disease Dynamics at University of California, Davis especially Nicole Gardner, Megan Moriarty, Pranav Pandit, Jessica Sanchez, and Cristin Young. Special thanks to the late Jack Ames who assisted with our environmental sampling and to whom the field of otter conservation owes a great debt. Special thanks to the late Matthew Blake who embodied the inclusionary culture that made working in the One Health Institute's Karen C. Drayer Wildlife Health Center special.

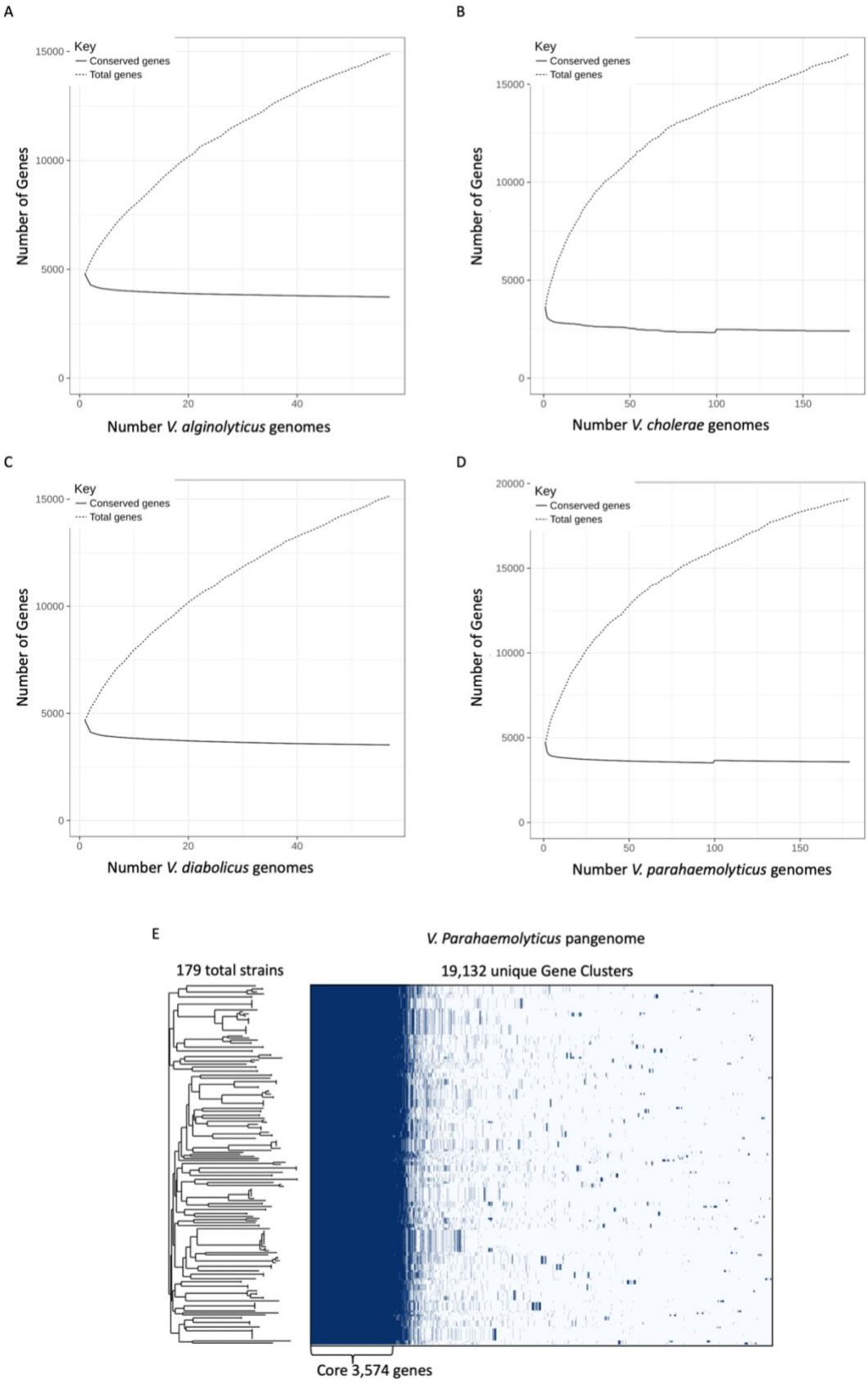
Data Availability Statement

Metadata for the 489 *Vibrio* spp. genomes collected in coastal Alaska, California, and Washington from 2000-2019 will be made available with accession numbers under bioproject PRJNA203445.

Supplementary Material



Supplemental Figure 1.1. Alluvial plot comparison of *Vibrio* species identification as labeled in the microbiology clinic versus using genomic methods for 489 isolates collected from various sources in coastal Alaska, California, and Washington between 2000-2019. While most isolates were corrected identified in the clinic (84%), the clinic methods were not able to identify *V. diabolicus* isolates.


A
Key
Conserved genes
Total genes
Number of Genes
Number V. alginolyticus genomes
B
Number V. cholerae genomes
C
Number V. diabolicus genomes
D
Number V. parahaemolyticus genomes
E
V. Parahaemolyticus pangenome
179 total strains
19,132 unique Gene Clusters
Core 3,574 genes

Supplemental Figure 1.2. Rarefaction curves for the pangenomes of A) *V. alginolyticus* (n=57), B) non-O1/O139 *V. cholerae* (n=177), C) *V. diabolicus* (n=57), and D) *V. parahaemolyticus* (n=179) collected from various sources in coastal Alaska, California, and Washington between 2000-2019. The total number of discovered genes in each pangenome continued to increase as additional genomes were added. E) Pangenome plot of *V. parahaemolyticus* (n=179) highlighting the small proportion of the total gene cluster pool (19,132) comprised of the core genome (3,574). Large accessory genomes were also observed for *V. alginolyticus*, *V. cholerae*, and *V. diabolicus* pangenomes. Blue blocks indicate the presence of a gene cluster (columns) in their respective genomes (rows).

Apparent expansion of virulent *Vibrio parahaemolyticus* in humans and sea otters

Peter J. Sebastian[1,2], Cory Schlesener[3], Barbara A. Byrne[4], Melissa Miller[2,5], Woutrina Smith[2], Francesca Batac[5], Kathy Burek-Huntington[6], Caroline E.C. Goertz[7], Natalie Rouse[6], Natalie Hunter[7], Bart C. Weimer[3*], Christine K. Johnson[1,2*]

[1]EpiCenter for Disease Dynamics, One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA
[2]Karen C. Drayer Wildlife Health Center, One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[3]Department of Population Health and Reproduction, 100K Pathogen Genome Project, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[4]Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[5]Marine Wildlife Veterinary Care and Research Center, California Department of Fish and Wildlife, Santa Cruz, CA, USA

[6]Alaska Veterinary Pathology Service, Eagle River, AK, USA

[7]Alaska SeaLife Center, Seward, AK, USA

*** Correspondence:**
Christine Kreuder Johnson: ckjohnson@ucdavis.edu

Bart C. Weimer: bcweimer@ucdavis.edu

Keywords: whole genome sequening$_1$, genomic epidemiology$_2$, vibriosis$_3$, *Vibrio parahaemolyticus*$_4$, sea otter$_5$, virulence$_6$, *Enhydra lutris*$_7$, thermostable direct hemolysin$_8$, T3SS$_9$, T6SS$_{10}$.

Formatted for submission to *Virulence*.

Abstract

Vibriosis is the most important public health threat from seafood consumption and marine recreation. Pathogenic *Vibrio* spp. employ virulence factors including hemolysins and secretion systems frequently detected in human cases, but virulence data in northern and southern sea otters (*Enhydra lutris keyoni* and *E. l. nereis,* respectively) are limited despite their potential as marine bioindicators. Genomic epidemiology was used to characterize virulence factors of *Vibrio* spp. genomes (n=570) including *V. alginolyticus* (n=55), *V. diabolicus* (n=52), non-O1/O139 *V. cholerae* (n=163), and *V. parahaemolyticus* (n=287) collected in North America (2000-2019). Virulence factors of *V. parahaemolyticus* were compared between isolation sources: bivalves, environment, humans, and southern and northern sea otters. Hemolysins (*tdh*, *trh*) and type III secretion system 2 (T3SS2) gene prevalences were lowest in environmental isolates, while *tdh* and T3SS2 gene prevalences were higher in human and northern sea otter isolates than those from southern sea otters. A hemolysin allele (*trh1*) was detected almost exclusively in human and sea otter isolates. Despite *V. parahaemolyticus* genomic diversity, detected genomic clusters were comprised of highly related and *tdh*$^+$/*trh*$^+$genomes from non-environmental sources including humans and sea otters. Observed pathology in *Vibrio* spp. positive sea otters frequently included septicemia, enteritis, and moderate to severe melena. Co-occurrence of T3SS2 and T6SS1 in *V. parahaemolyticus* was associated with pathological findings and ampicillin susceptible genotypes, suggesting a trade-off between virulence and antimicrobial resistance. Based on these findings, *V. parahaemolyticus* undergoes selection pressures resulting in apparent expansion of virulent strains infecting humans and sea otters.

Introduction

The genus *Vibrio* is diverse with approximately a dozen species implicated in human illness, and even more responsible for disease in marine fish and shellfish. Vibriosis accounts for approximately 80,000 human illnesses in the United States each year; symptoms can include gastroenteritis, skin and ear infections, and septicemia.[1] Species often associated with human vibriosis, including *V. alginolyticus*, non-O1/O139 *V. cholerae*, and *V. parahaemolyticus*, are frequently isolated from marine environments, yet these environmental isolates are usually considered minimally virulent.[2–4]

Virulence factors have been characterized to varying extents in pathogenic *Vibrio* species. *Vibrio parahaemolyticus* is most frequently implicated in vibriosis cases and can carry many well characterized virulence factors.[5] Two related *V. parahaemolyticus* toxins, thermostable direct hemolysin (*tdh*) and thermostable direct hemolysin-like (*trh*), are the primary virulence factors known to cause hemolysis and gastroenteritis in humans and other mammalian hosts.[6–8] Human isolates frequently carry either *tdh*, *trh*, or both, while environmental isolates rarely carry either.[4,6,9] Contradictory studies have reported high prevalences of *tdh* and *trh* from both environmental and shellfish sources.[10,11] Other virulence factors were discovered in attempts to understand why some environmental strains carry one or both hemolysins while some clinical strains are *tdh*/*trh* negative.

Additional *V. parahaemolyticus* virulence factors associated with pandemic strains include type III secretion system 2 (T3SS2), which co-occurs on the same pathogenicity island as *tdh* and *trh*,[12–14] and type VI secretion systems (T6SS), specifically T6SS1 which does not reside on the

same pathogenicity island.[6,15,16] Together T3SS2 and T6SS1 appear to have synergistic effects via adhesion (T6SS1) and host enterocytotoxicity (T3SS2), although the exact functions of T6SS1 are unknown.[16,17]

Similar virulence mechanisms may be exhibited in other *Vibrio* species. Virulence of non-O1/O139 *V. cholerae* strains are partially attributable to the presence of T3SS2.[18] Additionally, T6SS are also common to many gram-negative bacteria and a T6SS has been described for *V. cholerae*.[19] Virulence mechanisms of *V. alginolyticus* and *V. diabolicus* are poorly characterized compared to *V. parahaemolyticus*, and have largely relied on the detection of *V. parahaemolyticus* virulence factors including T3SS, T6SS, and rarely, homologous *tdh* or *trh* hemolysins.[20–22]

Virulence factors represent key host adaptations that *Vibrio* spp. utilize for human colonization and infection,[17,23] but few studies have investigated the role of *Vibrio* spp. virulence factors in pathogenicity to mammals native to marine ecosystems. *Vibrio* spp. were initially considered commensal marine mammal microflora because early reports in cetaceans and pinnipeds rarely reported *Vibrio*-associated pathology.[24] *Vibrio* spp. have since been described as opportunistic pathogens in marine mammals, with varied pathologies of gastroenteritis, septicemia, abscesses, and pneumonia attributed to *Vibrio* spp. infections in a spotted seal (*Phoca largha*), California sea lions (*Zalophus californianus*), harbor seals *(Phoca vitulina*), and northern elephant seals (*Mirounga angustirostris*). [25,26] Septicemia and gastroenteritis have been attributed to *Vibrio* spp. infections in both northern and southern sea otters (*E. l. kenyoni* and *E. l. nereis*, respectively); fatal infections have also been reported in southern sea otters.[27,28] Although detection of *Vibrio* spp. virulence factors is more limited in

marine mammal hosts, confirmed virulence factors include *tdh*$^+$ *V. parahaemolyticus* in bottlenose dolphins (*Tursiops truncates*),[29] *tdh*$^+$/*trh*$^+$ *V. parahaemolyticus* in Pacific harbor seals (*Phoca vitulina richardii*),[30] and isolation of five *tdh*$^+$ *V. parahaemolyticus* strains from northern and southern sea otters.[27,31]

Due to unique aspects of their biology,[28] sea otters may act as bioindicators and natural reservoirs for vibriosis.[32] Both northern and southern sea otters are frequent carriers of *Vibrio* spp. including *V. alginolyticus*, non-O1/O139 *V. cholerae*, and *V. parahaemolyticus*, and sea otters exhibit relatively high site-fidelity to their coastal environment, thus can represent localized *Vibrio* infection risks.[27,32–34] Sea otters also consume large volumes of benthic invertebrates known to concentrate pathogenic *Vibrio* spp., providing a shared source of infections for humans and sea otters.[11,35] *Vibrio* spp. exposure in sea otters is also associated with environmental factors, including exposure to high freshwater outflow and dense coastal human populations.[36]

Further investigation is needed to determine whether virulence factors of *Vibrio* spp. are also adapted for sea otter hosts and to infer the significance of virulence factor detection for sea otter disease. Improved molecular characterization of *Vibrio* spp. from multiple isolation sources is needed to optimize studies of the relatedness and shared transmission pathways of environmental, human and sea otter strains, and ultimately the utility of sea otter *Vibrio* spp. surveillance to protect sea otter and public health. Greater characterization of the co-occurrence of antimicrobial resistant and virulent genotypes from varied isolation sources is also needed given concerns of antibiotic efficacy and the potential implications for treatment of vibriosis in humans and sea otters.[33,37–39] In this study, whole genome sequencing and

population genomics with molecular epidemiology approaches were used to characterize virulence factors and assess genomic similarity of *Vibrio* spp. isolates from sea otters, bivalves, the coastal environment, and humans. *Vibrio parahaemolyticus* was emphasized due to the importance of this bacterium in human infections, and the broad availability of human-derived strains to compare with those from sea otters and bivalves.

We hypothesized that related and virulent *V. parahaemolyticus* with shared virulence factors and genomic similarity would be detected in sea otters and humans, and that infection with virulent strains would be associated with reports of pathology in necropsied sea otters. Our aims were to 1) characterize presence of virulence factors in *Vibrio* spp. isolates based on source, 2) determine the genomic similarity and shared virulence factor patterns of *V. parahaemolyticus* isolates from humans and sea otters and 3) assess whether virulence factor genotypes are associated with pathology in necropsied sea otters, and 4) given the lack of research examining co-occurrence of antimicrobial resistant and virulent genotypes,[38,39] investigate associations between antimicrobial resistance and virulence factors.

Methods

Sample Collection

General methods for sample collection and processing follow the methodology described in Chapter 1. Briefly, n=444 *Vibrio* spp. isolates (55 *V. alginolyticus*, 163 non-O1/O139 *V. cholerae*, 52 *V. diabolicus*, 161 *V. parahaemolyticus,* 3 *V. anguillarum*, 1 *V. harveyi*, 6 *V. metschnikovii*, and 3 *V. ziniensis*) stored at -80°C at the University of California, Davis - Veterinary Medical

Teaching Hospital were used for whole genome sequencing (WGS). *Vibrio* spp. isolated during 2000-2019 in California originated from previous projects including isolates from live-sampled and necropsied southern sea otters,[31,33,36,40] avian feces,[31,41] estuarine environmental samples, and bivalves (*Mytilus* spp. mussels and pismo clams, *Tivela stultorum*).[42] Additional *Vibrio* spp. isolates were collected from live-sampled and necropsied northern sea otters in Alaska from 2004-2015,[27,34] and 3 water samples from coastal Washington state in 2019. Previously sequenced *V. parahaemolyticus* genomes isolated from across the United States and Canada (2006-2007) from oysters (n=58) and human clinical sources (n=68) were also included to provide additional context, resulting in a total sample of 570 *Vibrio* spp. genomes, and 287 *V. parahaemolyticus* genomes.[2]

Whole Genome Sequencing

Isolates were plated on 5% sheep blood agar (Biological Media Services, University of California, Davis; Hardy Diagnostics, Santa Maria, CA) at 37°C with 5% CO_2 for 24 hours and gDNA was extracted from cultures using either the UCP Pathogen extraction kit with pre-lysis using Qiagen small lysis tubes (Qiagen, Hilden, Germany) or the Wizard genomic DNA purification kit (Promega Corporation, Madison, WI). WGS was performed by the Weimer laboratory at the University of California, Davis[2,49–53] as part of the 100K Pathogen Genome Project (http://www.genomes4health.org/), using Illumina HiSeq XTEN with PE 150 plus index reads (Illumina, San Diego, CA) and libraries constructed from 1 mg of gDNA with the HTP library preparation kit (Kapa Biosystems, Wilmington, MA).[43] Previously assembled *V.*

parahaemolyticus genomes from humans and oysters were re-assembled from public domain SRA files using the same pipeline as the newly sequenced genomes.[2]

TruSeq universal adapters and Illumina standards were trimmed from paired-end FASTQ files using Trimmomatic (v.0.39) under previously published parameters.[49,50] Quality of raw reads were examined using FastQC (v.0.11.9)[44] before assembly with Shovill (v.1.0.4) under standard settings[45]. Species classification was assigned with Bracken (v.2.6.1)[46] using the database generated from Kraken2 microbes[47] and a standard Bracken database formation (k-mer size = 35, 150 base length reads). Genome quality was assessed using FastQC and CheckM (v.1.1.2)[48] and all included genomes passed the following metrics: >90% completeness on CheckM, <5% contamination of CheckM, <300 contigs, and >20x genome coverage. Prokka (v.1.14.6)[49] was used for genome annotation with standard use parameters using the assembly input from Shovill. For *V. parahaemolyticus* genomes, genomic similarity was assessed using a minhash based approach with Sourmash (v.3.2.3) using a k-mer size of 31 and scaled to 100,000 k-mers/Mbp.[50–52] *V. parahaemolyticus* genomes were also sequence typed for compatibility with the prevailing multilocus sequence typing (MLST) scheme using the bioconda mlst package (v2.23.0; T. Seemann, https://github.com/tseemann/mlst).[53,54]

Virulence Factor and Antimicrobial Resistance Detection

ABRicate (bioconda v1.0.1; T. Seemann, https:github.com/tseeman/abricate) was used to search genomes for virulence factors from the Virulence Factor Database (VFDB file updated March 17, 2018)[55,56] and for antimicrobial resistance genes from ARG-ANNOT (file updated July 16,2019)[57] with a minimum cutoff of 70% gene identity and 70% gene coverage. Sequence

variants of *bla*$_{CARB}$, *tdh* and *trh* genes were called using snippy (bioconda version 4.6.0)[58] with gene reference sequences from VFDB for *tdh*, accession number KP836460 for *trh1*, and accession number KP836463 for *trh2*.[59] Predicted protein sequences for TRH, TDH, and bla$_{CARB}$ were compiled from Prokka.

Sea Otter Pathology

Mortality data gathered from northern and southern sea otter necropsies were used to characterize broad associations between *Vibrio* spp. isolation and sea otter pathology. Evidence for septicemia, melena, and enteritis were assessed; possible and probable enteritis was defined by observation of gross lesions including intestinal mural thickening and edema, intestinal pleating, or mucosal erosion and ulceration. Confirmation of enteritis was determined by histopathology. Pathological findings classified as confirmed, probable or possible were considered positive. For *V. parahaemolyticus* isolates collected from various compiled specimens in necropsied southern sea otters (n=53), mortality data were used to examine associations between the presence/absence of virulence factors and pathological findings recorded at gross necropsy.

Statistical Analysis

Univariate logistic regression was used to investigate associations between the isolate source (environment, bivalves, humans, northern sea otters, and southern sea otters) and prevalence of various virulence factors in *V. parahaemolyticus*. For southern sea otter isolates, chi-squared analyses, or one-factor (univariate) logistic regression for non-dichotomous

categorical variables were used to examine associations between virulence genes and potential risk factors including sex (male, female, unknown) and age (adult, juvenile, unknown). Jalview (v.2.11.2.0) was used to visualize alignment of virulence factors *tdh* and *trh*[60] and Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used for multiple sequence alignment of *tdh* and *trh* protein sequences. Antimicrobial resistance gene patterns of southern sea otters, northern sea otters, and environmental *Vibrio* spp. isolates were previously examined (Chapter 1). Three bla_{carb} alleles ($bla_{carb-18}$, $bla_{carb-20}$, and $bla_{carb-21}$) were associated with resistance to ampicillin while alleles with rare non-synonymous substitutions were associated with ampicillin susceptibility. Univariate and bivariate logistic regression was used to compare predicted ampicillin susceptible genotypes with presence of virulence factors. All other analyses were performed in R (4.0.3).

Results and Discussion

Virulence factors shared between *Vibrio* spp.

