Comparative Genomic Analysis and Evaluation of Control Strategies for Hypervirulent *Aeromonas hydrophila* in Channel Catfish (*Ictalurus punctatus*)

by

Cody R. Rasmussen-Ivey

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 3rd, 2019

Keywords: *Aeromonas hydrophila*, Comparative Genomics, Quantification Methods

Approved by

Mark R. Liles, Chair, Professor of Biological Sciences
Leslie R. Goertzen, Associate Professor of Biological Sciences
Eric J. Peatman, Associate Professor of Fisheries and Allied Aquaculture
Jeffrey S. Terhune, Associate Professor of Fisheries and Allied Aquaculture
Todd D. Steury, Associate Professor of Forestry and Wildlife Sciences

Abstract

This dissertation is a reflection of the insight gained through the genetic analyses of a hypervirulent pathotype of the Gram-negative pathogen, *Aeromonas hydrophila*, (v*Ah*) in the context of enzymatic, challenge, growth, and population-level data. After characterizing this pathogen of warmwater fishes through disease challenges, by histological descriptions, and by associations between genotype and phenotype, genetic profiles unique to v*Ah* were evident. One significant example was the *myo*-inositol catabolism pathway, which provided an additional support for the use of biological control agents that aim to inhibit v*Ah* proliferation by reducing the abundance of this carbon source in the environment.

To quantify the resultant microbiome shifts of introducing these agents, foundational research was performed with the aim of developing a novel tool (provisionally referred to as 'quantitative polybacterial polymerase chain reaction' or 'qpPCR') to evaluate mixed cultures of bacteria at the species or subspecies level, while providing quantitative data on microbial abundance without the biases associated with culture-based and common culture-independent methods. While qpPCR was designed to be the first technique capable of defining the composition of any microbiome, unresolved biases associated with any method reliant on the polymerase chain reaction and/or with the purification of low abundance amplicons prevents the application of this method from being a practical replacement for existing microbial quantification methods. Therefore, the inclusion of this chapter is intended to serve as an experimental foundation for the application of the conserved and divergent sequences validated

within this research for use in future research based on this method that uses real-time sequencing.

Chapter one presents an overview of known virulence factors, amidst the phylogenetic reshuffling of what was considered the dominant research strain for *A. hydrophila*, a primer on the use of *Bacillus* spp. probiotics and the enzyme phytase in aquaculture, and a discussion of methods for the quantification of mixed cultures, with an emphasis on the biases surrounding these methods. Chapter two presents a genetic and phenotypic characterization of the pathogen v*Ah* within the species *A. hydrophila*. Chapter three describes the preliminary development of the qpPCR method and proposes an alternate approach to circumvent pitfalls identified through extensive experimental evaluation. Finally, chapter four describes experiments in which the enzyme phytase and the probiotic *Bacillus velezensis* AP193 were used as feed additives to reduce disease due to v*Ah* through reductions of the anti-nutrient *myo*-inositol. Collectively, this research has mediated a separation of knowledge by discussing what remaining literature on virulence factors is applicable within *A. hydrophila*, characterized a hypervirulent pathotype that has significant economic relevance to the aquaculture industry within the United States as well as other countries with farmed fishes, formed a foundation for the development of new quantitative methods for complex sample analyses, and explores the significance of shifts in gut microbiota based on the introduction of the phytase enzyme and the phytase-producing probiotic to collectively provide a deeper understanding of the effects these agents have when introduced to aquaculture systems.

Acknowledgments

Foremost, I would like to thank my advisor, Professor Mark Liles, for his unparalleled guidance and constant support. Throughout my Ph.D., Mark demonstrated exceptional integrity as a researcher and as a mentor. Next, I would like to thank Associate Professor Les Goertzen for his intellectual contributions in developing each stage of this dissertation as well as his openness in discussing the theoretical underpinnings of biology. I would also like to thank Associate Professor Jeff Terhune who masterfully integrated the results of our computational- and lab-based research with practical knowledge of the aquaculture field. In addition, I would like to thank Associate Professor Eric Peatman for enabling me to explore the practical applications of my findings in *A. hydrophila* biological control and for providing significant contributions towards this dissertation. Furthermore, I would like to thank Associate Professor Todd Steury for accepting the responsibility of serving as the outside reader for this dissertation and for his valuable perspective that led me to apply statistical models to understand the underlying nature of biological systems.

In addition to the faculty members listed above, I would also like to thank the past and present members of Professor Mark Liles' lab. First, I would like to thank our lab manager, Nancy Capps, who was always helpful and became a true friend over the course of my degree. Second, I would like to collectively thank the graduate students of the Liles' lab who were exemplary role models, mentors, and colleagues. Third, I would like to thank the numerous undergraduates who I mentored for their contributions and for their genuine fascination with science.

Next, I would like to thank the members of my family. First, I would like to thank my mother, Verna Ivey, for her tireless dedication to my education and the incredible sacrifices she made as a single mother. Second, I would like to thank my uncle, Edward Maxwell, for being a pillar of distinct moral fiber as well as for his support. Finally, I would like to thank my aunt, Ruby Maxwell, for her unconditional support, friendship, love, and guidance.

Table of Contents

List of Tables

List of Illustrations

List of Equations

List of Abbreviations

aerA	Aerolysin A
AHH1	Extracellular Heat-Labile Hemolysin
AHL	N-Hexanoyl-L-Homoserine Lactones
AICc	Akaike's Information Criterion Corrected for Small Sample Sizes
AL	Alabama
ANI	Average Nucleotide Identity
APC	Aerobic Plate Count
ARDA	Amplified Ribosomal DNA Restriction Analysis
BHI	Brain heart infusion
BHL	N-(Butanoyl)-L-Homoserine Lactones
BLAST	Basic Local Alignment Search Tool
Cas9	CRISPR-associated endonuclease
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CFU	Colony Forming Units
CS1	Common Sequence 1
CS2	Common Sequence 2
DePCR	Deconstructed-PCR
DGGE	Denaturing Gradient Gel Electrophoresis
DMSO	Dimethyl Sulfoxide

dNTPs	Deoxynucleotides
DOP-PCR	Degenerate Oligonucleotide Primed
dsDNA	Double-Stranded DNA
EMA	Ethidium Monoazide
epr	Extracellular protease
FACS	Fluorescence-Activated Cell Sorting
FCR	Food Conversion Ratio
FISH	Fluorescence in situ Hybridization
FTU	Phytase Units
gDNA	Genomic DNA
GOLD	Genomes OnLine Database (Joint Genome Institute)
Hcp	Hemolysin Coregulated Protein
hlyA	Hemolysin A
HTH	Helix-Turn-Helix
LCM	Laser-Capture Microdissection
M9I	Myo-Inositol Infused Agar
MALBAC	Multiple Annealing And Looping Based Amplification Cycles
MAS	Motile *Aeromonas* septicemia
MDA	Multiple Displacement Amplification
MS-222	Tricaine mesylate
MuMIn	Multi-Model Inference
ML	Maximum Likelihood
MS	Mississippi

NCBI	National Center for Biotechnology Information
NMDS	Non-Multidimensional Scaling
NMPDR	National Microbial Pathogen Data Resource
non-v*Ah*	*Aeromonas hydrophila* that are not associated with epidemic outbreaks
OD	Optical Density
OTU	Operational Taxonomic Unit
PATRIC	Pathogen Resource Integration Center
PCR	Polymerase Chain Reaction
pla	Phospholipase A1
plc	Phospholipase C/Lecithinase
PLFA	Phospholipid-Derived Fatty Acids
PMA	Propidium Monoazide
qPCR	Quantitative Polymerase Chain Reaction
qPCR	Real-Time Quantitative PCR
qpPCR	Quantitative Polybacterial PCR
QueD	Queuosine Biosynthesis
RAST	Rapid Annotations using Subsystems Technology
RISA	Ribosomal ribonucleic acid Intergenic Spacer Analysis
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
rtPCR	Real-Time PCR
RUBRIC	Read Until Basecall Reference-Informed Criteria
ser	Serine Protease

SNPs	Single Nucleotide Polymorphisms
SOLiD	Sequencing by Ligation
SPC	Standard Plate Count
SSCP	Single Strand Conformation Polymorphism
ST251	Sequence type 251
T2SS	Type II secretion system
T3SS	Type III secretion system
T6SS	Type VI secretion system
T-RFLP	Terminal Restriction Fragment Length Polymorphism
TagA	ToxR-regulated lipoprotein
TGGE	Temperature Gradient Gel Electrophoresis
TSA	Tryptic soy agar
TSB	Tryptic Soy Broth
USA	United States of America
USDA-ARS	United States Department of Agriculture Committee of the Aquatic Animal Health Research Unit
UV	Ultraviolet Light
v*Ah*	Hypervirulent Aeromonas hydrophila
VgrG	Valine-Glycine Repeat Protein G

Chapter I

Literature Review of the Virulence Factors of *Aeromonas hydrophila*, a Primer on the use of the Enzyme Phytase and the Genus *Bacillus* to Degrade Phytate/Phytic Acid, an Overview of *Bacillus* spp. as Biological Control Agents in Aquaculture, and a Review of Common Biases Associated with Microbiome Analyses

1. Abstract: Literature Review of the Virulence Factors of *Aeromonas hydrophila*.

The ubiquitous "jack-of-all-trades", *Aeromonas hydrophila*, is a freshwater, Gram-negative bacterial pathogen under revision in regard to its phylogenetic and functional affiliation with other aeromonads. While virulence factors are expectedly diverse across *A. hydrophila* strains and closely related species, our mechanistic knowledge of the vast majority of these factors is based on the molecular characterization of the strains *A. hydrophila* AH-3 and SSU, which were reclassified as *A. piscicola* AH-3 in 2009 and *A. dhakensis* SSU in 2013. Individually, these reclassifications raise important questions involving the applicability of previous research on *A. hydrophila* virulence mechanisms; however, this issue is exacerbated by a lack of genomic data on other research strains. Collectively, these changes represent a fundamental gap in the literature on *A. hydrophila* and confirm the necessity of biochemical,

molecular, and morphological techniques in the classification of research strains that are used as a foundation for future research. This review revisits what is known about virulence in *A. hydrophila* and the feasibility of using comparative genomics in light of this phylogenetic revision. Conflicting data between virulence factors, secretion systems, quorum sensing, and their effect on *A. hydrophila* pathogenicity appears to be an artifact of inappropriate taxonomic comparisons and/or be due to the fact that these properties are strain-specific. This review audits emerging data on dominant virulence factors that are present in both *A. dhakensis* and *A. hydrophila* in order to synthesize existing data with the aim of locating where future research is needed.

2. Introduction: Literature Review of the Virulence Factors of *Aeromonas hydrophila*.

The ubiquitous bacterium *Aeromonas hydrophila* is a freshwater, facultatively anaerobic, chemoorganoheterotroph [5] and the etiologic agent of disease in amphibians, birds, fishes, mammals, and reptiles, with the most common forms of disease being gastroenteritis, septicemia, and necrotizing fasciitis [6-9]. Virulence in *A. hydrophila* is multifactorial, with disease resulting from the production and/or secretion of virulence factors, such as adhesins, cytotoxins, hemolysins, lipases, and proteases as well as the capacity to form biofilms, use specific metabolic pathways, and mediate virulence factor expression through quorum sensing [10-13]. The majority of experimental studies on identifying virulence determinants in *Aeromonas* spp. have been performed in the strain *A. hydrophila* SSU, which was later recognized to be affiliated with *A. dhakensis* on the basis of ANI and phylogeny comparisons [14]. Adding confusion to this complexity, the literature on *A. hydrophila* is riddled with conflicting reports on the molecular determinants of virulence attributed to this species because of changes in classification and

problems stemming from misidentification [15,16]. The purpose of this review article is to provide an updated view on what is known about virulence factors in the aftermath of reclassification of *A. hydrophila* SSU.

In 2002 Huys *et al.* recognized that some diarrheal isolates, while closely related to *A. hydrophila*, show atypical metabolic activities for urocanic acid (+), L-fucose (-), and L-arabinose (-). On these bases, these strains were classified into a subspecies known as *A. hydrophila* subsp. *dhakensis* [17]. Then, in 2013, *A. hydrophila* subsp. *dhakensis* was recognized to be synonymous to *A. aquariorum* and both were combined under the name *A. dhakensis*, a species that is functionally divergent from *A. hydrophila*, based on multilocus phylogenetic analyses and phenotypic characteristics [18]. Studies on the virulence factors expressed by the diarrheal isolate SSU, previously considered to be affiliated to *A. hydrophila* and now know to be *A. dhakensis*, are regarded as the seminal literature on molecular pathogenesis of *Aeromonas* [14]. Given the turbulent nature of classification within *Aeromonas* spp., this review aims to clarify which virulence factors have been characterized within current members of *A. hydrophila* (**Table 1**) by auditing the body of knowledge on the molecular understanding of these genes so that future research can progress from a more solid foundation.

3. Regulation of *Aeromonas* Virulence Determinants.

Cascades of genetic regulation that lead to situational expression of virulence factors are known to occur in *Aeromonas* spp., but these interactions remain a relatively uncharted area of research in phylogenetically confirmed *A. hydrophila* strains. For example, outbreaks of *A. hydrophila* are generally thought to be linked with changes in host susceptibility caused by environmental changes, such as hypoxic conditions and excessive nitrite levels in farmed fish, as well as increases in temperature, which are linked with the production of virulence factors, such

as cytotoxins and hemolysins [9,19,20]. To exploit changes in host susceptibility due to increases in temperature, *Aeromonas* spp. virulence factors have also evolved temperature-dependent expression [21,22]. For example, clinical strains of *A. hydrophila* can grow at temperatures greater than the isolate's optimal growth temperature of 28°C [23]; however, when temperatures increase to 37°C, protease activity decreases and cytotoxin and hemolysin activity increases [24]. In contrast, environmental isolates are well adapted to low temperatures and can grow uninhibited at temperatures as low as 4°C, a temperature that restricts growth of clinical isolates [19]. Some of the better studied regulatory effects are the linkage between quorum sensing and biofilm formation which was shown to not only mediate the expression of virulence factors, but also regulate cell density [9,25,26]. In addition, while polar flagella in *A. hydrophila* are constitutively expressed, there are well-described regulators that trigger lateral flagella expression such as surface contact and viscosity [27,28]. Another class of regulatory effects includes the upregulation of virulence factors through lysogenic conversion; however, to-date no experimental data has been published on this phenomenon within *A. hydrophila*. Considering the broad effects that these regulatory factors have on disease, experimental studies that resolve these interactions are fundamental to the advancement of knowledge for the field of *A. hydrophila* as a whole. A review of known virulence factors and the respective regulatory effects that have been evaluated in *Aeromonas* spp. and are genetically present within *A. hydrophila* are presented in **Figure 1**.

4.1. Secretion Systems: Type II Secretion System and Effector Proteins.

The widely-conserved type II secretion system (T2SS) is present in all known members of *A. hydrophila* and is integral in the extracellular secretion of a wide array of virulence factors including aerolysin, amylases, DNases, and proteases [29-32]. In *A. dhakensis* SSU, the T2SS

secretes what is perhaps the most potent virulence factor; the aerolysin-related cytotoxic enterotoxin Act [33]. While genes for this and other virulence factors that interact with the T2SS are present in current members of *A. hydrophila*, the contribution of this system to virulence remains unquantified [34].

4.2. Secretion Systems: Type III Secretion System and Effector Proteins.

Found in higher frequency in clinical isolates than in aquatic isolates [32,35], the type III secretion system (T3SS) functions as a molecular needle, injecting effector toxins into host cells [36-38]. Although no studies have been performed in members of *A. hydrophila* with publically accessible genomic data, the T3SS has been shown in *Aeromonas* spp. to be co-regulated by contact with host cells, cytotoxic enterotoxin Act, DNA adenine methyltransferase, flagella, lipopolysaccharides, DNA methylation, temperature, calcium/magnesium levels, and quorum sensing while requiring effectors to have the appropriate secretion signal [39-44]. Because of its strong association with the export of virulence factors by many *Aeromonas* spp., the experimental manipulation of genes that encode for subunits of this secretion system, which resulted in attenuation of virulence in the reclassified *A. piscicola* AH-3 (formerly *A. hydrophila*), may also result in the attenuation of *A. hydrophila* [44-49]. At the same time, genetic heterogeneity may prevent the translation of this research. For example, calcium chelation promotes T3SS/AexT expression in *A. piscicola* AH-3 and in *A. salmonicida* JF2267, but these effects are absent in *A. salmonicida* A229, *A. salmonicida* A449 and in *A. dhakensis* SSU [40,41,44,48]. On a molecular level, *A. salmonicidia* JF2267 was shown to lose its plasmid, which contains the T3SS genes, at 25°C whereas *A. salmonicidia* A449 conversely increases transcription of T3SS genes between 25-28°C [41]. Therefore, while the same system appears, they are different on a procedural level.

Within *A. hydrophila*, numerous studies linked the T3SS and its effector proteins with virulence. In *A. hydrophila* AH-1, an isolate of blue gourami (*Trichopodus trichopterus*) with publicly available nucleotide data (whole genome is not available), insertional mutagenesis of *aopB* (T3SS translocator) and *aopD* (integral T3SS transmembrane component) causes a reduction in cytotoxicity and an increase in phagocytosis because the T3SS is no longer able to translocate effector proteins [50]. Similarly, in *A. dhakensis* SSU, T3SS genes have been linked with virulence that include the T3SS-associated exoenzyme effector (AexU), which increases host evasion, degrades host actin, and is independently lethal [38,51]**.** AcrH is a chaperone that complexes with AopB and AopD [52]; *acrH* mutants are predicted to have attenuated virulence. Contextually, the *aexU* and *acrH* genes are present in a minority of *A. hydrophila* and no experimental studies have been performed to establish their respective roles in virulence. While no experimental manipulations were performed, a subsequent study compared clinical and environmental isolates of *A. hydrophila*, showing that T3SS structural genes *aopB* and *ascV* are most abundant in *A. hydrophila* disease isolates [53], a link with virulence that is supported by the attenuation of virulence in *A. piscicola* AH-3 *ascV* mutants [45]. Collectively, these results appear to indicate that the T3SS is a strong contributing factor for virulence of *Aeromonas* spp.. However, genomic analyses of pathogenic *A. hydrophila* isolates indicate that alternate secretory mechanisms may also be critical for pathogenesis given that hypervirulent isolates of *A. hydrophila* that infect farmed fish lack T3SS core components [32,54].

Previously described in *A. salmonicida* [39], the ADP-ribosylating toxin AexT is present in ~90% of *Aeromonas* spp. that have a T3SS and when this T3SS effector is abrogated in *A. piscicola* AH-3, a slight attenuation of virulence has been observed based on virulence assays for cytotoxicity and phagocytosis as well as fish and mice challenges [48]. The *aexT*-like gene *aexU*

shows a stronger contribution to virulence, with *aexU* mutants having an LD_{50} of 60% using 2-3 times the dose of wild-type *A. dhakensis* SSU [48,55,56].

4.3. Secretion Systems: Type VI Secretion System and Effector Proteins.

The type VI secretion system (T6SS) functions analogously to a phage tail, allowing injection of virulence factors into host cells via valine glycine repeat G (VrgG) proteins and hemolysin-coregulated protein (Hcp), which functions as an antimicrobial pore-forming protein when secreted or as a structural protein [57]. In *A. dhakensis* SSU, the transcriptional regulator VasH and the helical transmembrane protein VasK are linked with secretion of Hcp, with *vasH* and *vasK* mutants resulting in decreased anti-phagocytic activity and attenuated virulence in a septicemic mouse model which serves as a line of evidence that the T6SS is involved in the manifestation of disease [58], but similar to the disparate results of the T3SS, the T6SS is not obligatory for *A. hydrophila* virulence. For example, some members of the newly described hypervirulent *A. hydrophila* pathotype of freshwater fishes have a complete T6SS while others retain only 4/13 core components [1,32]. With a distribution in 26 out of 37 strains listed as *A. hydrophila* in GenBank, the T6SS's role in virulence may be specific to the mode of infection with bacteria that contain a complete T6SS having greater antimicrobial activity, but at the cost of stimulating host defenses. In other bacteria, the T6SS also plays a role in biofilm formation, and evasion of the host immune system, but future research is needed to assess the role(s) of the T6SS within *A. hydrophila*.

5. Biofilm Formation.

Biofilms provide bacteria with resistance to antimicrobial agents and host defenses [26,59]. *Aeromonas* spp. evolved multiple regulatory mechanisms for biofilm formation that are

intimately linked with the production of virulence factors. The quorum sensing response regulator of the reclassified isolate *A. piscicola* A1 (formerly *A. hydrophila*), *ahyRI*, produces LuxRI homologs, N-(butanoyl)-L-homoserine lactones (BHL) and N-hexanoyl-L-homoserine lactones (AHL); autoinducers that regulate cell division [25]. In *A. dhakensis* SSU Δ*ahyRI* mutants, T6SS effectors Hcp and Vgr are unable to be secreted which results in decreased biofilm formation [60]. Interestingly, some strains transcribe *ahyRI* (e.g. *A. hydrophila* ATCC 7966), but lack AHL/BHLs, which may indicate an alternate function of *ahyRI* that has yet to be described [61]. Similarly, the recently-characterized autoregulatory two-component signal transduction system QseBC is a widely conserved system within *A. hydrophila* and was first described in *A. dhakensis* SSU, as mutants with an inactive response regulator (QseB) have reduced swimming and swarming motility, form thicker biofilms, and secrete fewer virulence factors, which leads to attenuation of virulence. When the gene *aha0701h* is overexpressed in Δ*qseB* mutants, biofilm formation decreases, presumably due to dysregulation of genes *fleN* (regulates flagellar number) and *vpsT* (transcriptional response regulator) [62,63].

6. Flagella and Pili.

A. hydrophila isolates produce lateral flagella for surface movement/swarming and polar flagella for movement in suspension. Polar flagella production has been studied within *A. piscicola* AH-3, with mutations in *flaAB*, *flaH*, *fliA*, *fliM*, *maf-1*, and *flrC* abolishing production of polar flagella and resulting in decreased adherence and biofilm formation [64]. Considering that flagellar glycosylation was shown to be linked with the ability to form biofilms as well as adhere to Hep-2 cells, it is important to mention that there are notable differences within *Aeromonas* species. In addition to having only a single lateral flagellin, polar and lateral flagella are

glycosylated in *A. piscicola* AH-3 whereas *A. hydrophila* AH-1 has two lateral flagellins and only the polar flagellum is glycosylated. When pseudaminic acid biosynthesis genes *pseB* and *pseI* were mutagenized, the result was an inability to produce both polar and lateral flagella in *A. piscicola* AH-3, but only affected polar flagella production in *A. hydrophila* AH-1. Therefore, lateral flagella production was unaffected in glycosylation negative *A. hydrophila* AH-1 mutants [65]. Similarly, in the diseased eel isolate *A. hydrophila* W (no genome submitted), mutations in *flgE, flgN, flhA, fliJ, flmB, lafK*, and *maf-5* result in loss of lateral flagella, which causes decreased motility, biofilm formation, and mucosal adherence [66]. While polar and lateral flagella transcriptional hierarchies, regulation, and contribution to virulence are well-described in other species, as the date of this publication, no member of *A. hydrophila* with a publicly accessible genome has undergone genetic manipulations to evaluate the contribution of polar or lateral flagella for virulence.

The *A. hydrophila* gene cluster *tapABCD* is responsible for type IV pilus biogenesis and is an integral part of the extracellular secretory pathway. To test for function, the *A. hydrophila* Ah65 (genome unavailable) gene *tapD* gene was used to successfully complement a strain of *Pseudomonas aeruginosa* that lacks PilD (an orthologue of TapD) [67]. Another type IV pilus is the bundle-forming pilus, which is encoded by *bfp* and acts an important internal colonization factor for multiple species of *Aeromonas* (*A. hydrophila* Ah65 was observed expressing both *bfp* and *tap*) [68,69]. Taken with the observation that TapD is required for secretion of virulence factors, such as aerolysin and proteases, these genes appear to be fundamental for pathogenicity.

7. Structural Proteins, Phospholipids, and Polysaccharides.

Capsules, O-antigens, and S-layer proteins provide mechanisms to evade host defenses. Within *Aeromonas* spp., capsules also show anti-phagocytic activity, increase resistance to the complement system, and increase adherence [70,71]. O-antigens are a class of structurally diverse lipopolysaccharides that act as colonization factors. At 20°C O-antigen is produced by *A. piscicola* AH-3, but not at 37°C, resulting an O-antigen-deficient strains that are unable to colonize hosts and have reduced expression of T3SS components [44,72]. Across *A. hydrophila*, eight distinct O-antigen gene clusters are present, with all epidemic strains isolated from channel catfish (*Ictalurus punctatus*) sharing a homologous O-antigen gene cluster [73]. In *A. hydrophila* TF7 (genomic data unavailable), the S-layer protein gene (*ahsA*) encodes an external paracrystalline layer that is lost upon insertional mutagenesis of *spsD* (S-protein secretion) [74]. Another study of S-layer proteins in five pathogenic human and eel isolates of *A. hydrophila* (A19, AH290, E37, E40, and TW1; genomic data unavailable) shows that serogroups of *A. hydrophila* other than O:11 contain S-layer proteins O:14 and O:81 [75].

8. Hemolysins.

Hemolysins are a diverse group of multifunctional enzymes that play a central role in *A. hydrophila* pathogenesis [76,77]. The extracellular heat-labile hemolysin (AHH1) is the most abundant of several widely distributed hemolysins (AerA, AHH1, AhyA, and Asa1), with the most cytotoxic genotype being a synergistic combination of *aerA* and *ahh1* [78,79]. In *A. media* A6 (formerly *A. hydrophila*) Aerolysin A (*aerA*) and Hemolysin A (*hlyA*) comprise another two-component hemolytic system in which virulence is attenuated only when both *hlyA* and *aerA* activity is abolished [80]. In *A. dhakensis* SSU, the iron dependent, *fur* and *gidA*-regulated, enterotoxin Act is the most cytotoxic virulence factor of and a core gene within *A. hydrophila*,

with studies in *A. dhakensis* SSU demonstrating that Act induces multiple effects including hemolytic, cytotonic, and cytotoxic activities, but unlike other virulence factors exported via the T3SS or T6SS, Act is exported through the T2SS [47,81-85].

9. Collagenase, Serine Protease, Metalloprotease, Enolase, and Lipase.

A. hydrophila spp. express diverse degradative enzymes that can contribute to virulence including collagenase, elastase, enolase, lipase, metalloprotease, and serine protease. *A. piscicola* AH-3 contains a collagenase, which has sequence similarity to the open reading frame AHA_0517 of *A. hydrophila* ATCC 7966^{T}, and was shown to be cytotoxic to Vero cells, with loss of this enzyme resulting in a 5-15% increase in cell viability; however, this mutation did not result in complete reduction of cytotoxicity [86]. The *ahpAB* genes of *A. hydrophila* AG2 (genomic data unavailable) produce potent virulence factors: an extracellular protease that is not essential for virulence, but is present in the most virulent pathotypes ($AerA^{+}Alt^{+}Ahp^{+}$) along with a secreted elastase with caseinolytic and elastolytic activity that correlates with an LD_{50} 100 times more virulent than $AhpB^{-}$ mutants when assayed in rainbow trout (*Oncorhynchus mykiss*) [87-89]. Another extracellular protease (*epr*) was discovered in the soft-shell turtle isolate *A. hydrophila* AH1 and found to be present in the most common pathotype in diseased fishes ($Aer^{+}Alt^{+}Act^{+}EprCAI^{+}Ahp^{+}$) [90,91]. In the rainbow trout isolate *A. hydrophila* B32, a novel serine protease (*ser*) was found that exhibits cytotoxic properties and is thermostable, both of which are characteristics that differentiate this protease from known *A. hydrophila* α-hemolysins and β-hemolysins [92]. While four times less active than serine protease, the virulent *A. hydrophila* EO63 (genomic data unavailable) was shown to produce a thermostable metalloprotease with enzymatic activity on casein and elastin, an optimal pH of 8.0, and an LD_{50} of 3.5 μg/g [93].

Enolase, a secreted and surface-expressed glycolytic enzyme, was identified as a virulence factor in *A. dhakensis* SSU, based on binding to human plasminogen which leads to production of plasmin (degrades blood plasma proteins), with previous reports showing that enolase functions as a heat-shock protein and a regulator of transcription by binding host chromatin/cytoskeletal structures as well as being necessary for viability [94].

In general, lipases have diverse functions, but are linked with virulence in numerous pathogens [95]. An extracellular lipase (EC3.1.1.3) is produced by *A. piscicola* AH-3 (formerly *A. hydrophila*); however, the link between virulence and this gene is speculative in *A. hydrophila* [96]. Conversely, the heat-labile lipase Alt and the heat-stable lipase Ast are important cytotonic enterotoxins in the pathogenicity of *A. dhakensis* SSU, with both being able to cause significant fluid secretion, with only the previously described cytotoxic enterotoxin Act having a greater effect on fluid secretion [37,89]. Based on comparative genomics, Alt and Ast are core elements of *A. hydrophila*; however, no experiments have been performed within existing members of this species to characterize these toxins. Two additional lipases, phospholipase A1 (*pla*) and phospholipase C (*plc*), were explored in *A. piscicola* AH-3, with the finding that *pla* lacks a significant effect on virulence while *plc* (lecithinase) was cytotoxic and has LD_{50} values 10 times more virulent than phospholipase C-deficient mutants [97].

10. Other Virulence Factors.

The range of virulence factors encoded by *A. hydrophila* includes adherence proteins, catalysts, nucleases, and toxins that may be expressed differently depending upon the respective environment. The role of the adhesin *minD* in virulence is its ability to mediate mucosal adherence, increase biofilm formation, and facilitate cell division as well as motility [98]. The

enzyme 5-enolpyruvylshikimate 3-phosphate synthase (encoded by *aroA*) is required for folate availability and contributes to *A. hydrophila* AG2 (genomic data unavailable) viability in intraperitoneally injected rainbow trout, with *aroA* mutants no longer recoverable from fish internal organs because environmentally-derived folate is scarce [99]. Another element of host evasion is the nuclease encoded by the *ahn* gene of *A. hydrophila* J-1 which shows no significant change in hemolytic activity or growth *in vitro*; however, when Δ*ahn* mutants are introduced into fish and mice models, virulence is attenuated [100]. Another conserved gene of *A. hydrophila* that has only been characterized in *A. dhakensis* SSU is *vacB,* which encodes RNase R; an exoribonuclease with multiple functions that include permitting growth at 4°C (cold-shock protein) and supporting motility. Isogenic mutants of RNase R show a 70% attenuation in virulence [101].

Another virulence factor that is shown to increase host evasion for *A. dhakensis* SSU is the pore-forming RTX toxin RtxA that requires contact with host cells and is regulated by the *rtxACHBDE* operon so that production of RtxA coincides with regulation of other cytotoxins, such as aerolysins and hemolysins and acts to covalently cross-link host cytoskeletal-actin, resulting in host cells having a rounded phenotype that leads to apoptosis [102,103]. Another conserved virulence factor across all *A. hydrophila* strains in GenBank is the ToxR-regulated lipoprotein (TagA) of *A. dhakensis* SSU, which cleaves the complement C1-esterase inhibitor, thereby increasing serum resistance and decreasing erythrocyte lysis [104].

11. The Role of Horizontal Genetic Transfer in Virulence.

The introduction of virulence factors as well as their effects on the alternate regulation within *Aeromonas* spp. is a recurring theme of crucial importance, yet these elements remain

understudied. Plasmids are a confirmed source of multidrug resistance in *Aeromonas* spp. and have been shown to have the potential to be conjugally transferred between known human pathogens, such as *Acinetobacter baumannii* AYE and *A. hydrophila* [105]. In addition to being used as a "molecular map" to identify ancestral lineages, prophage that contain putative *cis*-acting elements and *trans*-acting factors were found to be conserved within hypervirulent strains of *A. hydrophila* which strongly implies that the differential regulation of virulence factors (and therefore the dramatic increase in virulence) may be caused by the lysogenic conversion of this conserved *A. hydrophila* lineage by these mobile genetic elements [54]. Of note, while *A. dhakensis* SSU contains the majority of virulence factors that are present within confirmed members of *A. hydrophila*, all other members of *A. dhakensis* with fully sequenced genomes appear to lack these genes **(Table 1)**. Results of numerous core genome phylogenies and average nucleotide identity analyses support the grouping of *A. hydrophila* SSU within *A. dhakensis* [1,15,16], if virulence factors are introduced or controlled by mobile genetic elements, then taxonomy and functionality demand separate analyses.