Virulence factors characterized for all 570 *Vibrio* spp. genomes included genes that were conserved across all tested species (Supplemental Table 2.1). Conserved virulence factors included genes involved in adherence (*ompU*, *ilpA*), quorum sensing (*luxS*), regulation (*htpB-hsp60*), catalase (*katB*), and other functions (*gmhA* and *kdsA*). Some virulence factors were only associated with Harveyi clade species (*V. alginolyticus, V. diabolicus, V. harveyi,* and *V. parahaemolyticus*), including the thermolabile hemolysin gene (*tlh*), adherence factor (*mam7*), catalase gene (*katA*), and T3SS1 genes. In a genomic comparison study using *V. alginolyticus, V. diabolicus, and V. parahaemolyticus* genomes isolated in Colombia, all carried various T3SS1

genes although *V. alginolyticus* genomes were missing some T3SS1 genes such as *vscU*, *vcrD*, *vscL*, *vscO*, and *vxsC* that were present in 100% of the Harveyi clade genomes used in this study.[61] The *gspG* gene, which encodes a major pseudopilin protein of the type II secretion system (T2SS),[62] was nearly ubiquitous across all sampled *Vibrio* spp., except for lower detection in *V. alginolyticus* (32.7%). The T2SS is present in many pathogenic and non-pathogenic gram-negative bacteria; its function is to translocate virulence-related proteins across membranes from microbial cells to host cells.[62]

Virulence factors of *V. cholerae*

All *V. cholerae* isolates carried toxin genes *hlyA*, *rtxB*, *rtxC*, and *rtxC* and 59.6% carried the *rtxA* gene. The *hlyA* and *rtxA* genes code for accessory toxins associated with cytolysis in non-O1/O139 strains.[63,64] Further sequence analysis of *rtxA* is warranted as allelic variants associated with environmental strains may contain premature stop codons.[65] In this study, *ace* and *zot* toxin genes were found in two avian fecal *V. cholerae* isolates; these toxins were previously associated with intestinal secretion and mucosal permeability in O1/O139 *V. cholerae* strains.[64] Additionally, one bivalve and 3 avian *V. cholerae* isolates carried adherence genes (*acfA*, *acfB*, *acfC*, and *acfD*) necessary for cholera toxin function.[66] While these isolates were negative for cholera toxin, detection of toxins and adherence genes in avian fecal isolates supports growing evidence that aquatic birds may be important vectors for virulent non-O1/O139 *V. cholerae*.[67–69]

One *V. anguillarum* genome and all the *V. cholerae* genomes carried an additional gene involved in quorum sensing (*cqsA*). The T6SS, which help *V. cholerae* evade phagocytes and enhance intestinal inflammation,[19] were also detected in *V. cholerae* and *V. anguillarum* genomes, while *vipB/mglB* were the only T6SS genes to be consistently detected in the Harveyi clade species as well. Reference sequences for T6SS genes in the VFDB are optimized specifically for *V. cholerae*,[56,70] and separate reference genes should be added to optimize T6SS detection in other *Vibrio* species. The T3SS genes were present in nearly half of *V. cholerae* genomes, including all northern sea otter *V. cholerae* isolates (n=5); these genes may enhance virulence and intestinal colonization of non-O1/O139 isolates,[18,71] and are linked to gastroenteritis in humans.[72] Two critical T3SS translocon genes (*vopB2* and *vopD2*), were significantly more likely to be detected in environmental isolates, when compared with those from southern sea otters ($p<0.005$, OR 14.3 (95% CI: 5.3, 38.6)), birds ($p<0.005$, OR 7.2 (95% CI: 2.9,17.9)), or bivalves ($p<0.005$, OR 23.9 (95% CI: 4.5,128.1)). The biological rationale behind the high observed prevalence of T3SS genes in environmental non-O1/O139 *V. cholerae* is unclear, especially because of the much lower prevalences of T3SS genes reported in other environmental non-O1/O139 *V. cholerae* populations.[73,74] Selection bias may have played a role in the detection of T3SS translocon genes, as all the environmental *V. cholerae* isolates were collected from a localized site in Big Sur, CA between May and June of 2010, and frequently shared high genomic similarity (Chapter 1). Sampling may have occurred when conditions were optimized for more virulent strains to outcompete avirulent or low-virulent environmental strains.

Virulence factors of *V. alginolyticus* and *V. diabolicus*

None of the detected virulence genes were unique to *V. alginolyticus* or *V. diabolicus*. The clinically relevant *tdh* and *trh* hemolysin genes of *V. parahaemolyticus* were not detected in *V. alginolyticus* nor *V. diabolicus*. While few previous studies have detected alleles of *tdh* or *trh* in *V. alginolyticus*[75–77] or *V. diabolicus*,[20] most have failed to detect either.[22,61,78,79] Although one study of *V. alginolyticus* isolated from Mexican seafood detected an extremely high prevalence of *tdh* (>86%), samples originated from the same market and included multiple isolates per sample.[75] *Vibrio alginolyticus*-associated pathology can include skin and ear infections, gastroenteritis, and sepsis in humans and marine wildlife, but *V. alginolyticus* virulence factors are less well characterized when compared with *V. parahaemolyticus*.[21] The role of hemolysins in *V. diabolicus* pathogenesis is even less studied and a large knowledge gap exists regarding the potential virulence of *V. diabolicus* in humans and marine mammals. [20,22,80] Future studies should target species-specific alleles and de novo virulence factor discovery for *V. alginolyticus* and *V. diabolicus,* because the existing virulence factor database relies heavily on characterization of *V. cholerae* and *V. parahaemolyticus* virulence genes.

Virulence factors of *V. parahaemolyticus*

Thermostable hemolysin *tdh* and thermostable hemolysin-like *trh* genes were detected in 39.7% and 57.8% of *V. parahaemolyticus* isolates, respectively. Homologs of hemolysin complex genes (*hlyA*, *hlyB*, *hlyC*, and *hlyD*) that are common in *V. cholerae* were also detected in 28.2% of *V. parahaemolyticus* genomes. Homologous T3SS2 genes were detected in both *V. cholerae*

and *V. parahaemolyticus* isolates, although some T3SS2 genes were only detected in *V. parahaemolyticus* (Supplemental Table 2.1). Three partially homologous variants of the T3SS2 gene complex (α, β, and γ) are present in *V. parahaemolyticus* and are transmitted through their respective pathogenicity islands (*Vp-PAI*). T3SS2α is associated with *tdh*$^+$*/trh*$^-$ strains, T3SS2β is associated with *tdh*$^-$*/trh*$^+$ strains, while T3SS2γ is a hybrid associated with *tdh*$^+$*/trh*$^+$ strains.[81–84] Only 11 isolates in this study were *tdh*$^+$*/trh*$^-$, including all 7 ST3 human isolates. While *tdh*$^+$*/trh*$^-$ T3SS2α strains are considered the most common pathogenic strains globally, the *tdh*$^+$*/trh*$^+$ T3SS2γ strains are most frequently linked to human clinical disease within North America.[81,84]

Known T3SS2 effector genes including *vopA/P*, *vopC*, *vopL*, *vopT*, and *vopZ,* were rarely detected (4.2-4.5%). However, T3SS2 effectors were characterized from T3SS2α strains, and are still poorly characterized in T3SS2β and T3SS2γ, although some putative homologs have recently been identified.[81,84,85] Prokka annotations were compared to previously published putative *vopA/P* homolog sequences[81] and all but 2 of the T3SS2 carrying *V. parahaemolyticus* had a *vopA/P* homolog. Thus, the majority of T3SS2 apparatuses in this study were likely to be fully functional, despite missed detection by the Virulence Factor Database. The two T3SS2$^+$ genomes without a putative *vopA/P* homolog (BCW_12146, BCW_12151) were also *tdh*$^-$ despite carrying the variant of *trh* (*trh1*) associated with *tdh*$^+$*/trh*$^+$ strains, which may indicate false negatives due to contig breaks or deletion of part of a *Vp-PAI* due to transposases.[86] These findings suggest that the VFDB[55,56] is only suitable for detection of T3SS2α effectors, and further study is needed to identify novel T3SS2 effector homologs and characterize their role in virulence.[81]

Genomic Relatedness of Virulence in *V. parahaemolyticus*

An all-against-all comparison of *V. parahaemolyticus* genome relatedness assessed the genomic similarity between isolates from various isolation sources (Figure 2.1). Isolates collected from sea otters were representative of the breadth of genomic diversity in *V. parahaemolyticus*. Some otter isolates exhibited genomic similarity with isolates from human and or bivalve sources originating from other states, suggestive of apparent widespread expansion and potentially shared transmission interfaces between hosts. One genomic cluster included three southern sea otter isolates from 2003, 2008, and 2012, one northern sea otter isolate from 2015, and 18 sequence type 36 (ST36) human strains collected from *V. parahaemolyticus* outbreaks across 10 states between October 2006 and November 2007.[2] Apparent expansion with broad geographical distribution can also occur in strains not associated with human infection, as evidenced by detection of ST12 in an oyster from Texas[2] with genomic similarity to 11 additional isolates collected from 2002 to 2019 from mussels (2, CA), sediment (2, CA), southern sea otters (6, CA), and a northern sea otter (1, AK).

Consumption of bivalve mollusks or other seafood could be a shared transmission contact interface for sea otter and human infections; some STs like ST12 and those in clonal complex 34 (CC34 includes ST32, ST34, ST324) are more supportive of this hypothesis than ST36, which was not detected in any of the bivalve isolates, although ST36 and other virulent strains may be highly specialized to successfully shift from a dormant state to colonization of human or sea otter hosts.[2,87] In particular, hemolysins and other genes on their associated pathogenicity

island may help virulent strains shift from dormancy to host colonization, thus leading to proliferation and potential expansion of hemolysin-carrying strains. Within three phylogenetic regions of high genomic similarity (red boxes in Figure 2.1), 71 of 87 isolates (81.6%) were *tdh*$^+$/*trh*$^+$, while only 32 of 199 isolates (16.1%) outside these three regions were *tdh*$^+$/*trh*$^+$. A similar expansion in hemolysin carrying strains was observed in China, where 99% of strains associated with human vibriosis outbreaks (defined as 2 or more cases) were *tdh* positive.[87]

Isolates that were untyped using the MLST scheme were common in environmental (65.5%) and southern sea otter (36.8%) sources, and much lower in human isolates (4.4%), indicating that the current MLST scheme may not have sufficient resolution to inform population dynamics of *V. parahaemolyticus* from novel isolation sources. One southern sea otter strain that was highly related to ST36 was untyped by the MLST scheme, thus the important relatedness between this strain and human outbreaks would have been missed using MLST alone. While sequence typing with MLST is frequently used to detect clinically relevant *V. parahaemolyticus* strains based on sequence variation in just 7 house-keeping genes,[53] core genome MLST[88] or whole genome phylogeny[89] can improve the accuracy of characterization over traditional MLST. The use of whole genome sequences to compare the similarity between genomes allowed us to identify sea otter strains that were closely related to isolates from human cases and better understand the population structure of untyped sea otter strains (Figure 2.1).

The large genomic diversity of *V. parahaemolyticus* was further corroborated by the relatively small size of the core genome within the pangenome (Supplemental Figure 2.1), which consisted of 3,559 gene clusters out of 25,491 total gene clusters detected (14%). The

genomes derived from human clinical and oyster strains re-analyzed here were previously described as creating a closed genome[2]; however, addition of our isolates from varied isolation sources resulted in an open genome. Sampling considerations such as isolation source, location, and selection bias should be considered when assessing openness of the pangenome, as some genome types may be either missed or overrepresented depending on the study design.

Virulence factors associated with isolation sources in *V. parahaemolyticus*

Effectors associated with T3SS2α (*tdh*$^+$/*trh*$^-$) were only detected in human (13.2-14.7%) and bivalve genomes (3.4%) except for one southern sea otter genome. Since effectors are poorly characterized and are still being discovered in T3SS2β and T3SS2γ, the lack of effectors detected by VFDB is unlikely to reflect their actual presence or absence.[81,84,90] Genes *vopB2* and *vcrD2* were chosen as representative markers for the presence of T3SS2, as most of the non-effector T3SS2 genes had the same prevalence as *vopB2* (62%), while *vcrD2* was present in slightly fewer isolates (58.5%) than *vopB2*. Except for one *tdh*$^-$/*trh*$^-$/*vopB2*$^+$ southern sea otter genome, detection of *vopB2* consistently coincided with the identification of *tdh* and/or *trh*, [83] thus *vopB2* may be a putative biomarker for hemolysin positive strains that is reliable for diagnostic screening. Meanwhile, the gene *hlyA* was only present on a subset of *trh*$^+$ genomes, primarily in genomes with *trh*$\Downarrow$ pseudogenes (57 of 57) or *trh2* variants (24 of 61), but only in one *trh1*$^+$ genome.

The relationship between isolation source and virulence factor genome content was characterized further for representative virulence factors of hemolysins and secretion systems

(Table 2.1). Environmental isolates were least likely to carry *tdh* or *trh*. Northern sea otter and human isolates were more likely to carry *tdh* than bivalve and southern sea otter isolates. Unexpectedly, *trh* prevalence did not significantly differ between isolates from humans and bivalves. A fine resolution examination of the *trh* allelic variants (Table 2.2 and Supplemental Table 2.2) revealed that *trh1* allelic variants were most associated with human isolates and were not present in bivalve isolates. Bivalve and northern sea otter isolates were instead associated with *trh*⋔ (*tdh*$^+$) isolates that may be less virulent due to a presumed non-functioning *trh1* pseudogene,[59,91] although the base sequence of ATA (isoleucine) can act as a potential start codon in bacterial species. During the discovery of *trh2*, it was noted that *trh1*$^+$ isolates caused stronger hemolysis and were more associated with human rather than environmental sources, although both strains should be considered potentially virulent.[92] Further refinement of *trh* gene detection to specific alleles may improve genomic characterization of virulent and avirulent strains.[93] Additionally, variation in the promoter region can be just as important as the gene itself.[94]

Human and northern sea otter isolates were more likely to carry the T3SS2 genes *vopB2* and *vcrD2* than southern sea otters, although environmental isolates were the least likely to carry T3SS2 genes. Our findings corroborate previous evidence which suggests that T3SS2 genes are relatively rare in environmental isolates.[95] However, environmental strains that carry T3SS2 genes are a cause for concern as they have been shown to be pathogenic to human cell lines.[96] The *hlyA* gene was significantly higher in bivalve and northern sea otter isolates than those from humans, and the association of *hylA* with *trh2* or *trh*⋔ suggests *hylA* may be transmitted on a *Vp-PAI* containing *trh2* or *trh*⋔. A recent study of a rare *trh*$^-$/*tdh*$^-$ strain (*Vp*-353) from

China claims to be the first to predict the presence of the α-hemolysin *hlyA* along with *hlyB*, *hlyC*, and *hlyD* using VFDB,[97] although analogous proteins of *E. coli hlyA* hemolysin had already been identified in *V. parahaemolyticus*.[98] This study highlights frequent detection of *hlyA* from North American sea otter and bivalve strains. Therefore, *hlyA* may not be the most reliable biomarker for virulent strains in humans, but should be assessed as a potential virulence factor in benthic invertebrates and marine mammals.

The VFDB includes T6SS genes for *V. cholerae*, although it still detected a *vipB/mglB* homolog associated with *V. parahaemolyticus*. The T6SS1 markers *vipB/mglB* were detected 3.97 (95% CI: 1.58, 9.96) times more frequently in human than environmental isolates. While T6SS2 is present in almost all *V. parahaemolyticus,* some strains associated with increased virulence carry a T6SS1 gene cluster on chromosome 1.[5,99] T6SS1 is believed to secrete bactericidal effectors that help *V. parahaemolyticus* compete against other bacteria and may have a synergistic effect with T3SS2.[100]

Prevalences of *hlyA*, *tdh*, *trh*, and T3SS2 genes were higher in northern than southern sea otters. In particular, the α-hemolysin *hlyA* was present in 76.2% of northern sea otter isolates compared to 25.5% of southern sea otter isolates. The habitat of northern sea otters includes marine waters that are typically below the 15ºC threshold for *V. parahaemolyticus* growth.[27,101] While *V. parahaemolyticus* can survive in a dormant state below 15ºC, colder environmental temperatures may act as a selection pressure towards strains suitable for infecting marine wildlife hosts, resulting in outbreaks of virulent strains carrying hemolysins and secretion systems. There appeared to be a potential sampling bias for northern sea otter isolates towards tdh^+/trh^+ genomes that were related to clonal complex 34 (n=8). The CC34 isolates included two

separate potential sea otter outbreaks. Four similar ST34 strains were collected within a five-kilometer distance near Cordova, AK; one strain was sampled in March 2010 while three additional strains were sampled two years later in April of 2012. A separate potential outbreak included three ST324 strains sampled within a ten-kilometer distance in July 2013 near Homer, AK, although another ST324 strain was detected in November 2009 near Kodiak, AK over 200 km away. Since both potential outbreaks were preceded by detection of a related strain years earlier suggest persistence of certain *V. parahaemolyticus* sequence types in Alaskan coastal ecosystems.

Climate change could result in broader *V. parahaemolyticus* expansion in Alaskan coastal waters, which are likely to be negatively impacted by rising sea surface temperatures and more frequent marine heatwaves.[102] Climate change-related extreme weather events have already been linked to disease outbreaks of *V. parahaemolyticus* in Alaska and Canada.[103,104] Interpretation of northern sea otter isolates is limited by the relatively low sample size of northern sea otter isolates and a potential sampling bias towards CC34 outbreaks (n=8).

Examination of virulence factor patterns and genome relatedness revealed that southern sea otters carry a diversity of presumed low-virulent and high-virulent *V. parahaemolyticus* with similarity to strains from diverse sources. Southern sea otters appear to be exposed to environmental strains that may be common intestinal commensals, as well as pathogenic strains from shellfish and humans. Continued surveillance of *V. parahaemolyticus* in southern sea otters can advance our understanding of the environmental determinants and often unidentified reservoirs of pathogenic *V. parahaemolyticus* outbreaks. Even though the majority of *V. parahaemolyticus* in southern sea otters was likely non-pathogenic or low-virulent, the

detection of isolates that clustered with some of the most frequent clinical sequence types in North America (ST36 and ST631) raises concern that *V. parahaemolyticus* is a potential southern sea otter pathogen.

Southern Sea otter pathology and virulence factors

Data were available for 24 *V. alginolyticus*, 12 *V. cholerae*, 11 *V. diabolicus*, and 53 *V. parahaemolyticus* isolates obtained from 94 southern sea otters necropsied between 2001-2012. *Vibrio* spp. isolates were obtained from the following specimens at necropsy: feces (n=38), blood or cardiovascular tissue (n=21), the lower and upper respiratory tract (n=11), mammary tissue and external wounds including an abscess (n=10), abdominal fluid or abdominal device implants (n=5), gastrointestinal tissues, lymph nodes, or reproductive organs (n=3 each), brain (n=2), and bone, liver, neoplasia, or pleural wall laceration (n=1 each). All isolates obtained from abdominal fluid or abdominal device implants and lymph nodes were obtained from sea otters with a final diagnosis of sepsis. Due to a limited sample of most specimen types, pathological findings were pooled for most analyses.

Enteritis, melena, and sepsis were common pathological findings in southern sea otters infected with *Vibrio* spp. (Table 2.3). Sepsis or systemic bacterial infection was listed as the primary cause of death for one *V. alginolyticus*-positive and 2 *V. parahaemolyticus*-positive southern sea otters. Of the four *Vibrio* species assessed in this study, *V. diabolicus* had the lowest association with reports of enteritis, moderate to severe melena, and sepsis in *Vibrio*-positive, necropsied southern sea otters. Although the pathogenicity of *V. diabolicus* is

unknown, this species is genetically similar to, and has similar virulence factor profiles to *V. alginolyticus*.

The top 5 primary causes of death among *Vibrio* spp.-positive sea otters were white shark (*Carcharodon carcharias*) bite wounds (n=22), acanthocephalan peritonitis (n=15), systemic protozoal infection (n=10), cardiomyopathy (n=8), and human-caused injuries, such as boat strike, gunshot, and net or line entanglement (n=8). Twenty-four isolates (8 *V. alginolyticus*, 2 *V. cholerae*, 3 *V. diabolicus*, and 11 *V. parahaemolyticus*) were obtained from sea otters with presumptive or confirmed white shark bite as a cause of death, of which 6 were from acute shark bite cases (1 *V. alginolyticus*, 2 *V. cholerae*, and 3 *V. parahaemolyticus*). The 18 *Vibrio* positive otters with subacute shark bites included 8 *V. alginolyticus* (n=6) or *V. diabolicus* (n=2) isolates from external tissues or deep lacerations(skin wounds, male genitalia, mammary tissue, or chest wall laceration). External wounds have been implicated in *V. alginolyticus* infections in humans;[1] shark bites may provide a similar route of transmission in sea otters by disrupting skin barriers that prevent *V. alginolyticus* colonization.

Acanthocephalan peritonitis is commonly associated with bacterial co-infection due to intestinal mucosal and mural damage; bacterial co-infections were reported in 61% of fatal acanthocephalan peritonitis cases in a recent southern sea otter mortality study, compared to only 4% for otters without acanthocephalan peritonitis.[28] Fatal acanthocephalan peritonitis was associated with 37 *Vibrio* spp. isolates (21 *V. parahaemolyticus*, 8 *V. alginolyticus*, and 4 each for *V. cholerae* and *V. diabolicus*). Because acanthocephalan peritonitis is a common cause of southern sea otter death, improved understanding of *Vibrio* spp. co-infection is warranted.

Because virulence factor patterns were most variable in *V. parahaemolyticus*, associations between *V. parahaemolyticus* virulence factors and reported southern sea otter pathology were examined (Table 2.4). Enteritis, melena, and sepsis were most prevalent in southern sea otters with *V. parahaemolyticus* isolates carrying both T6SS1 and T3SS2 genes. While enteritis, melena, and sepsis are all associated with *Vibrio parahaemolyticus* infections in humans,[1] it is hard to attribute the presence of these conditions specifically to *Vibrio parahaemolyticus* in sea otters based on current knowledge due to confounding from various primary, secondary, tertiary, and quaternary causes of death, bacterial co-exposure, the range of specimens assessed, and difficulty in assessing etiopathogenesis in carcasses with varying levels of autolysis. Mixed bacterial ~~isolates~~ cultures were common, especially for *Vibrio* spp. in combination with *Streptococcus* spp., *E. coli*, and *Clostridium* spp. Primary bacterial infections were reported in approximately 12% of sea otter mortalities [28], but bacterial sepsis was associated with 59% of the *Vibrio* spp. positive southern sea otters with mortality data. Future sea otter mortality studies should determine the extent to which *Vibrio* spp. infections are associated with sepsis in comparison to other opportunistic bacterial pathogens.

Due to limited sample sizes, no specific *V. parahaemolyticus* sequence types could be linked to higher risk of pathology in southern sea otters. For example, although ST36 is implicated in numerous vibriosis outbreaks in humans in North America, only one isolate similar to human ST36 strains had southern sea otter mortality data in the current study. For that animal, the primary cause of death was subacute shark bite, but possible enteritis, melena, and sepsis were also noted, which could have been attributable to *V. parahaemolyticus* infection. Further

sampling efforts are needed to better assess putative relationships between highly virulent *V. parahaemolyticus* and clinical disease or pathology in southern sea otters.

Northern Sea otter pathology and virulence factors

Thirty-three isolates (*V. parahaemolyticus* (n=18), *V. alginolyticus* and *V. diabolicus* (n=5 each), *V. cholerae* (n=4), and *V. anguillarum* (n=1)) were collected from various specimens of 27 necropsied northern sea otters with corresponding mortality data, although the associations between presence or absence of enteritis and septicemia were assessed in fewer isolates (n=32 and 31, respectively). Specimens from which the 33 isolates were obtained included gastrointestinal tissues (n=22), tonsil or lymph nodes (n=8), blood (n=2), and feces (n=1). Isolates from lymphatic tissues were all associated with a final diagnosis of bacterial sepsis. Co-infections with *Streptococcus* spp. were common (24 of 33 isolates). Cause of death in *Streptococcus* spp.-positive sea otters included septicemia (n=4), endocarditis (n=2), and meningoencephalitis (n=1). A previous report showed that "*Strep* syndrome" defined as endocarditis, meningoencephalitis, and/or septicemia due to *Streptococcus* spp., was the most frequent cause of northern sea otter mortality between 2002 to 2012.[34] In this study, other frequently observed causes of death included emaciation (n=5) and gastrointestinal disease (n=4). Further study is needed to determine the association of *Vibrio* spp. infections with specific causes of mortality in northern sea otters.