12. Discussion of *Aeromonas* spp. Virulence Factors.

In *Aeromonas* spp., as with all pathogens, disease is the result of complex molecular interactions between bacterium, environment, and host; however, the literature on *A. hydrophila* remains limited by the lack of experimental data on validated members of *A. hydrophila*. While there numerous virulence factors shared between members of *A. hydrophila*, *A. dhakensis* SSU, and *A. piscicola* AH-3 there are also key examples in the literature that show conflicting data between virulence factors, secretion systems, quorum sensing, and their effect on pathogenicity. This inconsistency is illustrated by the highly virulent catfish isolate *A. hydrophila* ML09-119

that acts as a primary pathogen when other members of this species act as secondary pathogens [106,107]. With that in mind, the thorough research conducted on *A. dhakensis* SSU by researchers, such as Dr. Ashok Chopra, still holds relevance for *A. hydrophila*, but future research should be mindful of the phylogenetic reclassification for strains AH-3 and SSU and that there may be significant differences in the molecular determinants of virulence for *A. hydrophila.*

To compare isolates of *A. hydrophila*, biochemical, morphological, and molecular techniques are required [16,108,109]. As of 2016, few strains exist that have enough supporting data to facilitate comparative studies. There are many sources of uncertainty when comparing *A. hydrophila* strains, including genetic heterogeneity, the lack of natural models of infection, and reclassification of bacterial strains as new data emerges. Future research should aim to couple typing techniques (e.g. genome sequencing) with experimental data on virulence determinants so that there is a clear phylogenetic context for these studies.

When considering known virulence factors, the definitive biological separation of *A. hydrophila*, *A. piscicola* AH-3, and *A. dhakensis* SSU has yet to be established. In *A. hydrophila* as in these other species, disease is the result of a molecular symphony, with each virulence factor contributing to a cumulative effect **(Figure 1)**. Research studying novel virulence factors and regulatory effects will help unveil the determinants that allow for infection and what differentiates *A. hydrophila* from other aeromonads. To better understand *A. hydrophila* pathogenesis it is imperative that future research develops natural models of infection, assesses the role of mobile genetic elements in virulence, and quantifies the interplay between virulence factors and host response in concert with molecular genetic approaches.

13. Phytase and *Bacillus* spp. in the Degradation of Phytate/Phytic Acid.

The presence of phytic acid or the salt phytate in plant-based animal feeds is considered an antinutrient because of its lack of bioavailable phosphorous. Upon exposure to the enzyme phytase, inorganic phosphorus (PO_4^{3-}) is released, which increases bioavailability for the host as well as for microbial life. While the enzyme phytase is widely used, the source of this enzyme influences its specific activity. For example, the bacterial isolated phytase from *Escherichia coli* has an enzymatic affinity for position six of the inositol ring (6-phytase) whereas the fungal isolate of phytase from *Aspergillus niger* has an enzymatic affinity for the third phosphorous group of inositol (3-phytase) [110]. To date, significant implementation of this enzyme has been evaluated to have beneficial effects in numerous aquaculture species, including, but not limited to, Nile tilapia (*Oreochromis niloticus*), carp (*Cyprinus carpio*), rainbow trout (*Oncorhynchus mykiss*), and channel catfish (*Ictalurus punctatus*) [111-115]. Furthermore, phytase 'superdosing' (e.g. 500-5,000 phytase units per kilogram [FTU/kg]) stemmed from these positive results, with the aim of rapid phosphohydrolsis of phytate/phytic acid [115-121]. In channel catfish, the 3-phytase enzyme has been studied extensively for its benefit of increasing food conversion ratios (FCR), its potential to replace dicalcium phosphate supplementation, and its correlation with increased bioavailable zinc, manganese, and phosphorous [112,122-124]. Similarly, although no phytase biosynthetic gene cluster is present within *Bacillus* spp., phytase activity was demonstrated through application to soy-based fish feed, resulting in reduced eutrophication of ponds, as evidenced by reductions in total phosphorus (19%) and in total nitrogen (43%), leading to substantial increased in growth [4]. While the use of phytase as a food additive appears positive, there is a general need for cataloging the effects of this introduction on a system-wide scale.

14. Overview of *Bacillus* spp. as Biological Control Agents in Aquaculture.

Probiotics within the genus *Bacillus* have shown striking results in both laboratory tests and clinical trials in the prevention and mediation of disease in animal and plant species [4,125,126]. These encouraging results have led to a better understanding of their antagonistic activity against pathogens and their often beneficial host interactions [4,125,127,128], the adoption of use as prophylactic and as therapeutic agents [129,130], and rapid commercialization in the global market [131]; however, this rapid acceptance is also the source of dysregulation within the United States, which has led to inaccurately labeled probiotics and incomplete studies in probiotic therapy[132]. While the breadth topics within this subject are beyond the thematic scope of this dissertation, this section aims to highlight notable benefits of using *Bacillus* spp. as probiotics as well as nuances within this field, the status of regulation within the United States, and the outlook of research focused on developing *Bacillus* spp. as probiotic agents.

The official definition of a probiotic (pro, meaning "for"; and bios, meaning "life"), according to the World Health Organization (WHO), is a "[l]ive microorganism[s] that when being administered in [an] appropriate dose…confer[s] benefit[s] of health to the receiver." In human consumption, *Bacillus* spp. probiotics are often delivered in dairy; such as in chitosan-coated alginate beads of yogurt or in kefir (derived from the Turkish word "Keyif"; "good feeling") [15], in pill form (lyophilized or encapsulated spores) [4-6], and in prebiotic fibers (dietary supplements that promote the colonization of the probiotic); however, in aquaculture, the primary method of delivery is through amending *Bacillus* spp. spores directly onto feed [4,133-135].

The founding concept of probiotics - that microbes modulate health - was conceived in the early 1900s by Ilya Ilyich Mechnikov who is generally credited with linking the responses of the human immune system with microbial activity. Since that time, the microbiomes have been accepted as integral to the development and health of their respective host organism, often

having measurable effects on development. For example, the probiotic *Bacillus velezensis* AP193 was recently shown to confer improved nutrient availability by degrading the anti-nutrient phytate/phytic acid, resulting in 32–40% increased growth in channel catfish (10 week study in ponds) [4]. Similarly, aquatic pathogens that include *Aeromonas hydrophila*, *Edwardsiella ictaluri*, *Edwardsiella tarda*, *Flavobacterium columnare*, *Saprolegnia ferax*, *Streptococcus iniae*, and *Yersinia ruckeri* are inhibited by numerous species within *Bacillus* spp. [126]. Building on our understanding of microbial activity, with over 140 journal articles published on this strain, the *B. velezensis* strain FZB42 originally isolated from the rhizosphere of a beet is considered the prototype for plant-growth-promoting and biocontrol rhizobacterial species and a prime example of the applicability of mechanistic studies [136,137].

Understanding the reasons that support a result, such as the antagonistic activity seen against common aquaculture pathogens, is fundamental in the long-term incorporation of these agents so that as disease-associated microbes inevitably become resistant to a probiotic's arsenal, alternate therapeutic agents/strategies could be applied with guidance. In a parallel example with antibiotic therapy, the effects of penicillin were discovered in 1897; however, scientific breakthroughs that led to understanding the mechanism of action did not occur until the mid-1900s. Initially, antibiotic therapy appeared infallible, but antibiotic resistance increased as dosing became prevalent [138]. With the continued misuse of antibiotics, selection had time to act, resulting in a putative rise in prevalence of microbes that are resistant to the antibiotics that were once used to treat these pathogens. The lesson from this example is that there are always underlying mechanisms that selection acts on which may render a therapeutic strategy ineffective. Therefore, continued research towards understanding these processes is imperative so that the development of novel treatments outpaces the development of resistance.

15. Review of Polybacterial Sequence Analysis and Quantification Methods.

Without exception, every culture-dependent and culture independent method that is used to either identify or quantify bacteria, contains biases that have sculpted our perception of microbial life. Culture-based methods are relatively consistent in that quantification of microbial samples requires growth of bacterial cells under the conditions provided for growth – the very principle that is relied on for analysis is also the source of bias. At the same time, because of the specificity necessary for bacterial growth and the range of applications that culture-based methods are applied, this approach has become one of the most diverse methods in microbiology, when considering the different media types and growth conditions established. While these traditional culture-based methods maintain a meaningful role in microbiological assessment, the advent of culture-independent methods began around the early 1990's and has continued to grow in both adoption of existing techniques by researchers as well as in the number of techniques developed to address new questions or address biases within the current 'gold standard' methods. This review encompasses the prominent methods of microbial quantification, with a specific emphasis on the biases of culture-dependent and culture-independent methods.

15.1. Culture-Dependent Methods.

The main biases of culture-dependent methods are the very same mechanisms that enables these tools to produce results. Because of the intrinsic nature of these biases, bacteria that are most readily quantified are also those that best meet the growth requirements of the specific media used for culturing, resulting in potentially skewed representations of abundance, unless the media is an extract of the source of interest and then there are other abiotic factors to consider. In

effect, culture-based methods are prone to reveal diversity that is more reflective of metabolic affinity or antibiotic resistance or simply generate isolates that are more adapted to laboratory conditions instead of their source of isolation. In general, culture-dependent methods are not sufficient discriminatory tools for resolving most microbiological challenges because they do not provide an accurate representation of the microbial populations from the source of isolation [139]. The core method for culture-based quantification is commonly referred to as a plate count, an aerobic plate count (APC), or a standard plate count (SPC). To quantify the microbiota, this method relies on serial diluting a sample onto petri dishes that contain one or more out of the hundreds of types of growth media available. Once plated, results from this method are a function of the growth rate of bacteria within the sample, taking an average of 8-24 hours before visible colonies form and counts may be counted, with acceptable counts ranging range from 30-300 colony forming units (CFU) per plate. The general types of culture media include:

- General culture media - provides amino acids; a carbon source; nitrogen; and water (e.g. tryptic soy agar).
- Minimal media - provides a low-energy carbon source; salts; and water (e.g. minimal salts media).
- Selective media - may contain a range of compounds that allows for the specific growth of a certain microbe(s) (e.g. MacConkey agar selects for Gram-positive bacteria).
- Differential media, which uses indicators to distinguish microbes (e.g. sheep's blood agar can be used to differentiate microbes based on hemolytic activity).

In addition to the previously described general media types, surveying microbial communities may require extracting nutrients from the sample's source of isolation to create more natural growth conditions. The general process for this is sterilization, either by filter or by

heat (e.g. in an autoclave), then resultant colonies will be produced only from bacteria added to the sample; however, filtering may remove particulates that are significant to the sample and heat may denature macromolecules or cause release of intracellular compounds. While removal of particulates may be undesirable, if using filtration to sterilize a sample, 0.2-0.22 micron filters are necessary to remove bacterial cells (0.45 micron filters were originally thought to do this until researchers found *Brevundimonas diminuta* penetrate the filter); however, this pore size is insufficient to separate bacteriophage from the sample [140-142].

In general, to survey a specific microbe or set of microbes, the synthesis of artificial media requires previous knowledge of the metabolic demands of the target microbes(s). In effect, the type of media used should match the research question being asked. For example, if a researcher was interested in solely identifying if the hypervirulent pathotype of *Aeromonas hydrophila* is present, then the researcher may use the minimal media, such as M9 media, which contains *myo*-inositol as the sole carbon source; however, if the researcher was interested in seeing if there may be co-infection between the hypervirulent *A. hydrophila* and a novel species of *A. veronii*, media that sustains both bacterial species is required, such as the more general media tryptic soy agar (TSA) [143]. Beyond metabolic requirements, information about oxygen, pH, and temperature tolerance as well as optimums are necessary pieces of information so that when designing conditions that permit growth *in vitro*, unless specifically designed to maximize growth rate, these conditions should aim to emulate the source of isolation or the target of interest.

With respect to metabolic biases, if an *r*-selected bacterial species (referencing the *r*/*K* selection theory) is grown under permissive conditions, these bacteria will have rapid metabolic activity, and reproduce quickly because, in general, these types of microbes use a broad range of nutrient sources. In contrast, bacteria that are considered *K*-selected (again referencing the *r*/*K*

selection theory) tend to be slow growing, reproduce at a lower rate, and have more specialized metabolic requirements and are therefore more difficult to quantify with culture-dependent methods [144]. These inherently different evolutionary approaches make quantifying samples that contain both *r*- and *K*-selected microbes inaccurate, if not impossible.

Expanding on the list of difficult to quantify bacteria, there are also microbes that fall with a phenomenon known as "The Great Plate Count Anomaly", which refers to microbial cells that can be viewed under a microscope and appear to be viable, but when culture attempts are made, these cells do not replicate. As such, these types of cells fall under the term 'viable, but not culturable', a term that was coined by Staley and Konopka in 1985 [145,146]. Note that this phenomenon is often referred to by a common misnomer that suggests that the target microbe is 'unculturable'. Although the immediate inability to culture a microbe is accurate, the inabilities of today are in not an absolute indicator that future efforts will not succeed. Instead, the more likely scenario is that there is insufficient information on the biological needs of the microbe that would permit culturing the bacteria of interest under artificial conditions. Depending on the sample, quantifying slow-growing microbes may appear to fit into this category because plate counts rely on visual space between colonies and, if the sample contains microbes that grow at a faster rate; disperse into large colonies; or are motile, quantification of slower-growing microbes may not be possible because these colonies may overlap. Therefore, the inability to transition from environmental to laboratory conditions may represent a significant source of bias well as a potential origin for confounding information within the vast amount of literature that relied on culture-based methods [147].

15.2. Culture-Independent Methods.

Expanding upon the diverse list of culture-based methods, culture-independent methods combine technologies from a host of fields to address equally diverse research questions and methodological biases with previously unattainable accuracy and efficiency. While the number of culture-independent methods is ever growing, several methods have risen to widespread use in the field of microbial ecology, such as 'real-time' quantitative polymerase chain reaction (real-time qPCR); cloned and direct sequencing of regions within the 16S (a sedimentation coefficient referred to as Svedberg that describes 16 x 10^{-13} seconds) ribosomal ribonucleic acid (rRNA) gene; microbiome sequencing; flow cytometry; various forms of spectroscopy; techniques that use gel electrophoresis (e.g. denaturing gradient gel electrophoresis [DGGE], temperature gradient gel electrophoresis [TGGE], single strand conformation polymorphism [SSCP], terminal restriction fragment length polymorphism [T-RFLP], rRNA intergenic spacer analysis [RISA], amplified ribosomal DNA (deoxyribonucleic acid) restriction analysis [ARDA], and phospholipid-derived fatty acids [PLFA]); fluorescence *in situ* hybridization [25]; and DNA microarrays. These methods all fill a specific niche; however, when used for microbial quantification and taxonomic resolution, each method fails to accurately capture both elements with meaningful levels of resolution.

15.3. Sample Storage Biases.

Sample storage represents one of the most formidable sources of bias when attempting to maintain a representative sample over extended periods of time. Ideally, genomic DNA (gDNA) has been purified of cellular byproducts, including destructive agents (e.g. nucleases), and chemical inhibitors (e.g. 2-mercaptoethanol, Ca2+, EGTA, SDS, and iodoacetate) have been included to prevent residual degradation. Furthermore, samples should be kept at -25°C to reduce

enzymatic degradation, out of ultraviolet light to reduce dimerization, out of solutions with high salt content to reduce precipitation, and out of both high and low pH solutions that would denature or hydrolyze the phosphodiester backbone of DNA [148-150]. After storage, before any additional analyses are performed, the next analytical process must be quantification.

15.4. DNA Quantification.

Like other methods discussed in this review, DNA quantification can have significant downstream effects by skewing inputs into subsequent methods; however, the notable distinctions of this method are that the main sources of error are from improperly calibrated standards, subsamples that are not representative of the sample and a lack of consistent sensitivity. Due to different efficiencies of quantifying DNA of different fragment lengths and compositions, there is presently no singular DNA quantification method that stands out as the leader in terms of accuracy, ease of use, and applicability [2], although PicoGreen has been referred to as the 'gold standard' [151]. Common approaches to quantify DNA include the use of intercalating dyes (QuBit, SYBR Green), real-time quantitative PCR (qPCR), titration (droplet digital PCR [ddPCR], PicoGreen, qPCR, and QuBit), and ultraviolet light [135] absorption (Nanodrop). From these approaches, the most commonly used DNA quantification methods employ the principle of titration: evaluate a standard or a set of standards to estimate the concentration of DNA (or other molecule of interest [referred to as the titrant]) within a sample. Titration-based methods subject the standard(s) to the same conditions as the sample; therefore the analytical method is calibrated for sample-specific variance, resulting in a more accurate measurement of the sample. For spectrofluorimetric methods that quantify double-stranded DNA (dsDNA), such as PicoGreen and QuBit, the average size of molecule measured is used estimate

the sample-wide concentration, with a sensitivity that ranges from picograms to nanograms and distinguishes DNA from RNA, protein, and other contaminants [151]. Similar in principle to the other titration methods that use average molecular size to estimate sample concentration, qPCR measures DNA by evaluating the abundance of a target sequence, which allows increased sensitivity (0.1-1.0 picograms) and estimates of DNA fragmentation [152]. Unlike the previous methods, the relatively new ddPCR library titration method does not rely on a bioanalyzer and, instead of taking an average of the molecular size, this method performs a count of the molecules per microliter, resulting in estimates of molarity that are accurate to the level of femtomolar per microliter [2]. Selection of the most applicable DNA quantification method will not only help ensure that both appropriate amounts of DNA are used, but also that the right downstream analytical method is selected.

15.5. PCR.

The polymerase chain reaction is a series of controlled temperature fluctuations that cause physical forces, coupled with specific reagents, to promote the biological process of replication *in vitro*. Used across numerous fields for an array of applications, the primary steps of PCR are relatively consistent and involve: a mixture of deoxynucleotides (dNTPs), primers (~20 bases of RNA or DNA), heat-stable polymerase, water/buffer, and double-stranded (dsDNA) being denatured at 94-95°C, followed by an annealing stage where primers bind to the complementary site within the dsDNA at 50-58°C, and an extension stage, where replication occurs, at 72°C. These three primary stages repeat for two linear cycles and are typically followed by an average of 30-50 cycles where exponential amplification occurs until the reaction is stopped or any reagent is depleted. The core biases of PCR are the result of

differential/inappropriate amplification because of primer/template binding properties, or the result of stochastic events. The primary issues with PCR are described below.

15.6. PCR: Template Bias.

PCR bias within multi-primer or multi-template reactions, is the result of compositional differences that translate to physiochemical inconsistencies and skewed amplicon production. Hypothetically, amplification will occur at the same rate if the nucleotide content (often referred to as G+C content characterized by the triple bonds formed by these nucleotides, which alters the strength of the complex) and the order is consistent across both template and amplicon. As multi-template interactions begin to diverge from this hypothetical mixture of excess primers to template, the resultant reaction between primers and templates also diverges. While low primer concentration is a dominant problem, the most significant stochastic effects have been shown to occur when the template concentration was low (<0.1 ng), resulting in significant differences in sample distributions. To compensate for this divergence, adjusting reaction conditions (e.g. melting temperature [T_m]) is necessary to maintain proportionally variant hybridization/amplification. If annealing conditions are restrictive (e.g. high $T_{m\ or}$ primer/template mismatches), then hybridization will not readily occur and the propensity for spontaneous amplification will increase.

15.7. PCR: Annealing Temperature and Duration.

When selecting favorable conditions for amplification of template DNA, one of the most immediate, yet overlooked biases comes from using generalized reaction conditions with suboptimal temperatures and/or inordinate annealing durations. PCR is an artificial process that

relies on temperature, instead of being driven by cellular processes, to create favorable conditions where primers will anneal to their respective target(s). As either of these conditions deviates from mimicking replicative processes, the probability of forming aberrant products increases. While calculations for annealing temperature have been adopted as a mandatory formulaic process of designing primer pairs as well as PCR temperatures, annealing duration has also been shown to serve a significant role in regions of DNA that are rich in guanine and cytosine (G+C). For these regions of >70% G+C content, the window for optimal annealing conditions is much smaller than in regions where nucleotides tend to approach more average abundances. Translating this information to PCR inputs, researchers showed that a region with ~78% G+C content will have an optimal annealing time of 3-6 seconds. If annealing time is set to 10 seconds, then the resultant amplicon length becomes more variable, forming a smear when imaged on an agarose gel because of variable amplicon length [153]. Although this example only highlights the need to properly calculate and validate PCR reaction conditions, the addition of betaine as well as dimethyl sulfoxide (DMSO) has been repeatedly shown to improve amplification at sites with high G+C content as well as in multiplexed reactions [154-157].

15.8. PCR: Reaching a 1:1 Ratio of Primer to Template.

Another factor that has been shown to skew amplification, is the tendency for primer/template pairs to reach an equal ratio. As the number of templates produced by PCR increases exponentially with every round of amplification and the amount of primers decreases proportionally, the presumption that the concentration of primers is at a level where templates are saturated becomes less probable, resulting in artifact (sequences that are not existent at the start of PCR and result from spurious nucleotide polymerization) formation during extended

thermocycling [158]. To improve the probability of uniform amplification, the number of PCR cycles should be limited so that amplicon production occurs at a collinear rate across templates and primers.

15.9. PCR: Heteroduplex Formation.

During the annealing phase of PCR, duplexes (one primer bound to a single DNA template) and homoduplexes (complementary gDNA or a complementary primer bound to a single DNA template) are normal formations; however, heteroduplexes (cross-hybridized heterologous sequences) may also form if conditions permit. If heteroduplexes are present, downstream methods that rely on cloning may have a persistent problem when the host mismatch repair system 'corrects' the heteroduplex because the cloning vector's repair enzyme cannot identify the parental strand via methylation, so, the presence of a homoduplex from either of the heterologous sequences is the result of chance. In effect, the aberrant sequence generated may be misidentified as the blueprint for repair. If analyzed by a fingerprinting method, such as DGGE or T-RFLP, this artifact may be produce a band and even be misinterpreted as a low-abundance representative of the sample [159]. To circumvent complications surrounding heteroduplex inclusion, multiple approaches have been proposed, including the purification of PCR products, followed by the addition of new reagents (primers, dNTPs, polymerase, buffer, and, water) to reduce the generation of heteroduplexes; the separation of heteroduplexes via polyacrylamide gel electrophoresis; the addition of T7 endonuclease, which targets single-stranded DNA (the heteroduplex mismatch forms a DNA 'bubble'); or reducing the number of cycles below the point where the presence of heteroduplexed DNA competes with primers for binding sites.

15.10. PCR: Chimeras.

Chimeric sequences – single pieces of DNA originating from multiple transcripts – are generated either by an incomplete amplicon changing template stands during replication to a second strand that is downstream or by the complexing of an incomplete amplicon that is now acting as a primer. Template switching does not occur based on sequence [158]. Instead, template switching describes the process in which sufficiently complementary stands remain bound during primer extension, which causes strand displacement by either hydrolysis of the complement strand or by physical interaction as DNA polymerase reaches the complement stand, resulting in extension occurring on the alternate strand and a potentially chimeric amplicon being formed. Considering that amplicon concentration increases with the number of PCR cycles, the probability of forming chimeric amplicons, resulting from the partial hybridization of two amplicons/templates, also increases. If a partially complete amplicon is produced that has sufficient binding affinity to an off-target site, this amplicon may also act as a site of DNA replication. Similar in nature to chimeras formed by template-switching, the same probability of chimeras generated by amplicons acting as primers also increases as the number of PCR cycles increases, therefore the same solution also applies – reducing the number of cycles reduces the probability of chimeric strand formation.

15.11. PCR: Knowing when to Stop.

To reduce the probability of PCR bias, caused by the generation of artifacts, the recurring solution is to reduce the number of PCR cycles; however, reducing the number of cycles, in turn, reduces the PCR product, which may not provide sufficient numbers of amplicons for downstream analyses (~20 ng/μL for visualization on an agarose gel, stained with ethidium

bromide, or ~20 nM per 10 μL for most 'next generation' sequencing platforms). For sample-specific information, several methods can be used to monitor the generation of PCR artifacts when used in concert with real-time PCR (rtPCR) or qPCR: DNA-binding dyes, HybProbes, hydrolysis probes, molecular beacons, and fluorescently labeled primers.

- DNA-binding dyes, such as SYBR Green, bind to dsDNA (as opposed to a specific sequence) and fluoresce, thereby reporting amplicon production as it occurs, which should hypothetically increase in fluorescence at an exponential rate that coincides with the total rate of PCR amplicon production [160].
- HybProbes are sequence-specific probes designed to complex with amplicons and emit a fluorescence signal (referred to as fluorescence resonance energy transfer [FRET]). Due to their sequence-specific nature, HybProbes can report the specific abundance of a predetermined sequence within the reaction (the target sequence is normally near the middle of the amplicon), which translates to the rate of amplicon production for a specific target, with possible mutations being detected if the melting curve is analyzed.
- Hydrolysis probes are a sequence-specific two dye system (dye and quencher) that report the production of the target sequence as *Thermus aquaticus*' (*Taq*) polymerase's 5' to 3' exonuclease activity cleaves probe. These probes contain a covalently bound fluorescent tag on the 5' end and a quencher on the 3' end. When bound to single-stranded DNA (ssDNA), *Taq* polymerase catalyzes the release of the fluorophore, which then removes the quenching effect because of an increase in distance from the quencher, and fluorescence can be read by the thermocycler.
- Molecular beacons are sequence-specific oligonucleotide probes (~25 nt [nucleotide]) in the hairpin conformation that consist of a loop (18- 30 nt), a stem (5-7 nt per strand), an

internal fluorophore that is covalently bound at the 5’ end, and a covalently bound quencher at the 3’ end. For this probe, the quencher’s action relies on spatial proximity to the fluorophore, when the hairpin remains closed, fluorescence doesn’t occur, however, when the stem binds to the target sequence, the distance between the fluorophore and the quencher increases and fluoresce occurs.

- Fluorescently labeled primers are formed by attaching a fluorescent tag at the 5’ end of the primer to monitor the rate of primer-template binding. Mechanistically, the fluorophore is quenched while in proximity to guanine, which results in a negative feedback system where the signal diminishes exponentially as amplicons are produced [161].

15.12. Fluorescence *in situ* Hybridization.

Fluorescence in situ Hybridization [25] is a molecular cytogenetic method (a method that capitalizes on chromosomal events, such as interphase and metaphase nuclei, to report information) that uses fluorescently tagged oligonucleotides that are designed to complex with specific sequences (most notably, sequences within the 16S rRNA gene). If the probe complexes with its target sequence, a fluorescent signature is released that can be quantified by flow cytometry. The advantages of this technique are that the target sequence can be specific to the point of 10-25 nucleotides (longer sequences decrease in specificity) enabling a high level of phylogenetic resolution and results that can be considered quantitative measurements of the target sequence within the sample. Furthermore, imagining technologies exist that allow for *in vivo* imaging. Although this specificity allows for phylogenetic resolution, this method requires pre-existing knowledge of the target sequence(s). Therefore, this tool is inappropriate in probing bacterial communities of unknown compositions but fills a much-needed role in fields that are

interested in tracking or quantifying microbes with nucleotide sequence data, such as differences in microbial abundance of known bacterial phyla across healthy and diseased individuals.

15.13. Terminal Restriction Fragment Length Polymorphism.

Like FISH, Terminal restriction fragment length polymorphism (T-RFLP) uses fluorescent probes to identify known sequences within a sample. During PCR, T-RFLP uses primers to amplify a region of the gene encoding the 16S subunit of the rRNA operon. After PCR, amplicons are digested and separated by fragment size by gel electrophoresis. Because this technique uses the same fundamental approach as FISH, the biases are similar, including a lack of phylogenetic resolution as well as associated PCR biases and 16S rRNA gene biases. The main advantage that T-RFLP has over FISH is the ability to produce relatively similar results at a lower cost per run, when combined with 16S rRNA clone library analysis, while still producing phylogenetic data.

15.14. DNA microarrays.

The DNA microarray is another approach that uses florescent labels, but instead of using gels to image the PCR product, a laser is used for detection, resulting in a faster rate of analysis than possible with related methods. If targets (sequences) are known, this can produce quantitative data as well as phylogenetic information to the level specified by the amplicon. While an improvement, with respect to information, there are issues that are pronounced in microarray analyses that include cross hybridization between multiple probes and a single sequence of interest, high detection thresholds, and similar issues with PCR bias as referenced previously. Although because of the aforementioned limitation, the true accuracy for microarrays

used in quantification studies has been the subject of debate, yet this method serves as a tool for the rapid detection of multiple species based on small variations in sequence, such as single nucleotide polymorphisms (SNPs).

15.15. Denaturing and Temperature Gradient Gel Electrophoresis.

The methods denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) use PCR to amplify sequences of similar length, most notably regions within the 16S subunit of the rRNA operon, which are then separated through either by the addition of a denaturant or by temperature, respectively. These methods are combined with gel electrophoresis where amplicons are suspended within a gel matrix and pulled through by an electric current that acts on the negatively charged backbone of DNA, resulting in separation based on sequence length as well as by nucleotide content. As the concentration of denaturant (e.g. temperature or urea or formaldehyde) increases over the polyacrylamide gel, the strands begin to denature and migration proceeds until the strands are near complete denaturation. The main benefits of this and similar methods is that the time-to-completion is relatively low, the length of bands provides preliminary results, and bands may be excised for downstream analyses. At the same time, because these approaches rely on PCR, they are also subject to the same biases and lacks true quantitative or qualitative results when analyzing mixed cultures.

15.16. Extraction Method Biases.

After a representative sample is collected, the next step in most analytical pipelines requires DNA to be extracted from the sample. Due to the vast number of sample types and differences between extraction kits, the core mechanisms of the main extraction methods will be

focused on, as opposed to any specific kit or company. The fundamental steps that are shared between DNA extraction methods include:

- Cell lysis - detergents/surfactants degrade lipids, protease degrades proteins, and RNase degrades RNA.
- Addition of concentrated salts to aggregate waste, such as lipids; proteins; and RNA.
- Separation of waste via centrifugation.
- A purification step where DNA is extracted. The main approaches to DNA purification are: ethanol precipitation, phenol-chloroform extraction, and column purification.

15.17. Ethanol Precipitation.

Ethanol precipitation uses the negative charge from the phosphate backbone of DNA, the polarity of water, and salts to form a hydration shell that encapsulates DNA. While there is evidence in support of changes in recovery percent based on incubation temperature, with lower temperatures increasing the viscosity and resulting in slower migration of DNA through the matrix, centrifugation conditions have the largest combined effect on percent recovery of DNA [162,163]. The next step includes the addition of 95% ethanol (EtOH), followed by an incubation period, with recovery of small fragments of DNA or low concentrations of DNA being greatest when incubation times are longer and recovery of large fragments of DNA or high concentrations of DNA being greatest when incubation times are shorter [162,164]. After centrifugation, a wash with 70% ethanol (or isopropanol, which requires a longer drying period) is performed that removes salts, followed by another round of centrifugation. After air-drying (over-drying causes denaturation of DNA), DNA is resuspended using either buffer or water.

15.18. Phenol-Chloroform Extraction.

Phenol-chloroform extraction is a liquid-liquid extraction method (separates compounds based on solubility) that uses miscibility to separate DNA from lysed cellular material. After adding an equal volume of phenol to chloroform to the sample, the mixture is vortexed and centrifuged. Because the phenol:chloroform mixture is immiscible with water, there are two layers: the aqueous phase that contains nucleic acids, salts, and sugars and rests on top because it has a lower density than the organic phase, phenol:chloroform, which rests on the bottom and contains lipids and proteins. The aqueous phase can be pipetted off at this point or additional rounds of exposure to phenol:chloroform, mixing, and centrifugation may be performed to increase DNA purity. For DNA extractions, the pH of the mixture must not be acidic because that would result in the migration of DNA to the organic phase, instead of remaining in the aqueous phase (note: RNA remains in the aqueous phase if the mixture is acidic).

15.19. Column-Based Purification.

In contrast to the liquid-liquid phenol:chloroform extraction method, column-based purification uses a solid phase extraction approach. In effect, DNA is bound to the solid phase of silica by either ethanol or isopropanol; chaotropic agents that disrupt hydrogen bonding between water and DNA. Once bound by the positively charged ions, a salt bridge is formed between the negatively charged phosphate backbone of DNA and the negatively charged silica matrix. After the column is centrifuged to filter waste, the column is washed with a buffer to remove residual waste from the silica column, then centrifuged to dry. At this stage, an elution buffer or water is added to the column, which resuspends and carries the purified DNA through the silica matrix.

15.20. DNA purification biases.

In addition to the inevitability that DNA purification will result in loss of gDNA/amplicons, inconsistencies between core purification methods, and even inconsistencies within the same purification method (kit-kit differences) [165-169], extracellular DNA, can persist anywhere between 3-60 days outside of bacterial cells in an aqueous solution and is considered a contaminant when trying to quantify viable cells [170]. Although the persistence of DNA is dependent on multiple factors, such as ultraviolet light; nucleases; or consumption by microbes, in an aqueous solution in minimal light, if the presence of extracellular DNA is not removed or viable cells are not separated from the sample during the extraction process by either using ethidium monoazide (EMA), microscopic separation, or flow cytometry [171-174], the residual environmental/extracellular DNA will act as an inclusion and be effectively treated as part of the viable cell population, introducing another bias into downstream analyses. Prior to extraction or PCR amplification, propidium monoazide [175] can be added to the sample. In effect, DNA will be covalently bound by PMA's azide group once exposed to bright, visible, light. Samples that undergo this process may now undergo the remainder of the workflow because DNA that is covalently bound will not readily amplify during PCR [176].

15.21. Microbiome Target: the 16S rRNA Gene.

The gene encoding the 16S subunit of the ribosomal RNA operon is the leading diagnostic target used in microbiome quantification. This component of the 30S prokaryotic ribosome was driven to popularity by the microbiologist Carl Woese, who used conserved regions within this gene as primer binding sites to create a map of the degenerate sequences therein, resulting in phylogenetically mapping microbes that are now known as archaea. Since

that time, researchers have used this region for its ease of amplification, relative abundance, and ever-growing databases on this gene. As a molecular target, this gene contains nine hypervariable regions that range from 30 to 100 bp within the secondary structure of the functional subunit. To observe this degeneracy contained within these ubiquitous sequences, over time, researchers have designed a total of 14 'universal primers' to capture archaeal and bacterial diversity. And, with the use of technologies including quantitative real-time PCR and a suite of other methods, the gene encoding the 16S subunit has served as the 'gold standard' for measuring microbial abundance for over a decade. While nearly as prolific in scientific literature as in nature, this gene also displays notable shortcomings that prevent its accuracy in either qualitative or quantitative analyses, a lack of consistency (with respect to gene copy number) and a lack of nucleotide diversity (resulting in a lack of phylogenetic resolution attainable from this gene) – clear indicators that this is a transitional target until a more appropriate region is discovered or method is invented.

Gene copy number ranges from 1-15 copies in bacteria [177], which if left unaddressed, biases all quantitative measurements because not only would you have skewed ratios [178] (e.g. if you had equal populations of a one rRNA operon copy bacteria 'A' and a 15 copy rRNA operon bacteria 'B', then there would appear to be exponentially more of bacteria 'B', even though there is a 1:1 ratio, because not only do you start with 15 times the template, because of the exponential nature of PCR, you will end up with 7.5×2^{cycles} more amplicons). Highlighting the significance of gene copy number, researchers evaluated three separate 16S rRNA gene standards by first performing 35 cycles of PCR, followed by a second PCR that ligated a fluorescent nucleotide derivative to the amplicon and digestion by restriction endonucleases. After imaging by gel electrophoresis, the specific quantities of amplicons were determined by

sequencing, which showed that the reannealing of gDNA progressively reduces the propensity for hybridization between target sequences and primers to occur, resulting in disproportional/reduced amplification. To alleviate this bias, pooling samples in triplicate increases the amount of template, while maintaining the relative concentration of templates, thereby standardizing the standards [179]; however, this solution is perpetuating an unsustainable practice. Considering that most rRNA operon copy numbers are unknown, researchers have proposed programs to estimate gene copy number [180]. To further complicate this issue, closely related species have also been recorded as having different rRNA copy numbers and, when sequenced, show conflicting phylogenetic placement [181-184]. Meaning, copies of the 16S rRNA gene may have greater variation within the same bacterium, than with the 16S rRNA gene sequence from another bacterium [185]. Therefore, the favor placed with the 16S rRNA gene has ultimately restricted the analytical capability of methods that rely on this sequence [185].

In addition to gene copy number, the sequence diversity between conserved regions in the 16S rRNA gene is not diverse enough to resolve taxa on a meaningful level. Based on the 97% sequence similarity standard for taxonomic resolution, this gene provides relatively little differentiation beyond the level of genus. In the same line, even several genera would not be classified under the same operational taxonomic unit (OTU) because they share identical 16S rRNA gene sequence regions [183]. To resolve this limitation, while keeping associated techniques viable, researchers have evaluated alternate targets.

Alternate genes to the 16S rRNA gene include *cpn60*, *rpoB*, *gyrA*, and *gyrB*. Each of these genes would resolve the copy number issue because they are all single copy genes. Furthermore, each of these genes provides greater taxonomic resolution, often to the species or sub-species level [183,186-188]. While this would resolve both copy number and gene diversity, the

degree of degeneracy within these genes presently exceeds the tolerance for PCR; however, novel methods aim to resolve these limitations so that more representative analyses may be performed [189].

15.22. Microbiome Sequencing.

Sequencing methods are presently in the transition from 'next generation' methods to 'third generation' methods, which move away from fragmentation and amplification to analysis of individual molecules. For direct sequencing of microbiomes, fragmentation methods surpass amplicon-based (typically 16S rRNA gene-based) methods because they are not as limited by the diversity within the region sequenced. While these methods provide greater resolution between taxa, direct sequencing also costs more than non-sequencing methods (e.g. agarose-based verification methods), is more time intensive, and *de novo* sequencing (i.e. shotgun sequencing) required the generation of a clone library followed by Sanger sequencing (the chain termination method). Although these methods have been replaced by sequencing methods with greater throughput that do not require the generation of clone libraries to generate species-level taxonomic information, there are a host of steps that require technological understanding of so that bias is not inadvertently introduced.

All sequencing technologies are inconsistent in reading regions with either extreme (high or low) G+C content, regions with runs of homopolymers (e.g. 5'-AAAAAAAAAAAAAA-3'), palindromic regions (e.g. 5'-ATTATTA-3'), inverted repeats (e.g. 5'-TTACGAAACGTAA-3'), and in the case of sequencing methods that require a cloning vector, gene sequences that produce elements that are detrimental or toxic to the host cell. These inconsistencies result in sequence data that has lower on-average coverage than regions with more average compositions of DNA,

creating sources of bias that, to date, the primary suggestion for correcting these errors is to use multiple sequencing platforms that have complementary biases, thereby cancelling each other out [190]. Beyond the lack of uniform read distribution, which prevents a direct translation of coverage to abundance within the sample, 'next generation' sequencing methods share the common necessity for library generation that includes fusion of an adapter to DNA so that the complex can bind to a fixed surface. Unfortunately, every manufacturer's protocol has sample preparation steps that introduce bias [191]. The dominant source of bias for maintaining representative sequencing data are the same biases introduced during normal PCR amplification, which have been shown to be reduced by using a higher fidelity polymerase (e.g. Kappa HiFi instead of Phusion); lowering extension temperatures; adding betaine; and for low copy targets, replacing PCR entirely with MDA. Note that the biases introduced by the use of 'barcodes' (strings of nucleotides with an identifiable sequence, ligated to amplicons during PCR), which enable multiple samples to be run in parallel, has been shown to be unpredictable based solely on the secondary structure of the modified barcoded primers [192].

Moving beyond the previously discussed biases of PCR, two new sources have been identified: fragmentation and size selection. Fragmentation introduces bias because of preferential breakage due to compositional differences in DNA and its associated molecules (e.g. methylation), which can be resolved by another fragmentation step after precipitation so that all fragments are within the sequencing platform's optimal size range [191,193]. Bias introduced by size selection is caused by differential heating of agarose gel, which can be resolved by melting the agarose gel slices at room temperature [191,194]. Furthermore, when the template concentration is low (0.1 ng), the propensity for the previously mentioned biases to occur increases, presumably due to stochastic effects [195]. To reduce the chance of this variance, input DNA concentration

should always be optimized for the specific sequencing platform being used, instead of simply increasing the number of replicates, which has shown no effect on the outcome of sequencing runs [196]. The methods for end-repair and adapter ligation have not been linked with skewing results. To determine bias within uniformly distributed or random data, the per-base or per-sequence bias can be calculated by comparing the relative coverage of the target base or sequence with the mean coverage of a reference base or sequence within the same sequencing run (**Equation 1**) [190]. Conceptually, this approach should be able to be combined with a titration-based approach in which multiple standards (that include average compositions of DNA as well as sequences with the most probably biases) are used, thereby correcting for method-based bias with a high degree of accuracy and accounting for sites that are either overrepresented or underrepresented. Diverging from general sequencing biases, the dominant culture-independent microbiome sequencing techniques are: single cell sequencing, shotgun metagenomics, and high-throughput sequencing.

Single cell sequencing methods do not produce population scale metrics, but these tools provide examples of alternate approaches to microbiome research that may, for example, enable researchers to eventually culture the 'unculturable' microbe. By the definition of single cell sequencing, these methods are required to analyze a single cell from within a sample. The most common methods that remove the cell from the sample are microfluidics and fluorescence-activated cell sorting (FACS), which are both highly accurate, but because they extract the cell from within its environment, the possibility of altering transcription or including contaminant material increases [197-200]. In contrast to the previous methods, laser-capture microdissection (LCM) does not remove cells from the sample (e.g. tissue) prior to analysis. Instead, LCM uses a laser to remove a section that contains the sample and non-contact methods (gravity, pressure,

and laser-induced forward transfer) to extract cellular material with minimal perturbation [201]. Once successful isolation has occurred, the next challenge is the limit on template DNA, so, several strategies have been implemented to address this: pure PCR-based amplification (degenerate oligonucleotide primed [DOP-PCR]), isothermal amplification (multiple displacement amplification [MDA]), and methods that are combination of PCR-based and isothermal methods (multiple annealing and looping based amplification cycles [MALBAC]). DOP-PCR uses a single primer with a Xhol restriction site at its 5' and 3' ends as well as a random hexamer between the ends. This method works by performing PCR under extremely low stringency (low annealing and extension temperature) so that amplification and tagging with the random hexamer sequence occurs readily throughout the target genome for 5-8 PCR cycles, then the stringency of conditions is increased for the remaining >25 cycles, producing amplicons that are anywhere between 200 and 1,000 bp. Although DOP-PCR creates a large amount of data, there can be frequent issues with its interpretation because the biases that are associated with PCR are exacerbated [202-204]. Out of the three methods, the most commonly used approach is MDA, which adopts the high-fidelity polymerase of Φ29 that uses strand displacement to produce additional copies of the genome. Because of the replicative nature of MDA, the amplicons generated by the phage polymerase early on will be exponentially more abundant than other regions that were amplified at a later cycle and tend to form chimeric sequences, which are reduced by the late inclusion of a specific endonuclease. [205-207]. The first commonly used hybrid method is MALBAC, which aims to increase loop formation of isothermal amplicons by using random primers and new thermocycling conditions so that new amplicons formation is restricted until uniform amplification has occurred up to this step [175]. Collectively, these methods provide

advanced levels of resolution, but are limited in their feasibility to represent populations or quantify them.

Shotgun sequencing or whole metagenome shotgun sequencing is an adapted sequencing method that takes a metagenomic sample and infers which microbes are present within the sample, based on *de novo* assembly. Because this method does not use a reference genome to assemble reads, this sequencing technology requires a high amount of coverage to assemble contigs (contiguous sequences). Reads are generated by shearing high molecular weight DNA at random, then constructing a library (historically, cloning sheared DNA into a vector was a necessary step), followed by pairwise end sequencing (also known as 'double barrel shotgun sequencing') and assembly. As a result of the biases of library generation discussed previously, the sequencing step is inherently biased, and, while this method traditionally struggled with 'linking' sequenced regions into complete genomes because the chain termination method has difficulty with repetitive sequences or palindromic sequences, these issues have been largely resolved by using different sequencing chemistries associated with 'next generation' (now referred to as 'high-throughput' technologies).

High-throughput sequencing technologies presently include: single-molecule real-time sequencing, which produces read lengths of 10-15 kbp at 87% read accuracy; ion semiconductor sequencing, which produces read lengths of <400 bp at 98% read accuracy; 454-pyrosequencing, which produces read lengths of <700 bp at 99.9% read accuracy; sequencing by synthesis, which produces varying read lengths, based on the specific platform, including 75-300 bp for the MiniSeq/NextSeq, 50-600 bp for the MiSeq, 50-500 bp for the HiSeq, and 300 bp for the HiSeq X, with an average of 99.9% read accuracy; and sequencing by ligation (SOLiD), which produces read lengths of 59+35 bp or 50+50 bp at 92-97% read accuracy for single reads and

99.96% for consensus sequences [208,209]. Given its present applicability to microbiome research, the Illumina-based platform known as the MiSeq has been thoroughly researched and shows a series of deviations that are significant to application in microbiome research. A term referred to as phasing or pre-phasing occurs when elongation fails or advances too quickly, resulting in an overall decrease in cluster signal that can be cumulative. Additionally, the adenine and cytosine are more likely to be read as substitution errors, which is caused by an indiscriminancy from the red laser or filter that distinguishes the two nucleotides. Furthermore, traditional quality scores (e.g. Phred scores) are not applicable to amplicon sequencing because low quality scores do not correlate well with error. Therefore, a high degree of coverage, with sequence data from both directions, is necessary to find the true sequence amongst error prone data [210]. Using these techniques to quantify entire microbiomes produces a massive amount of information that likely includes quantitative data about the population that is masked by the analyses leading to sequencing.

Once sequence data is generated, because of the immense amount of data generated from these methods, computational resources are required to process large datasets. At this stage, most prepackaged sequence assembly packages would process the data, but not account for the possibility that there was a range of starting templates that is unknown. As a result, data on low-copy number bacteria may be erroneously removed from the sequenced population. Also, the presence of ecologically-significant strains may be removed if they are present in a low abundance, with strains that are associated with the normal microbiome being identified as the correct read for the sequence. Furthermore, ecological studies often require systematic knowledge that may not translate well if the researcher lacks a background in either the field of

application or in bioinformatic/genomic analysis, so, the data interpretation must be performed by a researcher with sufficient systematic knowledge.

15.23. Discussion of Common Biases Associated with Microbiome Analyses.

Every method inherently skews the data it aims to interpret. By accounting for these biases, these 'artifacts of the technique' are removed and a more accurate interpretation of results is possible. Beyond understanding and accounting for these biases, when using method-based techniques or comparative approaches that rely on databases, cross-validation with independent methods and method-specific workarounds greatly reduce noise. Furthermore, with the advent of high throughput sequencing, advances in biocomputing, and the continued integration of interdisciplinary methods, advancing our knowledge of biological systems through the use of genomic, transcriptomic, and proteomic data are no longer restricted by the ability to generate or process 'big' data. Instead, major advances are now contingent on maintaining representative data.

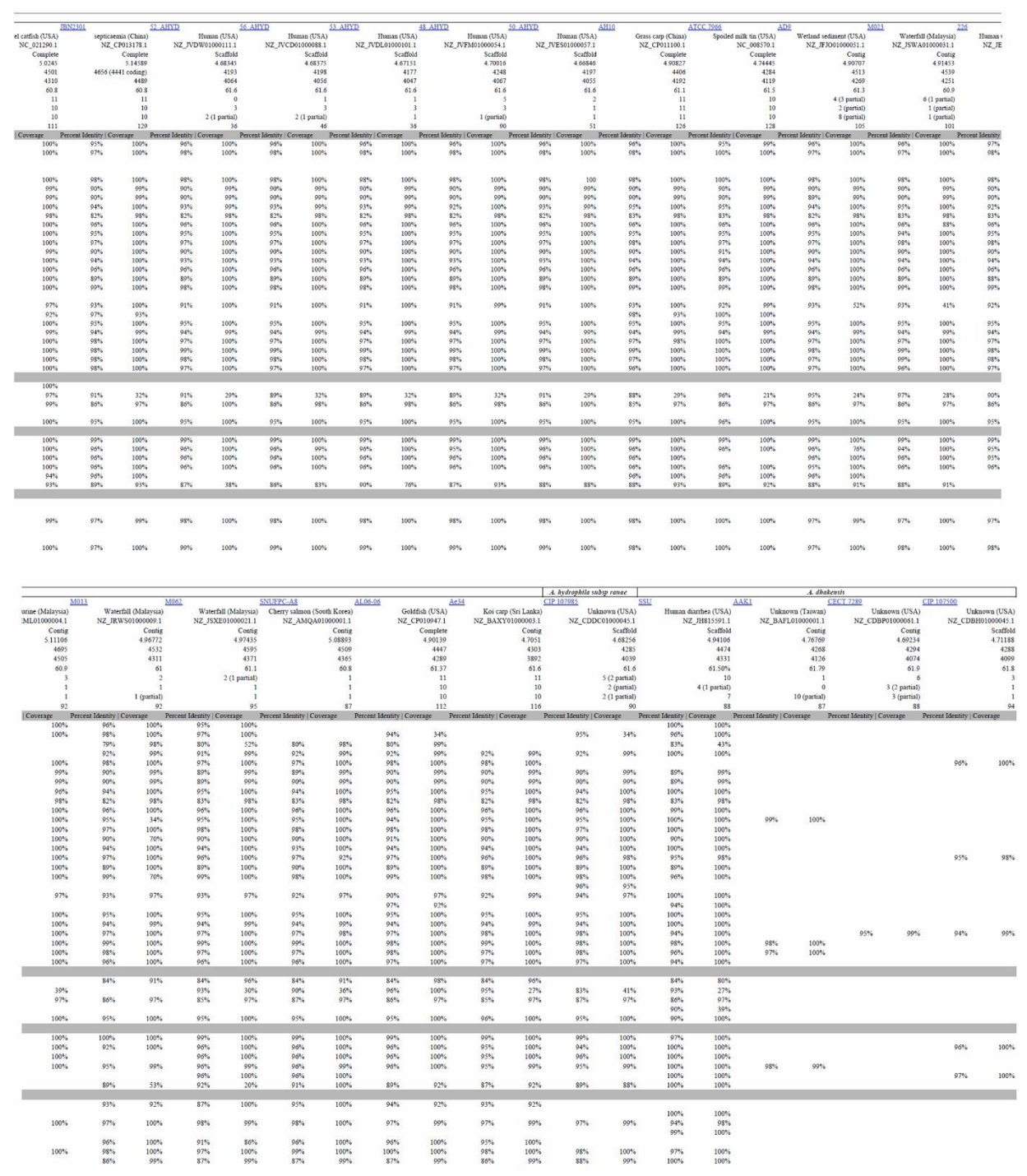
Statistical analysis programs, open source bioinformatic software, and sequencing kits all have one thing in common: they're prepackaged methods that provide a streamlined option that allows both experienced and novice users to process data. These easy-to-use methods serve an important role but are perilous because even the most ubiquitous method is not perfectly accurate. As molecular techniques are progressively implemented into the quantification of microbiomes, having a fundamental understanding of how these prepackaged methods inject bias into data is imperative when interpreting data and crucial to the ever-rapid advancement of microbial ecology.

Table 1. NCBI BLAST of all listed *A. hydrophila* with complete genomes in GenBank against putative virulence factors confirmed in *Aeromonas* spp.

				NF1		NF2		RB-AH		4AK4		BWH65		116		14		173
			Isolation source (Country)		Human (USA)		Human (USA)		Soil (Malaysia)		Industrial (China)		Human (USA)		Human bile (Malaysia)		(Malaysia)	
			NCBI reference sequence		NZ_JDWB01000013.1		NZ_JDWC01000001.1		NZ_JPEH01000001.1		NZ_CP006579.1		NZ_KQ087566.1		NZ_ANPN01000021.1		NZ_AOBM01000021.1	NZ_AC
			Genome completion		Contig		Contig		Contig		Complete		Scaffold		Contig		Contig	
			Genome size (Mbp)		4.80653		4.78711		5.08893		4.52799		4.44564		4.67704		4.67332	
			Predicted number of genes		4378		4417		4663		4074		4021		4226		4243	
			Coding sequences		4126		4146		4419		3917		3876		4077		4054	
			Percent G+C content		61.1		61.3		60.8		62		61.7		62		62	
			5S rRNA copy number		2		1		11		10		11		2		2	
			16S rRNA copy number		1		1		10		9		7 (2 partial)		1		1	
			23S rRNA copy number		1 (partial)		4 (partial)		10		10		8 (4 partial)		1		1	
			tRNA		93		87		121		104		81		91		88	
Enzymes	**Function**	**Strain referenced**	**Genbank ID**	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity
act	Cytotoxic enterotoxin Act	*A. dhakensis SSU*	NZ_JH815591.1	96%	100%			95%	99%									
aerA	Aerolysin	*A. hydrophila ATCC 7966*	NC_008570.1	98%	100%	96%	34%	96%	100%									
aexT	ADP-ribosylating toxin (AexT)	*A. hydrophila AH-3*	EF442031.1			80%	98%											
aexU	T3SS-associated exoenzyme effector U	*A. dhakensis SSU*	AGWR01000034.1	92%	99%	92%	99%	92%	99%									
ahh1	Extracellular hemolysin	*A. hydrophila ATCC 7966*	CP000462.1	98%	100%	98%	100%	97%	100%					96%	100%	96%	100%	96%
ahpB	Serine Protease	*A. hydrophila AG2*	AF193422.1	90%	99%	90%	99%	90%	99%	87%	99%	85%	99%	89%	99%	89%	99%	89%
ahyB	Elastase	*A. hydrophila AG2*	AF193422.1	90%	99%	90%	99%	90%	99%	87%	99%	85%	99%	89%	99%	89%	99%	89%
alt	Cytotonic enterotoxin (lipase)	*A. dhakensis SSU*	AGWR01000034.1	95%	99%	95%	100%	95%	100%	89%	96%	89%	98%	98%	100%	98%	100%	98%
apl-1	Nonhemolytic phospholipase C	*A. hydrophila JMP636*	U14011.1	83%	98%	82%	98%	82%	98%	85%	98%	84%	99%	83%	98%	83%	98%	83%
ast	Cytotonic enterotoxin	*A. dhakensis SSU*	AF419157.1	96%	100%			96%	100%					94%	55%	94%	55%	96%
dam	DNA adenine methyltransferase	*A. dhakensis SSU*	AGWR01000003.1	94%	100%	94%	100%	94%	100%	90%	98%	89%	100%	99%	100%	99%	100%	98%
eno	Enolase (surface-expressed)	*A. dhakensis SSU*	AGWR01000043.1	98%	100%	97%	100%	97%	100%	96%	99%	97%	100%	99%	100%	99%	100%	99%
eprA1	EprA1 (extracellular protease)	*A. hydropila AH1*	U93600.1	91%	100%	91%	100%	91%	100%	88%	100%	99%	100%	90%	100%	90%	100%	91%
gidA	Glucose-inhibited division protein	*A. dhakensis SSU*	NZ_JH815591.1	94%	100%	94%	100%	94%	100%	92%	100%	91%	100%	99%	100%	99%	100%	99%
hlyA	Pore-forming cytolysin (hemolysin)	*A. hydrophila A6*	U81555.1	96%	100%	96%	100%	96%	100%					95%	98%	95%	98%	95%
lip	Extracellular lipase	*A. hydrophila H3*	S65123.1	89%	100%	89%	100%	89%	100%	85%	100%	83%	100%	89%	100%	89%	100%	89%
mepA	Metalloprotease	*A. hydrophila AH-1*	AY841796.1	99%	100%	99%	100%	100%	100%	91%	100%	91%	100%	96%	100%	96%	100%	96%
plc	Phospholipase C	*A. hydrophila AH-3*	AF092034.1					100%	100%									
rtxA	RTX toxin (repeat in toxin A)	*A. dhakensis SSU*	NZ_JDWD01000006.1											97%	100%	99%	100%	98%
ser	Serine protease	*A. hydrophila ATCC 7966*	CP000462.1					97%	100%	86%	96%	87%	95%					
tagA	ToxR-regulated lipoprotein (TagA)	*A. dhakensis SSU*	DQ398103.1	95%	100%	95%	100%	95%	100%	80%	93%			97%	100%	97%	100%	97%
vacB	Exoribonuclease R	*A. dhakensis SSU*	EU380595.1	94%	100%	94%	99%	94%	99%	91%	99%	92%	99%	98%	100%	98%	100%	97%
AHA_1741	Collagenase	*A. hydrophila ATCC 7966*	NC_008570.1	96%	98%	97%	100%	98%	100%	89%	98%	94%	98%	95%	99%	95%	99%	95%
AHA_3147	Invasin	*A. hydrophila ATCC 7966*	NC_008570.1	98%	100%	99%	100%	99%	100%	82%	98%	85%	98%	98%	100%	98%	100%	97%
AHA_3217	Thermostable hemolysin	*A. hydrophila ATCC 7966*	NC_008570.1	98%	100%	97%	100%	98%	100%	81%	94%			96%	100%	98%	100%	96%
AHA_3493	Hemolysin III	*A. hydrophila ATCC 7966*	NC_008570.1	97%	100%	97%	100%	97%	100%	82%	98%			94%	100%	94%	100%	94%
Adhesion and invasion																		
flg/flt/laf	Lateral flagellin	*A. hydrophila AH-3*	DQ124694.1	84%	90%	84%	91%	84%	91%	81%	92%	80%	92%					
tapA	Type IV pilus subunit protein	*A. hydrophila Ah65*	U20255.1	88%	39%	93%	27%	92%	28%					94%	30%	94%	30%	95%
tapD	Type IV pilus biogenesis	*A. hydrophila Ah65*	U20255.1	86%	97%	87%	97%	86%	97%	86%	96%	82%	96%	86%	97%	86%	97%	85%
ahsA	S-layer protein (AhsA)	*A. hydrophila TF7*	L37348.1	89%	38%	88%	38%	99%	99%									
mtnD	Mucosal adherence (MinD)	*A. hydrophila W-01*	KP271930.1	95%	100%	95%	100%	95%	100%	91%	100%	90%	100%	98%	100%	98%	100%	99%
Quorum sensing & biofilm formation																		
ahyRI	Quorum sensing regulon activator (Lux	*A. hydropila A1*	X89469.1	99%	100%	99%	100%	100%	100%	87%	55%	86%	96%	97%	100%	97%	100%	97%
hcp2	Hcp-2 hemolysin-coregulated protein	*A. dhakensis SSU*	DQ667172.1	95%	100%	95%	99%	95%	100%	93%	100%	95%	99%	96%	100%	98%	100%	96%
luxS	S-ribosylhomocysteinase	*A. dhakensis SSU*	EU499388.1	96%	100%	96%	100%	95%	100%	93%	100%	94%	100%	98%	100%	98%	100%	99%
qseBC	Two-component-based QseBC quorum	*A. dhakensis SSU*	JF922013.1	96%	99%	96%	99%	96%	99%	85%	99%	85%	99%	98%	99%	98%	99%	98%
vasHK	Sigma 54-dependent transcriptional reg	*A. dhakensis SSU*	DQ667172.1	96%	100%	96%	100%			91%	100%	97%	100%	97%	100%	97%	100%	
vgrG-1	Valine-glycine repeat protein G	*A. dhakensis SSU*	JX646703.1	89%	93%	92%	45%	88%	93%	85%	93%	90%	100%	87%	93%	87%	93%	89%
Other																		
acrV	Antigen for host evasion	*A. hydrophila AH-1*	AY394563.2	93%	92%	94%	100%	99%	100%									
acrV	Component of the T3SS	*A. dhakensis SSU*	AY763611.1															
AHA_0617	Mucin-desulfating sulfatase	*A. hydrophila ATCC 7966*	NC_008570.1	98%	99%	97%	99%	97%	99%					95%	98%	95%	98%	95%
aopB	T3SS translocator structural protein	*A. dhakensis SSU*	AY763611.1															
aopB	T3SS translocator structural protein	*A. hydrophila AH-1*	AY394563.2	96%	100%	96%	100%	100%	100%									
aroA	3-phosphoshikimate 1-carboxyvinyltran	*A. hydrophila ATCC 7966*	CP010947.1	98%	100%	98%	100%	98%	100%	91%	100%	92%	100%	97%	100%	97%	100%	97%
ascV	lcrD homolog (T3SS structural protein)	*A. dhakensis SSU*	AY763611.1	87%	99%	86%	99%	87%	99%									

																					A. hydrophila	
		145		L14f		187		259		277		KOR1		J-1		NJ-35		AL09-71		pc104A		ML09-119
Isolation source (Country)	(Malaysia)		Human puss (Malaysia)		Lake water (Malaysia)		Human puss (Malaysia)		Human blood (Malaysia)		Human puss (Malaysia)		Mangrove plant (China)		Swordtail fish (China)		Diseased Crucian carp (China)		(USA)		Soil (USA)	Diseased chann
NCBI reference sequence	BN01000009.1		NZ_JEMK01000043.1		NZ_JWJQ01000010.1		NZ_AOBO01000007.1		NZ_AOBP01000010.1		NZ_AOBQ01000001.1		NZ_KQ464176.1		NZ_CP006883.1		NZ_CP006870.1		NZ_CP007566.1		NZ_CP007576.1	
Genome completion	Contig		Contig		Contig		Contig		Contig		Contig		Scaffold		Complete		Complete		Complete		Complete	
Genome size (Mbp)	4.78635		4.8641		4.68278		4.78391		4.70069		4.7901		4.78317		5.00081		5.27964		5.02386		5.02383	
Predicted number of genes	4314		4360		4169		4357		4253		4384		4277		4482		4741		4506		4501	
Coding sequences	4124		4183		4023		4202		4087		4214		4110		4274		4538		4306		4302	
Percent G+C content	61.6		61.5		61.7		61.6		61.7		61.6		61.8		60.9		60.5		60.8		60.8	
5S rRNA copy number	2		3		2 (1 partial)		2		4		2		11		11		11		11		11	
16S rRNA copy number	1		1		1		1		2 (partial)		1		9		10		10		10		10	
23S rRNA copy number	1		2 (partial)		1		1		1		1		10		10		10		10		10	
tRNA	95		93		85		95		91		95		113		110		102		111		111	
	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity	Coverage	Percent Identity
act		98%	100%	97%	100%					98%	100%	98%	100%	95%	100%	95%	100%	95%	100%	95%	100%	95%
aerA		96%	100%	95%	100%					96%	100%	96%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%
aexT						83%	42%	82%	43%	82%	43%	82%	43%									
aexU						98%	100%	98%	100%	98%	100%	98%	100%									
ahh1	100%	96%	100%	96%	100%	97%	100%	96%	100%	97%	100%	96%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%
ahpB	99%	89%	99%	89%	99%	89%	99%	89%	99%	89%	99%	89%	99%	90%	99%	90%	99%	90%	99%	90%	99%	90%
ahyB	99%	89%	99%	89%	99%	89%	99%	89%	99%	89%	99%	89%	99%	90%	99%	90%	99%	90%	99%	90%	99%	90%
alt	100%	97%	100%	98%	100%	97%	100%	98%	100%	97%	100%	98%	100%	94%	100%	94%	100%	94%	100%	94%	100%	94%
apl-1	98%	83%	98%	83%	98%	83%	98%	83%	98%	83%	98%	83%	98%	82%	98%	82%	98%	82%	98%	82%	98%	82%
ast	55%	96%	55%	96%	55%	97%	55%	96%	55%	96%	55%	96%	55%	96%	100%	96%	100%	96%	100%	96%	100%	96%
dam	100%	99%	100%	98%	100%	98%	100%	98%	100%	99%	100%	98%	100%	95%	100%	95%	100%	95%	100%	95%	100%	95%
eno	100%	99%	100%	99%	100%	99%	100%	99%	100%	99%	100%	99%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%
eprA1	100%	90%	100%	90%	100%	90%	100%	90%	100%	91%	100%	90%	100%	90%	100%	90%	100%	90%	100%	90%	100%	84%
gidA	100%	99%	100%	99%	100%	98%	100%	99%	100%	99%	100%	99%	100%	94%	100%	94%	100%	94%	100%	94%	100%	94%
hlyA	98%	95%	98%	95%	98%	95%	98%	95%	98%	95%	98%	95%	98%	96%	100%	96%	100%	96%	100%	96%	100%	96%
lip	100%	89%	100%	90%	100%	89%	100%	89%	100%	89%	100%	89%	100%	89%	100%	89%	100%	89%	100%	89%	100%	89%
mepA	100%	96%	100%	96%	100%	96%	100%	96%	100%	96%	100%	96%	100%	99%	100%	99%	100%	99%	100%	99%	100%	99%
plc																						
rtxA	100%	99%	100%	99%	20%	98%	100%	99%	100%	97%	100%	99%	100%			92%	97%	92%	97%	92%	97%	93%
ser												94%	100%	96%	100%	96%	100%	97%	92%	97%	92%	97%
tagA	100%	97%	100%	98%	100%	97%	100%	98%	100%	98%	100%	98%	100%	95%	100%	95%	100%	95%	100%	95%	100%	95%
vacB	100%	97%	100%	97%	100%	97%	100%	97%	100%	98%	100%	97%	100%	94%	99%	94%	99%	94%	99%	94%	99%	94%
AHA_1741	100%	94%	99%	94%	99%	94%	99%	95%	99%	95%	99%	94%	99%	98%	100%	98%	100%	98%	100%	98%	100%	98%
AHA_3147	100%	96%	100%	97%	100%	97%	100%	97%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%
AHA_3217	100%	97%	100%	96%	100%	96%	100%	96%	100%	96%	100%	97%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%
AHA_3493	100%	95%	100%	94%	100%	94%	100%	94%	100%	94%	100%	94%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%
flg/flt/laf						84%	80%	84%	91%	84%	80%	84%	84%	97%	100%	97%	100%					97%
tapA	26%	92%	29%	94%	25%	94%	23%	89%	32%	98%	21%	93%	31%	91%	97%	91%	32%	91%	31%	91%	31%	91%
tapD	97%	86%	97%	85%	97%	86%	97%	86%	97%	86%	97%	85%	97%	86%	99%	86%	97%	86%	95%	86%	95%	86%
ahsA						91%	38%	90%	39%	90%	39%											
mtnD	100%	99%	100%	98%	100%	98%	100%	98%	100%	98%	100%	98%	100%	95%	100%	95%	100%	95%	100%	95%	100%	95%
ahyRI	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	99%	100%	99%	100%	99%	100%	99%	100%	99%
hcp2	100%	95%	100%	96%	100%	98%	100%	99%	100%	98%	100%	99%	100%	96%	100%	96%	100%	95%	100%	95%	100%	96%
luxS	100%	96%	100%	99%	100%	96%	100%	97%	100%	99%	100%	98%	100%	96%	100%	96%	100%	96%	100%	96%	100%	96%
qseBC	99%	97%	99%	98%	99%	98%	99%	98%	99%	98%	99%	98%	99%	96%	100%	96%	100%	96%	100%	96%	100%	96%
vasHK				97%	100%	98%	100%	97%	100%	97%	100%	97%	100%	96%	100%	96%	100%					78%
vgrG-1	92%	90%	92%	92%	90%	91%	92%	91%	100%	92%	100%	93%	100%	89%	93%	89%	93%	89%	93%	89%	93%	89%
acrV																						
acrV						99%	100%			99%	100%	99%	100%									
AHA_0617	99%	94%	98%	95%	98%	94%	98%	94%	98%	95%	98%	95%	98%	97%	99%	97%	99%	97%	99%	97%	99%	97%
aopB						99%	100%	99%	100%	99%	100%	99%	100%									
aopB																						
aroA	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%	100%	97%
ascV						99%	100%	94%	100%	97%	100%											

Table 1 (Continued). NCBI BLAST of all listed *A. hydrophila* with complete genomes in GenBank against putative virulence factors confirmed in *Aeromonas* spp.