Possible, probable, or confirmed enteritis was observed in three of the five *V. alginolyticus*-positive northern sea otters. Septicemia was a common primary or contributing cause of death

(n=4) in *V. alginolyticus*-positive sea otters; septicemia in these sea otters was complicated by mixed bacterial infections, including *Streptococcus* spp. (n=2) and *Erysipelothrix rhusiopathiae* (n=1). Two necropsied northern sea otters with *V. alginolyticus* infection had open fractures which were a probable route of infection.

Four out of five *V. diabolicus*-positive northern sea otters exhibited probable or confirmed septicemia, while four exhibited probable or confirmed enteritis. Of the four *V. cholerae*-positive necropsied sea otters, two died of septicemia and *Streptococcus* spp. was also cultured from both animals. Enteritis was not reported in *V. cholerae*-positive northern sea otters. A *V. anguillarum*-positive northern sea otter that died of septicemia, exhibited enteritis at necropsy, but was also infected with *Streptococcus lutetiensis*.

Two genetically distinct *V.* parahaemolyticus strains were isolated from a hospitalized northern sea otter (July 25th and August 1st, 2005) prior to death (August 5th, 2005). Primary cause of death was determined to be acute enteritis associated with *V. parahaemolyticus*. Probable or confirmed enteritis was common in *V. parahaemolyticus*-positive sea otters (10 of 16 otters with enteritis data), which corresponds with well-recognized gastrointestinal pathology described in human *V. parahaemolyticus* infections.[1] Melena and abnormal feces were each observed once from *V. parahaemolyticus*-positive sea otters. Five of 16 *V. parahaemolyticus*-positive northern sea otters with septicemia data exhibited probable or confirmed septicemia.

Frequencies of *V. parahaemolyticus* sequence types in sea otters are likely to differ between Alaska and California sub-populations; similarly, the population structure of human and oyster-

sourced *V. parahaemolyticus* isolates is partially related to their geographical origins.[2] In addition, differences in prey species between sea otter subspecies may result in differential exposure risks to pathogenic strains.[35,105] Geographical differences in STs may also impact the prevalence of pathology associated with *V. parahaemolyticus*. Eight of 21 northern sea otter *V. parahaemolyticus* isolates from Alaska were related to clonal complex 34 strains (ST34 and ST324) isolated from human clinical and oyster sources, compared to only three of 106 isolates for southern sea otters from California. Two other sequence types attributed to human outbreaks, ST36 and ST631, were each isolated from a northern sea otter at necropsy; enteritis was observed in both sea otters although future study is needed to examine the associations between specific sequence types and pathology.

Virulence in sea otters might also depend on the allelic variants of virulence factors.[92] For example, the *trh1* variant has 84% homology with *trh2* and strains carrying *trh1* were hemolytic in human, rabbit, sheep, and calf erythrocytes, while *trh2*$^{+}$ strains exhibited weaker hemolysis that was present only in human and rabbit erythrocytes.[92] In this study, *trh1* positive isolates (either full length or truncated) were associated with septicemia in 2 of 3 northern sea otters and 5 of 6 southern sea otters. Meanwhile, *trh2* isolates were associated with septicemia in 2 of 6 northern sea otters and 7 of 9 southern sea otters, and *trh*Ψ isolates were associated with septicemia in 3 of 11 northern sea otters and 7 of 11 southern sea otters.

Traditionally, *Vibrio* infections in sea otters have been considered opportunistic infections that arise secondary to other causes of death.[28,31,33,40] Many otters likely carry low to no-virulence *Vibrio* spp. or exhibit secondary *Vibrio* spp. infections from opportunistic infections, although preliminary findings from this study suggest that some *Vibrio* spp. infecting northern

and southern sea otters could be virulent pathogens. These findings provide a first look at the associations between *Vibrio* spp. isolation and detection of virulence factors with northern and southern sea otter pathology that can inform subsequent efforts to determine disease causality.

Fitness Trade-off between virulence and antimicrobial resistance in *V. parahaemolyticus*

Antimicrobial resistance patterns were not previously reported for the n=126 human clinical and oyster *Vibrio* spp. genomes [2]. In the current study, all human clinical and oyster isolates were positive for *tet34* and *tet35*, the multidrug efflux pump genes *vexA*, *vexB, vexF, vexK,* and *vexH*, and a variant of the bla_{carb} gene. Rare antimicrobial resistance genes included *qnr-S5* found in one humanisolate, *qnrC* in three human isolates, *fosA* in one human and three oyster isolates, and *vexC* in one oyster isolate. Whether human isolates are more likely to carry quinolone resistance genes is unknown, but no *Vibrio* spp. isolates obtained from other sources in the current study were positive for *qnr* genes (Chapter 1).

Alleles of bla_{carb} predicted to be associated with ampicillin susceptibility were present in 35 of 68 (51.5%) human-sourced *Vibrio* isolates, and 15 of 58 (25.9%) oyster-sourced isolates. Predicted ampicillin susceptible genotypes across all *V. parahaemolyticus* isolates were positively associated with detection of T3SS2 genes, hemolysins, or the T6SS1 gene *vipB*/*mglB*, but not with the T3SS1 gene *vopB* in univariate analyses (Table 2.5). Since T3SS2 and T6SS1 are thought to have a synergistic effect on virulence, a multivariate logistic regression model was used to examine interactions between *vipB*/*mglB* and *vopB2* as predictors for ampicillin

susceptible genotypes. When examined together in a bivariate logistic regression model, *vopB2* alone was a significant predictor of ampicillin susceptible genotypes (OR 2.47, 95% CI 1.08, 5.68, p-value = 0.033), but co-occurrence with *vipB*/*mglB* modified the effect such that concurrent presence of both markers was significantly associated with predicted ampicillin susceptibility (OR 12.41, 95% CI 5.29, 29.11, p-value < 0.001). Thus, the genomes with both T3SS2 and T6SS1 secretion systems were the most likely to be predicted as ampicillin susceptible.

Translated gene sequences of bla_{carb} are imperfect predictors for ampicillin resistance and further study should combine phenotypic and genotypic methods of both AMR and virulence detection to better understand the trade-off between the two. Variable gene expression of both antimicrobial resistance genes and virulence factor genes may also alter the phenotypic patterns that can be observed for both AMR and virulence. Potential trade-offs in fitness occur with both increased virulence and increased antibiotic resistance and the relationship between the two may depend on the specific genes.[106,107] Environmental isolates can be exposed to antimicrobials as coastal systems receive effluents from human waste that include antimicrobial residues, making coastal waters and sediments likely reservoirs for antimicrobial resistance.[108,109] Apparent expansion of highly virulent *V. parahaemolyticus* STs, such as ST36, may have occurred inadvertently in ampicillin-susceptible strains, leading to divergent virulent but ampicillin-susceptible populations.

Conclusions

This study provides evidence that highly virulent and genomically similar *V. parahaemolyticus* strains can infect both humans and a putative marine bioindicator species, sea otters. While sea otters are exposed to diverse V. parahaemolyticus strains with both low and high-risk virulence factor profiles, the results presented reveal that virulent *V. parahaemolyticus* strains are broadly associated with similar pathology in sea otters as previously described in human cases including enteritis and septicemia. Similar strains in both sea otters and humans included internationally expansive sequence types, suggesting apparent expansion of virulent *V. parahaemolyticus*. Future studies can build upon the comparative epidemiological approach applied here to provide causal inferences further linking markers of virulence to morbidity and mortality of sea otters and humans. As natural hosts occupying coastal ecosystems, sea otters may be informative of *Vibrio* spp. virulence selection that could impact human health. Given predicted impacts of climate change on *Vibrio* spp. infection risks and shared transmission contact interfaces for humans and sea otters, the role of sea otters as bioindicators of Vibrio spp. infections should be prioritized. This study is the first to characterize multiple virulence factors of *V. parahaemolyticus* isolated from sea otters and suggests that apparent expansion of virulent strains impacts both sea otter and human health.

Literature Cited

1. Baker-Austin C, Oliver JD, Alam M, Ali A, Waldor MK, Qadri F, Martinez-Urtaza J. *Vibrio* spp. infections. Nat Rev Dis Primers 2018; 4:1–19.

2. Miller JJ, Weimer BC, Timme R, Lüdeke CHM, Pettengill JB, Bandoy DJD, Weis AM, Kaufman J, Huang BC, Payne J, et al. Phylogenetic and biogeographic patterns of *Vibrio parahaemolyticus* strains from North America inferred from whole-genome sequence data. Appl Environ Microbiol 2021; 87.

3. DePaola A, Hopkins LH, Peeler JT, Wentz B, McPhearson RM. Incidence of *Vibrio parahaemolyticus* in U.S. coastal waters and oysters. Appl Environ Microbiol 1990; 56:2299–302.

4. Baker-Austin C, Mcarthur J V., Tuckfield RC, Najarro M, Lindell AH, Gooch J, Stepanauskas R. Antibiotic resistance in the shellfish pathogen *Vibrio parahaemolyticus* isolated from the coastal water and sediment of Georgia and South Carolina, USA. J Food Prot 2016; 71:2552–8.

5. Silvester R, Alexander D, Sudha S, Harikrishnan M, Hatha M. Virulence features of *Vibrio parahaemolyticus*: a review. Impact of Climate Change on Hydrological Cycle, Ecosystem, Fisheries and Food Security 2022; :211–8.

6. Raghunath P. Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in *Vibrio parahaemolyticus*. Front Microbiol2015; 5.

7. Shirai H, Ito H, Hirayama T, Nakamoto Y, Nakabayashi N, Kumagai K, Takeda Y, Nishibuchii M. Molecular epidemiologic evidence for association of thermostable direct hemolysin (TDH) and TDH-related hemolysin of *Vibrio parahaemolyticus* with gastroenteritis. Infection and Immunity 1990; 58:3568–73.

8. Saito S, Iwade Y, Tokuoka E, Nishio T, Otomo Y, Araki E, Konuma H, Nakagawa H, Tanaka H, Sugiyama K, et al. Epidemiological evidence of lesser role of thermostable direct hemolysin (TDH)–related hemolysin (TRH) than TDH on *Vibrio parahaemolyticus* pathogenicity. Foodborne Pathog Dis 2015; 12:131–8.

9. Ceccarelli D, Hasan NA, Huq A, Colwell RR. Distribution and dynamics of epidemic and pandemic *Vibrio parahaemolyticus* virulence factors. Front Cell Infect Microbiol 2013; 3:97.

10. Gutierrez West CK, Klein SL, Lovell harles R. High frequency of virulence factor genes *tdh*, *trh*, and *tlh* in *Vibrio parahaemolyticus* strains isolated from a pristine estuary. Appl Environ Microbiol 2013; 79:2247–52.

11. Klein SL, Lovell CR. The hot oyster: levels of virulent *Vibrio parahaemolyticus* strains in individual oysters. FEMS Microbiol Ecol 2017; 93.

12. Park KS, Ono T, Rokuda M, Jang MH, Okada K, Iida T, Honda T. Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. Infect Immun 2004; 72:6659–65.

13. Ham H, Orth K. The role of type III secretion System 2 in *Vibrio parahaemolyticus* pathogenicity. Journal of Microbiology2012; 50:719–25.

14. Okada N, Iida T, Park KS, Goto N, Yasunaga T, Hiyoshi H, Matsuda S, Kodama T, Honda T. Identification and characterization of a novel type III secretion system in *trh*-positive *Vibrio parahaemolyticus* strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect Immun 2009; 77:904–13.

15. Salomon D, Gonzalez H, Updegraff BL, Orth K. *Vibrio parahaemolyticus* type VI secretion system 1 is activated in marine conditions to target bacteria, and is differentially regulated from system 2. PLoS One 2013; 8:e61086.

16. Jana B, Keppel K, Fridman CM, Bosis E, Salomon D. Multiple T6SSs, mobile auxiliary modules, and effectors revealed in a systematic analysis of the *Vibrio parahaemolyticus* pan-genome. mSystems 2022; 7:e00723-22.

17. Wang J, Zhan Y, Sun H, Fu X, Kong Q, Zhu C, Mou H. Regulation of virulence factors expression during the intestinal colonization of *Vibrio parahaemolyticus*. Foodborne Pathog Dis 2022; 19:169–78.

18. Zeb S, Shah MA, Yasir M, Awan HM, Prommeenate P, Klanchui A, Wren BW, Thomson N, Bokhari H. Type III secretion system confers enhanced virulence in clinical non-O1/non-O139 *Vibrio cholerae*. Microb Pathog 2019; 135:103645.

19. Crisan C V., Hammer BK. The *Vibrio cholerae* type VI secretion system: toxins, regulators and consequences. Environ Microbiol 2020; 22:4112–22.

20. Klein SL, Gutierrez West CK, Mejia DM, Lovell CR. Genes similar to the *Vibrio parahaemolyticus* virulence-related genes *tdh*, *tlh*, and *vscC2* cccur in other *vibrionaceae* species isolated from a pristine estuary. Appl Environ Microbiol 2014; 80:595–602.

21. Cai H, Yu J, Li Q, Zhang Y, Huang L, Cai H, Yu J, Li Q, Zhang Y, Huang L. Research progress on virulence factors of *Vibrio alginolyticus*: a key pathogenic bacteria of sepsis. In: Heat Illness and Critical Care [Working Title]. IntechOpen; 2022.

22. Song J, Liu X, Wu C, Zhang Y, Fan K, Zhang X, Wei Y. Isolation, identification and pathogenesis study of *Vibrio diabolicus*. Aquaculture 2021; 533:736043.

23. Pazhani GP, Chowdhury G, Ramamurthy T. Adaptations of *Vibrio parahaemolyticus* to stress during environmental survival, host colonization, and infection. Front Microbiol2021; 12.

24. Buck JD, Spotte S. The occurrence of potentially pathogenic *Vibrios* in marine mammals. Mar Mamm Sci 1986; 2:319–24.

25. Thornton SM, Nolan S, Gulland FMD. Bacterial isolates from California sea lions (*Zalophus californianus*), harbor seals (*Phoca vitulina*), and northern elephant seals (*Mirounga angustirostris*) admitted to a rehabilitation center along the central California coast, 1994-1995. Journal of Zoo and Wildlife Medicine 1998; 29:171–6.

26. Li M, Zhao L, Ma J, Zhao N, Luo J, Wang C, Chen L, Ma G, Wang Y, He H. *Vibrio vulnificus* in aquariums is a novel threat to marine mammals and public health. Transbound Emerg Dis 2018;

27. Goertz CEC, Walton R, Rouse N, Belovarac J, Burek-Huntington K, Gill V, Hobbs R, Xavier C, Garrett N, Tuomi P. *Vibrio parahaemolyticus*, a climate change indicator in Alaska marine mammals. In: Responses of Arctic marine ecosystems to climate change. Alaska Sea Grant, University of Alaska Fairbanks. 2013. page 41–52.

28. Miller MA, Moriarty ME, Henkel L, Tinker MT, Burgess TL, Batac FI, Dodd E, Young C, Harris MD, Jessup DA, et al. Predators, disease, and environmental change in the nearshore ecosystem: mortality in southern sea otters (*Enhydra lutris nereis*) from 1998–2012. Front Mar Sci 2020; 7:582.

29. Stewart JR, Townsend FI, Lane SM, Dyar E, Hohn AA, Rowles TK, Staggs LA, Wells RS, Balmer BC, Schwacke LH. Survey of antibiotic-resistant bacteria isolated from bottlenose dolphins Tursiops truncatus in the southeastern USA. Dis Aquat Organ 2014; 108:91–102.

30. Hughes SN, Greig DJ, Miller WA, Byrne BA, Gulland FMD, Harvey JT. Dynamics of *Vibrio* with virulence genes detected in Pacific harbor seals (*Phoca vitulina richardii*) off California: implications for marine mammal health. Microb Ecol 2013; 65:982–94.

31. Oates SC, Miller MA, Byrne BA, Chouicha N, Hardin D, Jessup D, Dominik C, Roug A, Schriewer A, Jang SS, et al. Epidemiology and potential land-sea transfer of enteric bacteria from terrestrial to marine species in the Monterey Bay region of California. J Wildl Dis 2012; 48:654–68.

32. Jessup DA, Miller M, Ames J, Harris M, Kreuder C, Conrad PA, Mazet JAK. Southern sea otter as a sentinel of marine ecosystem health. Ecohealth 2004; 1:239–45.

33. Brownstein D, Miller MA, Oates SC, Byrne BA, Jang S, Murray MJ, Gill VA, Jessup DA. Antimicrobial susceptibility of bacterial isolates from sea otters (*Enhydra lutris*). J Wildl Dis 2011; 47:278–92.

34. Burek Huntington KA, Gill VA, Berrian AM, Goldstein T, Tuomi P, Byrne BA, Worman K, Mazet J. Causes of mortality of northern sea otters (*Enhydra lutris kenyoni*) in Alaska from 2002 to 2012. Front Mar Sci 2021; 8:630582.

35. Johnson CK, Tinker MT, Estes JA, Conrad PA, Staedler M, Miller MA, Jessup DA, Mazet JAK. Prey choice and habitat use drive sea otter pathogen exposure in a resource-limited coastal system. Proceedings of the National Academy of Sciences 2009; 106:2242–7.

36. Miller MA, Byrne BA, Jang SS, Dodd EM, Dorfmeier E, Harris MD, Ames J, Paradies D, Worcester K, Jessup DA, et al. Enteric bacterial pathogen detection in southern sea otters (*Enhydra lutris nereis*) is associated with coastal urbanization and freshwater runoff. Vet Res 2010; 41.

37. Loo KY, Letchumanan V, Law JWF, Pusparajah P, Goh BH, Ab Mutalib NS, He YW, Lee LH. Incidence of antibiotic resistance in *Vibrio* spp. Rev Aquac 2020; 12:2590–608.

38. Ashrafudoulla M, Furkanur Rahaman Mizan M, Park H, Byun KH, Lee N, Park SH, Ha S Do. Genetic relationship, virulence factors, drug resistance profile and biofilm formation ability of *Vibrio parahaemolyticus* isolated from mussel. Front Microbiol 2019; 10:513.

39. Park K, Mok JS, Kwon JY, Ryu AR, Kim SH, Lee HJ. Food-borne outbreaks, distributions, virulence, and antibiotic resistance profiles of *Vibrio parahaemolyticus* in Korea from 2003 to 2016: a review. Fish Aquatic Sci2018; 21:1–10.

40. Kreuder C, Miller MA, Jessup DA, Lowenstine LJ, Harris MD, Ames JA, Carpenter TE, Conrad PA, Mazet JAK. Patterns of mortality in southern sea otters (*Enhydra lutris nereis*) from 1998–2001. J Wildl Dis 2013; 39:495–509.

41. Siembieda JL, Miller WA, Byrne BA, Ziccardi MH, Anderson N, Chouicha N, Sandrock CE, Johnson CK. Zoonotic pathogens isolated from wild animals and environmental samples at two California wildlife hospitals. J Am Vet Med Assoc 2011; 238:773–83.

42. Miller WA, Miller MA, Gardner IA, Atwill ER, Byrne BA, Jang S, Harris M, Ames J, Jessup D, Paradies D, et al. *Salmonella* spp., *Vibrio* spp., *Clostridium perfringens*, and *Plesiomonas shigelloides* in marine and freshwater invertebrates from coastal California ecosystems. Microb Ecol 2006; 52:198–206.

43. Miller B, van Kets V, van Rooyen B, Whitehorn H, Jones P, Ranik M, Geldart A, van der Walt E, Appel M, Kong N, et al. A novel, single-tube enzymatic fragmentation and library construction method enables fast turnaround times and improved data quality for microbial whole-genome sequencing. 2015.

44. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

45. Seemann T. Shovill—Assemble bacterial isolate genomes from Illumina paired-end reads. 2020.

46. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci 2017; 3:e104.

47. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol 2019; 20:1–13.

48. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015; 25:1043–55.

49. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014; 30:2068–9.

50. Brown CT, Irber L. sourmash: a library for MinHash sketching of DNA. The Journal of Open Source Software 2016; 1:27.

51. Hernández-Juárez LE, Camorlinga M, Méndez-Tenorio A, Flores Calderón J, Huang BC, Bandoy DR, Weimer BC, Torres J. Analyses of publicly available Hungatella hathewayi genomes revealed genetic distances indicating they belong to more than one species. 2021.

52. Flores-Valdez M, Ares MA, Rosales-Reyes R, Torres J, Girón JA, Weimer BC, Mendez-Tenorio A, De la Cruz MA. Whole genome sequencing of pediatric *Klebsiella pneumoniae* strains reveals important insights into their virulence-associated traits. Front Microbiol 2021; 12.

53. González-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus LA, DePaola A. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J Bacteriol 2008; 190:2831–40.

54. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 2018; 3.

55. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res 2022; 50:D912–7.

56. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res 2016; 44:D694–7.

57. Gupta SK, Padmanabhan BR, Diene SM, Lopez-rojas R, Kempf M, Landraud L, Rolain J-M. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014; 58:212–20.

58. Seemann T. Snippy: fast bacterial variant calling from NGS reads. 2015.

59. Nilsson WB, Turner JW. The thermostable direct hemolysin-related hemolysin (*trh*) gene of *Vibrio parahaemolyticus*: sequence variation and implications for detection and function. J Microbiol Methods 2016; 126:1–7.

60. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009; 25:1189–91.

61. Pérez-Duque A, Gonzalez-Muñoz A, Arboleda-Valencia J, Vivas-Aguas LJ, Córdoba-Meza T, Rodriguez-Rey GT, Díaz-Guevara P, Martinez-Urtaza J, Wiesner-Reyes M. Comparative genomics of clinical and environmental isolates of *Vibrio* spp. of Colombia: implications of traits associated with virulence and resistance. Pathogens 2021; 10:1605.

62. Korotkov K V., Sandkvist M, Hol WGJ. The type II secretion system: biogenesis, molecular architecture and mechanism. Nature Reviews Microbiology 2012 10:5 2012; 10:336–51.

63. Ethica SN, Hidayati N, Fuad H, Arham C, Ariyadi R, Purwaningrum E, Rahman KMZ. Detection of *rtxA* gene as a biomarker of seafood-borne pathogen *Vibrio cholerae* using in silico PCR assay. Squalen Bulletin of Marine and Fisheries Postharvest and Biotechnology 2020; 15:91–8.