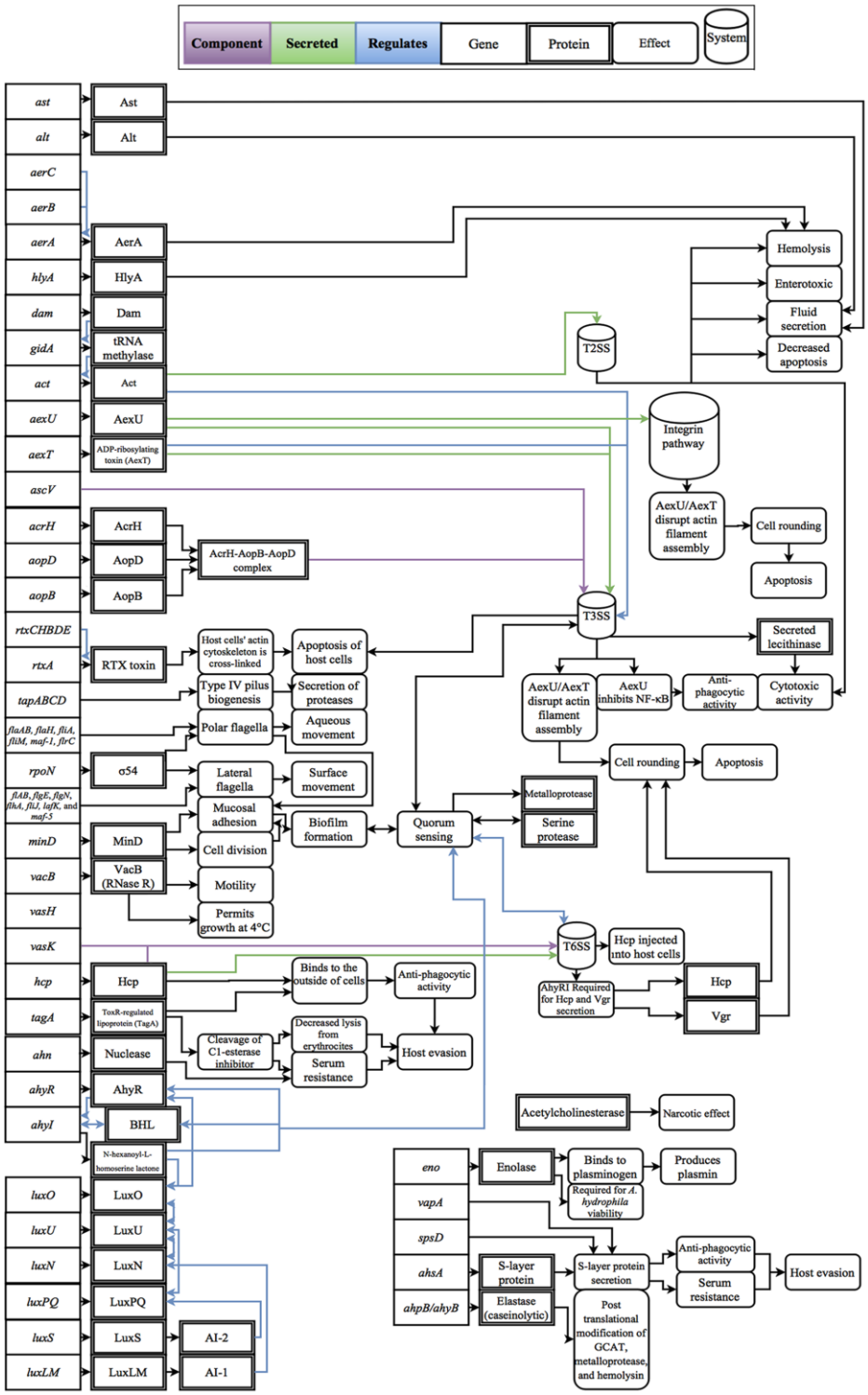


Figure 1. Diagram of the gene products, molecular interactions and functions implicated in *A. hydrophila* virulence. These interactions are based on the collective literature referenced in this proposal.

Equation 1. Calculation of per-site or per-sequence bias, with reference to results from the average composition within a sequencing run.

$$Coverage\ bias = \frac{\text{Site (nucleotide or sequence)coverage}}{\text{Coverage of equally distributed nucleotides of the same length}}$$

Chapter II

Classification of a Hypervirulent *Aeromonas hydrophila* Pathotype Responsible for Epidemic Outbreaks in Warm-Water Fishes

1. Abstract.

Lineages of hypervirulent *Aeromonas hydrophila* (v*Ah*) are the cause of persistent outbreaks of motile *Aeromonas* septicemia in warm-water fishes worldwide. Over the last decade, in the People's Republic of China and in the United States, this virulent lineage of *A. hydrophila* has resulted in annual losses of millions of tons of farmed carp and catfish. Multiple lines of evidence indicate that US catfish and Asian carp isolates of *A. hydrophila* affiliated with sequence type 251 (ST251) share a recent common ancestor. To address the genomic context for the putative intercontinental transfer and subsequent geographic spread of this pathogen, we conducted a core genome phylogenetic analysis on 61 *Aeromonas* spp. genomes, of which 40 were affiliated with *A. hydrophila* and 26 were epidemic strains. Phylogenetic analyses indicate that all ST251 strains form a coherent lineage affiliated with *A. hydrophila*. Within this lineage,

conserved genetic loci were identified that are unique within *A. hydrophila* as well as genes that are present in consistently higher copy numbers than in non-epidemic *A. hydrophila* isolates. In addition, results from analyses of representative ST251 isolates supports that multiple lineages are present within US v*Ah* isolated from Mississippi, whereas v*Ah* isolated from Alabama appear clonal. This is the first report of genomic heterogeneity within US v*Ah* isolates, with some Mississippi isolates showing closer affiliation with the Asian grass carp isolate ZC1 than with other v*Ah* isolated in the US. To evaluate the biological significance of the identified heterogeneity, comparative disease challenges were conducted with representatives of different v*Ah* genotypes revealing that isolate ZC1 yielded significantly lower mortality in channel catfish, relative to Alabama and Mississippi v*Ah* isolates. Like other Asian v*Ah* isolates, the ZC1 lineage contains all core genes for a complete type VI secretion system (T6SS). In contrast, more virulent US isolates retain only remnants of the T6SS (*clpB, hcp, vgrG,* and *vasH*) which may have functional implications. Collectively, these results characterize a hypervirulent *A. hydrophila* pathotype that affects farmed fish on multiple continents.

2. Introduction.

In countries across the world, aquaculture industries have been decimated by epidemics of a hypervirulent pathotype of *A. hydrophila* (v*Ah*) [211,212]. *A. hydrophila* is ubiquitous within warm-water environments and has a diverse host range (i.e., amphibians, birds, fishes, reptiles, and mammals) with equally diverse diseases that includes motile *Aeromonas* septicemia (MAS); a septicemic disease that produces internal and external lesions and can induce protrusion of the eyes (exophthalmos), followed by mortality within several hours after the manifestation of disease [18,211,213-215]. The first report of disease caused by the v*Ah* pathotype was from *A.*

hydrophila J-1, a strain that would later be categorized under sequence type 251 (ST251) during the 1989 outbreaks of MAS in China's Jiangsu Province where this bacterium was identified as the etiologic agent of mortalities in grass carp (*Ctenopharyngodon idella*). Outbreaks of MAS in farmed carp have persisted in China, resulting in economic losses estimated at 2,200 tons of dead fish per year [32,211,216,217]. Epidemics of ST251-associated MAS occurred within the same province in 2010, with v*Ah* isolates including *A. hydrophila* NJ-35 identified as the etiologic agent [218]. Another v*Ah* isolate from China was *A. hydrophila* ZC1, which was isolated from a grass carp exhibiting signs of hemorrhagic septicemia at an aquaculture farm within China's Guangdong Province [219]. In general, MAS has occurred each summer and resulted in significant losses for the Chinese aquaculture industry, with estimates of losses of at least five billion yuan per year (Prof. Hui Chen, personal communication).

In 2004, the first reported case of ST251-related MAS in the US arose when *A. hydrophila* S04-690 was isolated from a diseased channel catfish (*Ictalurus punctatus*) after an outbreak of MAS in a catfish farm within Washington County, Mississippi (MS), USA [220]. Beginning in 2009, v*Ah* strains were consistently recovered from recurring outbreaks of MAS in aquaculture ponds in western Alabama (AL) with a reported 2,000 tons of dead fish in the first year [212]. To date, this number has grown to exceed an estimated 10,500 tons with v*Ah* isolates representing the largest percentage (35%) of disease cases at the Alabama Fish Farming Center [221]. Although representative data on production losses attributed solely to v*Ah* are difficult to attain, the threat that v*Ah* represents to warm-water aquaculture industries is clear.

Comparative genomic analyses of v*Ah* strains, isolated from catfish in the US and carp in China, previously indicated these strains share a recent common ancestor [220]. Within this monophyletic clade, v*Ah* strains isolated from carp and catfish have unique phenotypes and

genotypes (L-fucose metabolism, an inducible prophage and the ability to use *myo*-inositol as a sole carbon source) that distinguish them from strains of *A. hydrophila* that are not associated with epidemic outbreaks (non-v*Ah*) [32,54,220]. In addition to these results, these studies also showed that MS v*Ah* strain S04-690 is more closely related to the carp isolate ZC1 than to other v*Ah* strains isolated within AL [220]. Since this study was published, MAS outbreaks in the MS delta region have increased, with annual losses approaching 50 tons. The purpose of this study was to characterize the v*Ah* pathotype by examining v*Ah* strains collected from farmed catfish in AL and MS in recent years and compare genome sequences of all available ST251 strains together with other *Aeromonas* spp. genomes available in GenBank by conducting a phylogenomic analysis, a study of the presence of putative virulence factors, a gene linkage model and of the relative capacity of selected strains to cause MAS.

3. Methods.

3.1. Bacterial strains: Disease Isolates and Catfish Challenge.

Channel catfish raised in production ponds in western AL demonstrating symptoms typical of MAS were collected in a moribund state and then submitted for diagnostic evaluation and necropsy at Auburn University. Liver and kidney tissues were sampled for aerobic bacterial cultures and preserved for histopathology in 10% neutral buffered formalin. Samples of tissue were homogenized in sterile phosphate buffered saline and portions of the homogenate streaked onto tryptic soy agar (TSA; Beckton Dickinson, Franklin Lakes, NJ) or brain heart infusion (BHI; Beckton Dickinson) for bacterial isolation. Pure cultures were identified as v*Ah* strains by the v*Ah*-specific qPCR method previously described and/or by utilization of *myo*-inositol as a sole carbon source [222]. From this sampling design, *A. hydrophila* isolates ML09-119, ML10-51K,

S04-690, S14-296, and S14-452 were cryogenically preserved (mixed with 50% glycerol, stored at -80°C) and subsequently used in the catfish immersion challenge.

3.2. Bacterial Strains: Comparative Genomics.

In total, 61 complete and draft *Aeromonas* spp. genomes were included in this study (**Table 1**). Genomes from isolates other than *A. hydrophila* were included based on previous research that recognized erroneous affiliation with *A. hydrophila* to serve as a reference point [16,223]. These genomes were retrieved from the US National Center for Biotechnology Information (NCBI) GenBank database and included two *A. caviae* isolates, one *A. dhakensis* isolate, one *A. enteropelogenes* isolate, one *A. media* isolate, one *A. molluscorum* isolate, one *A. taiwanensis* isolate, one *Aeromonas sp.* isolate, 26 *A. hydrophila* disease isolates, and 28 non-epidemic *A. hydrophila* isolates.

3.3. Genome Sequencing.

Strains were selected for Illumina sequencing based on the results of the v*Ah* genotype-specific PCR (see below) so representative isolates from each geographic area that experienced MAS outbreaks are present. Genome sequencing with 250 bp read-length using paired-end sequencing was performed on the Illumina MiSeq platform using the Nextera XT kit (Illumina, San Diego, CA) to prepare bar-coded fragment libraries according to the manufacturer's protocol, with sequence reads trimmed and assembled *de novo* using the CLC Genomics Workbench (Qiagen, Redwood City, CA) using default settings. A total of 14 v*Ah* strain draft genomes were generated for strains Ahy_Idx7_1, ALG15-098, IPRS-15-28, ML10-51K, S13-612, S13-700, S14-230, S14-296, S14-458, S14-606, S15-130, S15-242, S15-400, and S15-591. In addition to a standard

Illumina MiSeq run for v*Ah* strain S14-452 a NxSeq 20 kb mate pair library was constructed and sequenced using an Illumina MiSeq at the Lucigen Corporation (Middleton, WI). The *de novo* assembly resulting from the standard Illumina MiSeq sequences resulted in 13 contigs (80.15 average coverage) whereas the combination of these sequences together with the mate pair-derived sequences using *de novo* assembly with SPAdes (v3.5.0) resulted in a complete genome sequence.

3.4. Lineage-Specific PCR for Unique v*Ah* Genotypes.

Primers specific to v*Ah* genotypes were developed based on sequences unique to *A. hydrophila* strains ML09-119 (NC_021290), S14-452 (SAMN05256776) and ZC1 **(**SAMN02404465). To determine v*Ah* genotype affiliation for disease isolates in AL and MS, genomic DNA was isolated from each isolate (Gentra Puregene DNA isolation kit; Qiagen, Hilden, Germany) and used as a template in a 25-µl PCR that comprised of 13 µl of Econotaq Plus Green 2x MasterMix (Lucigen, Madison, Wisconsin, USA), 20 picomoles of each oligonucleotide primer and 50 ng of template gDNA. The samples were run on a C1000 Touch thermal cycler (BioRad, Hercules, California, USA) with an initial denaturation of 94^{o}C for 3 minutes followed by 35 cycles of 94^{o}C for 30 seconds, 58^{o}C for 30 seconds, and 72^{o}C for 1 minute, with a final extension at 72^{o}C for 5 minutes. Amplicons were electrophoretically resolved through a 1% agarose gel in the presence of ethidium bromide and visualized under ultraviolet light to confirm the presence of appropriate sized bands.

After genome sequencing of isolates identified by the previously described PCR method, new primers were subsequently designed that are specific to v*Ah* genotypes based on unique nucleotide sequences that were distinct, yet conserved within the respective v*Ah* genotype

(Table 2). Candidate sequences were identified by comparative analysis of annotations produced by the Rapid Annotations using Subsystems Technology [181] v2.0 server, followed by *in silico* evaluation using Geneious v. R9. With respect to binding locations, the v*Ah*-specific primer set targets a region within the gene that encodes serine protease and is predicted to produce amplicons in all known v*Ah* isolates; the JBN2301 lineage-specific primer set targets a genetic locus predicted to encode a hypothetical protein and is predicted to produce amplicons in only JBN2301 (indicated in blue in **Figure 1B**); the ML09-119 lineage-specific primers target a genetic locus predicted to encode MobA and is predicted to produce amplicons in 16 v*Ah* isolates (indicated in green in **Figure 1B**); the S04-690 lineage-specific primer set targets a genetic locus predicted to encode a hypothetical protein and is predicted to produce amplicons in only S04-690 (indicated in red in **Figure 1B**); the S14-452 lineage-specific primer set targets the COG3339 genetic locus and is predicted to produce amplicons in five v*Ah* isolates (indicated in purple in **Figure 1B**); and the ZC1 lineage-specific primers targeted a locus encoding a hypothetical protein (indicated in orange in **Figure 1B**), which is shared with other Asian isolates like J1 and NJ-35, but is not predicted to produce an amplicon in these other Asian carp isolates based on *in silico* analysis.

3.5. Histopathology of Channel Catfish in an Immersion Challenge Model.

Channel catfish were obtained as fry from the Warmwater Aquaculture Research Unit in Stoneville, MS and reared to experimental size in 340-liter troughs supplied with 26 ± 2°C dechlorinated municipal water under pathogen-free conditions. All animal experiments were approved by, and conducted in compliance with, regulations of the Institutional Animal Care and Use Committee of the Aquatic Animal Health Research Unit (USDA-ARS) in Auburn, Alabama.

Water temperature was maintained at 26 ± 2°C with a centralized heater. Prior to trials, heart, liver, head kidney, trunk kidney, spleen, brain, and skeletal muscle tissues were collected from 10 randomly sampled fish to verify fish were not presently infected with v*Ah*.

Two hundred catfish fingerlings, with a mean weight of 111 ± 47g and length of 19 ± 3 cm, were acclimated for 12 days in 56-liter glass aquaria (10 fish per tank, 3 tanks per isolate and 2 tanks mock infected) containing about 50-liter water prior to challenge. The immersion challenge was conducted using the recently described fin clip method [224]. At the time of infection, aquarium water was reduced to 15 liters per tank. To sedate animals for handling, fish were netted from individual aquaria and placed into a container filled with 20 liters of dechlorinated water containing 150 mg/L of buffered Tricaine-S (tricaine methanesulfonate; Western Chemical, Inc., Ferndale, WA). After fish were anesthetized, the adipose fin was clipped at its base and fish were returned to aquaria for recovery from anesthesia.

For the bacterial challenge, 100 mL of tryptic soy broth (TSB) containing approximately 3.0 x 10^9 CFU/mL of the respective *A. hydrophila* strains (ML-09-119, ML10-51K, S04-690, S14-296, S14-452, and ZC1) was added to each of three aquaria, with the resultant concentration of bacteria being ~2.0 x 10^7 CFU/mL within aquaria. Two tanks served as mock infected controls, receiving only 100 mL sterile TSB. After one-hour exposure, water flow to aquaria (0.5 L/min) was resumed. Fish mortality was monitored daily for seven days. At least 50% of dead fish were sampled for confirmation for the presence of v*Ah* in liver and kidney tissues using M9 minimal medium containing 0.3 % (w/v) *myo*-inositol (M9I) agar [222]. Moribund fish were removed from aquaria daily and surviving fish were euthanized by at least 15 min exposure to 300 mg/L buffered Tricaine-S solution, then necropsied. Heart, liver, head kidney, trunk kidney, spleen, brain, and skeletal muscle tissues were harvested and fixed in 10% buffered formalin for

histopathology. Tissues were also collected and used for bacterial identification and quantitation. Formalin fixed tissues were processed and embedded in paraffin. The tissues were cut in 4 micron sections, stained with hematoxylin and eosin and evaluated for microscopic lesions by light microscopy.

3.6. Core Genome Analyses.

A core genome was created using both coding and noncoding sequences of 61 genomes labelled as *A. hydrophila* within GenBank, some of which were mislabeled as they have other species affiliations (e.g. *A. dhakensis* SSU). Specifically, any contigs less than 10Kbp in size were first filtered from draft genomes in order to increase computational efficiency. Filtered data were then submitted as FASTA files to the multiple whole genome alignment tool Mugsy v1.2.3 [225] under default parameters. The resulting alignment was subsequently processed with GBLOCK v0.91b [226] in order to identify regions of high conservation across all isolates. Parameters for retention by GBLOCK are dictated by the input alignment and were: a minimal of 31 and 51 sequences for conserved and flanked positions, respectively, a maximum of 8 contiguous, but non-conserved positions, a minimal block length of 10, and one-half of the sequences allowed to possess gapped positions within a block. From the final alignment, a maximum likelihood [3] phylogeny for the 61 *Aeromonas* spp. isolates, including 54 isolates labeled in GenBank as *A. hydrophila*, was inferred using RAxML v8.2.8 [227] under the General Time Reversible model of evolution with estimated proportions of invariable sites and rate variation among sites (i.e. GTR+I+G) and 1,000 bootstraps to determine branch supports. Trees were visualized using Archaeopteryx v.beta 0.9901.

Following generation of a consensus sequence, the National Microbial Pathogen Data Resource (NMPDR) Rapid Annotations using Subsystems Technology [181] v2.0 server was used in conjunction with the SEED v2.0 algorithm to annotate the core genome and generate metabolic models [228,229]. These predictive models were evaluated using both protein-protein Basic Local Alignment Search Tool (e.g., BLASTp and BLASTx) algorithms through GenBank as well as the Joint Genome Institute's Genomes OnLine Database (GOLD) v5.0.

3.7. Calculating Average Nucleotide Identity.

To assess overall genetic similarity, the average nucleotide identity (ANI) comparison of 61 *Aeromonas* spp. genomes was evaluated using JSpecies (v1.2.1) and cross-validated with the Konstantinidis lab ANI calculator [230,231]. ANI values >96% indicated strains belong to the same species according to criteria used for the genus *Aeromonas* [15,16].

3.8. v*Ah* Differential Gene Identification.

To evaluate differences in gene content among the 61 genomes, data from the PAthogen Resource Integration Center [232] protein family sorter tool, RAST/SEED gene annotations, the Pathogen Host Interaction database (http://www.phi-base.org), and the Virulence Factors of Pathogenic Bacteria databases (http://www.mgc.ac.cn/VFs) were combined with copy number data for previously identified virulence genes [233]. To identify differential genes, results were evaluated by comparing predicted virulence-associated genes with closely and disparate related isolates. Notably, genes shared between v*Ah* and non-v*Ah* isolates or those unique to an individual strain were removed from downstream analyses. These data were subsequently coupled with screening of virulence factors and v*Ah*-associated genes using searches of the

NCBI GenBank database with MegaBLAST and BLASTn algorithms [3]. Thresholds for absence were specific to the respective gene and were restricted to mutations altering the predicted functional domains of proteins in which the protein sequence in question returned a functionally divergent protein. Once validated, these gene clusters and virulence factors were transformed for statistical analyses using Orange Data Mining software v.3.3.5 and/or R Studio v.0.99.896. After pre-processing in R Studio using packages MuMIn v.1.15.6, randomForest v.4.6-12, and kmeans v.0.1.1, heat maps of resultant data were generated using Orange Data Mining software. To analyze subclade differences, v*Ah* and non-v*Ah* strains with known virulence properties (n=25) were evaluated using k-means at 20 clusters (100% between sum of squares / total sum of squares). In addition to these analyses, the T346 Secretion System Hunter (version is not published) was used to identify type VI secretion system (T6SS)-associated gene clusters using Glimmer v3.02, and HMMER3 v3.1b2 [234].

4. Results.

4.1. Identifying New v*Ah* Isolates.

The S04-690 genome was previously found to be the genetic intermediate (raw genetic distance) between US catfish (represented by strain ML09-119) and Asian carp (ZC1) v*Ah* genotypes [220]. Subsequently, disease isolates of *A. hydrophila* from AL and MS were obtained in 2013-2015 from MAS outbreaks that were confirmed as v*Ah* by phenotypic (*myo*-inositol usage) and/or genomic tests (qPCR) [222]. A total of 38 v*Ah* isolates were identified from Mississippi from 2013-2015, of which 18 gave PCR products using the S14-452-specific primer set (data not shown**)**.

4.2. Core Genome Analysis.

Alignment of the complete and filtered draft genomes of the 61 *Aeromonas* spp. genomes via mugsy produced a matrix of 19,817,762 positions. Following processing with GBLOCK, the core genome of these 61 strains contained 32,401 blocks and a consensus of 3,776,490 bp. This included 388,235 complete, 120,049 variable, and 79,507 informative sites, with percent G+C composition of 62.6%. Notably, these conserved regions collectively have a higher percent G+C content than the 61-isolate percent G+C average of 61.1% (with a range of 60% for the genome of the type strain of *A. molluscorum* 848^T to 63.2% for the one of *A. taiwanensis* LMG24683^T; p-value = 0.003). Within complete sites, the average transition to transversion ratio was 1.437 for all sequence pairs, with a minimum of 0 transitions and 1 transversion in both *A. hydrophila* S04-690 and S13-700 as well as a maximum of 13 transitions and 2 transversions in the genomes of the strains labelled as *A. hydrophila* 113 and 14 in GenBank, but cluster with *A. dhakensis* (**Figure 1**).

4.3. Core Genome Phylogeny.

Conserved sequences from the core genome analysis were used to infer phylogenetic relationships among the 61 *Aeromonas* spp. isolates **(Figure 1A)**. The core genome phylogeny indicates with 100% bootstrap support that v*Ah* strains form a monophyletic group that is fundamentally distinct from other *A. hydrophila* **(Figure 1A)**. Building on previous research within these ST251 isolates, this phylogenetic analysis also revealed support for distinct clades among v*Ah* isolates. For example, *A. hydrophila* ZC1, isolated from a grass carp in China **(Figure 1B)**, is affiliated with v*Ah* strains isolated from catfish in MS (S14-452, S14-458, S15-130, S15-400, and S15-591). Each of these ZC1-affiliated strains, as well as the other strains

isolated from carp in China (i.e. J-1 and NJ-35), were found to contain at least 80% of the core proteins necessary for a complete T6SS, with ZC1 and S14-452 having two separate T6SS-associated gene clusters whereas J-1 and NJ-35 each have a single complete cluster (**Figure 2**). In contrast, v*Ah* isolates from catfish in AL as well as other MS isolates (i.e., S13-612, S13-700, S14-606, and S14-296) formed a distinct subclade that lack the majority of the core T6SS genes **(Figure 1B; Figure 2)**. Notably, while AL and MS v*Ah* strains lacked the majority of T6SS components, they consistently retain genes for valine-glycine repeat protein G (VgrG), a T4 bacteriophage tail-like hole forming protein [235]; hemolysin coregulated protein (Hcp), a repetitive tubular protein that is similar to the phage major tail protein gpV [236]; the chaperone protein clpB, a chaperone and ATPase that interacts with Hcp to translocate effectors [237]; and vasH, a putative transcriptional regulator [238] **(Figure 2)**.

4.4. Average Nucleotide Identity.

The average nucleotide identity values of the 61 *Aeromonas* spp. genomes was determined (**Figure 3).** High ANI values (>99%) were found for all v*Ah*-v*Ah* pairwise comparisons of *A. hydrophila*, supporting the core genome phylogeny (**Figure 1B**; **Figure 3**). In contrast, all v*Ah* comparisons with non-v*Ah* isolates possessed ANI values less than 97%. Building on previous results, 14 strains appear to have a discrepancy between the GenBank species assignation and the ANI species assignation **(Table 1)**.

4.5. Virulence of v*Ah* Strains in Channel Catfish.

Challenge of channel catfish with v*Ah* isolates from AL and MS resulted in ≥ 60% mortality **(Figure 4)**. Within these US v*Ah* isolates tested, there were no observed differences in

virulence among ML09-119, ML10-51K, S04-690, S14-296, and S14-452, based on Duncan's multiple range test (p-value > 0.05). The carp isolate ZC1 was less virulent than AL and MS isolates with only 26.7% mortality observed **(Figure 4)**. Most mortality (~96%) occurred within 48 hours post challenge for all isolates including ZC1. All dead fish (100%) sampled for confirmation were positive for the presence of v*Ah* in liver tissue. Control fish yielded no mortality from the mock challenge.

4.6. Pathology of Epidemic *A. hydrophila* Infections.

Cutaneous lesions observed in fish infected with v*Ah* by immersion challenge included extensive hyperemia over the pale ventrum of the fish, in tissues surrounding and within the mouth and at fin bases **(Figure 5A)**. There was also extensive hyperemia around the eyes and exophthalmos in some fish. Cutting into the muscle of the lateral body wall revealed multifocal to coalescing foci of congestion/hemorrhage. Gill lesions were variable; gills were pale in some fish and reddened in others. Internally, there was widespread hyperemia of abdominal organs as well as petechial and ecchymotic hemorrhages scattered over mesenteric tissues. The spleen was moderately to severely swollen and dark red **(Figure 5B)**. Head and trunk kidneys were moderately edematous and red and friable when harvested. The intestinal tract was mildly to moderately dilated and red **(Figure 5B)**. The liver was mildly to moderately swollen with slightly rounded edges **(Figure 5B)**. Glisson's capsule was peppered with variable numbers of petechial and ecchymotic hemorrhages. The atrial chamber of the heart was dilated and filled with blood.

Histopathologic lesions observed were strongly suggestive of a septic disease and included edema and necrosis in many internal organs. Splenic lesions included ellipsoidal

necrosis and congestion/hemorrhage in the splenic red pulp **(Figure 5C)**. In scattered necrotic ellipsoids variable numbers of short, rod-shaped bacteria were observed. In the liver, there was often necrosis of the acinar pancreatic tissue surrounding hepatic vessels **(Figure 5D)**. Scattered acinar cells were rounded and necrotic and extracellular zymogen granules could be observed in the necrotic exudate. Small numbers of inflammatory cells including macrophages, lymphocytes, and neutrophils were observable in scattered areas of pancreatic acinar necrosis. Multifocally within the hepatic parenchyma were variable sized foci of hepatocellular necrosis characterized by the presence of small aggregates of degenerating and necrotic hepatocytes. Hematopoietic cells in the renal interstitum and renal epithelial cells lining scattered renal tubules were undergoing degeneration and necrosis. In sections of head kidney, the tissue was edematous and congested/hemorrhagic. There was scattered degeneration and necrosis of red and white cell precursors. Intestinal lesions were minimal in fish infected in this fin clip model and were limited to mild congestion and hemorrhage of the vessels in the lamina propria and vessels of the muscularis and serosa. In some sections of heart there was mild necrosis of myofibers in the myocardium. Vessel of the brain were often moderately dilated and congested but the neuropil of the cerebrum, cerebellum and brain stem was normal. The epithelium of the gills was normal but branchial capillaries were sometimes dilated and congested with erythrocytes.

4.7. Differentiating Pathotypes.

In order to robustly define the v*Ah* pathotype-specific loci in this study, previous results on established *A. hydrophila* virulence factors as well as unique genes were used in combination with a clustering approach and a random forest decision tree to identify genes that may contribute to functional differences in virulence. This approach confirmed previously described

gene clusters for L-fucose and O-antigen biosynthesis as well as *myo*-inositol catabolism. Additionally, we identified predicted virulence factors conserved within all v*Ah* strains **(Table 3)** which includes virulence factors well known to be important in *A. hydrophila* pathogenesis [233]. Among predicted virulence factors conserved among v*Ah* strains, there were many genes uniquely associated with v*Ah* strains that were not present in other sequenced *A. hydrophila* strains, including L-serine dehydratase, N-acetylmannosamine kinase, N-acetylneuraminate lyase, and queuosine-arachaeosine **(Table 4)**. In a more comprehensive approach, all RAST/SEED predicted genes from the 41 confirmed *A. hydrophila* genomes (26 v*Ah* and 15 non-v*Ah*) included in this study were evaluated on the basis of linkage with the v*Ah* pathotype using exhaustive iterations of random forest modeling, which resulted in 26 genes uniquely associated with v*Ah* by either presence/absence or by differential copy number when compared to non-v*Ah* **(Figure 6)**.

5. Discussion.

The core genome analysis of epidemic *A. hydrophila* strains, from several US states and Chinese provinces (identified within the literature as ST251/v*Ah*), supports the genetic and functional unification of these hypervirulent bacteria within a monophyletic clade. These data are in agreement with previous reports based on single or multiple genetic loci from smaller numbers of v*Ah* strains [220,239]. Interestingly, this study found evidence for genomic heterogeneity among the sampled v*Ah* strains that may reflect geographic origin and/or host switching. For example, the isolates obtained from MS in 2013-2015 are affiliated with two different ST251/v*Ah* clades, specifically the Asian carp-affiliated clade and the US catfish-affiliated clade. In contrast, the v*Ah* isolates from diseased catfish in AL from 2009-2015 reflect a single, clonal clade. These

data suggest that v*Ah* has greater diversity within MS aquaculture ponds compared to those of AL. As previously hypothesized, this pattern would fit with a dissemination model in which carp or other fish or fish products from Asian source(s) were first introduced to the Mississippi delta region, after which a particularly more virulent v*Ah* lineage spread among farmed catfish resulting in the initial epidemic outbreaks within AL.

Genomic comparisons indicate that all members of the v*Ah* pathotype strains share unique genetic loci that may be a result of their close genetic affiliation and may also contribute to their pathogenesis. The previously identified v*Ah*-specific gene clusters of L-fucose and *myo*-inositol catabolism were also confirmed in this study. The use of *myo*-inositol as a sole carbon source is a rarely observed phenotype among *Aeromonas* species and, to our knowledge, has only been reported in v*Ah* strains and strains of *A. finlandiensis* [240]. This study identified additional genetic loci that are present in all sequenced v*Ah* strains and may have a contribution to virulence, such as L-serine dehydratase, a N-acetylmannosamine kinase, a N-acetylneuraminate lyase, a sialic acid transporter, a transcriptional regulator of pyridoxine metabolism, an archaeosine tRNA-ribosyltransferase, a gene product required for queuosine biosynthesis (QueD), an acriflavine resistance protein A, and an IS5 transposase and transactivator. For example, in *Campylobacter jejuni* a L-serine dehydratase is essential for gut colonization [241]. With regard to sialic acid, *Vibrio cholera* has been shown to use this system to evade the innate immune response [242], and promote the binding of cholera toxin to the host intestinal epithelium [243]. In addition, increased acriflavine resistance may provide a selective advantage to v*Ah* considering that acriflavine is a commonly used antiseptic in aquaculture [244].

Expanding on this approach, all annotated genes were subsequently evaluated for linkage with the v*Ah* pathotype, a method that removes the bias inherent in assuming that known

virulence factors are the source of genetic heterogeneity within *A. hydrophila*. This analysis revealed that 26 genes are synonymous with the v*Ah* pathotype either by presence/absence or by copy number **(Figure 6)**. Supporting the robust nature of this approach, previously described genes that contribute to virulence were also identified, including genes associated with the *myo*-inositol catabolic pathway. Of note, within this pathway a methylmalonate-semialdehyde dehydrogenase that is required for *myo*-inositol catabolism was identified in both v*Ah* and non-v*Ah* isolates; however, this enzyme is required for both *myo*-inositol and valine metabolism and there is no indication that non-v*Ah* strains have the genetic capacity for *myo*-inositol catabolism. In addition to these results, this method also identified known *Aeromonas* spp. virulence factors, such as a DNA adenine methylase that has been identified as a regulator of virulence and required for viability in *A. dhakensis* SSU [245]. Putative v*Ah* virulence factors were also identified, such as a hypothetical protein with a LuxR-like domain. LuxR has been previously shown to be a regulator of virulence factors and quorum sensing within *Aeromonas* spp. [246]. Lastly, other genetic loci were identified as being associated with v*Ah* strains, such as the type II/IV system secretin RcpA/CpaC that is putatively involved in flp pilus assembly [247], but to our knowledge the contribution of this secretin to *A. hydrophila* pathogenesis has yet to be experimentally determined.

The histopathologic examination of tissues from *A. hydrophila*-infected farmed fish show a wide range of severity related to internal lesions, with some fish exhibiting minimal lesions and others having widespread sepsis with necrosis of spleen, liver, renal tissue, intestine, and brain tissues **(Figure 5)** with subsequent high rates of mortality occurring throughout affected farms. For fish challenged in aquaria via intraperitoneal injection, rapid onset of mortality without these disease sequelae was observed [220]. In this study, fish were challenged with an immersion model

and exhibited significant clinical signs including cutaneous and ocular hemorrhaging, splenic and renal congestion, and hemorrhage with mild to moderate necrosis of internal organs. Despite the high genetic similarity of the strains (ANIs > 99%, **Figure 3**), strain ZC1 had reduced virulence (~27% mortality) when compared with ML09-119; ML10-51K; S04-690; S14-296; and S14-452 which caused ≥ 60 % mortality in channel catfish **(Figure 4)**. The reduced virulence of strain ZC1, relative to strain ML09-119, in carp and in catfish was previously observed when fish were challenged intraperitoneally [220]. Interestingly, comparatively few reports of mortalities due to MAS have come from the state of MS which may reflect a number of geographic differences as well as the heterogeneity of v*Ah* strains present within MS aquaculture ponds or differences in management practices and/or in environmental conditions.

To identify genetic elements that may contribute to increased virulence, genomic comparisons were conducted between ZC1 and the v*Ah* strains included in the immersion challenge. These analyses revealed all other v*Ah* isolates tested in the disease challenge, but not strain ZC1, contain the mobile element protein from the helix-turn-helix (HTH) superfamily (InsE multi-domain) as well as the phage antirepressor protein from the *antA* superfamily present in the *ibrAB* island of a Shiga-toxin converting prophage found within *Salmonella enterica* serovar *Typhi* and *Escherichia coli* O157:H7 [248]. The presence of these transcription factors within the highly virulent v*Ah* strains could imply increased virulence factor expression in these strains. Furthermore, v*Ah* isolates from Asian carp or the ZC1-affiliated clade from MS (including S14-452) were found to have a potentially functional T6SS, but the clade comprised solely of AL and MS catfish isolates (e.g. ML09-119, ML10-51K, S04-690, and S14-296) were found to consistently lack 9/13 core T6SS genes related to secretion of virulence factors such as Hcp that lead to antimicrobial activity, but also stimulate the host's immune response. While

ZC1 appears to contain all core components, it is important to recognize that *clpB* is not present within either of ZC1's T6SS gene clusters and therefore may be under different regulation than in NJ-35 or J-1.

While experimental studies are required to evaluate T6SS functionality, if the T6SS is entirely absent in the AL/MS v*Ah* subclade (e.g. ML09-119), it stands to reason that this loss would cause virulence to be attenuated, but the immersion model of disease suggests that these AL/MS disease isolates are highly virulent compared to strain ZC1. Since previous disease challenges with strain ZC1 indicate that it has reduced virulence relative to the AL/MS v*Ah* clade in both catfish and carp [220], this suggests that attenuated virulence observed in strain ZC1 is not solely attributable to host physiological differences. The v*Ah* strains in the AL/MS subclade encode only a subset of T6SS components related to secretion that include VgrG, Hcp, ClpB, and VasH. The absence of T6SS core genes, including *vca0107-0109, vca0111-0114, vca0118, vca0119, vca0121, vasK, clpV, vasF* and *vasA*, largely accounts for the ~16 kbp net difference in the genome sizes between strains ZC1 and ML09-119. Given the absence of these genes and their established role in virulence within other bacteria, compared to strains isolated in China (i.e. NJ-35, J-1 and ZC1), the v*Ah* strains in the AL/MS subclade are hypothesized to have reduced motility, increased intestinal adherence, reduced antimicrobial activity due to no longer secreting Hcp, and exhibit better evasion of the host's immune response due to no longer triggering a host response from proteins secreted via the T6SS, based on studies of T6SS in *Aeromonas dhakensis* SSU and in *Salmonella enterica* serovar Typhi [58,249]. Future research efforts should aim to elucidate the role of these T6SS components, and the effect of a reduced or rearranged T6SS, on v*Ah* virulence.

As new v*Ah* isolates emerge and our collective knowledge of v*Ah* genomic diversity grows, future research should investigate the spread of these pathogens as well as improve the design of more effective biosecurity strategies for the aquaculture industry. Toward this effort, we report primer sets that differentiate hypervirulent *A. hydrophila* from non-epidemic *A. hydrophila* as well as primers to distinguish between known v*Ah* lineages and propose that these tools be employed as a method of efficiently typing future isolates to track the spread and persistence of v*Ah*. This study supports the use of either the v*Ah*-specific qPCR assay or *myo*-inositol growth assay as valid primary methods for establishing an *A. hydrophila* isolate as being affiliated with this v*Ah* pathotype [222], after which the v*Ah* subclade-specific primer sets could be used to differentiate among the v*Ah* lineages. These molecular and phenotypic tests are vital tools that can be used to map the worldwide distribution of v*Ah* in carp, catfish, and other warm-water fish species (e.g. tilapia) and therefore the magnitude of threat that the v*Ah* pathotype represents and the challenges involved in developing effective MAS disease control methods. Disease control strategies should take into account the variability observed in this study among v*Ah* strains and evaluate the efficacy of vaccination or other control measures against a panel of strains that represent the known diversity of this highly virulent *A. hydrophila* pathotype. In conclusion, hypervirulent *A. hydrophila* within ST251 have emerged as pathogens of farmed warmwater fishes that are classified within the v*Ah* pathotype based on strong phylogenetic evidence that includes a core genome phylogeny and ANI values >99%; metabolic activities that are unique within this species, such as *myo*-inositol and sialic acid metabolism; a suite of conserved *Aeromonas* spp. virulence factors; 26 conserved genetic loci putatively linked with virulence; and the ability to induce motile *Aeromonas* septicemia, which is characteristically

followed by rapid mortality in multiple species of farmed fish. Collectively, these traits distinguish v*Ah* from non-epidemic *A. hydrophila* and define the v*Ah* pathotype.

Table 1. Bacterial genomes used in comparative genomic analyses. Strains are indicated as virulent *A. hydrophila* (v*Ah*) or other *Aeromonas* spp. based on their core genome-derived phylogenetic affiliation **(Figure 1)**.

Strain	Phenotype	Isolation source	GenBank species assignation	Species based on phylogeny and ANI	Accession	Reference
Ae398	Non-vAh	Human	*A. caviae*	*A. caviae*	SAMEA2272404	Beatson et al., 2011
YL12	Non-vAh	Compost	*A. caviae*	*A. caviae*	SAMN02870964	Lim et al., 2014
AAK1	Non-vAh	Clinical	*A. dhakensis*	*A. dhakensis*	SAMD00036618	Martínez-Murcia et al., 2008
1999lcr	Non-vAh	Clinical	*A. trota*	*A. molluscorum*	SAMN02732394	Dallagassa et al., Unpublished
116	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_ANPN00000000.1	Chan et al., 2011
14	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_AOBM00000000.1	Chan et al., 2011
173	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_AOBN00000000.1	Chan et al., 2011
187	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_AOBO00000000.1	Chan et al., 2011
226	Non-vAh	Clinical	*A. hydrophila*	*A. hydrophila*	NZ_JEML00000000.1	Chan et al., 2011
259	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_AOBP00000000.1	Chan et al., 2011
277	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_AOBQ00000000.1	Chan et al., 2011
4AK4	Non-vAh	Industrial	*A. hydrophila*	*A.* sp. nov.	NZ_CP006579.1	Gao et al., 2013
AD9	Non-vAh	Soil	*A. hydrophila*	*A. hydrophila*	NZ_JFJO00000000.1	Lenneman and Barney, 2014
Ae34	Non-vAh	Koi carp	*A. hydrophila*	*A. hydrophila*	NZ_BAXY00000000.1	Jagoda et al., 2014
AH10	Non-vAh	Grass carp	*A. hydrophila*	*A. hydrophila*	NZ_CP011100.1	Xu et al., 2013
AL06-01	Non-vAh	Bluegill	*A. hydrophila*	*A. hydrophila*	SAMN01085623	Hossain et al., 2013
AL06-06	Non-vAh	Goldfish	*A. hydrophila*	*A. hydrophila*	NZ_CP010947.1	Tekedar et al., 2015
AL10-121	Non-vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	NZ_LRRW00000000.1	Hossain, 2012
AL97-91	Non-vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN04967787	Hossain, 2012
ATCC7966^T	Non-vAh	Milk tin	*A. hydrophila*	*A. hydrophila*	NC_008570.1	Seshadri et al., 2006
BWH65	Non-vAh	Clinical	*A. hydrophila*	*A. caviae*	NZ_LESK00000000.1	Earl et al., 2015
E1	Non-vAh	Clinical	*A. hydrophila*	*A. hydrophila*	SAMN01886638	Grim et al., 2013
E2	Non-vAh	Clinical	*A. hydrophila*	*A. hydrophila*	SAMN01886639	Grim et al., 2013
GA97-22	Non-vAh	Rainbow trout	*A. hydrophila*	*Aeromonas* spp.	SAMN01085627	Hossain, 2012
HZM	Non-vAh	Soil	*A. hydrophila*	*A. caviae*	SAMN02596469	Chua et al., 2015
KOR1	Non-vAh	Mangrove	*A. hydrophila*	*A. dhakensis*	NZ_LJOE00000000.1	Yin et al., 2015
MN98-04	Non-vAh	Tilapia	*A. hydrophila*	*A. hydrophila*	SAMN04967900	Hossain, 2012
RB-AH	Non-vAh	Soil	*A. hydrophila*	*A. hydrophila*	NZ_JPEH00000000.1	Rheault et al., 2015
S14-230	Non-vAh	Tilapia	*A. hydrophila*	*A. hydrophila*	SAMN05292364	This study
SSU	Non-vAh	Clinical	*A. hydrophila*	*A. dhakensis*	NZ_AGWR00000000.1	Ribeiro et al., 2012
TN97-08	Non-vAh	Bluegill	*A. hydrophila*	*A. hydrophila*	NZ_LNUR00000000.1	Hossain, 2012
YL17	Non-vAh	Compost	*A. hydrophila*	*A. dhakensis*	NZ_CP007518.2	Lim et al., 2016
WS	Non-vAh	Water sample	*A. media*	*A. media*	SAMN02472129	Chai et al., 2012
848^T	Non-vAh	Wedge-shells	*A. molluscorum*	*A. molluscorum*	SAMN02471397	Spataro et al., 2013
LMG24683^T	Non-vAh	Unknown	*A. taiwanensis*	*A. taiwanensis*	SAMEA2752407	Colston et al., 2014
MDS8	Non-vAh	Dairy sludge	*Aeromonas* sp.	*A. dhakensis*	SAMN02472124	Raychaudhuri et al., 2013
Ahy_Idx71	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05292361	This study
AL09-71	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	NZ_CP007566.1	Pridgeon et al., 2014
AL09-79	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	NZ_LRRV00000000.1	Hossain, 2012
ALG15-098	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223361	This study
IPRS15-28	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223362	This study
J-1	vAh	Crucian carp	*A. hydrophila*	*A. hydrophila*	NZ_CP006883.1	Pang et al., 2015
JBN2301	vAh	Crucian carp	*A. hydrophila*	*A. hydrophila*	NZ_CP013178.1	Yang et al., 2016
ML09-119	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	NC_021290.1	Liles et al., 2011
ML09-121	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	NZ_LRRX00000000.1	Hossain, 2012
ML09-122	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	NZ_LRRY00000000.	Hossain, 2012
ML10-51K	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223363	This study
NJ-35	vAh	Crucian carp	*A. hydrophila*	*A. hydrophila*	NZ_CP006870.1	Pang et al., 2015
PB10-118	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN01085622	Hossain, 2012
pc104A	vAh	Soil	*A. hydrophila*	*A. hydrophila*	NZ_CP007576.1	Pridgeon et al., 2014
S04-690	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN02404466	Hossain et al., 2014
S13-612	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05292362	This study
S13-700	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05292363	This study
S14-296	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05292365	This study
S14-452	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05256776	This study
S14-458	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223364	This study
S14-606	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05292366	This study
S15-130	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223365	This study
S15-242	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223366	This study
S15-400	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223367	This study
S15-591	vAh	Channel catfish	*A. hydrophila*	*A. hydrophila*	SAMN05223368	This study
ZC1	vAh	Grass carp	*A. hydrophila*	*A. hydrophila*	SAMN02404465	Hossain et al., 2014

*Strains are indicated as virulent A. hydrophila (vAh) or other Aeromonas spp. based on their phylogenetic affiliation (**Figure 1**).*

Table 2. Oligonucleotide primers specific to vAh and five genetically distinct v*Ah* lineages.

Lineage	Direction	Sequence	Region amplified	Amplicon (bp)
JBN2301	Forward	5'-CTGGCGCAAGATGGCATTAC-3'	Hypothetical protein	598
	Reverse	5'-CCCCCTTCTCTTCAGCCTTG-3'		
ML09-119	Forward	5'-GGTGGTGGCCCATGACGGTG-3'	MobA	246
	Reverse	5'-GTGGCTCCTTGTCCGGTCGC-3'		
S04-690	Forward	5'- CGCTGAAAACAACTCGCACA-3'	Hypothetical protein	461
	Reverse	5'- CAAAGCTCTGGCTCGATTGC-3'		
S14-452	Forward	5'-GCCGGTCGGCCCTTTATCG-3'	COG3339	245
	Reverse	5' GAGTCGTACGCGCGTTGTGC -3'		
ZC1	Forward	5'-GCAATTCTGCGGTCACTTCTCG-3'	Hypothetical protein	400
	Reverse	5'-AGCGTACCGTCTCGTCGATATG-3'		
vAh	Forward	5'-AGCATCACCAGCGTTGGCCC-3'	Serine protease	746
		5'-GCCGGGCTGAACTTCCGCAT-3'		

Table 3. Predicted virulence factors that are conserved within v*Ah* strains (not unique to), based on a comparison of significant BLASTn hits between v*Ah* isolates against the VFDB (virulence factors with additional results are marked with an asterisk and are available in supplementary data).

Putative virulence factor	Gene	Uniprot ID	GI	Reference bacterium
3-oxoacyl-acyl carrier protein synthase II	*fabB*	A0A0H2V610	77416726	*Escherichia coli* O6:H1
Acriflavine resistance protein AcrB	*acrB*	P31224	25009252	*Escherichia coli* K12
Aerolysin/hemolysin/cytolytic enterotoxin	*ahh*	Q06303	89276735	*Aeromonas hydrophila* AH-1
Asparaginyl-tRNA synthetase	*asnS*	Q56112	16502162	*Salmonella enterica* serovar Typhi CT18
Cephalosporinase; class C beta lactamase	*ampC*	Q8KU09	21311545	*Aeromonas caviae* CIP 74.32
Enterochelin esterase	*fes*	A0A0H2V760	26106962	*Escherichia coli* O6:H1
Ethanolamine utilization protein EutN	*eutN*	B7LTU3	984388511	*Escherichia fergusonii* ATCC 35469
Flagellar motor switch protein FliN	*fliN*	A0A0H3QVI1	674744044	*Pseudomonas aeruginosa* Stone 130
General secretion pathway protein PulF	*pulF*	P15745	149305	*Klebsiella pneuoniae oxytoca* UNF5023
Protein translocase subunit SecA	*secA*	Q8YJG2	672757090	*Brucella melitensis* biotype 1
Rod shape-determining protein MreB	*mreB*	P0A9X4	557273544	*Escherichia coli* K12
Sodium/proline symporter proline permease	*putP*	P07117	131658	*Escherichia coli* K12
Transcriptional activator NtrC	*ntrC*	O86057	5731350	*Herbaspirillum seropedicae* DCP286A
Twitching motility protein PilU	*pilU*	G3XCX3	15595593	*Pseudomonas aeruginosa* PAO1

Table 4. Virulence factors that are unique to vAh strains, based on a comparison of significant BLASTn hits between *A. hydrophila* isolates against the RAST/SEED database.

Subsystem	Role	GI
Glycine/serine Utilization	L-serine dehydratase	958619257
Inositol catabolism	5-deoxy-glucuronate isomerase	958618826
Inositol catabolism	5-keto-2-deoxygluconokinase	827371814
Inositol catabolism	Epi-inositol hydrolase	612156152
Inositol catabolism	Inositol transport system ATP-binding protein	958621246
Inositol catabolism	Inositol transport system permease protein	958620586
Inositol catabolism	Inositol transport system sugar-binding protein	657060685
Inositol catabolism	Inosose dehydratase	507222178
Inositol catabolism	Myo-inositol 2-dehydrogenase 1	656991783
Inositol catabolism	Myo-inositol 2-dehydrogenase 2	827371809
Inositol catabolism	Transcriptional regulator of the myo-inositol catabolic operon	958618669
Sialic Acid Metabolism	N-acetylmannosamine kinase	827373367
Sialic Acid Metabolism	N-acetylneuraminate lyase	1043232173
Sialic Acid Metabolism	Predicted sialic acid transporter	446588390
Sialic Acid Metabolism	Sugar isomerase involved in processing of sialic acid	958620857
Pyridoxin Biosynthesis	Predicted transcriptional regulator of pyridoxine metabolism	16078013
Phage DNA synthesis	DNA adenine methyltransferase, phage-associated	67483065
Phage capsid proteins	Phage capsid scaffolding protein	516389014
Phage capsid proteins	Phage major capsid protein	507220251
Phage lysis modules	Phage lysin, 1,4-beta-N-acetylmuramidase	511291760
Phage packaging machinery	Phage portal protein	958618794
Phage packaging machinery	Phage terminase small subunit	759443491
Phage packaging machinery	Phage terminase, large subunit	958620694
Queuosine-Archaeosine Biosynthesis	Queuosine biosynthesis QueD, PTPS-I	1043232409
Queuosine-Archaeosine Biosynthesis	archaeosine tRNA-ribosyltransferase type 5	507221161

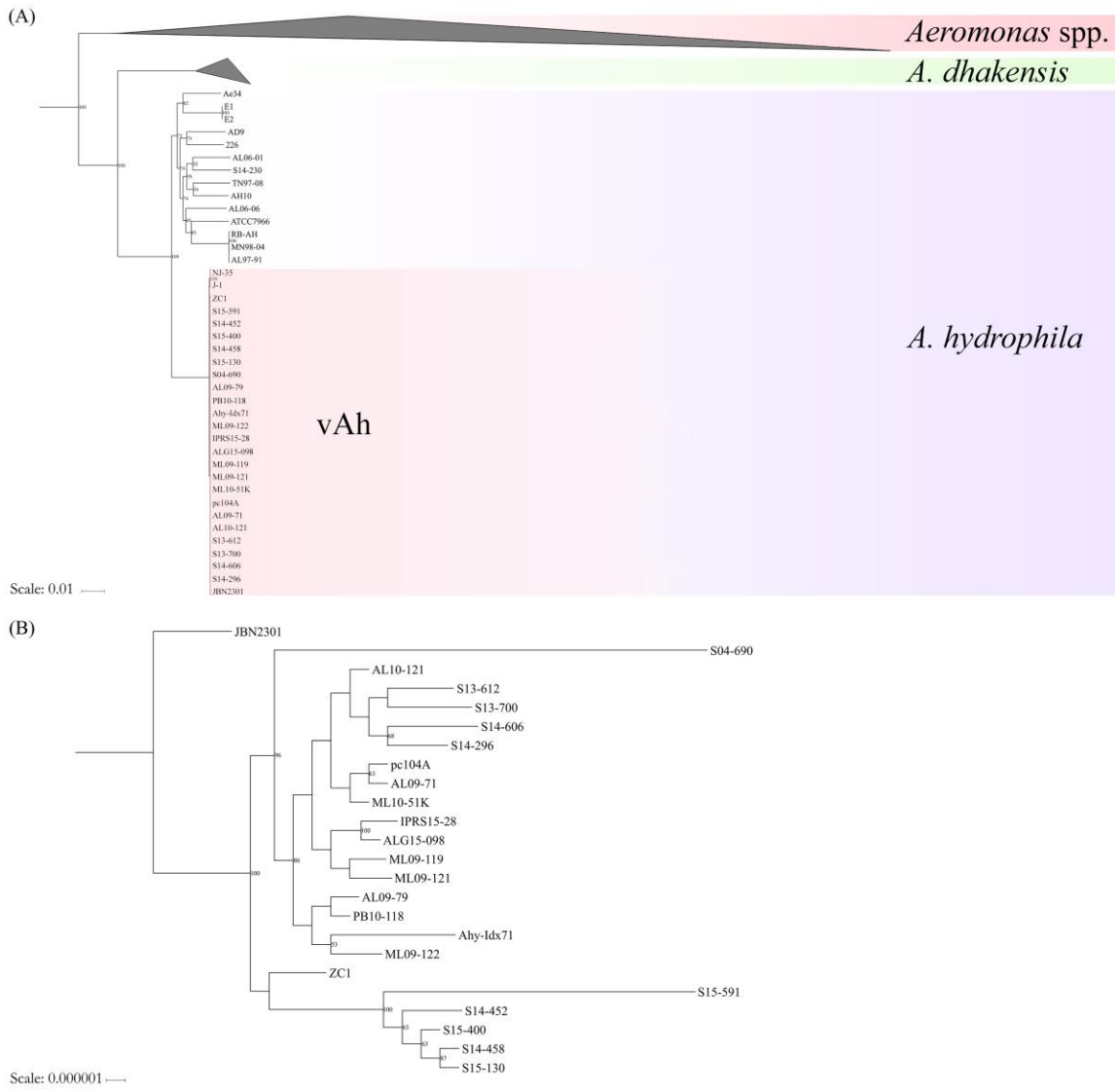


Figure 1. Maximum likelihood [3] phylogeny of (Panel A) *Aeromonas* spp. and (Panel B) v*Ah* isolates based on the core genome of 3.78 Mb conserved among these bacterial strains.

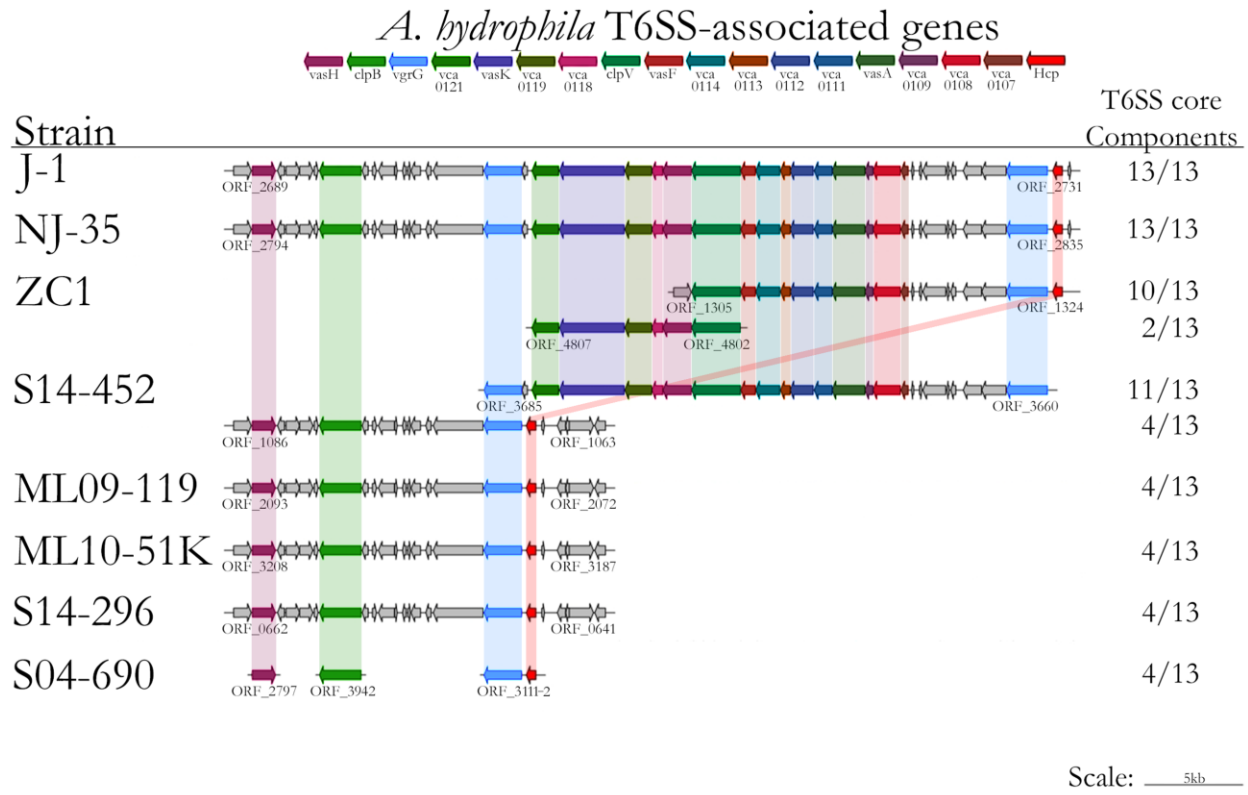


Figure 2. Type VI secretion system gene prediction using the T346 Secretion System Hunter, with results including strains included in the immersion catfish challenge (ML09-119, MNL10-51K, S04-690, S14-296, S14-452, and ZC1) and representatives from Chinese strains (J-1, NJ-35, and ZC1).

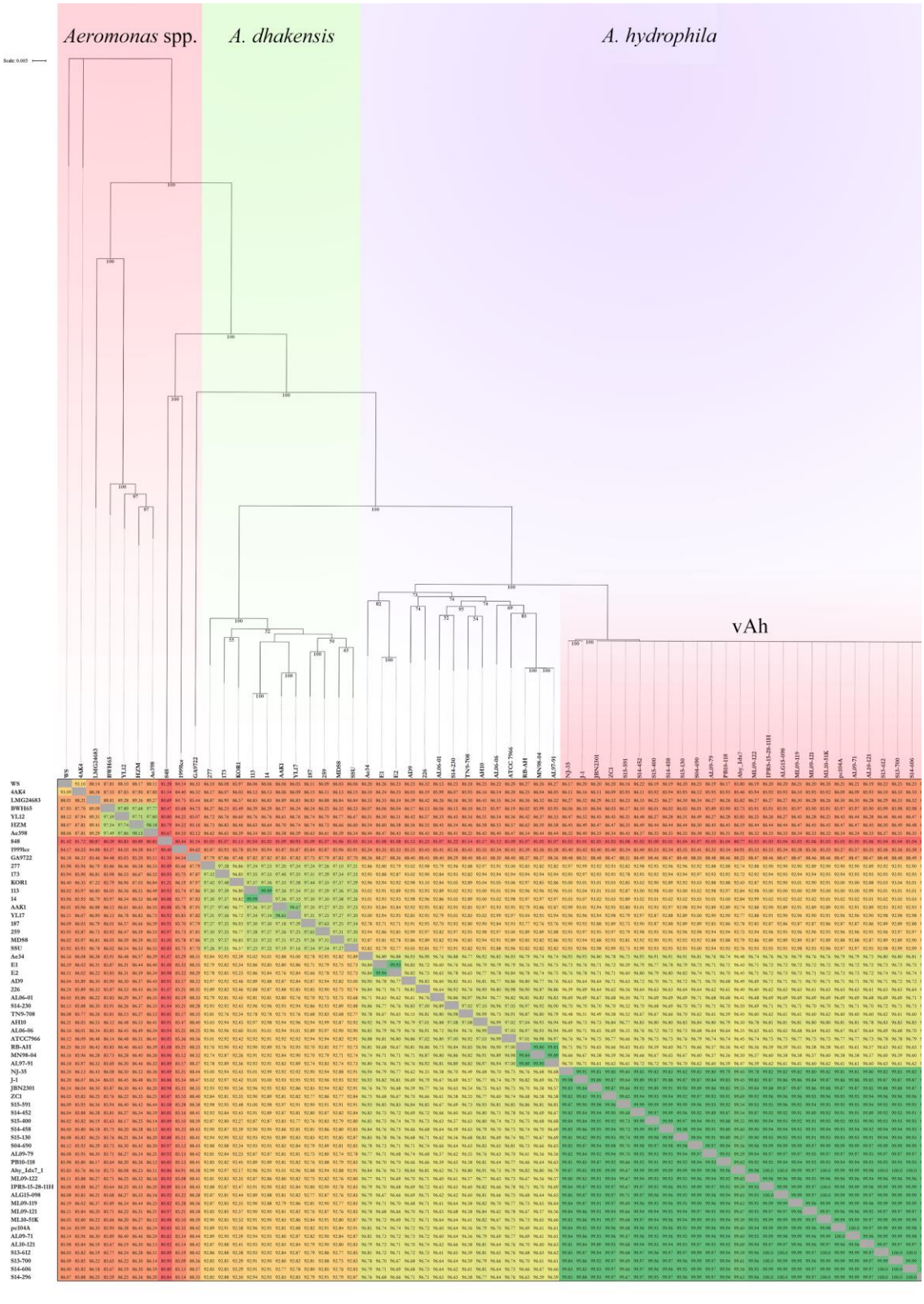


Figure 3. Average nucleotide identities (ANI) among *A. hydrophila* strains and their associated cladogram based on a core genome phylogeny (Figure 1). The pairwise ANI values are color-coded according to their percent identity.