64. Ramamurthy T, Nandy RK, Mukhopadhyay AK, Dutta S, Mutreja A, Okamoto K, Miyoshi SI, Nair GB, Ghosh A. Virulence regulation and innate host response in the pathogenicity of *Vibrio cholerae*. Front Cell Infect Microbiol 2020; 10:520.

65. Dolores J, Satchell KJF. Analysis of *Vibrio cholerae* genome sequences reveals unique *rtxA* variants in environmental strains and an *rtxA*-null mutation in recent altered EL tor isolates. mBio 2013; 4.

66. Valiente E, Davies C, Mills DC, Getino M, Ritchie JM, Wren BW. *Vibrio cholerae* accessory colonisation factor AcfC: a chemotactic protein with a role in hyperinfectivity. Scientific Reports 2018 8:1 2018; 8:1–11.

67. Laviad-Shitrit S, Izhaki I, Arakawa E, Halpern M. Wild waterfowl as potential vectors of *Vibrio cholerae* and *Aeromonas* species. Tropical Medicine and International Health 2018; 23:758–64.

68. Laviad-Shitrit S, Izhaki I, Halpern M. Accumulating evidence suggests that some waterbird species are potential vectors of *Vibrio cholerae*. PLoS Pathog 2019; 15:e1007814.

69. Halpern M, Senderovich Y, Izhaki I. Waterfowl—the missing link in epidemic and pandemic cholera dissemination? PLoS Pathog 2008; 4:e1000173.

70. Dong TG, Ho BT, Yoder-Himes DR, Mekalanos JJ. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. Proc Natl Acad Sci U S A 2013; 110:2623–8.

71. Chaand M, Miller KA, Sofia MK, Schlesener C, Weaver JWA, Sood V, Dziejman M. Type Three Secretion System Island-Encoded Proteins Required for Colonization by Non-O1/Non-O139 Serogroup Vibrio cholerae. Infect Immun 2015; 83:2862.

72. Miller KA, Tomberlin KF, Dziejman M. *Vibrio* variations on a type three theme. Curr Opin Microbiol 2019; 47:66.

73. Hasan NA, Ceccarelli D, Grim CJ, Taviani E, Choi J, Sadique A, Alam MC, Siddique AK, Bradley Sack R, Huq A, et al. Distribution of virulence genes in clinical and environmental *Vibrio cholerae* strains in Bangladesh. Appl Environ Microbiol 2013; 79:5782–5.

74. Ceccarelli D, Chen A, Hasan NA, Rashed SM, Huq A, Colwell RR. Non-O1/non-O139 *Vibrio cholerae* carrying multiple virulence factors and *V. cholerae* O1 in the Chesapeake Bay, Maryland. Appl Environ Microbiol 2015; 81:1909–18.

75. Hernández-Robles MF, Álvarez-Contreras AK, Juárez-García P, Natividad-Bonifacio I, Curiel-Quesada E, Vázquez-Salinas C, Quiñones-Ramírez EI. Virulence factors and antimicrobial resistance in environmental strains of *Vibrio alginolyticus*. International Microbiology 2016; 19:191–8.

76. Gargouti A, Ab-Rashid MNK, Ghazali M, Mitsuaki N, Haresh K, Radu S. Detection of *tdh* and *trh* toxic genes in *Vibrio Alginolyticus* strain from mantis shrimp (*Oratosquilla Oratoria*). J Nutr Food Sci 2015; 5:405.

77. Ellingsen AB, Olsen JS, Granum PE, Rørvik LM, González-Escalona N. Genetic characterization of *trh* positive *Vibrio* spp. isolated from Norway. Front Cell Infect Microbiol 2013; 3.

78. Xie ZY, Hu CQ, Chen C, Zhang LP, Ren CH. Investigation of seven *Vibrio* virulence genes among *Vibrio alginolyticus* and *Vibrio parahaemolyticus* strains from the coastal mariculture systems in Guangdong, China. Lett Appl Microbiol 2005; 41:202–7.

79. Hossain S, Wickramanayake MVKS, Dahanayake PS, Heo GJ. Occurrence of Virulence and Extended-Spectrum β-Lactamase Determinants in Vibrio spp. Isolated from Marketed Hard-Shelled Mussel (Mytilus coruscus). Microbial Drug Resistance 2020; 26:391–401.

80. Nava-Soberanes EB, Mendieta-Condado ER, Garcés-Ayala F, Araiza-Rodríguez A, Aguilera-Perez P, Alfaro-López A, Ibarra-Buelna AC, Montes-Colima NA, López-Martínez I, Hernández-Rivas L, et al. Draft genome sequence of first *Vibrio diabolicus* in Mexico strain InDRE-D1-M1, an emergent threat. J Infect Dev Ctries 2021; 15:1750–4.

81. Eggert S. Identification of type three secretion system effectors linked to clinical and environmental prevalence of *Vibrio parahaemolyticus*. Thesis 2020.

82. Matsuda S, Hiyoshi H, Tandhavanant S, Kodama T. Advances on *Vibrio parahaemolyticus* research in the postgenomic era. Microbiol Immunol 2020; 64:167–81.

83. Noriea NF, Johnson CN, Griffitt KJ, Grimes DJ. Distribution of type III secretion systems in *Vibrio parahaemolyticus* from the northern Gulf of Mexico. J Appl Microbiol 2010; 109:953–62.

84. Xu F, Gonzalez-Escalon N, Drees KP, Sebra RP, Cooper VS, Jones SH, Whistler CA. Parallel evolution of two clades of an Atlantic-endemic pathogenic lineage of *Vibrio parahaemolyticus* by independent acquisition of related pathogenicity islands. Appl Environ Microbiol 2017; 83.

85. Prabhakaran DM, Patel HR, Sivakumar Krishnankutty Chandrika S, Thomas S. Genomic attributes differ between *Vibrio parahaemolyticus* environmental and clinical isolates including pathotypes. Environ Microbiol Rep 2021.

86. Kamruzzaman M, Bhoopong P, Vuddhakul V, Nishibuchi M. Detection of a functional insertion sequence responsible for deletion of the thermostable direct hemolysin gene (*tdh*) in *Vibrio parahaemolyticus*. Gene 2008; 421:67–73.

87. Yang C, Li Y, Jiang M, Wang L, Jiang Y, Hu L, Shi X, He L, Cai R, Wu S, et al. Outbreak dynamics of foodborne pathogen *Vibrio parahaemolyticus* over a seventeen year period implies hidden reservoirs. Nature Microbiology 2022 7:8 2022; 7:1221–9.

88. Gonzalez-Escalona N, Jolley KA, Reed E, Martinez-Urtaza J. Defining a core genome multilocus sequence typing scheme for the global epidemiology of *Vibrio parahaemolyticus*. J Clin Microbiol 2017; 55:1682.

89. Jesser KJ, Valdivia-Granda W, Jones JL, Noble RT. Clustering of *Vibrio parahaemolyticus* isolates using MLST and whole-genome phylogenetics and protein motif fingerprinting. Front Public Health 2019; 7:66.

90. Plaza N, Urrutia IM, Garcia K, Waldor MK, Blondel CJ. Identification of a family of *Vibrio* type III secretion system effectors that contain a conserved serine/threonine kinase domain. mSphere 2021; 6.

91. Bechlars S, Jäckel C, Diescher S, Wüstenhagen DA, Kubick S, Dieckmann R, Strauch E. Characterization of *trh2* harbouring *Vibrio parahaemolyticus* strains isolated in Germany. PLoS One 2015; 10.

92. Kishishita M, Matsuoka N, Kumagai K, Yamasaki S, Takeda Y, Nishibuchi2 M, Shirai H, Ito H, Hirayama T, Nakamoto Y, et al. Sequence variation in the thermostable direct hemolysin-related hemolysin (*trh*) gene of *Vibrio parahaemolyticus*. Appl Environ Microbiol 1992; 58:2449–57.

93. Leoni F, Talevi G, Masini L, Ottaviani D, Rocchegiani E. *Trh* (*tdh* −/*trh* +) gene analysis of clinical, environmental and food isolates of *Vibrio parahaemolyticus* as a tool for investigating pathogenicity. Int J Food Microbiol 2016; 225:43–53.

94. Nakaguchi Y, Nishibuchi M. The promoter region rather than its downstream inverted repeat sequence is responsible for low-level transcription of the thermostable direct hemolysin-related hemolysin (*trh*) gene of *Vibrio parahaemolyticus*. J Bacteriol 2005; 187:1849–55.

95. Caburlotto G, Gennari M, Ghidini V, Tafi M, Lleo MM. Presence of T3SS2 and other virulence-related genes in *tdh*-negative *Vibrio parahaemolyticus* environmental strains isolated from marine samples in the area of the Venetian Lagoon, Italy. FEMS Microbiol Ecol 2009; 70:506–14.

96. Caburlotto G, Lleò MM, Hilton T, Huq A, Colwell RR, Kaper JB. Effect on human cells of environmental *Vibrio parahaemolyticus* strains carrying type III secretion system 2. Infect Immun 2010; 78:3280–7.

97. Zha F, Pang R, Huang S, Zhang J, Wang J, Chen M, Xue L, Ye Q, Wu S, Yang M, et al. Evaluation of the pathogenesis of non-typical strain with α-hemolysin, *Vibrio parahaemolyticus* 353, isolated from Chinese seafood through comparative genome and transcriptome analysis. Mar Pollut Bull 2023; 186:114276.

98. Pazhani GP, Ravichandiran V, Veerapandian M, Ramamurthy T. Bioinformatics on unproven secretion systems of *Vibrio Parahaemolyticus*. Scientific Journal of Biomedical Engineering and Biomedical Science 2017; 1:020–5.

99. Yu Y, Yang H, Li J, Zhang • Peipei, Wu B, Zhu B, Zhang Y. Putative type VI secretion systems of *Vibrio parahaemolyticus* contribute to adhesion to cultured cell monolayers. Arch Microbiol 2012; :827–35.

100. Yu Y, Fang L, Zhang Y, Sheng H, Fang W. VgrG2 of type VI secretion system 2 of *Vibrio parahaemolyticus* induces autophagy in macrophages. Front Microbiol 2015; 6:168.

101. Baker-Austin C, Trinanes J, Gonzalez-Escalona N, Martinez-Urtaza J. Non-cholera *Vibrios*: the microbial barometer of climate change. Trends Microbiol2017; 25:76–84.

102. Cheung WWL, Frölicher TL. Marine heatwaves exacerbate climate change impacts for fisheries in the northeast Pacific. Sci Rep 2020; 10.

103. McLaughlin JB, DePaola A, Bopp CA, Martinek KA, Napolilli NP, Allison CG, Murray SL, Thompson EC, Bird MM, Middaugh JP. Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. N Engl J Med 2005; 353:1463–70.

104. Taylor M, Cheng J, Sharma D, Bitzikos O, Gustafson R, Fyfe M, Greve R, Murti M, Stone J, Honish L, et al. Outbreak of *Vibrio parahaemolyticus* associated with consumption of raw oysters in Canada, 2015. Foodborne Pathog Dis 2018; 15.

105. LaRoche NL, King SL, Rogers MC, Eckert GL, Pearson HC. Behavioral observations and stable isotopes reveal high individual variation and little seasonal variation in sea otter diets in southeast Alaska. Mar Ecol Prog Ser 2021; 677:219–32.

106. Hennequin C, Robin F. Correlation between antimicrobial resistance and virulence in *Klebsiella pneumoniae*. European Journal of Clinical Microbiology and Infectious Diseases 2016; 35:333–41.

107. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evol Appl 2015; 8:273–83.

108. Kumarage PM, De Silva LADS, Heo GJ. Aquatic environments: a potential source of antimicrobial-resistant *Vibrio* spp. J Appl Microbiol 2022.

109. Da Silva L V, Ossai S, Chigbu P, Parveen S. Antimicrobial and genetic profiles of *Vibrio vulnificus* and *Vibrio parahaemolyticus* isolated from the Maryland coastal bays, United States. Front Microbiol 2021; 12.

Figures

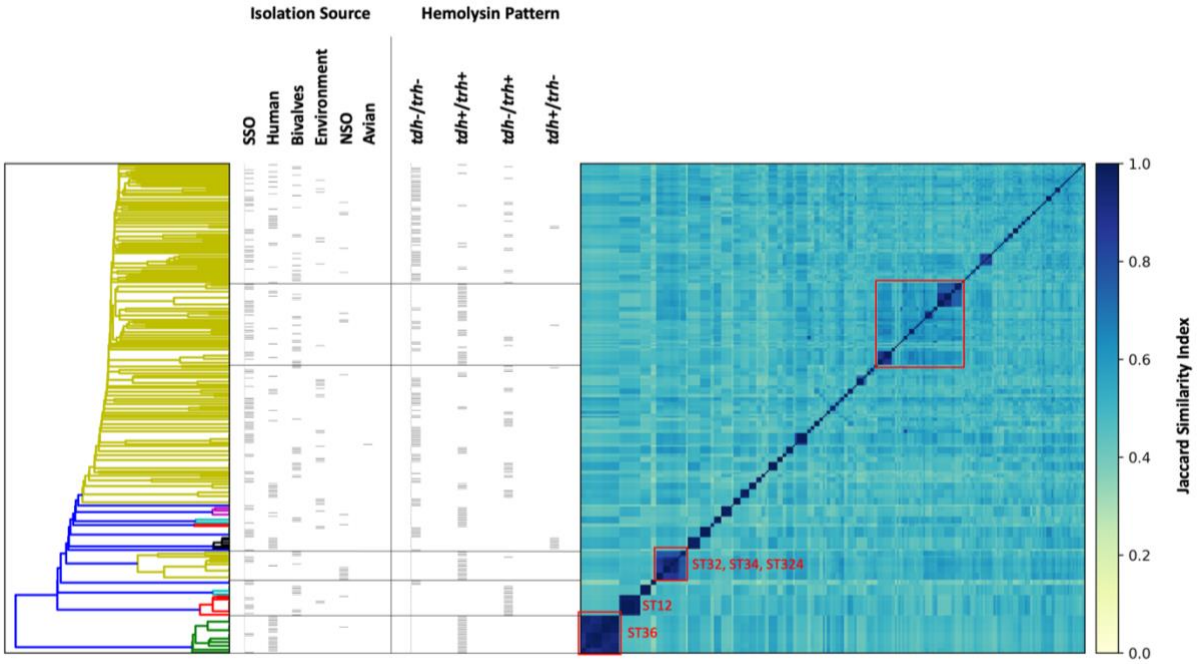


Figure 2.1. All-against-all comparative analysis of the similarity between 287 *V. parahaemolyticus* genomes collected from 2000 through 2019 from various sources across North America. Genomes are annotated by source from most frequent to least frequent and by hemolysin gene patterns. Darker blue colors indicate greater similarity based on the Jaccard Similarity Index (JSI). Sequence types mentioned in the article are labeled. Three regions comprised primarily of *tdh*$^+$/*trh*$^+$ genomes are highlighted (red boxes). From bottom left to top right, the first region is comprised by a genomic cluster of primarily ST36 isolated only from sea otters and humans. The second region is comprised by sequence types included in clonal complex 34 (ST32, ST34, ST324) and includes isolates from southern and northern sea otters (SSO and NSO, respectively), bivalves, and a human. The third region contains varied typed and untyped strains.

Tables

Table 2.1. Univariate logistic regression models comparing the likelihood of *V. parahaemolyticus* genomes carrying select virulence factors based on isolation source. Isolates were collected from 2000 through 2019 across various North American sites. Odds ratios were calculated using environmental isolates as the reference group; isolation sources that differed significantly ($p<0.05$) from environmental isolates are marked with an asterisk (*). The last model compares the likelihood of carrying both *vopB2* and *vipB*/*mglB* between isolation sources. NSO= northern sea otter. SSO= southern sea otter.

Gene	Source	Pos.	Total	%	OR	95% CI	p-value
hlyA	Environment	3	29	10.3	REF	REF	REF
	Human	12	68	17.6	1.86	0.48, 7.15	0.368
	Bivalves	23	62	37.1	5.11	1.39, 18.78	0.014*
	NSO	16	21	76.2	27.73	5.82, 132.10	<0.001*
	SSO	27	106	25.5	2.96	0.83, 10.57	0.094
tdh	Environment	2	29	6.9	REF	REF	REF
	Human	42	68	61.8	21.81	4.78, 99.44	<0.001*
	Bivalves	23	62	39.7	7.96	1.73, 36.62	0.008*
	NSO	14	21	66.7	27	4.94, 147.63	<0.001*
	SSO	33	106	31.1	6.1	1.37, 27.19	0.018*
trh	Environment	6	29	20.7	REF	REF	REF
	Human	43	68	63.2	6.59	2.37, 18.37	<0.001*
	Bivalves	44	62	71.0	9.37	3.27, 26.85	<0.001*
	NSO	20	21	95.2	76.67	8.49, 691.95	<0.001*
	SSO	53	106	50.0	3.83	1.44, 10.17	0.007*

Table 2.1 continued.

Gene	Source	Pos.	Total	%	OR	95% CI	p-value
vcrD2	Environment	6	29	20.7	REF	REF	REF
(T3SS2)	Human	50	68	73.5	10.65	3.74, 30.36	<0.001*
	Bivalves	46	62	74.2	11.02	3.81, 31.92	<0.001*
	NSO	19	21	90.5	36.42	6.58, 201.67	<0.001*
	SSO	47	106	44.3	3.05	1.15, 8.11	0.025*
vopB2	Environment	6	29	20.7	REF	REF	REF
(T3SS2)	Human	52	68	76.5	12.46	4.32, 35.92	<0.001*
	Bivalves	46	62	74.2	11.02	3.81, 31.92	<0.001*
	NSO	20	21	95.2	76.67	8.49, 691.95	<0.001*
	SSO	54	106	50.9	3.98	1.50, 10.56	0.006*
vipB/mglB	Environment	10	29	34.5	REF	REF	REF
(T6SS)	Human	43	68	63.2	3.97	1.58, 9.96	0.003*
	Bivalves	20	62	32.3	0.9	0.36, 2.30	0.833
	NSO	13	21	61.9	3.09	0.96, 9.92	0.058
	SSO	43	106	40.6	1.3	0.55, 3.06	0.553
vopB2* + *vipB	Environment	1	29	3.4	REF	REF	REF
	Human	38	68	55.9	35.47	4.57, 275.51	<0.001*
	Bivalves	14	62	22.6	8.17	1.02, 65.39	0.048*
	NSO	13	21	61.9	45.5	5.15, 402.20	<0.001*
	SSO	16	106	15.1	4.98	0.63, 39.18	0.127

Table 2.2. Univariate logistic regression models comparing the associations between three *trh* gene variants and isolation source. Isolates were collected from 2000 through 2019 across various North America sites. Odds ratios were calculated using environmental isolates as the reference group and isolation sources that differed significantly ($p<0.05$) from environmental isolates are marked with an asterisk (*). The *trh*⟱ is presumed to be a pseudogene due to the methionine to isoleucine substitution in the translated sequence resulting in loss of the traditional start codon (NA = length not applicable). The *trh1* allele was most associated with humans while *trh*⟱ was most associated with bivalves and northern sea otters. NSO= northern sea otter. SSO= southern sea otter.

Gene	Length	Source	Pos.	Tot.	%	OR	95% CI	p-value
trh1	Full	Environment	1	29	3.4	REF	REF	REF
		Human	24	68	35.3	15.27	1.95, 119.32	0.009*
		Bivalves	0	62	0	NA	NA	NA
		NSO	2	21	9.5	2.95	0.25, 34.85	0.391
		SSO	9	106	8.5	2.60	0.32, 21.39	0.375
trh⟱	NA	Environment	1	29	3.4	REF	REF	REF
		Human	5	68	7.4	2.22	0.25, 19.89	0.475
		Bivalves	21	62	33.9	14.34	1.83, 112.70	0.011*
		NSO	11	21	52.4	30.80	3.52, 269.63	0.002*
		SSO	18	106	17	5.73	0.73, 44.79	0.096
trh2	Full	Environment	4	29	13.8	REF	REF	REF
		Human	10	68	14.7	1.08	0.31, 3.76	0.907
		Bivalves	19	62	30.6	2.76	0.84, 9.04	0.093
		NSO	6	21	28.6	2.50	0.61, 10.32	0.205
		SSO	17	106	16	1.19	0.37, 3.87	0.768

Table 2.3. Associations between key pathological findings and *Vibrio*-spp. detection from various tissues in southern sea otters from California that were necropsied from 2001 through 2012. Reports of possible or confirmed enteritis were pooled as enteritis for analysis. Enteritis, melena, and sepsis were frequently observed in *Vibrio* spp.-positive southern sea otters.

Species	Enteritis (%)	Melena (%) Any Severity	Melena (%) Mod.-Severe	Sepsis (%)	Total
V. alginolyticus	12 (50)	20 (83.3)	12 (50.0)	15 (62.5)	24
V. cholerae (non-O1)	10 (83.3)	7 (58.3)	5 (41.7)	5 (41.7)	12
V. diabolicus	4 (36.4)	7 (63.6)	3 (27.3)	3 (27.3)	11
V. parahaemolyticus	30 (56.6)	35 (66)	22 (41.5)	36 (67.8)	53

Table 2.4. Associations between key pathological findings and *V. parahaemolyticus* detection from various tissues in southern sea otters from California that were necropsied from 2001 through 2012. Reports of possible or confirmed enteritis were pooled as enteritis for analysis. The top portion of the table (above the thick line) includes single factor comparisons; the bottom portion of the table compares pathological findings with the putative interaction between T3SS2 (*vopB2*) and T6SS1 (*vipB*).