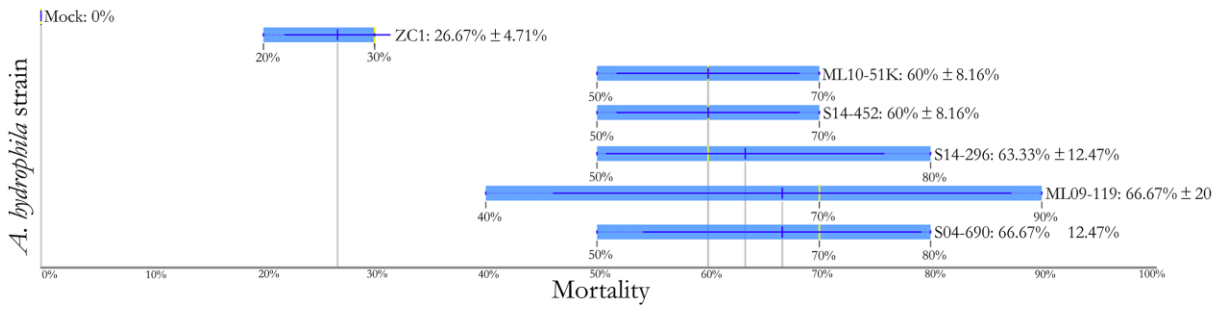


Figure 4. Comparative assessment of the relative virulence of v*Ah* isolates in channel catfish using one hour immersion exposure with fin clip (ANOVA=7.628, *P*-value=0.001).

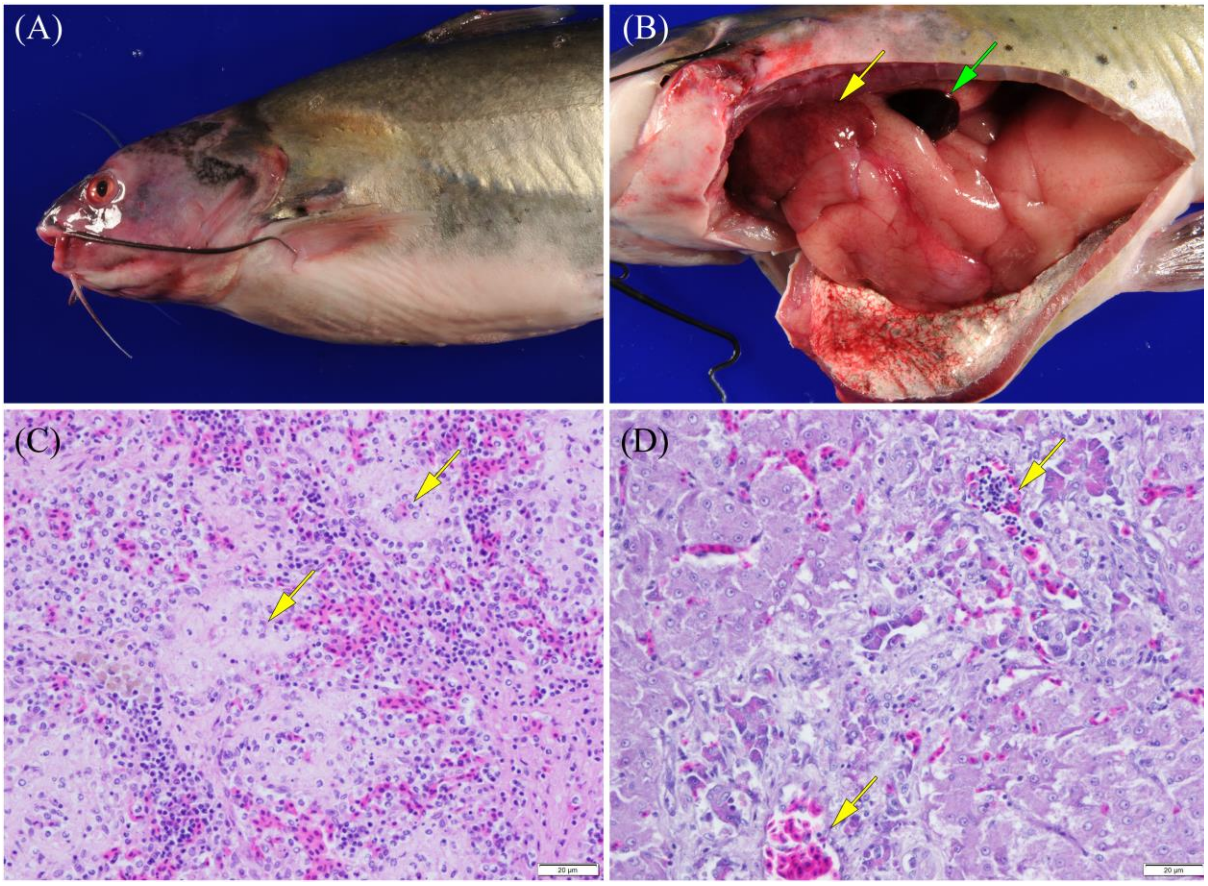


Figure 5. Photographs of channel catfish infected by vAh showing **(A)** external surfaces that are exhibiting congestion/hemorrhage around the head/pectoral fin and within the eye and **(B)** the celomic cavity that has internal organs moderately congested and enlarged, a congested/hemorrhagic spleen (green arrow), and multifocal pale foci corresponding to areas of necrosis (yellow arrow) scattered over the liver (photographs courtesy of Dr. Wes Baumgartner, Mississippi State University) as well as photomicrographs of a channel catfish infected by vAh strain ML09-119 showing **(C)** a section of spleen with splenic ellipsoids (arrows) that are edematous and ellipsoidal arteries that are lined by degenerating as well as necrotic endothelial cells and **(D)** a section of liver with edema and necrosis of pancreatic acinar tissue surrounding branches of the hepatic portal vein (arrows).

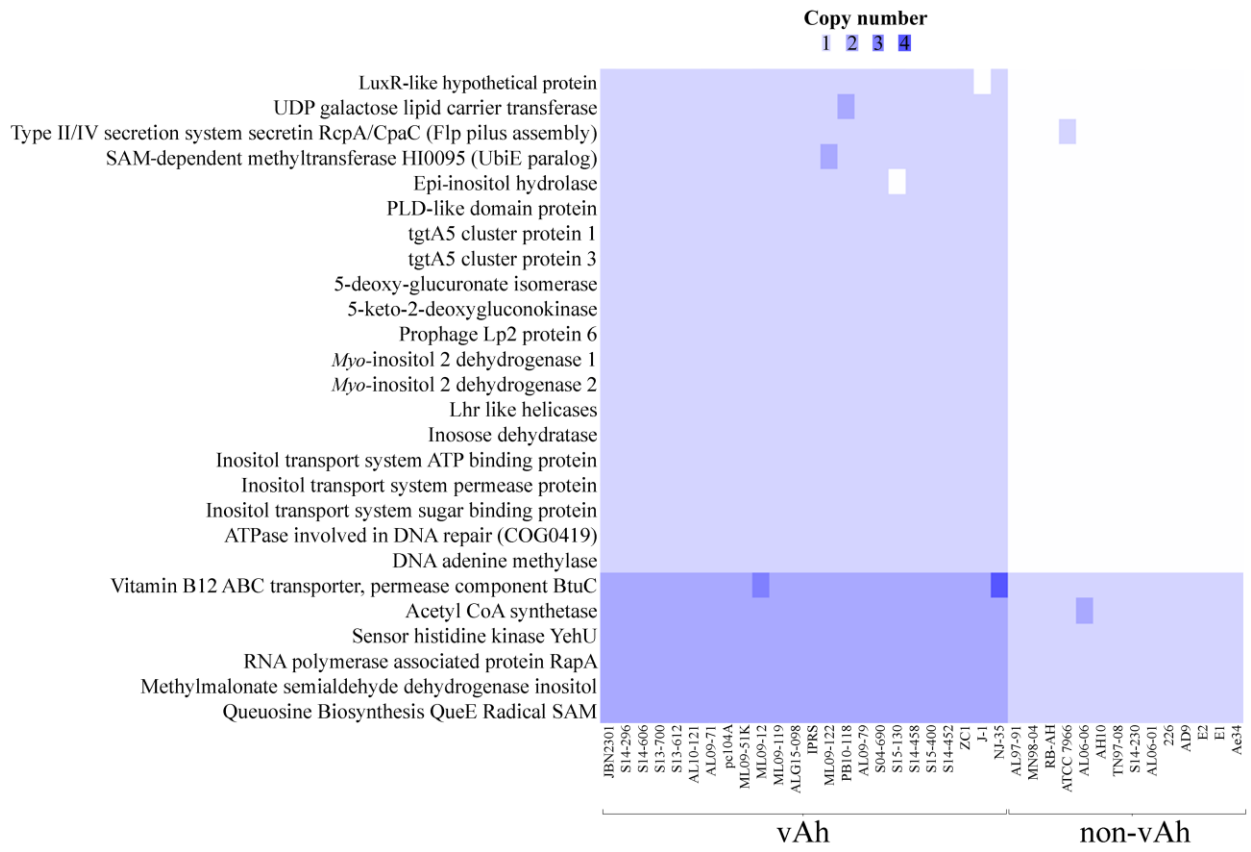


Figure 6. Comparative whole genome predicted gene-based analysis of all confirmed v*Ah* (n=26) and non-vAh isolates (n=15).

Chapter III

Quantitative Polybacterial PCR: Evaluation of Single Copy Genes as Representative Markers for the Quantification and Taxonomic Identification of Mixed Bacterial Samples

1. Abstract.

When challenged with the intricacies of analyzing complex microbial assemblages, no analytical method to-date has been able to provide accurate counts of microbial abundance, while also providing species or subspecies levels of phylogenetic resolution. The dominant molecular target of these studies is the gene encoding the 16S subunit of the prokaryotic ribosome - a gene prized for its highly conserved regions and abundance, with copy numbers ranging from 1-16 per genome. While these traits once served as the foundation on which microbiome analyses were made possible, polymerase chain reaction (PCR) template bias (the result of differences in starting template) and a lack of phylogenetic resolution (a consequence of analyzing sites without significant differences in variable regions) have fueled widespread demand for a more robust method. To reduce PCR biases and generate accurate quantitative and taxonomically informative data, we combined the newly described degenerate PCR (DePCR) method with the

use of the single copy gene *gyrB* to evaluate the practical constraints of using these molecular targets as proxies for measuring cellular abundance. Generation of primers capable of capturing the diversity of all bacterial life required alignments of all complete *gyrB* gene sequences available in the National Center for Biotechnology Information (NCBI) GenBank database (~158,000), resulting in the identification of seven conserved regions and the identification of taxonomically informative sites between these regions. The result of these alignments was a 'universal' primer cocktail for bacteria that demonstrated both *in situ* PCR amplification of every bacterium with a complete genome sequence from NCBI's GenBank and, in its present iteration, this method has shown promising results in the quantification of ecologically and genetically diverse bacteria across a range of concentrations and compositions. This research catalogs the foundational stages of development towards a novel quantification method with the goal of producing true-to-life estimates of microbial abundance, with taxonomic resolution at the level of species or subspecies.

2. Introduction.

Researchers use nucleotide diversity within the 16S subunit ribosomal RNA (rRNA) operon to characterize both composition and taxonomy of microbiomes through polymerase chain reaction (PCR) and sequencing-based methods (e.g. whole genome shotgun sequencing). These methods empower scientists to multiplex hundreds of microbiome samples in a single sequencing run, while providing high base coverage of ribotype diversity in each sample; however, when used to observe microbiome dynamics, these approaches are hindered by distorted representations of microbial relative abundance caused by primer-template interactions during PCR [189] and by variable copy numbers of the rRNA operon which confound quantitative

results [250]. To address these distortions in PCR-based microbial community surveys, we implemented the newly designed deconstructed-PCR (DePCR) method, which reduces primer bias using the following modified workflow:

- Step 1: During the first two cycles of PCR, amplification is linear and, barring stochasticity, amplicons are produced in a 1:1 ratio with the deoxyribonucleic acid [251] template [251]. To minimize the possibility of template-primer interactions not occurring, these cycles must take place under permissive binding conditions; including an abundance of degenerate primers that contain linker sequences specific to primers added during the exponential phase so that template-primer interactions reach saturation.
- Step 2: Before the exponential stage of PCR, degenerate primers and all PCR reagents are removed from the reaction either by magnetic bead purification or by enzymatic removal/degradation.
- Step 3: After purification, amplicons are resuspended in a new master mix comprised of deoxynucleotides (dNTPs), water, buffer, polymerase, and a set of primers designed to target linker sequences within amplicons generated during the first two cycles of PCR.
- Step 4: Exponential amplification, with an additional 28 cycles of PCR.
- Step 5: Amplicon sequencing (Illumina MiSeq).
- Step 6: Downstream analyses (e.g. read mapping or Kraken2 taxonomic profiling).

DePCR addresses core issues of PCR-based quantification methods by standardizing template-primer interactions, resulting in calculable exponential amplification. In other words, if template abundance can be inferred after sequencing of amplicons, then the relative abundance of each template within the sample can also be inferred; however, an accompanying solution is

required to circumvent bias introduced by inherent differences in 16S rRNA operon copy number. The overarching goal of this research was to combine the DePCR method with the use of genes present in a single copy per genome to enable the calculation of relative bacterial abundance (genomic copy number is used as a proxy for cellular abundance).

Towards this aim, a number of single copy house-keeping genes in bacteria, including *cpn60*, *gyrA*, and *gyrB* [252,253] were evaluated as potential targets. Akin to how natural selection has conserved the presence of these genes to a single copy per genome, these genes also contain regions of nucleotide conservation. For quantitative purposes, this quality is the cornerstone for being useful as a proxy for cellular abundance.

Beyond quantitative measures, being able to phylogenetically resolve bacteria at biologically informative levels (e.g. the level of species/subspecies) is the next desirable characteristic for microbiome analyses. Phylogenetic resolution with the 16S rRNA gene is consistent at the level of family, but inferences aimed at deeper levels of resolution (genus and species) are often not supported by the nucleotide differences on which they are inferred. In addition to highly conserved regions, single copy genes also contain highly variable regions, translating to greater nucleotide diversity and therefore greater taxonomic resolution than the 16S rRNA gene. For example, based on *in silico* analyses of the *gyrB* gene in *Aeromonas* spp., bacteria are commonly differentiated at the level of species and, in some taxa, this taxonomic resolution extends to the subspecies level [1]. With the development of DePCR, which allows for greater degrees of primer degeneracy, and advances in reduction of primer dimers (e.g. including bovine thrombin as a PCR additive), the differences that once excluded single copy genes from use in microbiome methods may now promote their use as ideal molecular targets [189,254].

While the experimental focus of this research is on PCR-based methods, the conserved sites validated within this research were also tested in simulations for future research that employs Oxford Nanopore's (Oxford, UK) real-time sequencing technologies (e.g. MinION), based on the concept that individual strands of DNA can be sequenced or rejected by changes in polarity. In practice, each DNA strand be read once and only if the sequence contains the single copy gene(s) specified, resulting in potentially greater quantitative and taxonomic accuracy. While exhaustive validation is beyond the scope of this chapter, preliminary results are described. In summary, this research explores the practical constraints of a tool designed to use the ubiquity of the single copy gene *gyrB* as an analytical target for the analysis of any sample that contains bacterial genomic DNA (gDNA), then proposes an alternate, direct sequencing, framework to improve upon these limitations and extend this methodology to other domains of life.

3. Methods.

3.1. Candidate Gene Selection.

The single copy gene *gyrB* was initially selected based on the ability to taxonomically resolve the hypervirulent pathotype of the aquatic Gram-negative bacterium, *A. hydrophila*, from its less-virulent counterparts (subspecies taxonomic resolution); however, upon deeper phylogenetic analyses, this gene appears to have similar taxonomic resolution in other bacterial genera [251]. Within *Aeromonas* spp., the previously generated phylogenetic mapping of the *gyrB* gene was updated to include all new members of the genus (and to clarify species-level assignations based on core genome phylogenetic analyses discussed within Chapter 2) using the Jukes Cantor genetic distance model, neighbor-joining for tree building, no outgroup, resampling

with replacement, and 1,000 bootstrap replicates, after an alignment with drive5's multiple sequence aligner, MUltiple Sequence Comparison by Log- Expectation (MUSCLE; v3.8.425; https://www.drive5.com/muscle/) [255,256] (**Figure 1**). In support of previous findings [251], this phylogeny indicated that the *gyrB* gene provided taxonomic resolution at the level of genus and often at the level of species and/or subspecies within the genus *Aeromonas*. Based on analyses using an identical approach, the level of degeneracy found within other single copy genes (e.g. *gyrA*, *rpoB*, and *cpn60*) precludes them for solitary use within the qpPCR framework; however, their implementation in concert could increase taxonomic resolution and also serve as a method for cross validation of a *gyrB* gene-based approach. Although preliminary analyses support this idea, the viability of this combinatorial approach was not explored experimentally within this study (data available upon request).

3.2. *gyrB* Gene Database Generation and Validation.

All genes were downloaded from NCBI's GenBank nucleotide, gene, and genome databases that contained at least one of the following labels: '*gyrB*', 'gyrase subunit B', or 'DNA topoisomerase type II'. Identical entries (e.g. the *gyrB* gene was pulled from both the nucleotide and genome databases for the same isolate) and incomplete entries (entries were considered incomplete if they had a length under 1,000 nt) were removed manually. In total, ~158,000 entries were selected as complete *gyrB* gene sequences and used in subsequent analyses (note: this method retained a number of partial gene sequences, but the alignment method is additive [not subtractive], so, additional downstream filtering was necessary, but potentially conserved sites were not excluded).

3.3. Alignment of *gyrB* Gene Sequences.

Identification of conserved sites that would eventually be developed as degenerate primers was performed by a series of multiple sequence alignments with Osaka University's MAFFT software (v.7.409; https://mafft.cbrc.jp/alignment/software/) encompassing all sequences from the curated *gyrB* gene database described above. To generate a reference for comparison, MUSCLE was also used; however, due to the size of this dataset, both alignment methods required multiple, separate, alignments to be performed. For these analyses, a total of 50 subgroups were created by assigning a random number to each sequence, then sequences were assigned to each group based on this tag (each subgroup contained a group-wide limit of 5,000 sequences [based on a computational ceiling]). After this pre-processing, alignments of each subgroup were performed with MAFFT and with MUSCLE ("-maxiters 2" option).

After consensus sequences were evaluated for outliers, which was defined as potential non-*gyrB* gene sequences identified through NCBI's BLASTn algorithm where no *gyrB* gene hits were present external to the original sequence (the megablast algorithm was tested, but not used because it uses a formula that is too restrictive), filtering was performed. After filtering, all subgroups with outliers were realigned, then a second alignment was performed between consensus sequences. Phylogenetic mapping of each alignment was performed using a method identical to the *Aeromonas* genus phylogeny, as described above. Mapping was performed to the deepest taxonomic level with a cut-off bootstrap support value of 90. While the complete phylogenetic map based on all *gyrB* genes was produced, the figure is too large to be included as a figure; however, this phylogenetic map as well as other taxonomic levels and filtered phylogenetic maps are available upon request.

3.4. Linear Amplification Primer Design: *in silico* Validation of Primer Binding.

After each alignment was validated, regions conserved at the 25%, 50%, 75%, and 100% identity levels were used to create candidate degenerate primers. Evaluation of these primers was performed *in silico* against all complete bacterial genome sequences in NCBI's genome database as well as the curated database, using the Sequence Manipulation Suite's Primer Map software (https://www.bioinformatics.org/sms2/primer_map.html) as well as Geneious v. R9 with a maximum mismatch setting of two bases (the average permissible number of mismatches tolerable by DePCR) and a band prediction interval of between 100-1,000 bases. Primers were excluded if they exhibited promiscuous binding, failed to bind to all members of each database, and/or did not conform to the length limitations of Illumina MiSeq sequencing (est. 550 bp; discussed in detail below). Primers that conformed to these characteristics were selected for experimental validation; however, due to the prohibitively high degree of degeneracy of using a single primer pair, a primer cocktail approach was eventually selected to reduce per-primer degeneracy, resulting in a maximum degeneracy of 2,304 (*gyrB*F34; **Table 1**) and a band of ~515 bp (excluding linker sequences and barcodes). While this unprecedented degree of degeneracy did not appear to skew the results of this research, a key limitation of this approach is hypothesized to be an inability to reach saturation between each primer-template. Given this inevitability, primer concentrations within the primer cocktail were adjusted based on the degeneracy so that the final concentration would be at or above levels required for primer-template saturation. With this workaround band formation decreased (data available upon request) due to putative primer dimerization because of the copious numbers of primers necessary to reduce degeneracy while maintaining individual primer concentrations, but returned when the bovine thrombin was added, which reduces the formation of secondary structures (note:

bovine thrombin was selected over DMSO, bovine serum albumin, and other PCR additives due to its fidelity to the core reaction and its comparable efficiency at concentrations 18-178 lower than these alternatives) [254].

3.5. Primer Synthesis and Screening.

Primers used in this study (**Table 1**) were synthesized by Eurofins Genomics (KY, USA; https://www.eurofinsgenomics.com/). Each degenerate primer was produced in a separate reaction. Initial validation was performed with primers that lacked the common sequence linkers (CS1 and CS2; **Table 2**), primers within the *gyrB* primer cocktail that did and did not have binding affinity to the known template (testing specificity), and the complete set of primers using a panel of PCR conditions (gradient PCR, touchdown PCR, numbers of cycles, extended extension time, primer concentration, input gDNA concentration, and the capacity to amplify both single templates as well as complex template compositions; results available upon request). After consistent amplification was achieved across a range of known gDNA templates (detailed below) with this extremely degenerate primer set, the overall melting temperature was determined to be 58°C, with individual melting temperatures ranging from 47.8 - 74.3°C. Alone, the degeneracy tested in this reaction would be restrictive; however, with the range of melting temperatures, a host of permissive reaction conditions were evaluated to ensure maximal binding (data available upon request), with the presumption that aberrant products produced from these conditions would be removed before downstream reactions allowed to proceed. Primers referenced for the remainder of this experiment included the CS1/CS2 linker sequences and were used for exponential binding with barcoded primers to identify samples after Illumina sequencing.

3.6. Selection of Bacteria for Experimental Validation.

Bacterial species were selected that represented both ecological diversity and a range of ribosome copy numbers. Based on these criteria, the ZymoBIOMICS Microbial Community DNA Standard (ZYMO Research, USA) was used as an externally validated DNA standard. Although their quantification method was not disclosed, the relative genomic abundance of these bacteria was artificially created by ZYMO Research. In effect, DNA was quantified after DNA extractions were performed on pure cultures, then combined into an equimolar library (with the exception of the non-bacterial isolates). In addition to encompassing a wide range of bacterial species (*Bacillus subtilis*, *Enterococcus faecalis*, *Escherichia coli*, *Lactobacillus fermentum*, *Listeria monocytogenes*, *Pseudomonas aeruginosa*, *Salmonella enterica*, and *Staphylococcus aureus*), the mock community also contained gDNA from two yeast isolates that would not be targeted with this method (*Saccharomyces cerevisiae* and *Cryptococcus neoformans*) (**Table 3**). To supplement the mock community, the bacterial species *Aeromonas hydrophila* and *Terriglobus roseus* were selected to explore the capacity for qpPCR to resolve low abundance templates as well as increase the range of 16S rRNA gene copy numbers represented for future comparisons between qpPCR and 16S rRNA gene sequencing methods. For each bacterial isolate, gDNA was isolated from pure cultures, with each species grown to their respective log phase growth in their preferred standard media, as measured by optical density (OD600).

3.7. Purification and Quantification of DNA.

Isolation of gDNA for all bacteria, excluding gDNA obtained from the ZYMO mock community, was performed by isolating pure cultures, then by growing each species in its preferred

media until log-phase growth was reached. After this growth rate was reached, gDNA was extracted using the E.Z.N.A.® Bacterial DNA Kit (Omega Bio-tek, Inc., Georgia, USA). After extraction, gDNA was quantified using the Qubit broad-spectrum assay kit in a Qubit fluorometer (Invitrogen, ThermoFisher Scientific, Carlsbad, California, USA). Resultant DNA was diluted to a working concentration of 50 nanograms of gDNA per microliter (ng/µL) in nuclease-free water in preparation for downstream application.

3.8. Evaluation of Standard vs. Touchdown PCR for Improved Amplification.

For the testing of degenerate primers, two PCR protocols were evaluated with additional rounds of amplification to visualize the results through gel electrophoresis, with each reaction mix consisting of 50 ng of template gDNA suspended in 1 µL of nuclease-free water, 12.5 µL of Econo*Taq*® Plus Green 2x Master Mix (Lucigen, Madison, Wisconsin, USA), and 20 picomoles of each degenerate primer, with additional amounts of degenerate primer included based on degeneracy (e.g. a twofold degenerate primer would have twice the amount of primer added), and 0.5 µL of bovine thrombin to reduce primer dimerization, resulting in a total volume of 25 µL per reaction (note: volume of water was proportionally reduced to account for the increased volume of the primer cocktail; **Table 4**).

Standard PCR cycling parameters consisted of an initial denaturation of 94°C for 3 minutes, then 30 cycles of 94°C for 30 seconds, 58°C for 30 seconds, and 72°C for 1 minute. Touchdown PCR cycling parameters consisted of an initial denaturation of 94°C for 3 minutes; then 30 cycles of 94°C for 30 seconds, 68°C for 30 seconds (-1°C per cycle), and 72°C for 1 minute. Amplicons were resolved by gel electrophoresis using 20 µL of sample, with 20 µL of the Lucigen 1 kbp ladder (Lucigen Corporation, Wisconsin, USA) on a 1% agarose gel through

exposure to 120 volts for 2 hours. After band separation, gels were stained with ethidium bromide for 10 minutes and destained for 5 minutes in deionized water. Gel images were produced by 5 second exposure to ultraviolet light, using the AlphaImager gel imaging station and accompanying software (ProteinSimple, San Jose, California, USA).

3.9. Linear Amplification Primer Design: Independent Validation of Primer Binding.

Validation of each primer pair was performed by evaluating the results of standard and touchdown PCR on an Eppendorf Mastercycler Gradient S, with an extended number of cycles (n = 30). Reaction composition was modified from the previous concentrations to account for a reduction in the total concentration of primers within the mix, resulting in a greater proportion of water (total reaction volume remained 25 µL). Standard PCR cycling parameters consisted of an initial denaturation of 94ºC for 3 minutes; then 30 cycles of 94ºC for 30 seconds, 58ºC for 30 seconds, and 72ºC for 1 minute. Touchdown PCR cycling parameters consisted of an initial denaturation of 94ºC for 3 minutes; then 30 cycles of 94ºC for 30 seconds, 68ºC for 30 seconds (-1ºC per cycle), and 72ºC for 1 minute. Amplicons were resolved by gel electrophoresis using 20 µL of sample, with 20 µL of the Lucigen 1 kbp ladder on a 1% agarose gel through exposure to 120 volts for 2 hours. After band separation occurred, gels were stained with ethidium bromide for 10 minutes and destained for 5 minutes in deionized water. Gel images were produced by 5 second exposure to ultraviolet light, using the AlphaImager gel imaging station and accompanying software.

3.10. Evaluation of Amplicon Purification on Downstream Amplicon Formation.

Amplicons that were produced from the first two cycles of the standard PCR reaction described above were purified using MagBind RX-PLUS magnetic beads (Omega Bio-tek, Inc., Georgia, USA) in accord with the manufacturer's protocol, with one exception: to address the removal of additional primers contained within the post- reaction PCR mix, two additional rounds of purification were performed, totaling four rounds of purification. To evaluate the effect that amplicon purification has on total amplification, portions of a single PCR was sampled after 2, 4, 6, 8, 10, 20, and 30 cycles. Each sample was then split, with one sample being purified and the other forgoing purification. After this step, DNA concentration was quantified by Qubit fluorometric quantification as described above and a second PCR reaction was performed using standard PCR cycling parameters (an initial denaturation of 94^{o}C for 3 minutes; then 28 cycles of 94^{o}C for 30 seconds, 58^{o}C for 30 seconds, and 72^{o}C for 1 minute). Based on these results, ten cycles with the *gyrB* primer cocktail were selected to offset this reduction with the aim of generating a preliminary dataset that would guide future iterations of this method.

3.11. Sequence-Based Quantitative Evaluation of qpPCR: Sample Design.

The goal of quantifying amplicons through next-generation sequencing was to experimentally evaluate the quantitative and taxonomic accuracy of using a single copy gene as a proxy for genomic abundance across a range of concentrations and sample compositions in the framework detailed below. To test concentration accuracy, *E. coli* and *T. roseus* were combined in the following amounts (tested in triplicate 25 µL reactions described below):

- *E. coli* and *T. roseus* in a 1:10 ratio; 1 ng *E. coli* DNA and 10 ng *T. roseus* DNA
- *E. coli* and *T. roseus* in a 10:1 ratio; 10 ng *E. coli* DNA and 1 ng *T. roseus* DNA
- *E. coli* and *T. roseus* in a 1:20 ratio; 1 ng *E. coli* DNA and 20 ng *T. roseus* DNA

- Zymo mock community with *T. roseus* spiked in a 1:10 ratio; 1 ng of ZYMO mock community DNA and 1 ng *T. roseus* DNA
- Zymo mock community; 10 ng of ZYMO mock community DNA

3.12. Step One: PCR Thermocycling Conditions.

Step one thermocycling conditions were based on the results of the preliminary screening and observation of significant amplicon reductions after amplicon purification. Thermocycling conditions were extended to 10 cycles of PCR, with a T_m of 58^{o}C to generate sufficient numbers of amplicons, with the caveat that downstream results would include the corresponding PCR bias. Reactions were conducted on an Eppendorf Mastercycler Gradient S, with reactions comprised of 11 - 21 ng of template gDNA suspended in µL of nuclease-free water, 12.5 µL of Econo*Taq* Plus Green 2x MasterMix, and 20 picomoles of each degenerate primer, with additional amounts of degenerate primer included based on degeneracy (e.g. a twofold degenerate primer would have twice the amount of primer added), resulting in a total volume of 25 µL per reaction (**Table 4**). As described above, these foundational tests were aimed to evaluate the qpPCR method on a basic level through sequence data analyses. Therefore, the number of cycles was extended from two cycles to ten cycles of 94^{o}C for 30 seconds, 58^{o}C for 30 seconds, and 72^{o}C for 1 minute, after an initial denaturation of 94^{o}C for 3 minutes. After two cycles of PCR, the resultant amplicons should include the linker sequences CS1/CS2 and be in a 1:1 ratio with the starting template (**Figure 2**); however, after the additional eight cycles of amplification, template bias and copy number bias are expected to have a measurable distortion on downstream quantification.

3.13. Step Two: Amplicon Purification.

Amplicons that were produced from the first stage of qpPCR were purified using MagBind RX-PLUS magnetic beads (Omega Bio-tek, Inc., Georgia, USA) in accord with the manufacturer's protocol, with one exception: to address the removal of additional primers contained within the post-reaction PCR mix, two additional rounds of purification were performed, totaling four rounds of purification. After purification, DNA concentration was quantified by Qubit fluorometric quantitation as described above (data available upon request).