Virulence Gene	Enteritis (%)	Melena (%)	Mod. to Severe Melena (%)	Sepsis (%)	Total
hlyA^{+}	10 (62.5)	12 (75)	10 (62.5)	10 (62.5)	16
hlyA^{-}	20 (54.1)	23 (62.2)	12 (32.4)	26 (70.3)	37
tdh^{+}	10 (62.5)	14 (87.5)	10 (62.5)	11 (68.8)	16
tdh^{-}	20 (54.1)	21 (56.8)	12 (32.4)	25 (67.6)	37
trh^{+}	16 (66.7)	20 (76.9)	15 (57.7)	19 (73.1)	26
trh^{-}	14 (51.9)	15 (55.6)	7 (25.9)	17 (63)	27
vipB^{+}	12 (57.1)	17 (80.1)	9 (42.9)	16 (76.2)	21
vipB^{-}	18 (56.3)	18 (56.3)	13 (40.6)	20 (62.5)	32
vopB2^{+}	16 (61.5)	20 (76.9)	15 (57.7)	19 (73.1)	26
vopB2^{-}	14 (51.9)	15 (55.6)	7 (25.9)	17 (63)	27
vipB^{+}/*vopB2*$^{+}$	5 (62.5)	7 (87.5)	5 (62.5)	7 (87.5)	8
vipB^{-}/*vopB2*$^{+}$	11 (61.1)	13 (72.2)	10 (55.6)	12 (66.7)	18
vipB^{+}/*vopB2*$^{-}$	7 (53.8)	10 (76.9)	4 (30.8)	9 (65)	13
vipB^{-}/*vopB2*$^{-}$	7 (50)	5 (35.7)	3 (21.4)	8 (57.1)	14

Table 2.5. Univariate (above thick line) and bivariate (below thick line) logistic regression models comparing the likelihood of bla_{carb} alleles associated with ampicillin-susceptibility in *V. parahaemolyticus* isolates collected from 2000 through 2019 across various tissue sources and North American sites based on presence of various virulence factors. Ampicillin-resistant bla_{carb} alleles of $bla_{carb\text{-}18}$, $bla_{carb\text{-}20}$, and $bla_{carb\text{-}21}$ were associated with phenotypic ampicillin resistance and alleles with non-synonymous substitutions were associated with ampicillin susceptibility (Chapter 1). Odds ratios were calculated using virulence factor negative as the reference groups. Virulence factors that were significantly associated with an ampicillin-susceptible bla_{carb} genotype are marked with an asterisk (*). The bivariate logistic regression model included an interaction between a T3SS2 marker (*vopB2*) and a T6SS1 marker (*vipB/mglB*). Presence of *vopB2* was associated with ampicillin-susceptible bla_{carb} alleles, and this effect was modified by the presence of *vipB/mglB* such that genomes with both T3SS2 and T6SS1 markers were most likely to harbor an ampicillin-susceptible bla_{carb} genotype.

Virulence Factor	Ampicillin-Susceptible bla_{carb} Alleles (n= 287)		
	OR	95% CI	p-value
hlyA$^+$	1.91	1.13, 3.23	0.015*
tdh$^+$	3.04	1.85, 5.01	<0.001*
trh$^+$	3.36	1.98, 5.72	<0.001*
vopB$^+$ (T3SS1)	1.45	0.61, 3.43	0.403
vcrD2$^+$ (T3SS2)	4.73	2.70, 8.28	<0.001*
vopB2$^+$ (T3SS2)	4.48	2.52, 7.97	<0.001*
vipB/mglB$^+$ (T6SS1)	3.04	1.85, 5.00	<0.001*
vipB$^-$/ *vopB2*$^-$	REF	REF	REF
vipB$^+$/ *vopB2*$^-$	1.36	0.50, 3.67	0.543
vipB$^-$/ *vopB2*$^+$	2.47	1.08, 5.68	0.033*
vipB$^+$/ *vopB2*$^+$	12.41	5.29, 29.11	<0.001*

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Peter J. Sebastian- Contributed to the conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, data visualization, and the writing and editing of the manuscript.

Cory Schlesener- Contributed to the conceptualization, data curation and visualization, formal analysis, methodology, and manuscript review.

Barbara A. Byrne- Contributed to the conceptualization, supervision of data generation and curation, investigation, resources, and manuscript review.

Melissa Miller- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Woutrina Smith- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Francesca Batac- Contributed to the data curation and manuscript review.

Kathy Burek-Huntington- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Caroline E.C. Goertz- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Natalie Rouse- Contributed to the data curation and manuscript review.

Natalie Hunter- Contributed to the data curation and manuscript review.

Bart C. Weimer- Contributed to the conceptualization, methodology, supervision of data generation and curation, project administration, resources, and manuscript review.

Christine K. Johnson- Contributed to the conceptualization, funding acquisition, supervision of data generation and curation, project administration, resources, and manuscript review.

Funding

P. Sebastian was supported by the National Institutes of Health under grant number T32OD011147. Additional funding was provided by an Oiled Wildlife Care Network grant.

Acknowledgments

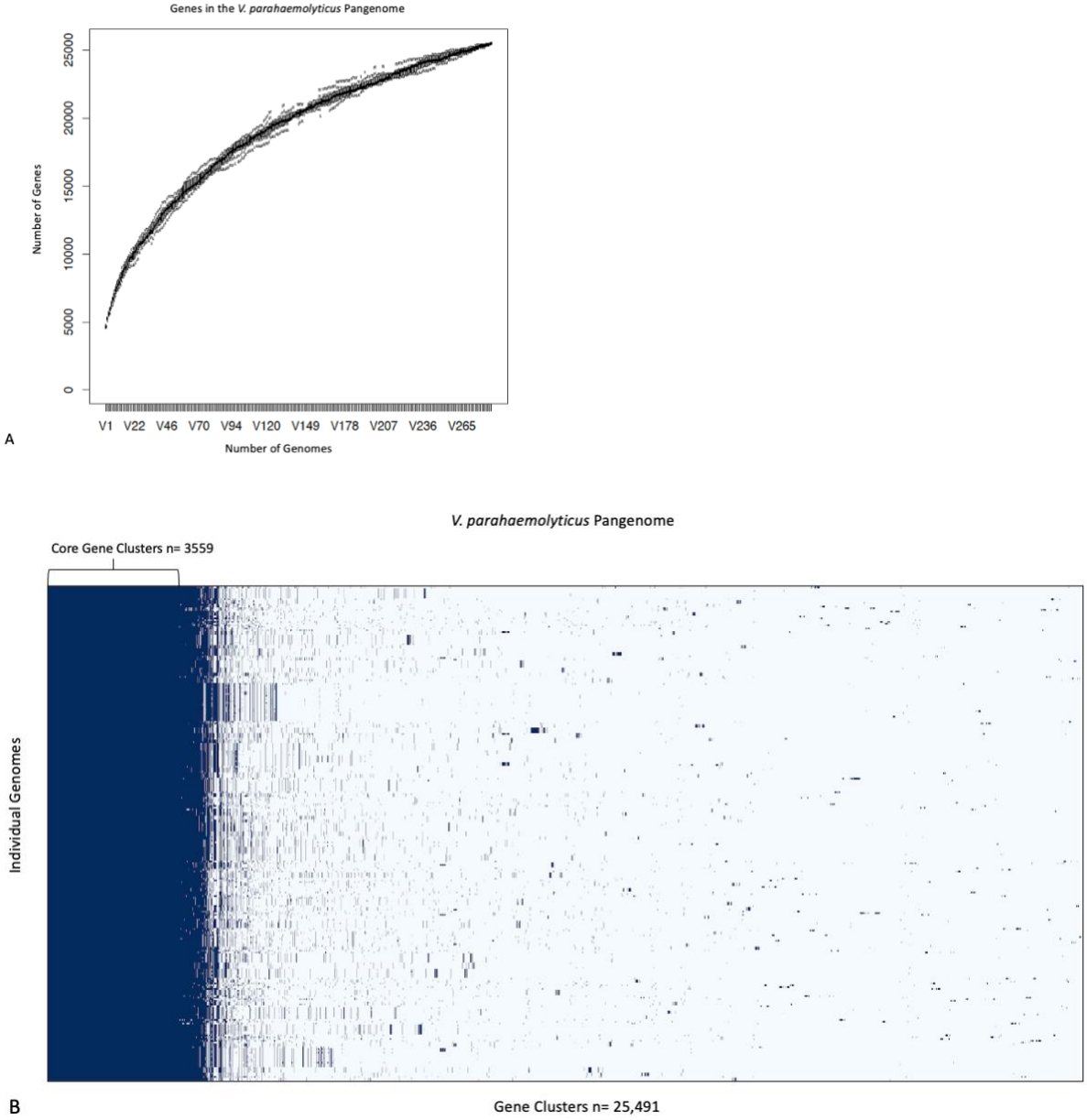
Thanks to the previous collaborative work between the Weimer lab and Jessica Jones' team at FDA including John J. Miller for providing the genomic comparison between sea otter isolates and broader human and shellfish isolates from across North America. Many thanks to the multitude of individuals and organizations that made this project possible (by organization): University of California, Davis: Brian Bird, Danielle Carrade-Holt, Patricia Conrad, Tracey Goldstein, Stori Oates, Jennifer Siembieda, Brett Smith, Robyn Stoddard; California Department of Fish and Wildlife; U.S. Geological Survey: Tim Tinker, Joe Tomoleoni; U.S. Fish and Wildlife Service: Joel Garlich-Miller, Verena Gill, Kristin Worman; Monterey Bay Aquarium: Mike

Murray, Marissa Young; Alaska SeaLife Center: Pam Tuomi; Sea Otter Savvy: Michelle Staedler; Washington Department of Fish and Wildlife: Katherine Haman. Additional thanks to Carol Huang and DJ Darwin Bandoy and the Epicenter for Disease Dynamics at University of California, Davis including but not limited to Nicole Gardner, Megan Moriarty, Pranav Pandit, Jessica Sanchez, and Cristin Young. Special thanks to the late Jack Ames who captained the boat during Elkhorn Slough sampling and helped pioneer the field of otter conservation. Thanks to the late Matthew Blake for making Karen C. Drayer Wildlife Health Center a special place to work.

Data Availability Statement

Whole genome sequence reads with corresponding metadata will be uploaded under bioproject PRJNA203445. Metadata for the 126 genomes from human cases and oysters is available in Miller et al. 2021.

Supplementary Material



Supplemental Figure 2.1. The open pangenome of n=287 *V. parahaemolyticus* genomes collected between 2000-2019 from various North America locations and isolation sources. A) A rarefaction curve for *V. parahaemolyticus* indicates that more genes were discovered as the number of genomes increased.

Supplemental Figure 2.1 continued.

B) Pangenome plot of *V. parahaemolyticus* highlights the small core genome (n=3,559 gene clusters) in relation to the total gene cluster pool (n=25,491). Various patterns of accessory gene presence were observed indicating a large genetic diversity within the sequenced *V. parahaemolyticus* that is unlikely to represent the full genetic diversity of the species.

Supplemental Table 2.1. Prevalence of virulence factors by isolation source in *Vibrio* spp. (n=570) collected between 2000-2019 from various sampling locations across North America. Environ= isolates collected from environmental sources, primarily seawater and marine sediments. Shell= shellfish from oysters, mussels, and clams. NSO = northern sea otters, SSO= southern sea otters.

		alginolyticus count (%)						*diabolicus* count (%)						*cholerae* count (%)					
Class	Gene	Avian	Environ	NSO	Bivalves	SSO	Total	Avian	Environ	NSO	Bivalves	SSO	Total	Avian	Environ	NSO	Bivalves	SSO	Total
		(n=0)	(n=7)	(n=6)	(n=3)	(n=39)	(n=55)	(n=1)	(n=14)	(n=5)	(n=12)	(n=20)	(n=52)	(n=50)	(n=52)	(n=5)	(n=12)	(n=44)	(n=163)
Toxin	*Ace*													2 (4)					2 (1.2)
	hlyA													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	hlyB, C, D																		
	rtxA													29 (58)	30 (57.7)	5 (100)	6 (50)	27 (61.4)	97 (59.5)
	rtxB, C, D													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	tdh																		
	tlh		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	trh																		
	zot													2 (4)					2 (1.2)
Adherence	*acfA,B,C,D*													3 (6)			1 (8.3)		4 (2.5)
	ompU		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	ilpA		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	mam7		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	*tcp**													3 (6)			1 (8.3)		4 (2.5)
T2SS	*gspG*		1 (14.3)	3 (50)		14 (35.9)	18 (32.7)	1 (100)	14 (100)	5 (100)	12 (100)	19 (95)	51 (98.1)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
T3SS1	*exsA*		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	exsD		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	sycN		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	tyeA		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vcrD		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vcrG		7 (100)	6 (100)	3 (100)	38 (97.4)	54 (98.2)	1 (100)	14 (100)	4 (80)	12 (100)	20 (100)	51 (98.1)						
	vcrH		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vcrR		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vcrV		7 (100)	6 (100)	3 (100)	38 (97.4)	54 (98.2)	1 (100)	11 (78.6)	4 (80)	9 (75)	17 (85)	42 (80.8)						
	vecA		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	virG																		
	vopB		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vopD		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vopN		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vopQ		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vopR		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vopS		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	VPA0450																		
	vscB		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscC		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscD, G		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscF, H, I		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscJ		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscK		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscL		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscN		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscO		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscP																		
	vscQ		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscR		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscS		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscT		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscU		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscX		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vscY		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	vxsC		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						

Supplemental Table 2.1 Continued.

		parahaemolyticus count (%)							*anguillarum* (%)			*harveyi* (%)	*metschnikovii* (%)	*ziniensis* (%)
Class	Gene	Avian	Environ	NSO	Bivalves	SSO	Clinical	Total	Environ	NSO	Total	SSO	Environ	Environ
		(n=1)	(n=29)	(n=21)	(n=62)	(n=106)	(n=68)	(n=287)	(n=2)	(n=1)	(n=3)	(n=1)	(n=6)	(n=3)
Toxin	*Ace*													
	hlyA		3 (10.3)	16 (76.2)	23 (37.1)	27 (25.5)	12 (17.6)	81 (28.2)						
	hlyB, C, D		3 (10.3)	16 (76.2)	23 (37.1)	27 (25.5)	12 (17.6)	81 (28.2)						
	rtxA													
	rtxB, C, D													
	tdh		2 (6.9)	14 (66.7)	23 (37.1)	33 (31.1)	42 (61.8)	114 (39.7)						
	tlh	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	trh		6 (20.7)	20 (95.2)	44 (71)	53 (50)	43(63.2)	166 (57.8)						
	zot													
Adherence	*acfA,B,C,D*													
	ompU	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	1 (50)	1 (100)	2 (66.7)	1 (100)		3 (100)
	ilpA	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	3 (100)
	mam7	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	*tcp**													
T2SS	*gspG*	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	3 (100)
T3SS1	*exsA*	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	exsD	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	sycN	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	tyeA	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vcrD	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vcrG	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vcrH	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vcrR	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vcrV	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	vecA	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	virG	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	vopB	1 (100)	26 (89.7)	20 (95.2)	62 (100)	84 (79.2)	67 (98.5)	260 (90.6)				1 (100)		
	vopD	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	vopN	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vopQ	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vopR	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vopS	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	VPA0450	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscB	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscC	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscD, G	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscF, H, I	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscJ	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscK	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	vscL	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscN	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscO	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscP	1 (100)	29 (100)	21 (100)	62 (100)	104 (98.1)	68 (100)	285 (99.3)						
	vscQ	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)						
	vscR	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscS	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscT	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscU	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscX	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vscY	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		
	vxsC	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)				1 (100)		

Supplemental Table 2.1 Continued.

		alginolyticus count (%)						*diabolicus* count (%)						*cholerae* count (%)					
Class	Gene	Avian	Environ	NSO	Bivalves	SSO	Total	Avian	Environ	NSO	Bivalves	SSO	Total	Avian	Environ	NSO	Bivalves	SSO	Total
		(n=0)	(n=7)	(n=6)	(n=3)	(n=39)	(n=55)	(n=1)	(n=14)	(n=5)	(n=12)	(n=20)	(n=52)	(n=50)	(n=52)	(n=5)	(n=12)	(n=44)	(n=163)
T3SS2	*vcrD2*													20 (40)	43 (82.7)	5 (100)	2 (16.7)	13 (29.5)	83 (50.9)
	vopA/vopP																		
	vopB2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vopC																		
	vopD2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vopL													16 (32)	33 (63.5)	4 (80)	2 (16.7)	8 (18.2)	63 (38.7)
	vopT																		
	vopZ																		
	VPA1337													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	VPA1340													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	VPA1350													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	VPA1351													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	VPA1352													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	VPA1353													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	VPA1363													7 (14)	5 (9.6)			1 (2.3)	13 (8)
	vscC2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vscI2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vscN2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	13 (29.5)	83 (50.9)
	vscQ2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vscR2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	13 (29.5)	83 (50.9)
	vscS2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	13 (29.5)	83 (50.9)
	vscT2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vscU2													20 (40)	43 (82.7)	5 (100)	2 (16.7)	11 (25)	81 (49.7)
	vtrA													20 (40)	43 (82.7)	4 (80)	2 (16.7)	8 (18.2)	77 (47.2)
Quorum Sensing	*cqsA*													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	luxS		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
Motility	*pilT*		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	19 (95)	51 (98.1)	1 (2)	3 (7)		1 (8.3)		5 (3.1)
T6SS	*clpB/vasG*													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	hcp-2								1 (7.1)				1 (1.9)	50 (100)	52 (100)	4 (80)	12 (100)	43 (97.7)	161 (98.8)
	icmF/vasK													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	vasA, B,C,D,E,H,I, J, L													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	vasF													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	VCA0109/TssE													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	VCA0122													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	vgrG2													41 (82)	39 (75)	3 (60)	8 (66.7)	36 (81.8)	127 (77.9)
	vgrG3													41 (82)	37 (71.2)	4 (80)	10 (83.3)	35 (79.5)	127 (77.9)
	vipA/mglA													50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	vipB/mglB				1 (33.3)	8 (20.5)	9 (16.4)	1 (100)	4 (28.6)	4 (80)	4 (33.3)	8 (40)	21 (40.4)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
Regulation	*htpB_hsp60*		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
Catalase	*katA*		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)						
	katB		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	13 (92.9)	5 (100)	12 (100)	20 (100)	51 (98.1)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
Other	*ddhA*													2 (4)	2 (3.8)			1 (2.3)	5 (3.1)
	fcl													7 (14)	1 (1.9)			1 (2.3)	9 (5.5)
	gmhA		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	kdsA		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	lpxC		7 (100)	6 (100)	3 (100)	39 (100)	55 (100)	1 (100)	14 (100)	5 (100)	12 (100)	20 (100)	52 (100)	50 (100)	52 (100)	5 (100)	12 (100)	44 (100)	163 (100)
	manB														1 (1.9)				1 (0.6)
	rfaD																		
	cap8E					1 (2.6)	1 (1.8)												
	gmd					1 (2.6)	1 (1.8)							7 (14)	1 (1.9)			1 (2.3)	9 (5.5)

Supplemental Table 2.1 Continued.

		parahaemolyticus count (%)							*anguillarum (%)*			*harveyi (%)*	*metschnikovii (%)*	*ziniensis (%)*
Class	Gene	Avian	Environ	NSO	Bivalves	SSO	Clinical	Total	Environ	NSO	Total	SSO	Environ	Environ
		(n=1)	(n=29)	(n=21)	(n=62)	(n=106)	(n=68)	(n=287)	(n=2)	(n=1)	(n=3)	(n=1)	(n=6)	(n=3)
T3SS2	*vcrD2*		6 (20.7)	19 (90.5)	46 (74.2)	47 (44.3)	50 (73.5)	168 (58.5)						
	vopA/vopP				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vopB2		6 (20.7)	20 (95.2)	46 (74.2)	54 (50.9)	52 (76.5)	178 (62)						
	vopC				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vopD2				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vopL				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vopT				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vopZ				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	VPA1337				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	VPA1340				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	VPA1350				2 (3.2)	1 (0.9)	10 (14.7)	13 (4.5)						
	VPA1351				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	VPA1352				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	VPA1353				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	VPA1363													
	vscC2		6 (20.7)	20 (95.2)	46 (74.2)	54 (50.9)	52 (76.5)	178 (62)						
	vscJ2				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vscN2		6 (20.7)	20 (95.2)	46 (74.2)	57 (53.8)	52 (76.5)	181 (63.1)						
	vscQ2				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vscR2		6 (20.7)	20 (95.2)	46 (74.2)	54 (50.9)	52 (76.5)	178 (62)						
	vscS2		6 (20.7)	20 (95.2)	46 (74.2)	54 (50.9)	52 (76.5)	178 (62)						
	vscT2				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
	vscU2		6 (20.7)	20 (95.2)	46 (74.2)	53 (50)	52 (76.5)	178 (62)						
	vtrA				2 (3.2)	1 (0.9)	9 (13.2)	12 (4.2)						
Quorum Sensing	*cqsA*									1 (100)	1 (33.3)			
	luxS	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	3 (100)
Motility	*pilT*	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	1 (50)		1 (33.3)	1 (100)		
T6SS	*clpB/vasG*								2 (100)	1 (100)	3 (100)			
	hcp-2								2 (100)	1 (100)	3 (100)			3 (100)
	icmF/vasK								2 (100)	1 (100)	3 (100)			
	vasA, B,C,D,E,H,I, J, L								2 (100)	1 (100)	3 (100)			
	vasF								2 (100)	1 (100)	3 (100)			
	VCA0109/TssE								2 (100)	1 (100)	3 (100)			
	VCA0122													
	vgrG2								1 (50)	1 (100)	2 (66.7)			
	vgrG3													
	vipA/mglA								2 (100)	1 (100)	3 (100)			
	vipB/mglB		10 (34.5)	13 (61.9)		43 (40.6)	46 (67.6)	132 (46)	2 (100)	1 (100)	3 (100)	1 (100)		
Regulation	*htpB_hsp60*	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	3 (100)
Catalase	*katA*	1 (100)	29 (100)	21 (100)	61 (98.4)	106 (100)	67 (98.5)	285 (99.3)				1 (100)		
	katB	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)		6 (100)	3 (100)
Other	*ddhA*													
	fcl		1 (3.4)	5 (23.8)		9 (8.5)	4 (5.9)	26 (9.1)						
	gmhA	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	3 (100)
	kdsA	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	3 (100)
	lpxC	1 (100)	29 (100)	21 (100)	62 (100)	106 (100)	68 (100)	287 (100)	2 (100)	1 (100)	3 (100)	1 (100)	6 (100)	
	manB													
	rfaD					1 (0.9)		1 (0.3)	2 (100)	1 (100)	3 (100)	1 (100)		3 (100)
	cap8E		1 (3.4)			2 (1.9)		5 (1.7)						
	gmd		2 (6.9)	10 (47.6)		20 (18.9)	25 (3.7)	70 (24.4)						

Supplemental Table 2.2. Presence of *trh* variants by isolation source in *trh*$^+$ *V. parahaemolyticus* isolates collected from 2000-2019 across various North American sites. Human *trh*$^+$ strains were most likely to carry the full *trh1* gene while shellfish and sea otters of both subspecies were most likely to carry either a pseudogene (*trh*⟘) or the full *trh*2 gene. NSO = northern sea otter. SSO= southern sea otter.

trh variant	Human	Environ.	NSO	Shellfish	SSO	Total
*trh*1 full (189)	24	1	2	0	9	36
*trh*1 153AA	3	0	1	0	4	8
*trh*1 52AA	1	0	0	0	4	5
trh⟘	5	1	11	21	18	56
*trh*2 full	10	4	6	19	17	56
*trh*2 65AA	0	0	0	2	0	2
*trh*2 174AA	0	0	0	2	1	3

Genomic comparison of closely related *Vibrio alginolyticus* and an emerging potential zoonotic pathogen, *Vibrio diabolicus*

Peter J. Sebastian[1,2], Cory Schlesener[3], Barbara A. Byrne[4], Melissa Miller[2,5], Bart C. Weimer[3*], Christine K. Johnson[1,2*]

[1]EpiCenter for Disease Dynamics, One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA
[2]Karen C. Drayer Wildlife Health Center, One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[3]Department of Population Health and Reproduction, 100K Pathogen Genome Project, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[4]Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[5]Marine Wildlife Veterinary Care and Research Center, California Department of Fish and Wildlife, Santa Cruz, CA, USA

***Correspondence:**
Christine K. Johnson
ckjohnson@ucdavis.edu
Bart C. Weimer
bcweimer@ucdavis.edu

Keywords: genome$_1$, vibriosis$_2$, *Vibrio diabolicus*$_3$, *Vibrio alginolyticus*$_4$, virulence, emerging pathogen (Min.5-Max. 8)

Formatted for submission to *Nature Communications*.

Genomic comparison of closely related *Vibrio alginolyticus* and an emerging potential zoonotic pathogen, *Vibrio diabolicus*

Abstract

The Harveyi clade contains some of the most pathogenic *Vibrio* species responsible for vibriosis in humans and diverse wildlife species from invertebrates to marine mammals. Two frequently implicated species are *Vibrio parahaemolyticus* and *Vibrio alginolyticus*, although another emerging species, *Vibrio diabolicus*, is closely related to and has been previously misidentified as *V. alginolyticus*. Despite the similarities, genomic comparisons between both species are scarce and have been impeded by limited sequence data and discordance between classification methods, highlighting the need for more comparative genomic sequencing and standardized approaches to enable more reliable diagnostics. In the current study, the reported species identities of public domain Harveyi clade genomes were reassessed with genomic taxonomic classification and 29 of the 150 genomes reported as *V. alginolyticus* were identified as other species. Reclassification of the reported species identified 28 previously mislabeled *V. diabolicus* genomes (37 total *V. diabolicus*), including 10 isolates from humans. Genome-wide association studies identified species-specific gene clusters for *V. alginolyticus* and *V. diabolicus* as targets to improve species resolution over current multilocus sequence analysis and typing schemes. Genes were identified that associated with isolation from either humans or sea otters (*Enhydra lutris*) and can be utilized to interrogate novel mechanisms of host adaptation and virulence or differentiate host specific strains. This study represents the largest published genomic comparison between *V. diabolicus* (n=88), *V. alginolyticus* (n=163), and *V. parahaemolyticus* (n=287) to date and can inform differentiation of potentially pathogenic strains of *V. diabolicus* and *V. alginolyticus* from humans and sea otters.

Introduction

The *Vibrionaceae* family is composed of over 190 species and 51 distinct clades, many of which are known to cause disease in marine wildlife or humans.[1] The causative agents of vibriosis, or non-cholera *Vibrio* spp. infections, are the leading public health threat from consumption of raw and undercooked seafood and recreational use of coastal waters.[2] Some of the most frequently implicated *Vibrio* spp. derive from the Harveyi clade, including established pathogens of humans, seafood, and marine mammals: *Vibrio parahaemolyticus, Vibrio alginolyticus, and Vibrio harveyi*.[3–5] Harveyi clade species outbreaks are also linked to environmental factors including temperature; during low temperatures *Vibrio* spp. can enter a viable but non-culturable state and can quickly spread during favorable conditions.[2,6] Climate changes such as rising sea surface temperatures and more frequent marine heatwaves have already enhanced *Vibrio* spp. expansion and will likely facilitate *Vibrio* spp. infections in humans and marine mammals in the coming decades.[7–10]

As the risk of vibriosis increases over time, accurate taxonomic classification of the Harveyi clade will be increasingly critical to separate pathogenic and non-pathogenic species in humans, aquaculture, and wildlife. Most Harveyi clade classification schemes lack the resolution to identify isolates accurately and consistently at the species level. Classification of the Harveyi clade is complicated by under-sampling of emerging species, high intraspecies genomic diversity, and discordance between species identification methods.[4,11] One such under-sampled and emerging species is *Vibrio diabolicus*, whose pathogenicity to humans and marine wildlife is poorly studied.[11]

The first characterized *V. diabolicus* was isolated from a deep-sea hydrothermal vent annelid but has since been isolated from at least 2 humans, vertebrate and invertebrate seafood including a diseased grouper, and from coastal environments, with a global distribution.[11–13] While *V. diabolicus* is under-studied and under-sequenced in comparison to closely related *V. alginolyticus* and *V. parahaemolyticus*, recent sampling efforts suggest that it is ubiquitous in coastal environments and is a putative human and marine animal pathogen.[13,14] Whole genome sequencing of *Vibrio* spp. uncovered *V. diabolicus* isolates from northern (*kenyoni*) and southern (*nereis*) sea otters (*Enhydra lutris*) which were clinically misidentified and which were associated with sea otter pathology (Chp 1 and 2), although the genetic mechanisms for pathogenicity and virulence of *V. diabolicus* are largely unknown.

Taxonomic classification of *V. diabolicus* has been difficult in part due to low resolution identification methods such as Matrix Assisted Laser Desorption Ionization -Time of Flight (MALDI-ToF), and Multi-Locus Sequence Analysis (MLSA) schemes that fail to accurately separate *V. alginolyticus* and *V. diabolicus*.[4,15] MLSA schemes may require frequent updates when emerging species are added because they rely on universal primers of “housekeeping” genes present in all species in the clade.[1,4,16] A recent genomic comparison using over one thousand genes for alignment discovered that multiple genomes labeled as *V. alginolyticus* should be reclassified as *V. diabolicus,* and that the previously separate species *Vibrio antiquarius* is a synonym for *V. diabolicus*.[17] Adding to the taxonomic confusion, the recently discovered *V. chemaguriensis*, which has not been validated by the International Code of Nomenclature of Bacteria, is most likely a synonym for *V. diabolicus* based on the most recently proposed *Vibrio* taxonomy.[1,18,19]

Based on the frequent changes in Harveyi taxonomy that have impacted the classification of *V. diabolicus* within the last decade,[1,16,17,20] we hypothesized that public domain genomes would include misidentified *V. diabolicus* isolates. Our first aim was to re-classify publicly available genomes and highlight the advantages of a genome-wide method of species identification and typing over MLSA and Multi-locus Sequence Typing (MLST) typing for differentiating *V. diabolicus* from closely related species. Second, we utilized species-specific genes to identify additional genetic markers that could be useful in future clinical species classification efforts between three highly related species, *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus*. Finally, we identified host-associated genetic markers for *V. alginolyticus* and *V. diabolicus* and illustrated how both species have adapted differentially to sea otter and human hosts.

Methods:

Sample Collection

Sample collection and processing of *Vibrio* spp. isolates are previously described (Chapter 1 and 2). Briefly, n=268 *Vibrio* spp. isolates (55 *V. alginolyticus*, 52 *V. diabolicus*, and 161 *V. parahaemolyticus*) stored at -80°C at the University of California, Davis - Veterinary Medical Teaching Hospital were collected between 2000-2019 in Alaska, California, and Washington from various sources. Most of the isolates were collected across the central coast of California from avian feces (n=2),[21,22] live-sampled and necropsied southern sea otters (n=165),[22–25] bivalves (n=19),[26] and environmental samples (water, sediment, algae, seagrass, and kelp swabs) from a June 2019 investigation of *Vibrio* spp. in Elkhorn Slough, California (n=47). Additional sampling

included live and dead northern sea otters in Alaska from 2004-2015 (n=32),[8,27]and water samples from July 2019 off the coast of Protection Island, Washington (n=3).

Whole genome sequencing and genome assembly

Isolates were plated on 5% sheep blood agar (Biological Media Services, University of California, Davis; Hardy Diagnostics, Santa Maria, CA) at 37°C with 5% CO_2 for 24 hours. DNA was extracted with either the QiaAMP UCP Pathogen extraction kit with mechanical pre-lysis using small pathogen lysis tubes (Qiagen, Hilden, Germany) or the Wizard genomic DNA purification kit (Promega Corporation, Madison, WI) followed by genomic DNA quality and yield assessment with the 2200 TapeStation and genomic DNA ScreenTape (Agilent Technologies, Santa Clara, CA).[28] Whole genome sequencing of *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* isolates was performed as part of the 100K Pathogen Genome Project managed by the Weimer laboratory at the University of California, Davis (http://www.genomes4health.org/).[28–32] Library construction was performed with 1 µg fragmented gDNA using the HTP library preparation kit (Kapa Biosystems, Wilmington, MA) followed by WGS performed on Illumina HiSeq XTEN with PE 150 plus index read (Illumina, San Diego, CA).[33,34] Trimming of adapters or Illumina standards was performed using Trimmomatic (v0.39),[35] and quality of sequencing reads was assessed using FastQC (v0.11.9)[36] before assembly using Shovill (v1.0.4).[37] The quality of assembled genomes was assessed using FastQC and CheckM (v1.1.2)[38] to ensure >90% completeness, <5% contamination, <300 contigs, and >20x genome coverage. Gene annotation was performed using Prokka (v1.14.6)[39] and genomic species identification was determined using Kraken2 (v2.0.8)[40] with a RefSeq microbial genomes database (downloaded May 5, 2021) and Bracken (v2.6.1).[41]

The NCBI genome browser was screened on May 18th, 2022 for whole genome sequence SRA files with submitted species names within the Harveyi clade using the following search terms: {txid717610[Organism:exp] AND "wgs"[Strategy] AND "illumina"[Platform] AND "paired"[Layout]}. Samples with inconsistent file formats or lack of SRA availability were excluded. A total of 3,442 SRA files were downloaded including those submitted to NCBI as *V. parahaemolyticus* (n=3,147), *V. alginolyticus* (n=150), and *V. diabolicus* (n=10). Public domain genomes were re-assembled from raw sequencing reads following the same genome assembly pipeline. An alluvial plot was constructed to compare the species classification submitted on NCBI with the re-assessed genomic species classification across Harveyi clade species using the ggplot2 R (v4.0.3) package. Public domain genomes with genomic species identities of *V. alginolyticus* (n=108; 49 human, 14 environmental, 9 seafood, 31 pipefish Kiel-Fjord ecotype, and 5 other/unlabeled), *V. diabolicus* (n=36; 1 avian, 10 human, 9 environmental, 4 seafood, and 12 other/unlabeled), or *V. parahaemolyticus* that was part of a published North American genome set from humans (n=68) and oysters (n=58), were included with the genomes from the described sampling efforts for further downstream analysis.[42] The 7-loci MLST scheme for *V. parahaemolyticus* typing was applied to all three species using the bioconda mlst program (v2.23.0; T. Seemann, https://github.com/tseemann/mlst), to assess utility as a MLSA scheme for discrimination between all three species.[43]

Genome and Pangenome Analyses

A k-mer based (31 k-mer scaled to 100,000 k-mers/Mbp) all-against-all MinHash sketch was generated using Sourmash (v3.2.3) to confirm appropriate clustering of genomes by their

genomic species classifications.[44,45] Genomes that shared a Jaccard Similarity Index score of 0.9 or higher were considered near identical to identical. A core pangenome analysis was performed using Roary (v3.12.0) using the default 95% or higher identification threshold for unique gene clusters.[46] The core gene phylogeny and pangenome heat map were visualized online using Phandango (v1.3.0).[47] Genes sharing the same Prokka annotation as the 7 loci of the *V. parahaemolyticus* MLST scheme[43] or the 8 loci *Vibrio* spp. MLSA scheme[16] were compiled and visualized with the core gene phylogeny online using Phandango.

Pangenome-association studies (pan-GWAS) were performed using Scoary v.1.6.16.,[48] which calculates a series of Fisher's exact tests to identify gene clusters associated with a binary phenotype of interest. Specifically, pan-GWAS identified gene clusters associated with *V. alginolyticus* (compared to *V. diabolicus* and *V. parahaemolyticus* combined) and genes associated with *V. diabolicus* (compared to *V. alginolyticus* and *V. parahaemolyticus* combined). From the pan-GWAS, shortened lists of gene clusters that were present in at least 98% of the target species and present in less than 5% of the remaining species were compiled. Additional comparisons within *V. alginolyticus* or *V. diabolicus* were performed to identify gene clusters associated with human sourced genomes (vs. non-human sourced) and sea otter sourced genomes (vs. non-sea otter sourced). Genomic features were considered significantly associated with a phenotype based on a naïve p-value and a Benjamini-Hochberg adjusted p-value <0.05. Due to the clonality of 31 *V. alginolyticus* Kiel-Fjord ecotype genomes and their significantly smaller genome sizes biasing pan-GWAS analyses, [49] analyses involving *V. alginolyticus* first removed 30 of the 31 Kiel-Fjord ecotype genomes. Scoary outputs were

parsed through eggnog-mapper (v2.1.6) to improve the functional annotation of significant gene clusters beyond Prokka annotations.[50]

Results and Discussion

Reported versus genomic species identification

Comparison of the reported species identities with genomically determined species identities of 3,442 genomes from Harveyi clade species revealed that 77 genomes were incorrectly classified (Figure 3.1). *V. diabolicus* genomes were underreported as 28 out of 37 genomes that we confirmed as *V. diabolicus* had been incorrectly classified and were most frequently mislabeled as *V. alginolyticus*. The accuracy of reported species identification was lowest for isolates labeled as *V. alginolyticus* (80.7%; 121 of 150). Genomes reported as *V. parahaemolyticus* were largely correctly identified (98.8%; 3,110 of 3,147) as were the collective genomes reported as other Harveyi clade species (95.2%; 138 of 145). Multiple causes for incorrect species identification were observed including mixed or contaminated sequencing, low coverage, low-quality sequencing, and taxonomic changes post-publication. Data entry errors were presumed to be minimal, although there was evidence of 18 contaminated or mixed species in the examined genomes. Frequent recommendations for taxonomic changes within the Harveyi clade have occurred in recent years, highlighting the need to revisit previously published genomes that have utilized outdated classification methods.[1,16,17,20,51] Incorrect species identification can also result from use of lower resolution speciation methods such as MLSA, as failure to discriminate between *V. alginolyticus* and *V. diabolicus* using various MLSA schemes has been previously observed.[4,15]

The correction of species identities greatly improved epidemiologic and ecologic interpretations for *V. diabolicus*, which can now provide a clearer picture of putative routes of global dissemination and pathogenic potential. The initial discoveries of *V. diabolicus* in 1991 (and synonymy with *V. antiquarius* in 1999) was from deep sea vents in the East Pacific rise region.[12,52] The oldest available genome appeared to have been isolated from cured hides of an unspecified animal from South Africa in 1973, but no corresponding publication or additional information to confirm its origins was available. Assuming it originated from the East Pacific rise region, *V. diabolicus* may have spread both East and West across the Pacific as evidenced by lab strains in Japan (1994-1996) and shellfish and sea otters in California (2001-2013). Selection bias in sampling efforts may have resulted in missed detection during prior spatiotemporal windows, although use of 'molecular clock' analyses could prove beneficial to support hypotheses for historical *V. diabolicus* transmission waves. *Vibrio diabolicus* isolates were first detected in Alaska in 2013-2015, suggesting that *V. diabolicus* may be expanding due to climate change into traditionally unsuitable marine environments in a manner similar to *V. parahaemolyticus*.[8] Further studies are needed because suitable environmental parameters such as temperature and salinity for *V. diabolicus* growth are not adequately described.

A genome of *V. diabolicus* originally identified as *V. alginolyticus* that was collected from a diseased fish in Greece appears to be the first reported strain from Europe, although this strain was not included in this study due to no publicly available raw sequencing reads.[53] *Vibrio diabolicus* genomes were sequenced from the Eastern United States, United Kingdom, and middle east (Amman, Jordan) since at least 2015, including 10 genomes from human clinical cases. *Vibrio diabolicus* infections may be a concern for aquaculture as it has been isolated from

shrimp in recent years (2019-2020), although at least one strain of *V. diabolicus* (Ili) may act as a probiotic to protect shrimp from acute hepatopancreatic necrosis disease (AHPND) caused by *V. parahaemolyticus*.[54–56]

Genomic comparison between *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus*

Whole genome phylogeny and genomic comparison,[44,57] showed that all three species separated based on genomic identities, further supporting the validity of the re-assessed genomic species identities (Figure 3.2). Of the three species, *V. diabolicus* and *V. alginolyticus* shared the most genomic similarity, while *V. parahaemolyticus* and *V. alginolyticus* were the least similar. *Vibrio diabolicus* was previously shown to be highly related to *V. alginolyticus* based on average nucleotide identity (ANI ~92%), which was close to the threshold for species delineation (<95%).[17,49] The previously reported Kiel-Fjord ecotype of *V. alginolyticus* isolated from pipefish (*Syngnathus typhle*) comprised a distinct genomic cluster within the larger *V. alginolyticus* cluster.[49,58] Kiel-Fjord ecotype *V. alginolyticus* genomes were significantly smaller with fewer accessory genes than other *V. alginolyticus*, *V. diabolicus*, or *V. parahaemolyticus* genomes.

Isolates from humans and sea otters were represented across the genomic diversity of both *V. alginolyticus* and *V. diabolicus*. Some sequence types associated with human *V. parahaemolyticus* outbreaks, such as ST3 and ST36, exhibit clonal expansion.[42,59–61] Apparent expansion of highly related *V. parahaemolyticus* strains from both sea otters and humans was previously detected within our genome set (Chapter 2). Evidence for similar expansion of *V.*

alginolyticus or *V. diabolicus* outside of the Kiel-Fjord ecotype was minimal, although increased sequencing of *V. alginolyticus* and *V. diabolicus* genomes may further identify highly related genomic clusters. The largest genomic cluster of *V. alginolyticus* included 6 genomes from the USA isolated from humans, southern sea otters, and estuarine water. Small genomic clusters of *V. diabolicus* were also observed, including 5 of the 6 Japanese isolates, and 4 of the USA estuarine water isolates sampled on the same day.

Highly related genomes may reflect persistence of clonally expansive strains or could represent selection bias due to spatiotemporally localized sampling efforts. Removal of redundant genomes could be beneficial for broad population dynamic comparisons to reduce the risk of selection bias but may also miss the ability to identify clonal expansion. A recent study identified only 469 (of 1103) non-redundant *V. parahaemolyticus* strains from various sources based on a threshold of <2000 core genome single nucleotide polymorphisms,[60] which indicates that more than half of the genomes were closely related to one or more of the other genomes. Future genomic sampling of *V. alginolyticus* and *V. diabolicus* should include varied locations, especially in Africa and Australia, as sampling to date has been heavily skewed towards the United States and Europe. Our dataset included genomes of *V. alginolyticus* and *V. diabolicus* from a coastal marine mammal (sea otters), which could provide unique insight into the dynamics of *Vibrio* spp. infections in nearshore marine environments, given their potential for *Vibrio* spp. infections and the role of sea otters as sentinels of coastal environment health.[5,24,27,62,63]

Pangenome and Species-specific markers

The shared pangenome of *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* was extremely diverse and comprised a pool of 49,465 gene clusters. The combined pangenome analysis of all three species identified separate pangenome regions including a shared core genome, core genomes shared between 2 out of 3 species, and species-specific core genomes (Figure 3.3). Despite high within-species diversity, there were clear genetic differences between the 3 *Vibrio* species indicating divergent evolution between all three species. There were 643 gene clusters that were highly associated with *V. alginolyticus* and 477 gene clusters highly associated with *V. diabolicus*. Species-specific gene clusters could be utilized as markers to improve or confirm species identification. Despite utilizing two separate gene annotators, many of these gene clusters are annotated as hypothetical proteins, indicative of the lack of molecular studies within *V. alginolyticus* and *V. diabolicus*. Future studies have potential to greatly advance the molecular characterization of both species by interrogating species-specific gene clusters. While whole genome methods are capable of the finest species resolution, lower resolution methods including MLSA remain popular because they require a limited number of loci that can be analyzed using PCR products for potentially lowered costs.[1] Presence or absence of genes annotated to gene names used in the 7 loci *V. parahaemolyticus* MLST scheme or the 8 loci MLSA scheme were plotted alongside the pangenome phylogeny, which revealed that multiple presumed housekeeping genes exhibited sequence divergence at a <95% identity between variants, suggesting apparent allelic variation (Figure 3.4). Sequence divergence of MLSA and MLST loci occurred both within and between species; in some cases, multiple genes with a shared annotation were present in the same genome, which may suggest

gene duplication. A commonly utilized genetic marker, *recA*, had 11 annotated gene products with <95% shared identity including an allele of *recA* that was primarily, but not completely restricted to *V. alginolyticus*. The *recA* loci is used in multiple MLSA schemes and the *V. parahaemolyticus* MLST scheme,[4,43] despite frequently undergoing recombination events that can affect its usefulness for species identification and typing.[64] Presence of the *V. alginolyticus* associated *recA* allele in both *V. diabolicus* and *V. parahaemolyticus* in this study further reflect concerns regarding its consistency as a marker. Other MLSA scheme loci with more than one annotated gene included *ftsZ*, *mreB*, *rpoA*, and *topA*.

All the *V. alginolyticus* and *V. diabolicus*, and 22.6% of the *V. parahaemolyticus* isolates were non-typeable using the *V. parahaemolyticus* MLST scheme. Most of the MLST loci (*dnaE*, *pntA*, *pyrC*, *tnaA*) could not be typed in any *V. alginolyticus* or *V. diabolicus* genomes, while *recA*, *dtdS*, *gyrB* exhibited limited utility in both species. A majority of the untyped *V. parahaemolyticus* belonged to genomes from California of either environmental (n=19/29, 65.5%) or southern sea otter sources (n=39/106, 36.8%) highlighting the need for higher resolution methods when working with understudied and underrepresented sampling sources. Inaccuracy of MLSA/MLST schemes can arise because they utilize drastically less sequence (0.1%) compared to whole genome alignment.[64] Novel MLSA and MLST schemes could instead utilize a greater number of gene clusters that are both unique to each species and ubiquitous within each species.

Host-associated selection

Diverse pangenomes with a large accessory gene pool, as observed in this study, are frequently attributed to horizontal gene transfer and niche adaptation.[65] The previously described Kiel-Fjord ecotype provides an extreme example of niche adaptation, with selection of a very specific *V. alginolyticus* genomic profile in pipefish.[49,58] Comparisons of *V. alginolyticus* and *V. diabolicus* genomes isolated from humans and non-humans, and isolates from sea otters and non-sea otters were used to examine potential virulence-related niche adaptation in mammalian hosts. Human *V. alginolyticus* genomes were either positively or negatively associated with 116 different gene clusters, while sea otter *V. alginolyticus* were associated with 209 gene clusters. No *V. diabolicus* gene clusters were significantly associated with human or sea otter hosts after Benjamini-Hochberg adjustment, but due to the lower sample size of *V. diabolicus* genomes, gene clusters with significant naïve p-values and large effect sizes worthy of further investigation were reported. Reports of genomic sequencing of *V. diabolicus* have been extremely limited, although new sequences continue to be published.[66–68] Future studies can utilize potential source-specific genes as targets to investigate host adaption and virulence.