3.14. Step Three: Post-Purification Thermocycling Conditions.

Standard and touchdown PCR protocols were explored as well as differences in amplification based on amplicon concentration; however, standard and touchdown PCR showed no significant difference in amplicon production after 28 cycles (data available upon request). Therefore, standard PCR cycling parameters were used, which consisted of an initial denaturation of 94^{o}C for 3 minutes; then 28 cycles of 94^{o}C for 30 seconds, 58^{o}C for 30 seconds, and 72^{o}C for 1 minute; and a final extension at 72^{o}C for 5 minutes. Reactions were comprised of 50 ng of purified amplicons (suspended in nuclease-free water; volume of purified amplicons included was increased to reach a total of 50 ng of DNA per reaction), 12.5 µL of Econo*Taq* Plus Green 2x MasterMix, and 20 picomoles of each non-degenerate primer, which contains complementary linker sequences to CS1/CS2 and a Fluidigm Access Array™ paired end barcode (San Francisco, USA), resulting in a total volume of 25 µL per reaction (**Table 5**). Based on these preliminary tests, a standard PCR protocol was selected, and all amplicons retained after purification were used.

3.15. Step Four: Pre-Sequencing Amplicon Verification.

After PCR with the degenerate *gyrB* primer cocktail, amplicon purification, and a second PCR with non-degenerate primers, amplicons were resolved by gel electrophoresis using 20 μL of sample, with 20 μL of the Lucigen 1kbp ladder (Lucigen Corporation, Wisconsin, USA) on a 1% agarose gel through exposure to 120 volts for 2 hours. After band separation occurred, gels were stained with ethidium bromide for 10 minutes and destained for 5 minutes in deionized water. Gel images were produced by 5 second exposure to ultraviolet light, using the AlphaImager gel imaging station and accompanying software (ProteinSimple, San Jose, California, USA). In addition to amplicon validation by comparisons against a ladder, restriction digests were also used to check for correct amplicon sequences, with the nuclease *TaqI* was selected because of its binding preferences, based on *in silico* analyses using the RestrictionMapper web-based tool (http://www.restrictionmapper.org) (**Table 6**); however, no digestion was observed in any qpPCR-generated amplicons, including replicate qpPCR runs (enzyme activity was verified on amplicons generated from 16S rRNA gene amplification; **Figure 3**).

3.16. Step Five: Sequencing Conditions.

After touchdown PCR with increased PCR cycles, amplicon preparation was performed with the Illumina Nextera sample preparation protocol. Sequencing conditions were based on protocol for 16S rRNA gene metagenomic studies (the user's manual can be obtained at https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). Sequencing of amplicons was performed with a 2 x 250 bp (v2 reagents)

kit in a pooled MiSeq run, which has a upper limit for read length that limited the total number of reads per sample, with the number of reads produced ranging from ~700 reads (not interpretable) to ~260,000 reads per sample. Considering that read coverage is the core metric intended for use as a proxy to determine the numeric abundance of amplified templates (~515 bp after trimming barcodes and linkers), consistency may be resolved by employing alternate methods with longer read lengths. Within the multiplexed reaction, samples with inconsistent numbers of reads (defined as a total difference greater than 50%) or too few reads (defined as less than 2,000 total mapped reads) were removed from downstream analyses.

3.17. Step Six: Processing Illumina MiSeq Reads.

Raw sequence files from amplicon sequencing were trimmed based on barcode as well as read length below target amplicon size (515 ± 50 nt) were removed with CLC Genomics v. 3.6.5. Preliminary validation of reads was performed through read mapping to each respective genome for each sample composition, using the same software, with a minimum length match of 95% and identity of 95%. After reads with low quality and/or non-target length were removed, FASTQ files were imported for processing with Johns Hopkins' Kraken 2 bioinformatics pipeline (v. 2.0.7-beta; https://ccb.jhu.edu/software/kraken2/; note: FASTQ file formats require the argument "-fastq-input").

In Kraken2, the default database, which is comprised of complete bacterial, archaeal, and viral genomes in the National Center for Biotechnology Information's (NCBI) Reference Sequence database (RefSeq), was evaluated as well as the database containing the *gyrB* gene sequences used to generate the alignments described above. Kraken2 parameters included the k-mer length of 35 and minimizer length of 31, with the assignation threshold set to "-confidence

0.95" (assignations were only performed if 95% of k-mers mapped to the lowest common ancestor). Resultant analyses were transformed with a python script for figure generation and statistical comparison in the software R (R Foundation for Statistical Computing, Vienna, Austria).

3.18. Calculation of Genomic Copy Number.

Including exponential amplification (n = 28 cycles; totaling 30 PCR cycles), a single *gyrB* gene would produce template producing 1.07374×10^9 amplicons (**Equation 1** and **Equation 2**). Based on a general threshold of ~20 ng of DNA for visual detection when stained on agarose with ethidium bromide, the minimum template requirement to visualize results at the end of a 30-cycle reaction is 4.237×10^{10} copies or approximately ~39 genomes/single copy genes and their resultant amplicons. For sequencing reactions on the Illumina MiSeq, a general starting threshold for successful reads is 20 nM of product in 10 μL. This translates to a hypothetical minimum threshold for detection with most sequencing platforms of 1.204×10^{11} copies at the time of sequencing or approximately 112 genomes/single copy genes, based on an average sequence length of 515 nt. Given the wide range of starting concentrations of gDNA in environmental and clinical samples, and therefore resultant template within samples, multiple sequencing reactions may be necessary. While most sequencing methods are robust and have the capacity to provide adequate coverage for all templates, as described in the sample design section, we explored amplicon concentrations that are largely divergent to evaluate these hypothetical thresholds.

In summary, the number of amplified templates sequenced is hypothesized to be a function of post-PCR values related to the concentration of single copy genes per microliter that

are contained within the volume added to the sequencing reaction. For the purpose of this study, all calculations are performed under the assumption that bias was consistent across all sample types for each method performed (excluding qpPCR and the additional eight non-linear cycles of PCR). As such, numeric abundance of post-sequencing templates per microliter will serve as a proxy to identify the initial concentrations of these single copy genes per microliter of sample and therefore also serve as a quantitative measurement for the initial concentrations of microbes based on the following equation:

3.19. Quantitative Evaluation of qpPCR: Statistical Analyses.

All statistical analyses were performed with the software R (R Foundation for Statistical Computing, Vienna, Austria). Significance for quantification accuracy was performed with a one-way analysis of variance [257] and Tukey's test for post-hoc analysis (statistical significance required a p-value < 0.05).

4. Results.

4.1. Gel Electrophoresis Detection Thresholds.

The hypothetical minimal microbial concentration quantifiable by agarose gel is reliant on the production of sufficient numbers of amplicons. For visual verification on a 1% agarose gel, samples are exposed to 120 volts for 2 hours, stained with ethidium bromide for 10 minutes, destained for 5 minutes in deionized water, and imaged after ~5 second exposure to ultraviolet light, using the AlphaImager gel imaging station and accompanying software. If the total number of amplicons produced after both linear and exponential cycles of PCR was below ~20 ng/μL, a band was inconsistent (data available upon request).

4.2. The Effect of Purification After Linear PCR Cycles.

After purification, the total DNA concentration was expected to decrease; however, the severity of these reductions led to an inhibitory effect on total amplicon production, even after the second PCR (28 cycles) with non-degenerate primers (CS1/CS2-tagged). In addition to demonstrating the need for abundant starting template, based on a series of PCR with standard concentrations, as described above, visible banding was present only after increasing the number of cycles with the *gyrB* primer cocktail to 10, followed by 28 PCR cycles with the CS1/CS2 barcoded primers; resulting in a total of 38 cycles (**Figure 4**). Although this approach improved yields, this method was also presumed to introduce distortion. Therefore, an alternate approach was explored.

With the aim of increasing final amplicon production, the number of PCR cycles with the non-degenerate CS1/CS2-tagged primer mix was increased from 28 cycles to 38 cycles as well as to 48 cycles across a range of starting template concentrations (0.035 – 3.72 ng of post-purification amplicons). Unfortunately, there was no difference based on DNA quantification of purified post-non-degenerate PCR amplicons (Qubit), nor was there a visual difference past 40 cycles of PCR (**Figure 5**). It is important to note that while increasing gDNA concentrations is possible with practically unlimited sample availability, metagenomic samples often are present in low concentrations and in low volumes, with no possibility for resampling. Towards evaluating this method in a practical context, extending the number of non-degenerate PCR cycles beyond 50 was not pursued because of the numerous biases associated with increased PCR cycles. While starting gDNA requirements and amplicon loss due to the purification method remain clear constraints of the qpPCR method, if purification is uniform, downstream relative abundance should reflect the increased number of PCR cycles with degenerate primers. Therefore, while a

1:1 ratio would no longer be maintained between sample abundance downstream relative abundance, the number of cycles for the initial, degenerate, PCR was increased to 10 for the sequence-based quantification tests described below.

4.3. Amplicon Verification via Gel Electrophoresis.

Each bacterial isolate assayed thus far produced the predicted ~620 bp amplicon (~515 nt *gyrB* sequence, plus the 47 nt and 58 nt Fluidigm Access Array barcodes with the CS1 and CS2 sequences). Beyond amplification using the *gyrB* primer cocktail, to confirm that amplicon was produced because of non-degenerate primer/template amplification, additional verification was performed using each primer separately that was predicted to bind to the target site in *A. hydrophila* ML09-119. No visually aberrant amplification occurred when the primers were introduced separately – abundant amplification occurred only when primers were combined (**Figure 5**).

Without increasing the number of cycles that employ degenerate *gyrB* gene primers to ~10, depending on the starting concentration of template, visually apparent bands may not be visible after 28 cycles of exponential amplification with the non-degenerate primers. Therefore, when evaluating successful amplification of samples that contain a low-abundance of gDNA/targeted single copy genes, additional cycles may be necessary to produce a sufficient number of amplicons or samples should be concentrated by pooling identical PCRs or by increasing the volume, then eluting with a minimal volume of buffer.

Using the PCR conditions described above, *A. hydrophila* ML09-119 gDNA was used as a template to identify a lower threshold for detection based on primer concentration, with and without magnetic bead purification. Based on a PCR with only degenerate primers across a range

of primer concentrations (0.07 – 20 µM per primer [increased proportionally based on degeneracy]), the minimum primer concentration is 5 µM per primer within the *gyrB* primer cocktail. Interestingly, without magnetic bead purification, amplicons are formed with per primer concentrations as low as 0.96 µM (**Figure 7**).

4.4. Band Validation with Restriction Enzymes.

Fragmentation patterns predicted by restriction digest mapping software did not match the patterns when visualized with gel electrophoresis for each bacterial isolate tested (*B. japonicum*, *R. denitrificans*, *S. coelicolor*, *E. coli*, and *A. hydrophila*) with the *gyrB* primer cocktail, but matched banding pattern for 16S rRNA gene predictions. Even with extended incubation, extensive PCR product purification (six rounds of purification via MagBind magnetic beads), *TaqI* failed to cut any PCR product produced using the qpPCR method (**Figure 3**). While this remains unresolved, one possible explanation for this observation is that the presence of binding sites on residual degenerate and non-degenerate primers may have created a masking effect, thereby exhausting the net fragmentation performed by the restriction enzyme and necessitating an alternate method for the removal of primers.

4.5. MiSeq Sequence Results.

Sequencing of 2 x 250 bp reads, using the v2 Reagents run type, produces a hypothetical 15,000,000 clusters within a pooled in a MiSeq run, resulting in over 100,000 reads per sample; however, while this suggests that it is possible to get 20,000,000 or more reads per round of cycling, the number of reads produced by this sequencing platform was greatly below that estimate (data available upon request). This could be attributable to the number of reads

generated through Illumina MiSeq sequencing being a function of the number of amplicons contained within a sample (low post-qpPCR amplicon concentrations, resulting from loss during amplicon purification) as opposed to a byproduct of hitting the upper limits of the 2 x 250 bp read sequencing technology, but this was not investigated beyond comparing pre-sequencing DNA concentrations with (data available upon request). After excluding samples with fewer than 2,000 paired end reads, trimming was performed on a subset of the samples, followed by read mapping and Kraken2 processing.

4.6. Read Mapping of MinION Results.

Read mapping was performed on samples that produced over 2,000 paired end reads for each triplicate, with at least 20X coverage (n = 15/15 samples). The reference sequences included all *gyrB* gene sequences used in alignment (n = ~158,000 *gyrB* gene sequences), with successful mapping being defined as 50 nt overlap between paired ends, with 99% identity and 99% length of the read mapping to the reference sequence. From the initial dataset, tests that produced over 2,000 reads with these parameters across each replicate included tests of quantitative accuracy that used a range of concentrations of *T. roseus* and *E. coli* (**Figure 8A**, **Figure 8B**, and **Figure 8C**), of the capacity of this method to quantify and identify the ZYMO mock community (**Figure 8E**), and of a mixed approach where *T. roseus* was spiked into a reaction with the ZYMO mock community at 75 times the concentration of each genome within the mock community (a 10:1 ratio of total DNA; **Figure 8D**). Note that the data presented are representative counts of relative count of each mapped paired end read to its respective genome and are not absolute values of cellular abundance because the number of reads sequenced were below the hypothetical thresholds described above.

4.8. Kraken2 Processing of MinION Results.

Kraken2 mapping was performed on the same samples processed by read mapping (n = 15/15 samples). Mapping was performed with the default database and with the database of *gyrB* gene sequences described above, with results mapped to the deepest taxonomic level supported by statistical significance. Evaluation of quantitative accuracy relied on k-mer counts of the same groups: 10 ng *T. roseus* DNA and 1 ng of *E. coli* DNA (**Figure 9A)**, the ZYMO mock community (**Figure 9C**), and *T. roseus* spiked into a reaction with the ZYMO mock community at 10 times the concentration of the mock community (a 10:1 ratio of total DNA; **Figure 9B**). Although k-mers are somewhat comparable to reads from a mapping perspective, k-mer counts are plagued with false-positive results (programs, such as KrakenUniq, aim to alleviate this bias) [258]. As a potential result of this mapping strategy, the majority of mapped k-mers were excluded because of aberrant identification, resulting in several groups that were successfully called with the read mapping approach being excluded. Even with strong removal of incorrectly mapped k-mers, these results did not match the hypothetical estimated concentrations or the read mapping results.

5. Discussion.

16S rRNA gene amplicon sequencing is the predominant method for microbiome quantification and taxonomic analyses. While this approach has been enormously useful in forming a foundation for understanding microbiome dynamics, as the field of molecular ecology has advanced, novel approaches are required to produce data with improved resolution capable of deeper investigations of microbiome composition. In this research, we describe first steps

towards this aim by defining the practical limitations of combining DePCR, next-generation sequencing, and single copy genes as a molecular target to improve phylogenetic resolution and quantitative accuracy, while reducing methodological bias associated with PCR-based analyses.

To generate a functional replacement for the 16S subunit of the prokaryotic ribosome, multiple alignments of all complete bacterial *gyrB* genes (n = ~158,000) were used to identify conserved regions within this gene as well as to identify regions within this gene that contain sequence diversity. From these alignments, seven highly conserved regions were identified and investigated as potential targets for the development of degenerate primers, based on total degeneracy, probability of off-target amplification, and distance from other conserved sites for viability in 2 x 250 bp paired amplicon-sequencing. After *in silico* amplification patterns were evaluated, candidate degenerate primers were screened *in vitro* for optimal thermocycling conditions by surveying viable melting temperatures for this complex primer pool (gradient PCR), by identifying the effect of touchdown PCR with additional rounds of amplification (template fidelity at the cost of introducing PCR bias with additional cycles), by assessing changes in target amplification based on a range extension times, by altering primer concentrations to determine points of potential primer-template saturation, by altering gDNA/template concentrations to capture potential for bias (low abundance templates, input gDNA concentration, and the capacity to amplify both single templates as well as samples with mixed compositions) by identifying concentration thresholds for template DNA and for primers, and by estimating primer binding/pairing preferences.

Next, the efficiency of degenerate primer removal and the reduction in template/amplicons after magnetic bead-based purification was explored. Based on evaluation of the number of rounds of purification, a single round of purification was sufficient to remove

~99% of degenerate primers; however, an additional three rounds of purification were required to remove products from non-specific amplification to below 0.01 ng/µL. To resolve this pitfall, enzymatic removal (ExoSAP-IT™; Thermo Fisher Scientific; MA, USA) and band excision were explored as alternate methods to extract amplicons produced during the linear stages of PCR; however, a total of ten PCR cycles with the *gyrB* primer cocktail were eventually selected to maintain post-purification concentrations at or above minimum template requirements for the subsequent PCR cycles (non-degenerate), with the aim of generating a preliminary dataset that would guide future iterations of this method. The effects of this workaround are present within evident in downstream comparative analyses and may contribute to a lack of quantitative accuracy but appear to be somewhat offset by total loss during purification.

After implementing the workaround during purification, post-purification amplicons were resuspended in a new PCR with non-degenerate primers designed to bind to identical linker sequences generated from PCR cycles with degenerate primers, thereby resulting in uniform primer-template interactions. While polymerase activity of Econo*Taq*® (Lucigen; WI, USA) was assumed to be consistent across all template compositions during amplification, future research could improve taxonomic resolution by employing a polymerase with a proof-reading activity. The minimum number of PCR cycles necessary for quantification with the Illumina MiSeq was identified (dependent on post-purification DNA concentration) and implemented to reduce the potential for chimera formation associated with each round of PCR.

After sequencing of amplicons on the Illumina MiSeq, two methods were used to quantify and taxonomically resolve the composition of samples. The first method was read mapping, with reads mapping to the *gyrB* gene database used to create the alignment. Overall, the results through read mapping of the concentration tests (**Figure 8A-C**) mirrored hypothetical

estimates; however, additional replication is necessary to infer statistical significance without relying on methods of statistical inference, such as bootstrapping. While the relative abundance of *T. roseus* spiked into the ZYMO mock community (**Figure 8D**) reflected hypothetical estimates with similar accuracy to the concentrations tests described above, the concentrations measured by read mapping of qpPCR amplicons from the ZYMO mock community consistently deviated from the company's estimates **(Figure 8D** and **Figure 8E**). Under the assumption that the advertised DNA concentrations of the mock community are accurate, this result may be caused by low DNA input and/or template bias introduced with the additional (n=8) cycles of PCR with degenerate primers. While each genome included within this study contained successfully mapped reads; however, the number of reads mapped to *L. fermentum* and *P. aeruginosa* were consistently low (≤ 2 reads per sample), making quantitative measurements and taxonomic assignations infeasible because of a lack of sample size and therefore statistical power. The second method used to process qpPCR reads was k-mer mapping using the Kraken2 pipeline. Overall, the quantitative accuracy with this approach for samples for each dataset was highly variable. In fact, the taxonomic resolution of the default database led to off-target matches in over ~90% of all k-mers, with on-target matches improving only marginally with a custom database generated from *gyrB* genes used in this study.

As an alternate approach to employing the conserved sequences in a PCR-based method, the results of these alignments could be used as search strings in the 'read until' real-time sequencing approach (Rang, 2018 #258). In effect, the majority of the reads would contain the single copy gene. If additional single copy genes/regions are included, then the overall power for classification and cross-validation of quantitative accuracy 1also increases. Even before experimental validation of this approach, there are clear limitations, such as the relatively high

basecalling error rate (~38%) and difficulty sequencing DNA with high Guanine and Cytosine base content, which would result in inappropriately called bases and distorted data [259-261]. While there would be constrains because of the sequencing method, the generally fragmented nature of DNA isolated form metagenomic samples would benefit from this approach by increasing the probability of sequencing the gene(s)/region(s) of interest. Furthermore, if restriction enzymes are introduced with an affinity for regions external to the gene(s)/region(s) of interest, the net efficiency of this sequencing/quantification/classification method could be improved. For example, if the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-associated endonuclease (Cas9) system was used to target promoter regions, then the net efficiency of this approach would increase, therefore enabling more exhaustive sequencing of the sample.

This research represents a proof-of-concept designed to refine qpPCR. In its present iteration, qpPCR appears to be predominantly limited by total throughput, with the rate limiting step being template abundance after inefficient removal during size-selective purification. Although quantitative data roughly represented starting template, based on relative read abundance, this method requires additional revision(s) before practical application. Two potential approaches to these revisions include exploring alternate methods of purification, as described above, and/or adaptation of the conserved sequences within a modified ‘read until’ program with Oxford Nanopore’s real-time sequencing technologies. Once either of these approaches demonstrate consistent results with samples containing known compositions of DNA, comparative analyses with the respective ‘gold standards’ of microbial ecology should be used to evaluate the robustness of each method across a range of DNA concentrations and sample

complexities. In summary, this research evaluated the practical constraints of using single copy genes as informative targets in the PCR-based method, qpPCR.

Table 1. Sequences for the degenerate primer cocktail based on a comprehensive alignment of *gyrB* gene sequences (note: sequences listed do not contain CS1 and CS2, which are listed in **Table 2**).

Primer	Sequence (5'→ 3')	Degeneracy	T_m(°C)
gyrBF1	ACACTGACGACATGGTTCTACAAAATGAGTTCCCCCCTC	1	66
gyrBF2	ACACTGACGACATGGTTCTACAACGTGTATGCCGCAGTG	1	67
gyrBF3	ACACTGACGACATGGTTCTACAATCTGATAACCTCAAAG	1	63
gyrBF4	ACACTGACGACATGGTTCTACACARGGAGGAACGCATCT	2	67
gyrBF5	ACACTGACGACATGGTTCTACACTCGGTGGGCGTCATCT	1	69
gyrBF6	ACACTGACGACATGGTTCTACACWAKGRGGGMSKCATTY	128	67
gyrBF7	ACACTGACGACATGGTTCTACAGAAGGCGGTACCCATGW	2	68
gyrBF8	ACACTGACGACATGGTTCTACAGAAGGKGGAACHCATKT	12	66
gyrBF9	ACACTGACGACATGGTTCTACAGAAGGTGGAACACATGA	1	66
gyrBF10	ACACTGACGACATGGTTCTACAGAAGGYGGAACKCAYSW	32	67
gyrBF11	ACACTGACGACATGGTTCTACAGAAGGYGGBACRCACGT	12	69
gyrBF12	ACACTGACGACATGGTTCTACAGACGGCGGCACBCMBHT	54	71
gyrBF13	ACACTGACGACATGGTTCTACAGACGGKGGTACSCACYT	8	69
gyrBF14	ACACTGACGACATGGTTCTACAGACGGYGGTACTCACCT	2	68

gyrBF15	ACACTGACGACATGGTTCTACAGAGGGWGGHACBCAYGW	72	68
gyrBF16	ACACTGACGACATGGTTCTACAGAGGGWGGMASMCAYSW	128	68
gyrBF17	ACACTGACGACATGGTTCTACAGAGGGYGGYACVCACGA	12	70
gyrBF18	ACACTGACGACATGGTTCTACAGAKGGSGGTACDCATBT	36	67
gyrBF19	ACACTGACGACATGGTTCTACAGAMGGHGGBACBCATCT	54	68
gyrBF20	ACACTGACGACATGGTTCTACAGARGGHGGAACYCATST	24	67
gyrBF21	ACACTGACGACATGGTTCTACAGARGGVGGVRCBCACGA	108	70
gyrBF22	ACACTGACGACATGGTTCTACAGARGGYGGAACWCACGA	8	68
gyrBF23	ACACTGACGACATGGTTCTACAGASGGGGGSRCGCATSW	32	71
gyrBF24	ACACTGACGACATGGTTCTACAGATGGAGGAACACACGA	1	67
gyrBF25	ACACTGACGACATGGTTCTACAGATGGCGGTACNCACCT	4	68
gyrBF26	ACACTGACGACATGGTTCTACAGATGGKGGWRYSCAYBT	192	68
gyrBF27	ACACTGACGACATGGTTCTACAGATGGTGGAACACACSW	4	67
gyrBF28	ACACTGACGACATGGTTCTACAGATGGTGGTACMCACCT	2	67
gyrBF29	ACACTGACGACATGGTTCTACAGATGGTGGWACVCACKT	12	67
gyrBF30	ACACTGACGACATGGTTCTACAGATGGWGGKACTCAYNW	64	66
gyrBF31	ACACTGACGACATGGTTCTACAGATGGYGGAACNCATKT	16	67
gyrBF32	ACACTGACGACATGGTTCTACAGAWGGHGGAACACACTT	6	66

gyrBF33	ACACTGACGACATGGTTCTACAGAWGGSGGCACBCACVW	72	70
gyrBF34	ACACTGACGACATGGTTCTACAGAWGGTGGNDBNCAYKW	2304	67
gyrBF35	ACACTGACGACATGGTTCTACAGAYGGMGGVACBCATHT	108	68
gyrBF36	ACACTGACGACATGGTTCTACAGAYGGSGGWACKCACTT	16	68
gyrBF37	ACACTGACGACATGGTTCTACAGAYGGYGGHDSHCAYNT	1728	68
gyrBF38	ACACTGACGACATGGTTCTACAKWWGGYGGAACMCATDT	96	66
gyrBF39	ACACTGACGACATGGTTCTACAKWWGGYGGKACCCATYT	64	67
gyrBF40	ACACTGACGACATGGTTCTACATATAGTGGCGGGCATGC	1	67
gyrBF41	ACACTGACGACATGGTTCTACATGGTGGTGTCCGCCAAC	1	69
gyrBF42	ACACTGACGACATGGTTCTACATKGTGKKGTCSKCCRAM	128	68
gyrBF43	ACACTGACGACATGGTTCTACATTGTGTGGTCGTCCGAA	1	67
gyrBR1	TACGGTAGCAGAGACTTGGTCTCAGCAGAGTCACCTTCVAC	3	68
gyrBR2	TACGGTAGCAGAGACTTGGTCTCAGCCGAATCGCCTTCCAC	1	69
gyrBR3	TACGGTAGCAGAGACTTGGTCTCAGCGCGGGGTCCTTTTCC	1	71
gyrBR4	TACGGTAGCAGAGACTTGGTCTCAGCKGAGTCYCCCTCTAC	4	68
gyrBR5	TACGGTAGCAGAGACTTGGTCTCAGCKSWATCTCCCTCNAC	32	68
gyrBR6	TACGGTAGCAGAGACTTGGTCTCBGCSSWKTCGCCCTCSAC	96	72
gyrBR7	TACGGTAGCAGAGACTTGGTCTCCGCAGAGTCACCCTCCAC	1	70

gyrBR8	TACGGTAGCAGAGACTTGGTCTCCGCDSWRTCVCCCTCCAC	72	71
gyrBR9	TACGGTAGCAGAGACTTGGTCTCCGCGGCGTTCCGGTCGCC	1	74
gyrBR10	TACGGTAGCAGAGACTTGGTCTCCGCRGAGTCACCTTCCAC	2	70
gyrBR11	TACGGTAGCAGAGACTTGGTCTCDGCMSWATCACCYTCBAC	144	68
gyrBR12	TACGGTAGCAGAGACTTGGTCTCGGCCGARTCGCCCTCCAC	2	72
gyrBR13	TACGGTAGCAGAGACTTGGTCTCGGCVGAGTCHCCYTCGAC	18	70
gyrBR14	TACGGTAGCAGAGACTTGGTCTCHGCTGARTCWCCTTCRAC	24	67
gyrBR15	TACGGTAGCAGAGACTTGGTCTCHGCVGAATCWCCYTCWAC	72	67
gyrBR16	TACGGTAGCAGAGACTTGGTCTCHGCWGAGTCACCCTCGAC	6	69
gyrBR17	TACGGTAGCAGAGACTTGGTCTCKGCAGAGTCTCCCTCAAC	2	68
gyrBR18	TACGGTAGCAGAGACTTGGTCTCKGCCGAGTYVCCYTCCAC	24	70
gyrBR19	TACGGTAGCAGAGACTTGGTCTCKGCTGAGTCWCCTTCCAC	4	68
gyrBR20	TACGGTAGCAGAGACTTGGTCTCMGCTGAGTCTCCCTCAAC	2	68
gyrBR21	TACGGTAGCAGAGACTTGGTCTCMGCWSWRTCACCCTCAAC	32	68
gyrBR22	TACGGTAGCAGAGACTTGGTCTCNGCVGAGTCCCCTTCCAC	12	70
gyrBR23	TACGGTAGCAGAGACTTGGTCTCRGCAGAGTYACCCTCAAC	4	68
gyrBR24	TACGGTAGCAGAGACTTGGTCTCRGCMGAATCYCCTTCGAC	8	69
gyrBR25	TACGGTAGCAGAGACTTGGTCTCSGCMGAGTCACCYTYTAC	16	68

gyrBR26	TACGGTAGCAGAGACTTGGTCTCSGCWGAGTCHCCTTCCAC	12	69
gyrBR27	TACGGTAGCAGAGACTTGGTCTCTGCAGAGTCACCCTCAAC	1	68
gyrBR28	TACGGTAGCAGAGACTTGGTCTCTGCAGAGTCTCCCTCCAC	1	69
gyrBR29	TACGGTAGCAGAGACTTGGTCTCTGCCGAGTCGCCCTCCAC	1	71
gyrBR30	TACGGTAGCAGAGACTTGGTCTCTGCCGAGTCGCCCTCGAC	1	71
gyrBR31	TACGGTAGCAGAGACTTGGTCTCTGCCGAGTCWCCCTCGAC	2	70
gyrBR32	TACGGTAGCAGAGACTTGGTCTCTGCGCTATCGCCTTCCAC	1	69
gyrBR33	TACGGTAGCAGAGACTTGGTCTCTGCNGAATYMCCYTCBAC	96	68
gyrBR34	TACGGTAGCAGAGACTTGGTCTCTGCRSWATCHCCCTCDAC	72	68
gyrBR35	TACGGTAGCAGAGACTTGGTCTCTGCVSWRTCTCCYTCTAC	48	67
gyrBR36	TACGGTAGCAGAGACTTGGTCTCVGCAGAATCHCCCTCCAC	9	69
gyrBR37	TACGGTAGCAGAGACTTGGTCTCVGCDSWRTCCCCCTCMAC	144	70
gyrBR38	TACGGTAGCAGAGACTTGGTCTCYGCGGMRTCHCSSTCSAC	192	71
gyrBR39	TACGGTAGCAGAGACTTGGTCTCYGCMGAATCGCCCTCRAC	8	70
gyrBR40	TACGGTAGCAGAGACTTGGTCTGTAGAGAGTCTTTGTGCAG	1	66
gyrBR41	TACGGTAGCAGAGACTTGGTCTGTCGAGCGCGCCCTTGCGG	1	73
gyrBR42	TACGGTAGCAGAGACTTGGTCTGTTGAGGGAAATTCAGCTG	1	66
gyrBR43	TACGGTAGCAGAGACTTGGTCTGTTGAGGGAGACTCAGCAG	1	68

Table 2. Common sequence linkers used to equally bind to amplicons from the initial PCR with degenerate primers. Note that each linker sequence is accompanied by a barcode used in Illumina MiSeq amplicon tagging and indexing (discussed in-text). Primers containing the combination of barcode and conserved sequence are collectively referenced as non-degenerate primers.

Common sequence 1 (CS1)

5'-(BARCODE)-ACACTGACGACATGGTTCTACA-3'

Common sequence 2 (CS2)

5′-(BARCODE)-TACGGTAGCAGAGACTTGGTCT-3′

Table 3. Composition of the ZYMO mock community DNA standard, based on DNA concentration. Genome copy number is a function of genomic DNA concentration divided by genome size.

	Theoretical Composition (%)	
Species	**gDNA Relative Abundance**	**Genome Copy Number**
Pseudomonas aeruginosa	12	6.1
Escherichia coli	12	8.5
Salmonella enterica	12	8.7
Lactobacillus fermentum	12	21.6
Enterococcus faecalis	12	14.6
Staphylococcus aureus	12	15.2
Listeria monocytogenes	12	13.9
Bacillus subtilis	12	10.3
Saccharomyces cerevisiae	2	0.6
Cryptococcus neoformans	2	0.4

Table 4. Reaction mix for qpPCR linear stage reactions.

Reagent	**Individual Volume**
EconoTaq 2x Master Mix	12.5 µL
Bovine Thrombin (2.5 µg/ mL)	0.5 µL
Nuclease-Free H20	1.5 µL
Primer Cocktail (250 nM concentration per level of degeneracy [primer specific])	9.5 µL
gDNA (>20 ng per ~5 Mbp genome)	1.0 µL

Table 5. Reaction mix for a 25 µL qpPCR exponential stage PCR (note: volume of water will change depending on concentrations of primers and input gDNA).