Multiple toxin-antitoxin (TA) systems, common in many bacteria species, exhibited a host preference for genomes from sea otters. TA systems, of which there are at least 6 described in *Vibrio* spp., have many putative roles which include maintenance of a viable but non-culturable state, stabilization of genomic and plasmid DNA, protection from phages, colonization, and biofilm formation.[69–72] Genes annotated to four distinct TA systems were highly associated with *V. alginolyticus* isolates from sea otters compared to *V. alginolyticus* isolates from non-otter sources (Table 3.1). A similar but non-significant pattern was observed for *V. diabolicus* isolates from sea otters. TA system genes may be shared via horizontal gene transfer, as evidenced by

an environmental strain of *V. diabolicus* (JBS-8-11-1) carrying a fitness island with 13 TA system genes.[73] A similar fitness island that allows *V. alginolyticus* or *V. diabolicus* to better compete in marine mammal hosts could explain the presence of sea otter-associated TA systems. Further molecular characterization studies should investigate the host-specific adaptation of *Vibrio* spp. TA system genes and their potential impact on virulence.

Variants of multiple cold shock protein (*csp*) genes were also strongly associated with *V. alginolyticus* from sea otters, and a similar but non-significant trend was observed with *V. diabolicus*. Cold shock response is vital to survival in harsh conditions, thus *csp* genes are frequently included in the core genome.[74–76] However, non-core *csp* genes may confer fitness that extends beyond cold tolerance. Despite their names, *csp* genes can have additional functions, including regulation of biofilm formation and expression of Type VI Secretion Systems.[74,76,77] Understudied *csp* genes of *V. alginolyticus* and *V. diabolicus* associated with sea otter hosts could have evolved specific functions to improve host specific gut colonization in sea otters but not humans, or help commensal *Vibrio* strains in sea otters survive transitions to and from marine environments.

Variants of the putative virulence factor of *V. alginolyticus*, *apx1B* were also associated with human or sea otter hosts. The *apx1B* gene is involved in repeat-in-toxin (rtx) translocation and is part of the Type I Secretion System (T1SS).[78–80] Almost all *V. alginolyticus* genomes carried a variant of *apx1B*, but separate variants were associated with isolates from sea otters and humans. The host-associated allelic variation of this gene could indicate that Rtx toxin secretion is host-adapted to improve gut colonization differentially in humans and sea otters.

Finally, a gene involved in biofilm formation, *csgG*, was associated with *V. alginolyticus* from human sources. The CsgF-CsgG protein complex functions as a secretion channel for curli fibers, which promote biofilm formation.[81] While *csgG* is characterized as a *V. parahaemolyticus* gene, this *csgG* variant cluster was nearly exclusive to *V. alginolyticus*.

The two most frequently studied virulence factors of *V. parahaemolyticus*, the thermostable direct hemolysin (*tdh*) and thermostable direct hemolysin-like hemolysin (*trh*), have rarely been detected from *V. alginolyticus* or *V. diabolicus*.[82–84] In this study, *tdh* was not detected in *V. alginolyticus* and *V. diabolicus* genomes, but one *V. diabolicus* genome (SRR10099733) carried *trh*. Virulence mechanisms of *V. alginolyticus* and *V. diabolicus* are understudied and rely heavily on knowledge from *V. cholerae* and *V. parahaemolyticus*, especially for annotated genes in the virulence factor database.[14,82,83,85–87] A genome-wide association study approach, as used in this study, has the ability to detect novel or understudied virulence genes and gene variants agnostic to what is reported in better characterized species.

Conclusions

Following previous work that showed 4 genomes that were initially reported as *V. alginolyticus* were correctly identified as *V. diabolicus*,[17] this study further confirmed that *V. diabolicus* genomes were frequently misclassified as *V. alginolyticus* and other *Vibrio* species. *Vibrio diabolicus* has been systematically underreported due to these misclassifications; proper identification of *V. diabolicus* genomes improved upon the limited epidemiologic and ecologic interpretations of the species and provided additional evidence that *V. diabolicus* is an overlooked, emerging human pathogen. While MLSA/MLST schemes provide a quick and low-

cost method to identify species and categorize relatedness between strains, interrogation of an 8-loci MLSA scheme and the 7-loci *V. parahaemolyticus* MLST scheme revealed sequence divergence of some loci genes, which may lead to a failure to appropriately classify strains. Additional species-specific genes and allelic variants identified here provide potential genetic markers for generating improved species classification. In addition, genome-wide association studies can help identify novel virulence factors or antimicrobial resistance genes in understudied marine species such as *V. alginolyticus* and *V. diabolicus* without relying on databases comprised of genes from related but better characterized species such as *V. parahaemolyticus*. Genome-wide associations were able to identify putative virulence genes that help characterize ways in which *V. alginolyticus* and *V. diabolicus* have adapted to various hosts, including humans and marine bioindicators. Genes putatively related to host-adapted virulence in these species worthy of further investigation include TA systems, cold-shock proteins, secretion systems, and genes involved in biofilm formation.

Due to predicted increases in *Vibrio* spp. infection due to changing climates,[9] accurate identification of species and classification of high-risk strains is needed to advance human and marine wildlife health. Genomic epidemiology analyses and tools presented here can help inform species classification and differentiation of virulent strains in poorly characterized *Vibrio* species. In addition, this study represents the largest published genomic comparison for three clinically important *Vibrio* species; *V. diabolicus*, an emerging potential pathogen of humans and marine wildlife, *V. alginolyticus*, an established but understudied pathogen, and *V. parahaemolyticus*, a relatively well studied cause of human and animal vibriosis.

References

1. Jiang, C. *et al. Vibrio* clade 3.0: new *Vibrionaceae* evolutionary units using genome-based approach. *Curr Microbiol* **79**, 1–15 (2022).

2. Baker-Austin, C. *et al. Vibrio* spp. infections. *Nat Rev Dis Primers* **4**, 1–19 (2018).

3. Mougin, J. *et al.* Rapid identification of *Vibrio* species of the Harveyi clade using MALDI-TOF MS profiling with main spectral profile database implemented with an in-house database: Luvibase. *Front Microbiol* **11**, 2494 (2020).

4. Culot, A. *et al.* Isolation of Harveyi clade *Vibrio* spp. collected in aquaculture farms: How can the identification issue be addressed? *J Microbiol Methods* **180**, 106106 (2021).

5. Miller, M. A. *et al.* Predators, disease, and environmental change in the nearshore ecosystem: mortality in southern sea otters (*Enhydra lutris nereis*) from 1998–2012. *Front Mar Sci* **7**, 582 (2020).

6. Yoon, J. H. & Lee, S. Y. Characteristics of viable-but-nonculturable *Vibrio parahaemolyticus* induced by nutrient-deficiency at cold temperature. *Critical Reviews in Food Science and Nutrition* vol. 60 1302–1320 Preprint at https://doi.org/10.1080/10408398.2019.1570076 (2020).

7. Cheung, W. W. L. & Frölicher, T. L. Marine heatwaves exacerbate climate change impacts for fisheries in the northeast Pacific. *Sci Rep* **10**, (2020).

8. Goertz, C. E. C. *et al. Vibrio parahaemolyticus*, a climate change indicator in Alaska marine mammals. in *Responses of Arctic marine ecosystems to climate change. Alaska Sea Grant, University of Alaska Fairbanks* 41–52 (2013). doi:10.4027/ramecc.2013.03.

9. Baker-Austin, C., Trinanes, J., Gonzalez-Escalona, N. & Martinez-Urtaza, J. Non-cholera *Vibrios*: the microbial barometer of climate change. *Trends in Microbiology* vol. 25 76–84 Preprint at https://doi.org/10.1016/j.tim.2016.09.008 (2017).

10. Gulland, F. M. D. *et al.* A review of climate change effects on marine mammals in United States waters: Past predictions, observed impacts, current research and conservation imperatives. *Climate Change Ecology* **3**, 100054 (2022).

11. Song, J. *et al.* Isolation, identification and pathogenesis study of *Vibrio diabolicus*. *Aquaculture* **533**, 736043 (2021).

12. Raguénès, G., Christen, R., Guezennec, J., Pignet, P. & Barbier, G. *Vibrio diabolicus* sp. nov., a new polysaccharide-secreting organism isolated from a deep-sea hydrothermal vent polychaete annelid, *Alvinella pompejana*. *Int J Syst Bacteriol* **47**, 989–995 (1997).

13. Dias, J. I. P. Diversity of tetracycline-resistant *Vibrio* spp. in an estuarine environment: *V. diabolicus* genome case study. (Universidade D Coimbra, 2020).

14. Pérez-Duque, A. *et al.* Comparative genomics of clinical and environmental isolates of *Vibrio* spp. of Colombia: implications of traits associated with virulence and resistance. *Pathogens* **10**, 1605 (2021).

15. Rahman, M. S. *et al.* Vibrio trends in the ecology of the venice lagoon. *Appl Environ Microbiol* **80**, 2372–2380 (2014).

16. Sawabe, T. *et al.* Updating the *Vibrio* clades defined by multilocus sequence phylogeny: Proposal of eight new clades, and the description of *Vibrio tritonius* sp. nov. *Front Microbiol* **4**, (2013).

17. Turner, J. W. *et al.* Comparative genomic analysis of *Vibrio diabolicus* and six taxonomic synonyms: a first look at the distribution and diversity of the expanded species. *Front Microbiol* **9**, 1893 (2018).

18. Ghosh, A. & Bhadury, P. *Vibrio chemaguriensis* sp. nov., from Sundarbans, Bay of Bengal. *Curr Microbiol* **76**, 1118–1127 (2019).

19. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)* **2020**, (2020).

20. Urbanczyk, H., Ogura, Y. & Hayashi, T. Taxonomic revision of Harveyi clade bacteria (family *Vibrionaceae*) based on analysis of whole genome sequences. *Int J Syst Evol Microbiol* **63**, 2742–2751 (2013).

21. Siembieda, J. L. *et al.* Zoonotic pathogens isolated from wild animals and environmental samples at two California wildlife hospitals. *J Am Vet Med Assoc* **238**, 773–783 (2011).

22. Oates, S. C. *et al.* Epidemiology and potential land-sea transfer of enteric bacteria from terrestrial to marine species in the Monterey Bay region of California. *J Wildl Dis* **48**, 654–668 (2012).

23. Miller, M. A. *et al.* Enteric bacterial pathogen detection in southern sea otters (*Enhydra lutris nereis*) is associated with coastal urbanization and freshwater runoff. *Vet Res* **41**, (2010).

24. Brownstein, D. *et al.* Antimicrobial susceptibility of bacterial isolates from sea otters (*Enhydra lutris*). *J Wildl Dis* **47**, 278–292 (2011).

25. Kreuder, C. *et al.* Patterns of mortality in southern sea otters (*Enhydra lutris nereis*) from 1998–2001. *J Wildl Dis* **39**, 495–509 (2013).

26. Miller, W. A. *et al. Salmonella* spp., *Vibrio* spp., *Clostridium perfringens*, and *Plesiomonas shigelloides* in marine and freshwater invertebrates from coastal California ecosystems. *Microb Ecol* **52**, 198–206 (2006).

27. Burek Huntington, K. A. *et al.* Causes of mortality of northern sea otters (*Enhydra lutris kenyoni*) in Alaska from 2002 to 2012. *Front Mar Sci* **8**, 630582 (2021).

28. Kong, N., Ng, W., Lee, V., Weimer, B. C. & Kelly, L. *Production and analysis of high molecular weight genomic DNA for NGS pipelines using Agilent DNA extraction kit (p/n 200600)*. http://100kgenome. (2013).

29. Chen, P. *et al.* Comparative genomics reveals the diversity of restriction-modification systems and DNA methylation sites in *Listeria monocytogenes*. *Appl Environ Microbiol* **83**, e02091-16 (2017).

30. Weimer, B. C. 100K Pathogen Genome Project. *Genome Announc* **5**, (2017).

31. Kong, N. *et al.* Draft genome sequences of 1,183 *Salmonella* strains from the 100K Pathogen Genome Project. *Genome Announc* **5**, e00518-17 (2017).

32. Lüdeke, C. H. M., Kong, N., Weimer, B. C., Fischer, M. & Jones, J. L. Complete Genome Sequences of a Clinical Isolate and an Environmental Isolate of Vibrio parahaemolyticus. *Genome Announc* **3**, (2015).

33. Miller, B. *et al. A novel, single-tube enzymatic fragmentation and library construction method enables fast turnaround times and improved data quality for microbial whole-genome sequencing*. *Kapa Biosystems Application Note* (2015) doi:10.13140/RG.2.1.4534.3440.

34. Lawton, S. J. *et al.* Comparative analysis of Campylobacter isolates from wild birds and chickens using MALDI-TOF MS, biochemical testing, and DNA sequencing. *Journal of Veterinary Diagnostic Investigation* **30**, 354–361 (2018).

35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

36. Andrews, S. *FastQC: a quality control tool for high throughput sequence data*. (2010).

37. Seemann, T. *Shovill—Assemble bacterial isolate genomes from Illumina paired-end reads*. *GitHub: San Francisco, CA, USA* (2020).

38. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).

39. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

40. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 1–13 (2019).

41. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* **3**, e104 (2017).

42. Miller, J. J. *et al.* Phylogenetic and biogeographic patterns of *Vibrio parahaemolyticus* strains from North America inferred from whole-genome sequence data. *Appl Environ Microbiol* **87**, (2021).

43. González-Escalona, N. *et al.* Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *J Bacteriol* **190**, 2831–2840 (2008).

44. Brown, C. T. & Irber, L. sourmash: a library for MinHash sketching of DNA. *The Journal of Open Source Software* **1**, 27 (2016).

45. Bandoy, D. D. R. & Weimer, B. C. Biological machine learning combined with campylobacter population genomics reveals virulence gene allelic variants cause disease. *Microorganisms* **8**, (2020).

46. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

47. Hadfield, J. *et al.* Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**, 292–293 (2018).

48. Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* **17**, 1–9 (2016).

49. Chibani, C. M., Roth, O., Liesegang, H. & Wendling, C. C. Genomic variation among closely related *Vibrio alginolyticus* strains is located on mobile genetic elements. *BMC Genomics* **21**, 1DUMM (2020).

50. Cantalapiedra, C. P., Herṇandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825–5829 (2021).

51. Hoffmann, M., Monday, S. R., Fischer, M. & Brown, E. W. Genetic and phylogenetic evidence for misidentification of *Vibrio* species within the Harveyi clade. *Lett Appl Microbiol* **54**, 160–165 (2012).

52. Hasan, N. A. *et al.* Deep-sea hydrothermal vent bacteria related to human pathogenic *Vibrio* species. *Proc Natl Acad Sci U S A* **112**, E2813–E2819 (2015).

53. Castillo, D. *et al.* Draft genome sequences of *Vibrio alginolyticus* strains V1 and V2, opportunistic marine pathogens. *Genome Announc* **3**, (2015).

54. Restrepo, L. *et al.* Microbial community characterization of shrimp survivors to AHPND challenge test treated with an effective shrimp probiotic (*Vibrio diabolicus*). *Microbiome* **9**, 1–20 (2021).

55. Nava-Soberanes, E. B. *et al.* Draft genome sequence of first *Vibrio diabolicus* in Mexico strain InDRE-D1-M1, an emergent threat. *J Infect Dev Ctries* **15**, 1750–1754 (2021).

56. Ramirez, M. *et al.* The probiotics *Vibrio diabolicus* (Ili), *Vibrio hepatarius* (P62), and *Bacillus cereus sensu stricto* (P64) colonize internal and external surfaces of *Penaeus vannamei* shrimp larvae and protect it against *Vibrio parahaemolyticus*. *Aquaculture* **549**, 737826 (2022).

57. Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence comparisons with sourmash. *F1000Res* **8**, (2019).

58. Chibani, C. M., Hertel, R., Hoppert, M., Liesegang, H. & Wendling, C. C. Closely related *Vibrio alginolyticus* strains encode an identical repertoire of caudovirales-like regions and filamentous phages. *Viruses 2020, Vol. 12, Page 1359* **12**, 1359 (2020).

59. Xu, F. *et al.* Sequence type 631 *Vibrio parahaemolyticus*, an emerging foodborne pathogen in North America. *J Clin Microbiol* **55**, 645 (2017).

60. Yang, C. *et al.* Recent mixing of *Vibrio parahaemolyticus* populations. *The ISME Journal 2019 13:10* **13**, 2578–2588 (2019).

61. Boyd, E. F. *et al.* Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. *BMC Microbiol* **8**, 1–14 (2008).

62. Jessup, D. A. *et al.* Sea otters in a dirty ocean. *J Am Vet Med Assoc* **231**, 1648–1652 (2007).

63. Jessup, D. A. *et al.* Southern sea otter as a sentinel of marine ecosystem health. *Ecohealth* **1**, 239–245 (2004).

64. Gunasekara, C. W. R., Rajapaksha, L. G. T. G. & Wimalasena, S. H. M. P. Comparative analysis unravels genetic recombination events of Vibrio parahaemolyticus recA gene. *Infection, Genetics and Evolution* **107**, 105396 (2023).

65. Shapiro, B. J. *et al. Population Genomics of Early Events in the Ecological Differentiation of Bacteria*. vol. 17 (2012).

66. Teixeira, J. S. *et al.* Draft genomes of 75 *Vibrio* isolates from retail mollusks collected in Canada between 2014 and 2018. *Microbiol Resour Announc* **11**, (2022).

67. Teixeira, J. S. *et al.* Draft genomes of 83 *Vibrio* isolates from fresh Canadian mollusks. *Microbiol Resour Announc* **11**, (2022).

68. Teixeira, J. S. *et al.* Draft genomes of 92 *Vibrio* isolates from warm-water shrimps imported into Canada between 2009 and 2019. *Microbiol Resour Announc* **11**, (2022).

69. Song, X., Lin, Z. & Yuan, W. Toxin–antitoxin systems in pathogenic Vibrio species: a mini review from a structure perspective. *3 Biotech* **12**, 1–12 (2022).

70. De Jonge, N. *et al.* Structural and thermodynamic characterization of *Vibrio fischeri CcdB*. *Journal of Biological Chemistry* **285**, 5606–5613 (2010).

71. Zheng, C. *et al.* Identification and characterization of the chromosomal *yefM-yoeB* toxin-antitoxin system of *Streptococcus suis*. *Scientific Reports 2015 5:1* **5**, 1–13 (2015).

72. Budde, P. P., Davis, B. M., Yuan, J. & Waldor, M. K. Characterization of a *higBA* toxin-antitoxin locus in *Vibrio cholerae*. *J Bacteriol* **189**, 491–500 (2007).

73. Klein, S., Pipes, S. & Lovell, C. R. Occurrence and significance of pathogenicity and fitness islands in environmental *Vibrios*. *AMB Express* **8**, 1–7 (2018).

74. Datta, P. P. & Bhadra, R. K. Cold shock response and major cold shock proteins of *Vibrio cholerae*. *Appl Environ Microbiol* **69**, 6361 (2003).

75. Zhu, C. *et al.* Genomic and transcriptomic analyses reveal distinct biological functions for cold shock proteins (VpaCspA and VpaCspD) in *Vibrio parahaemolyticus* CHN25 during low-temperature survival. *BMC Genomics* **18**, 1 (2017).

76. Nam, D., Motegi, W., Ishimori, K. & Uchida, T. Heme binding to cold shock protein D, CspD, from *Vibrio cholerae*. *Biochem Biophys Res Commun* **624**, 151–156 (2022).

77. Townsley, L., Sison Mangus, M. P., Mehic, S. & Yildiz, F. H. Response of *Vibrio cholerae* to low-temperature shifts: CspV regulation of type vi secretion, biofilm formation, and association with zooplankton. *Appl Environ Microbiol* **82**, 4441–4452 (2016).

78. Yoo, H. J., Lee, S. & Ryu, D. Y. Role of the ApxIB/ApxID exporter in secretion of the ApxII and ApxIII toxins *in Actinobacillus pleuropneumoniae*. *Korean J Vet Res* **60**, 225–228 (2021).

79. Benz, R. Channel formation by RTX-toxins of pathogenic bacteria: Basis of their biological activity. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1858**, 526–537 (2016).

80. Linhartová, I. *et al.* RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* **34**, 1076–1112 (2010).

81. Yan, Z., Yin, M., Chen, J. & Li, X. Assembly and substrate recognition of curli biogenesis system. *Nature Communications 2020 11:1* **11**, 1–10 (2020).

82. Hernández-Robles, M. F. *et al.* Virulence factors and antimicrobial resistance in environmental strains of Vibrio alginolyticus. *International Microbiology* **19**, 191–198 (2016).

83. Ren, C. *et al.* Distribution and pathogenic relationship of virulence associated genes among *Vibrio alginolyticus* from the mariculture systems. *Mol Cell Probes* **27**, 164–168 (2013).

84. Klein, S. L., Gutierrez West, C. K., Mejia, D. M. & Lovell, C. R. Genes similar to the *Vibrio parahaemolyticus* virulence-related genes *tdh*, *tlh*, and *vscC2* cccur in other *vibrionaceae* species isolated from a pristine estuary. *Appl Environ Microbiol* **80**, 595–602 (2014).

85. Cai, H. *et al.* Research progress on virulence factors of *Vibrio alginolyticus*: a key pathogenic bacteria of sepsis. in *Heat Illness and Critical Care [Working Title]* (IntechOpen, 2022). doi:10.5772/INTECHOPEN.108206.

86. Xie, Z. Y., Hu, C. Q., Chen, C., Zhang, L. P. & Ren, C. H. Investigation of seven *Vibrio* virulence genes among *Vibrio alginolyticus* and *Vibrio parahaemolyticus* strains from the coastal mariculture systems in Guangdong, China. *Lett Appl Microbiol* **41**, 202–207 (2005).

87. Xiong, P. *et al.* Virulence-related genes of *Vibrio alginolyticus* and its virulence in mice. *Wei Sheng Wu Xue Bao* **54**, 80–88 (2014).

Figures

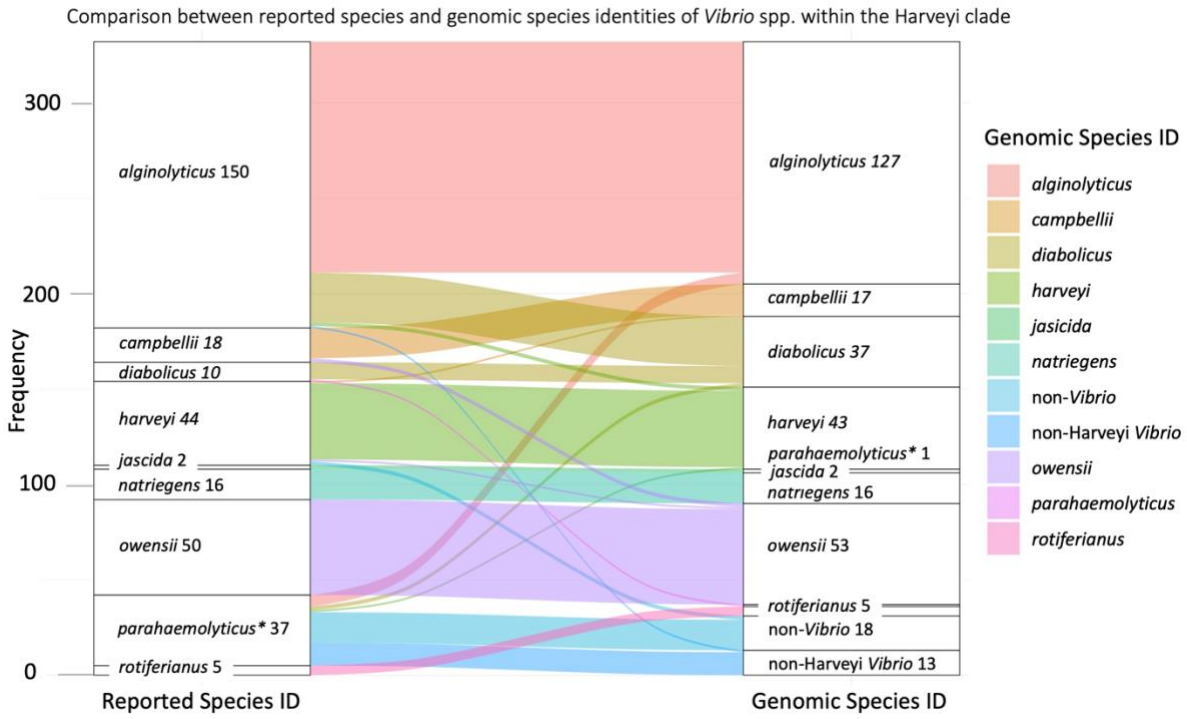


Figure 3.1. Alluvial diagram comparison between the reported species identities of 332 out of 3,442 NCBI uploaded whole genome sequence raw read files and the determined genomic species identities of Harveyi clade *Vibrio* species. *The 3,110 concordant *V. parahaemolyticus* isolates were not included in the flow chart. The most frequent misclassification involved *V. diabolicus* genomes reported on NCBI as *V. alginolyticus*.

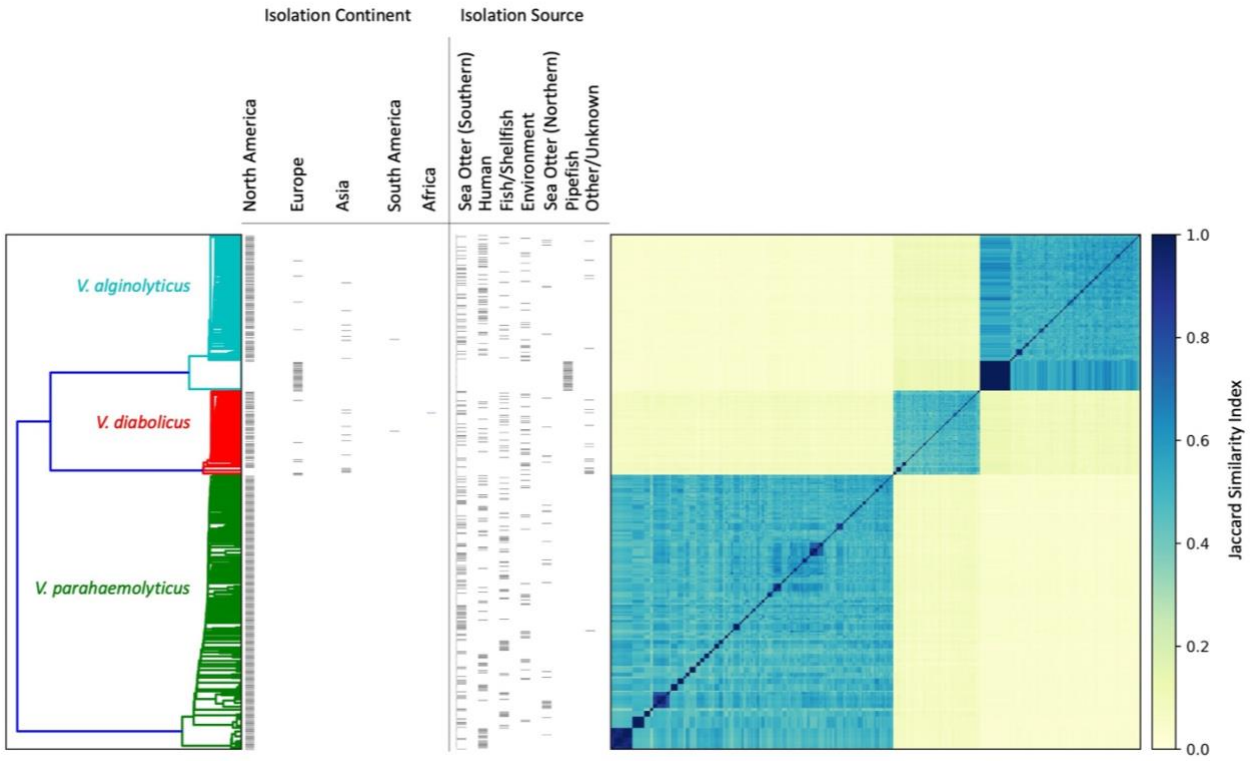


Figure 3.2. All-against-all Comparative analysis of the genomic similarity between *V. alginolyticus* (n=163), *V. diabolicus* (n=88), and *V. parahaemolyticus*(n=287) genomes collected from various sources globally. Darker blue colors indicated greater similarity between genomes based on the Jaccard Similarity Index. All three species appropriately clustered into distinct non-overlapping regions of genomic similarity. The highest similarity between species was observed for *V. alginolyticus* and *V. diabolicus*. Genomes were labeled based on isolation source and continent where the isolate was obtained. Colors on the non-rooted phylogenetic tree indicate species (*V. alginolyticus*; blue, *V. diabolicus;* red, *V. parahaemolyticus;* green).

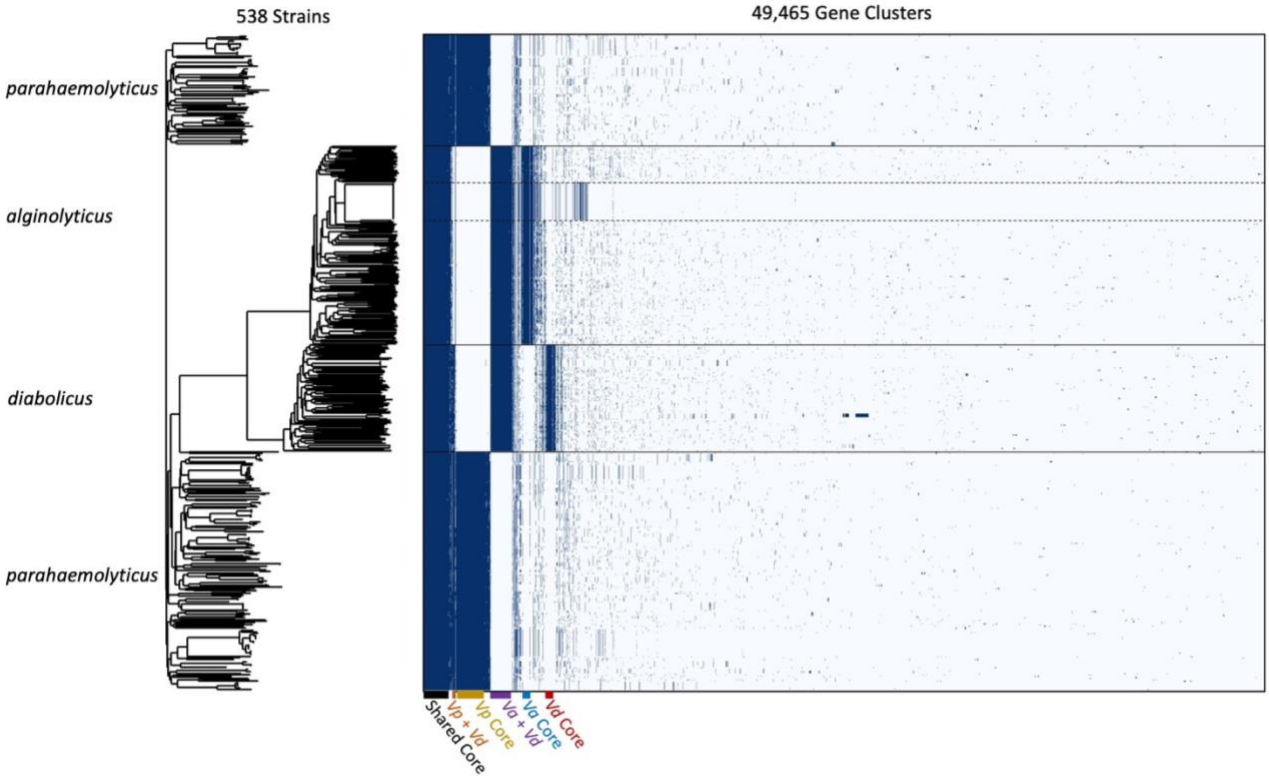


Figure 3.3. Combined pangenome for three closely related species *V. alginolyticus* (n=163), *V. diabolicus* (n=88), and *V. parahaemolyticus* (n=287) collected from various sources globally. A 95% identity cutoff of the predicted amino acid sequences was used, which resulted in an extremely diverse gene pool (49,465 unique gene clusters). Species are separated on the heatmap of gene cluster presence by solid black lines while the Kiel-Fjord ecotype found in *V. alginolyticus* from pipefish are separated by dashed black lines. Core gene clusters shared across all species, 2 species, or individual species are color-coded at the bottom.

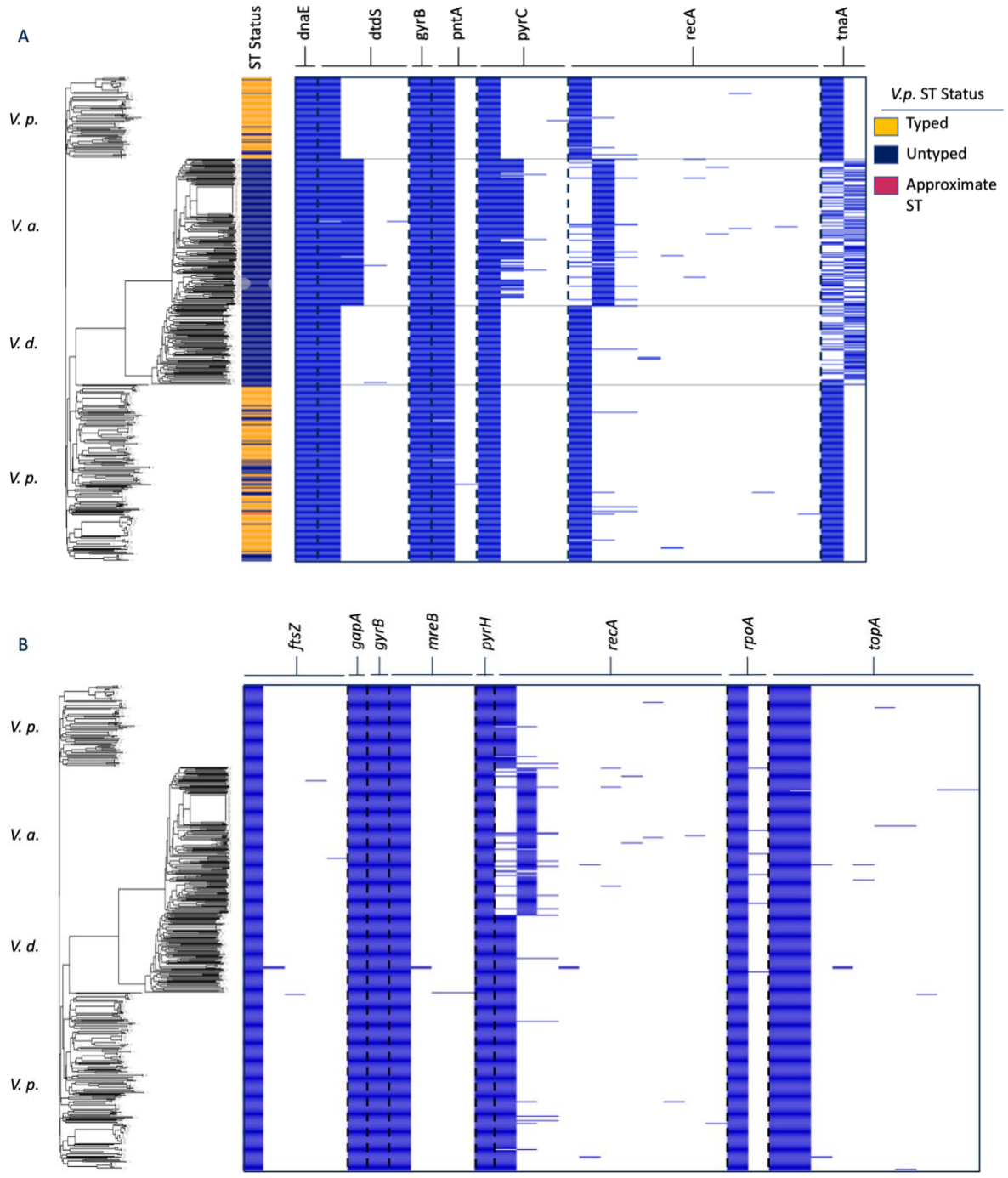


Figure 3.4. Heatmaps for presence (blue) or absence (white) of "housekeeping" genes used in MLST and MLSA schemes across *V. alginolyticus* (n=163), *V. diabolicus* (n=88), and *V. parahaemolyticus* (n=287) genomes collected from various sources globally. A) Heatmap for the genes with annotations matching the 7 loci *V. parahaemolyticus* MLST scheme.[43] B) Heatmap for the genes with annotations matching an 8 loci *Vibrio* spp. MLSA scheme.[16]

Figure 3.4 continued. Dashed black lines separate genes annotated to different loci while columns between dashed black lines include all gene variants with annotation to the labeled loci. Genes were segregated into allelic variants using a 95% identity cutoff of predicted amino acid sequences. The heatmaps and genome phylogeny were visualized and reproduced using Phandango.[47]

Tables

Table 3.1. Toxin genes from Toxin-Antitoxin (TA) systems in *V. alginolyticus* and *V. diabolicus* which were negatively associated with human, and/or positively associated with sea otter sources. The TA genes in *V. diabolicus* were significantly associated with sea otters based on the naïve p-value but were not significantly (n.s.) associated after Benjamini-Hochberg adjustment due to the relatively small sample size of *V. diabolicus* genomes, and the large number of genes tested.

Gene	Species	Source Comparison	Odds Ratio (95% CI)	Naïve p	B-H Adjust p
ccdB	*alginolyticus*	Human v Non-Human	0.23 (0.11, 0.5)	<0.001	0.028*
	alginolyticus	Otter v Non-Otter	43 (9.73, 190.07)	<0.001	<0.001*
	diabolicus	Otter v Non-Otter	4.5 (1.45, 13.98)	0.012	n.s
higB	*alginolyticus*	Human v Non-Human	0.19 (0.08, 0.42)	<0.001	0.016*
	alginolyticus	Otter v Non-Otter	30.67 (10.58, 88.93)	<0.001	<0.001*
	diabolicus	Otter v Non-Otter	7.78 (1.82, 33.21)	0.005	n.s.
parE group 1160	*alginolyticus*	Human v Non-Human	0.14 (0.06, 0.35)	<0.001	0.007*
	alginolyticus	Otter v Non-Otter	28.31 (10.53, 76.06)	<0.001	<0.001*
	diabolicus	Otter v Non-Otter	7.78 (1.82, 33.21)	0.005	n.s
parE group 6999	*alginolyticus*	Human v Non-Human	0.05 (0.01, 0.42)	<0.001	0.025*
	alginolyticus	Otter v Non-Otter	11.98 (4.07, 35.27)	<0.001	<0.001*
parE group 13663	*alginolyticus*	Human v Non-Human	0.18 (0.07, 0.47)	<0.001	0.036*
	alginolyticus	Otter v Non-Otter	15.3 (6.26, 37.38)	<0.001	<0.001*
yoeB	*alginolyticus*	Human v Non-Human	0.22 (0.10, 0.49)	<0.001	0.028*
	alginolyticus	Otter v Non-Otter	17.73 (7.04, 44.65)	<0.001	<0.001*
	diabolicus	Otter v Non-Otter	4.47 (1.36, 14.68)	0.020	n.s

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Peter J. Sebastian- Contributed to the conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, data visualization, and the writing and editing of the manuscript.

Cory Schlesener- Contributed to the conceptualization, data curation and visualization, formal analysis, methodology, and manuscript review.

Barbara A. Byrne- Contributed to the conceptualization, supervision of data generation and curation, investigation, resources, and manuscript review.

Melissa Miller- Contributed to the supervision of data generation and curation, investigation, resources, and manuscript review.

Bart C. Weimer- Contributed to the conceptualization, methodology, supervision of data generation and curation, project administration, resources, and manuscript review.

Christine K. Johnson- Contributed to the conceptualization, funding acquisition, supervision of data generation and curation, project administration, resources, and manuscript review.

Funding

P. Sebastian and partial research costs were supported by the National Institutes of Health under grant number T32OD011147. Additional funding was provided by an Oiled Wildlife Care Network grant.

Acknowledgments

We would like to thank the tremendous efforts of the individuals and organizations involved in sample collection and processing (by organization): University of California, Davis: Brian Bird, Danielle Carrade-Holt, Patricia Conrad, Tracey Goldstein, Stori Oates, Jennifer Siembieda, Brett Smith, Woutrina Smith, Robyn Stoddard; California Department of Fish and Wildlife: Francesca Batac; U.S. Geological Survey: Tim Tinker, Joe Tomoleoni; U.S. Fish and Wildlife Service: Joel Garlich-Miller, Verena Gill, Kristin Worman; Monterey Bay Aquarium: Mike Murray, Michelle Staedler, Marissa Young; Alaska Veterinary Pathology Services: Kathy Burek-Huntington, Natalie Rouse; Alaska SeaLife Center: Carrie Goertz, Natalie Hunter, Pam Tuomi; Washington Department of Fish and Wildlife: Katherine Haman; Sea Otter Savvy: Michelle Staedler. We would also like to thank previous collaborators with the Weimer Lab from Jessica Jones' team at the FDA whose previously published *V. parahaemolyticus* genomes were utilized in this study. Special thanks to Carol Huang from the Weimer Laboratory and the 100K Pathogen Project for your expertise and laboratory contributions to genome sequencing preparation. Thanks to past and present members of both the Weimer laboratory working group and the Epicenter for Disease Dynamics at University of California, Davis who provided support and scholarly

discussion especially Nicole Gardner, Megan Moriarty, Pranav Pandit, Jessica Sanchez, and Cristin Young. We would like to also honor the contributions of the late Jack Ames for his assistance during environmental sampling and his lasting contributions to sea otter conservation and the late Matthew Blake for fostering an inclusive community at the One Health Institute's Karen C. Drayer Wildlife Health Center.

Data Availability Statement

Whole genome sequence reads with corresponding metadata will be uploaded under bioproject PRJNA203445. Metadata for the 126 genomes from human cases and oysters is available in Miller et al. 2021.

CONCLUSION

The three chapters presented in this dissertation utilize a One Health approach to comparative genomic epidemiology of *Vibrio* spp. infections in sea otters and humans. These studies advance the field by informing on the relatedness between *Vibrio* spp. strains across humans, sea otters, and a shared coastal environment. In addition, this dissertation research provides epidemiological insights into *Vibrio* spp. gene functions relating to antimicrobial resistance (AMR), virulence, and host-adaptation that can benefit human and wildlife health.

In chapter 1, *Vibrio* spp. isolates from sea otters and the coastal ecosystem shared similar AMR genes, including genes associated with resistance to ß-lactams, tetracyclines, and less frequent detection of genes associated with resistance to fluoroquinolones, aminoglycosides, chloramphenicol, and sulfonamides. There was discordance between phenotypic and genomic AMR patterns that indicate current *Vibrio* spp. AMR prediction requires more than gene detection, but prediction can be improved through identification of alleles, as exemplified by ampicillin resistance in *V. parahaemolyticus*. The results of chapter 1 suggest that both phenotypic and genomic methods in tandem are currently needed to determine the underlying mechanisms of *Vibrio* spp. AMR.

Chapter 2 characterized the genetic markers of virulence of *Vibrio* spp. in sea otters, humans, bivalves, and the coastal environment. Virulence factors and genomic similarity were present in both sea otters and human *V. parahaemolyticus* infections indicating that strains related to human outbreaks also infect sea otters. Genomic similarity between human and sea otter *Vibrio* spp. strains implies both host species have shared transmission contact interfaces

for *Vibrio* spp. infections. Chapter 2 utilized pathological findings in necropsied sea otters to further show the potential of *Vibrio* spp. to cause similar disease in sea otters as in human infections.

A major discovery from the *Vibrio* spp. sequenced throughout this dissertation was the identification of *V. diabolicus* isolated from sea otters and their coastal ecosystems, which was misidentified as *V. alginolyticus* previously using clinical identification methods. In chapter 3, whole genome methods of species identification were used that could differentiate closely related *Vibrio* spp. species and Harveyi clade species incorrectly reported from public domain genomes. Whole genome sequencing methods utilized in chapter 3 can improve the identification of the closely related *V. alginolyticus*, *V. diabolicus*, and *V. parahaemolyticus* species. Compared to multilocus sequence typing, which uses limited loci, a genome-wide association approach was used to identify species-specific genetic markers for *V. alginolyticus* and *V. diabolicus* which can assist in creating quick and fine resolution species identification. Currently available virulence gene databases rely heavily on *V. parahaemolyticus* genes; thus, a genome-wide association approach identified genes associated with either human or sea otter hosts which did not rely on incomplete databases. Newly identified genes provide a tool for future investigations to discover host-adaptations of *V. alginolyticus* and *V. diabolicus* that improve sea otter and human colonization.

Current genomic methods of antimicrobial resistance and virulence detection are limited by the quality and completeness of available databases. The appropriateness of gene databases to predict antimicrobial resistance and virulence of less characterized bacterial species, like *Vibrio* spp., is often unknown. Together these three chapters highlight the limitations of gene

databases for inferring the phenotypes of antimicrobial resistance and virulence in *Vibrio* spp. Whole genome approaches used here improve recognition of the current discordance between phenotype and genotype. Genomic studies were also shown to be beneficial for investigation of allelic variation, synergy of co-occurring genes, and identification of novel markers. This dissertation provides proof of concept for combining the One Health approach with genomic epidemiology to characterize *Vibrio* spp. in sea otters and humans in their shared coastal ecosystem.