Reagent	**Individual Volume**
EconoTaq 2x Master Mix	12.5 µL
Nuclease-Free H20	9.5 µL
CS1 Primer w/Barcodes (250 nM final concentration)	1.0 µL
CS2 Primer w/Barcodes (250 nM final concentration)	1.0 µL
gDNA (>20 ng per ~5 Mbp genome)	1.0 µL

Table 6. Predicted restriction digestion activity for the *gyrB* gene amplicon (excluding linker sequences and barcodes).

Enzyme	Binding Site	Overhang	Frequency	Cut Positions
TaqI	TCGA	5'	2	218, 299
TaqI	TCGA	5'	3	15, 198, 234
TaqI	TCGA	5'	5	16, 178, 472, 550, 631
TaqI	TCGA	5'	1	288
TaqI	TCGA	5'	1	156

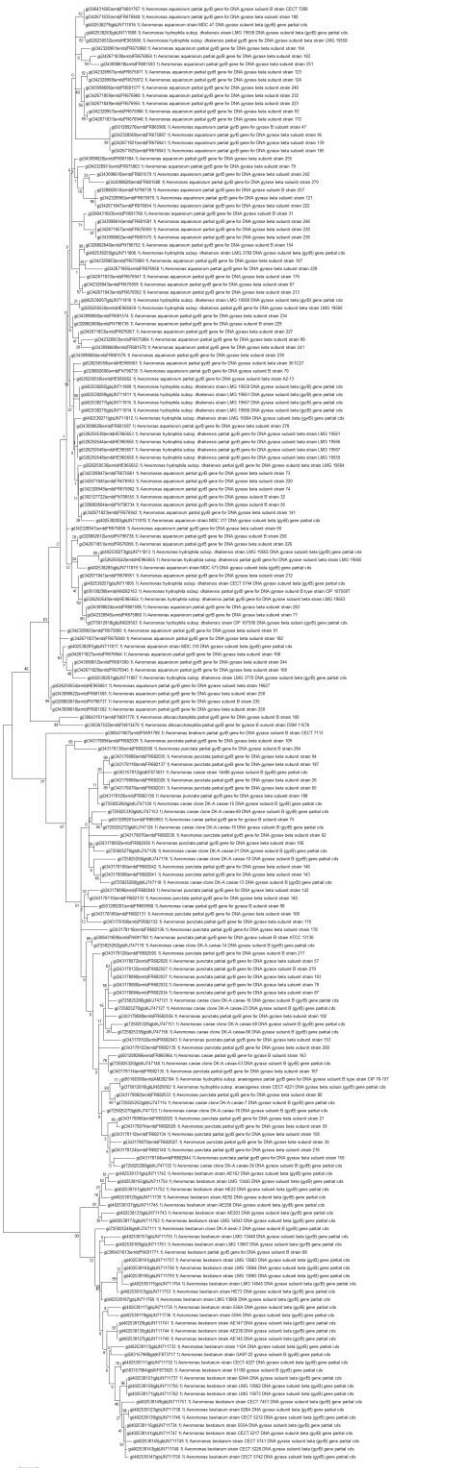


Figure 1. Phylogenetic map of *Aeromonas* spp. based on the *gyrB* gene. Bootstrap values of 90 were used as the threshold for taxonomic separation.

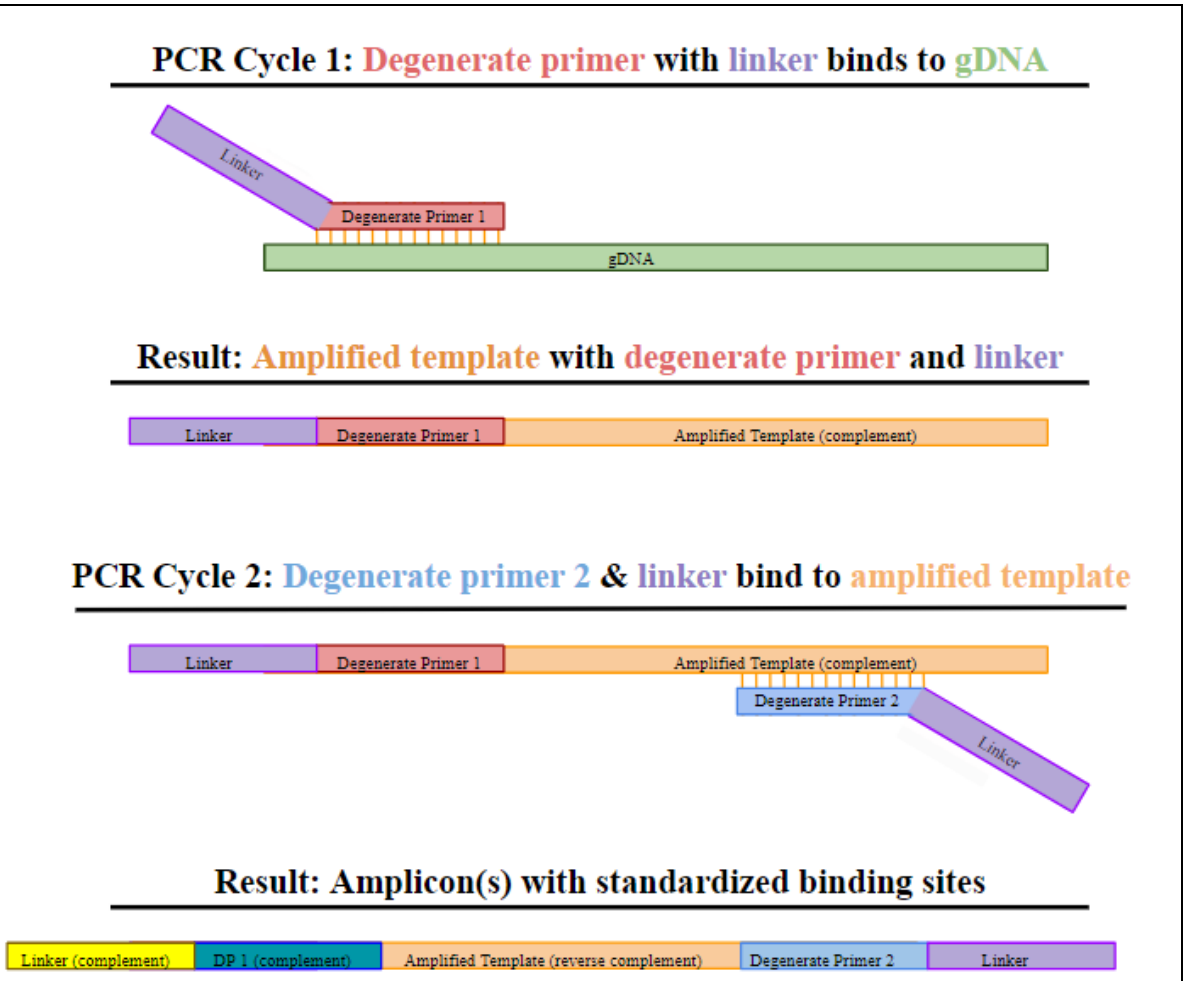


Figure 2. Conceptual schematic of binding for the first two cycles of PCR and the resultant amplicon composition.

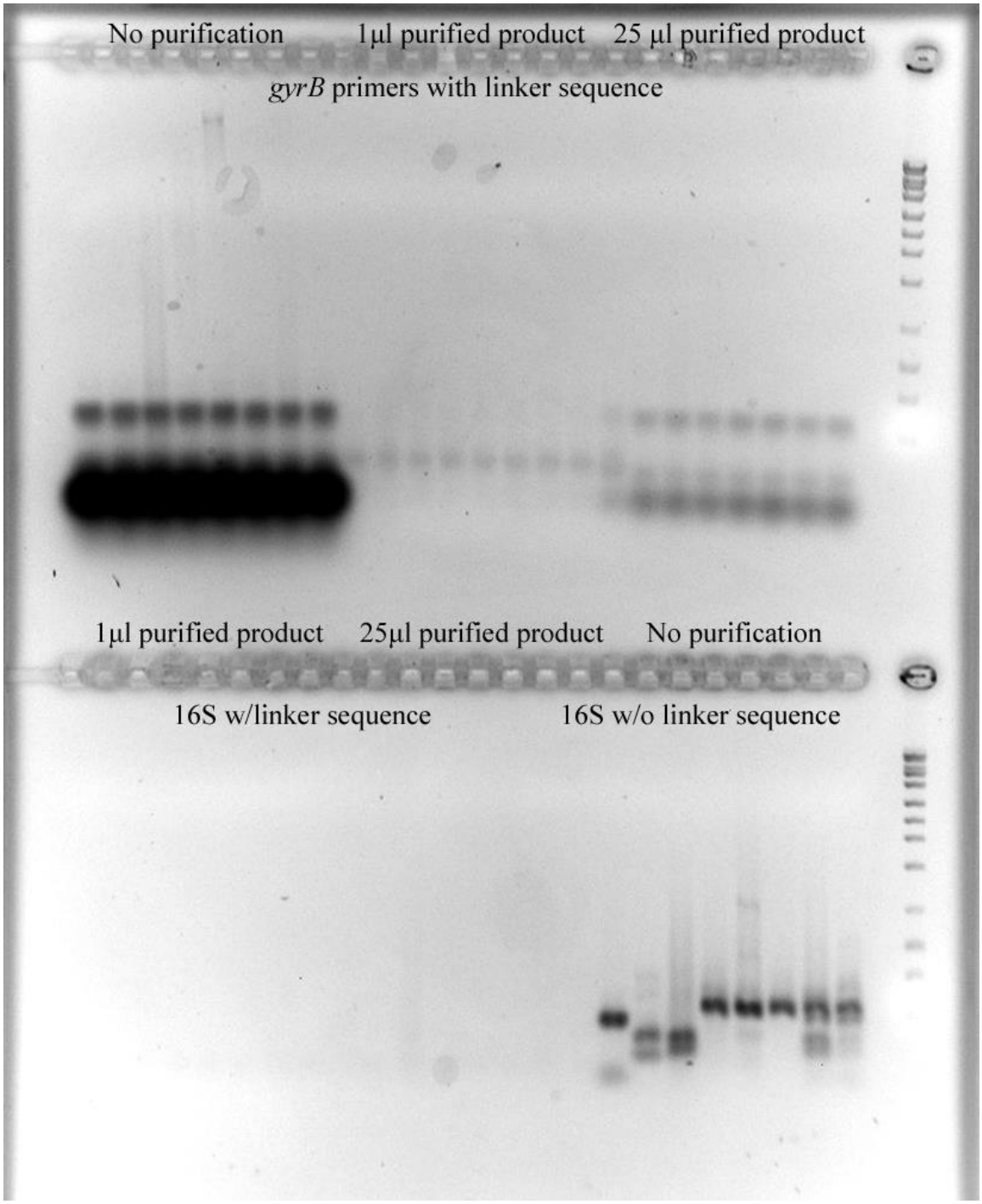


Figure 3. Overnight restriction digestion of amplicons produced with a set of 16S rRNA gene primers (27F and 907R) as well as the qpPCR primer cocktail, each with or without the non-binding linker sequence, with *TaqI*; DNA ladder used was the Lucigen 1 kb ladder.

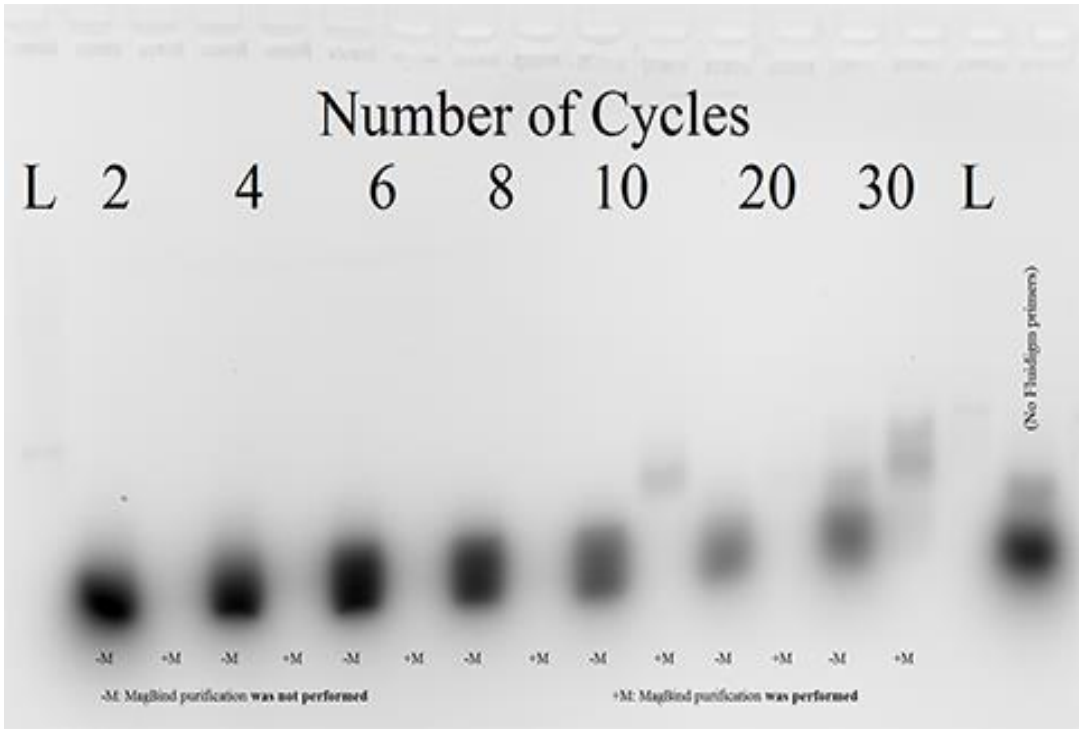


Figure 4. Amplicons resolved via gel electrophoresis where M+ represents MagBind magnetic bead purification being performed and M- represents no purification after the number of PCR cycles listed above, followed by 28 cycles of PCR with DePCR primers described below; DNA ladder (L) used was the Lucigen 1 kb ladder.

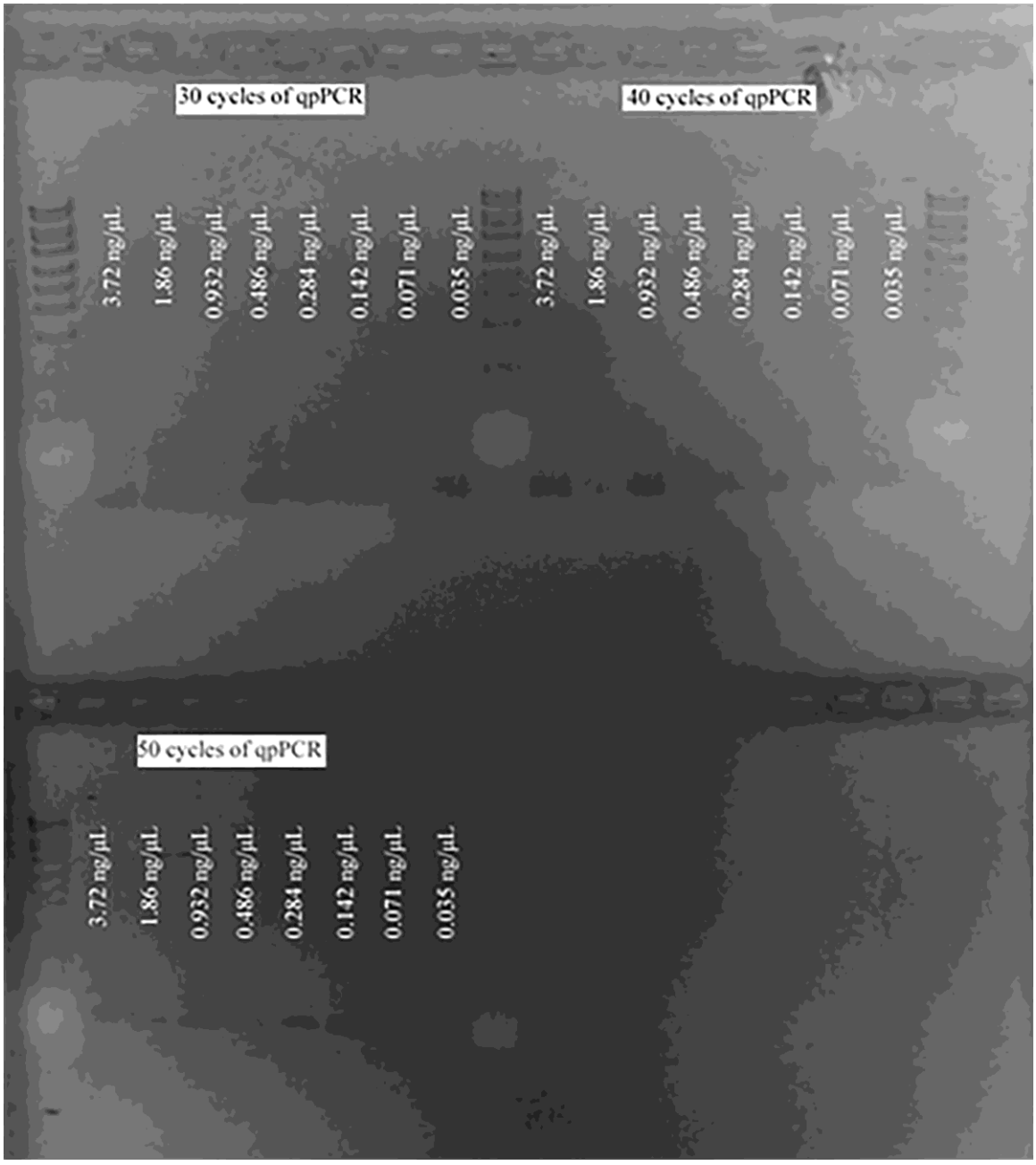


Figure 5. Gel electrophoresis of a gDNA concentrations, with two PCR cycles using the degenerate *gyrB* primer cocktail and 28, 38, and 48 PCR cycles with the non-degenerate CS1/CS2-tagged primers; DNA ladder used was the Lucigen 1 kb ladder.

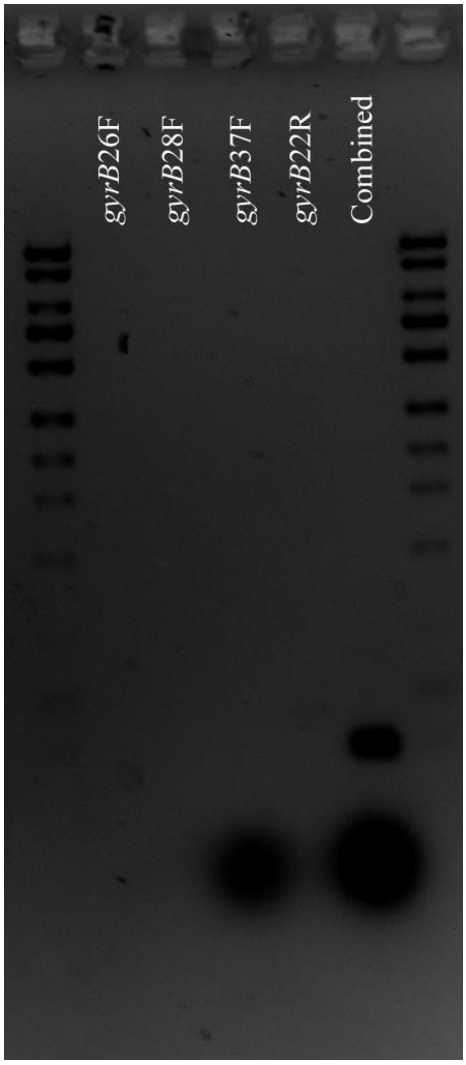


Figure 6. Example of evaluation of binding success for primer combinations. Primers *gyrB*26F (lane 2), *gyrB*28F (lane 3), *gyrB*37F (lane 4), and *gyrB*22R (lane 5) are shown, with amplification only occurring after a complementary primer pair was included (lane 6); DNA ladder used was the Lucigen 1 kb ladder.

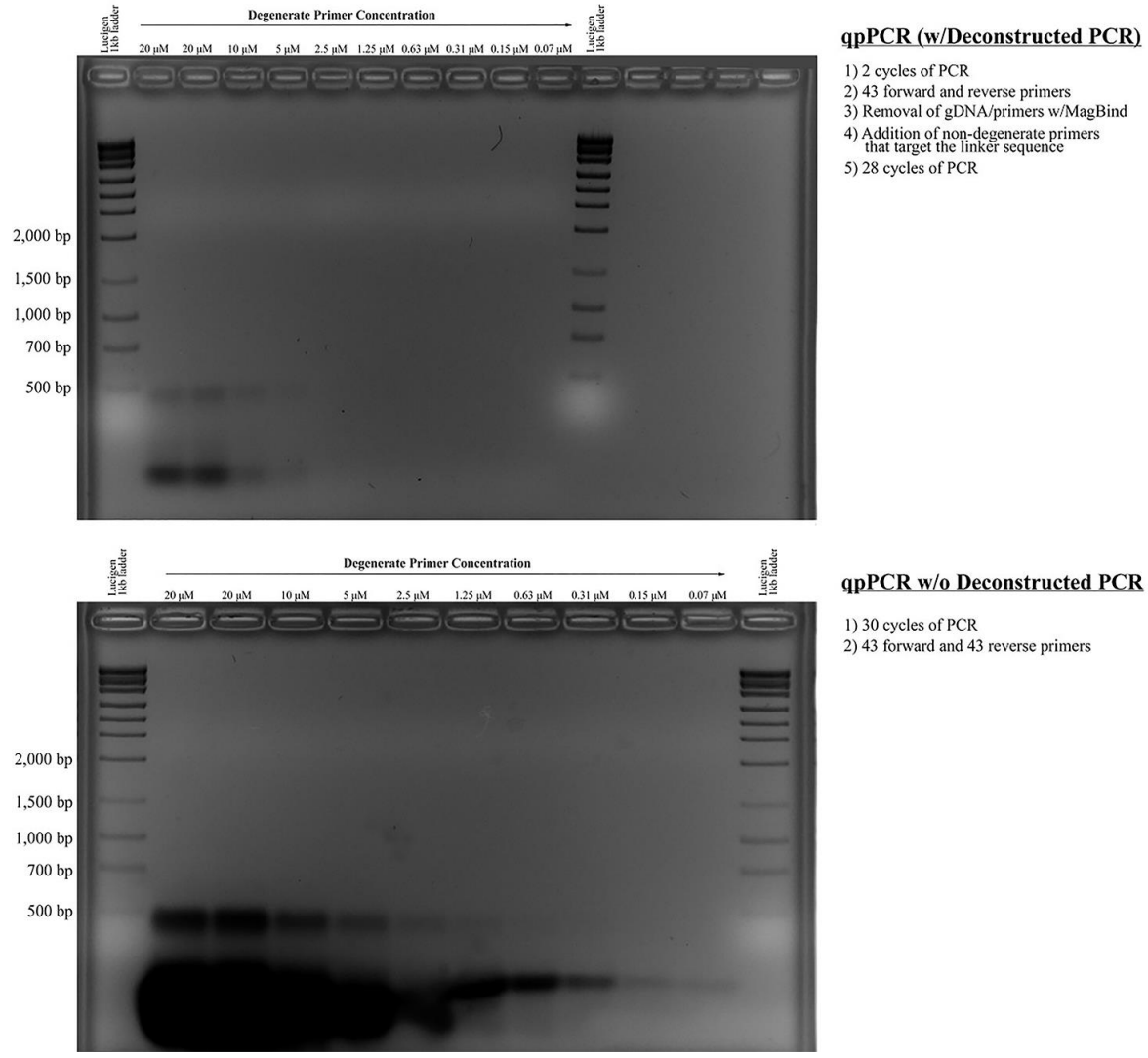


Figure 7. Evaluation of the minimal degenerate *gyrB* primer cocktail concentration required for qpPCR, with and without magnetic bead purification.

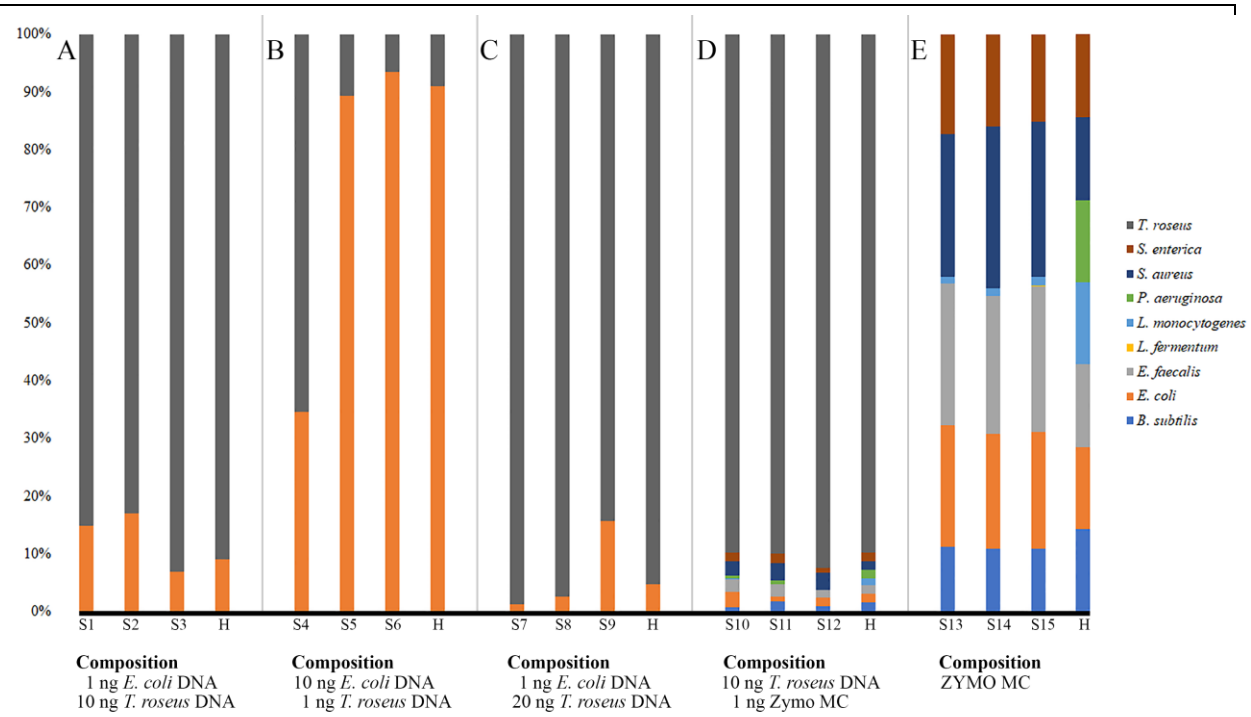


Figure 8. Read mapping of qpPCR results for (A) 1 ng of *E. coli* DNA and 10 ng of *T. rosesus* DNA, (B) 10 ng of *E. coli* DNA and 1 ng of *T. rosesus* DNA, (C) 1 ng of *E. coli* DNA and 20 ng of *T. rosesus* DNA, (D) 1 ng of the ZYMO mock community DNA standard and 10 ng of *T. rosesus* DNA, (E) 10 ng of the ZYMO mock community DNA standard. Samples (S) were performed in triplicate; hypothetical (H) outputs were based on input DNA concentrations as verified by Qubit.

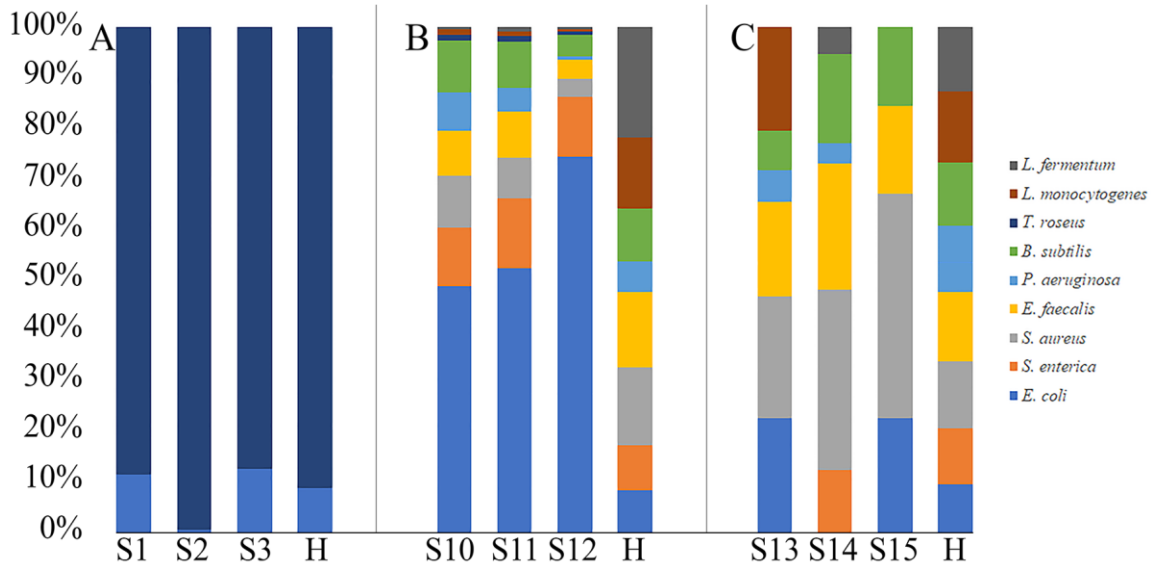


Figure 9. Kraken2 k-mer mapping of qpPCR results for (A) 1 ng of *E. coli* DNA and 10 ng of *T. rosesus* DNA, (B) 1 ng of the ZYMO mock community DNA standard and 10 ng of *T. rosesus* DNA, (C) 10 ng of the ZYMO mock community DNA standard. Samples (S) were performed in triplicate; hypothetical (H) outputs were based on input DNA concentrations as verified by Qubit.

Equation 1. Calculation of PCR product based on input copy number and number of cycles; equation assumes no random interference and no limiting reagents.

$$PCR\ product\ =\ (template\ copy\ number)\ \times\ 2^{(PCR\ cycles)}$$

Equation 2. Copy number of a given sequence, based on substance properties.

$$Copy\ number \approx \frac{(volume\ of\ reaction\ \mu L) \times \left(concentration \frac{ng}{\mu L}\right) \times \left(6.0221 \times \frac{10^{23} molecules}{mole}\right)}{(molecular\ weight\ of\ sequence) \times\ 10^{9} \frac{nanograms}{gram}}$$

Chapter IV

Evaluation of the Enzyme Phytase and the Phytase-expressing Probiotic *Bacillus velezensis* in the Prevention of Disease Due to Hypervirulent *Aeromonas hydrophila*

1. Abstract.

This chapter is a summary of two studies aimed to evaluate the effects on the survival of channel catfish and on the microbial community of the catfish gut when the enzyme phytase or the phytase-producing enzyme *Bacillus velezensis* AP193 are amended to soy-based feed. Towards this aim, two complementary studies were performed to first identify these interactions, then to contextualize these results in a production setting. The first study consisted of a fin-clip disease challenge in aquaria, which showed significant reductions in mortality due to the hypervirulent pathotype of *Aeromonas hydrophila* (v*Ah*) in both fish given feed amended with the enzyme phytase (23%) as well as fish given feed amended with the probiotic (56%). In addition to detailing reductions in mortality, this challenge also provides evidence that including these agents as feed additives results in a microbial shift, which may contribute to host resistance against v*Ah*. Furthermore, this challenge also provides a potentially significant line of evidence in that virulence is attenuated when inositol hexakisphosphate/phytic acid is degraded, even

when v*Ah* is introduced into aquaria at 1.8 x 10^9 colony forming units per milliliter. In this approach, the benefits conferred based on nutrient advantage were reduced, which allowed examination of the previously postulated role of this carbon source in transcriptional regulation of virulence factors. The second study consisted of identical treatments/controls in production ponds, which served to evaluate the net effect these approaches have when subjected to practical conditions and when used in concert with other biological control agents. Collectively, this research describes a multi-faceted approach to evaluating the effect that introducing the enzyme phytase or the probiotic *B. velezensis* AP193 has on catfish survival and the catfish gut microbiome, while also providing new lines of evidence towards understanding the natural mode of infection for this ubiquitous freshwater pathogen.

2. Introduction.

Since 2009, catfish producers in the US have lost millions of pounds of market-sized channel catfish (*Ictalurus punctatus*) each year due to outbreaks of a hypervirulent strain of *Aeromonas hydrophila* (v*Ah*). Early phenotypic characterizations by API 20E assay, produced from Dr. Mark R. Liles' lab, indicated that members of the hypervirulent pathotype of *A. hydrophila* had a unique biochemical profile, including the ability to ferment inositol – a well-documented signaling molecule that has enabled intracellular replication in macrophages within numerous pathogenic bacteria, including *Salmonella enterica* serovar Typhimurium; *Clostridium perfringens*; and *Bacillus subitilis* [262,263]. This anomaly later proved to hold significance because only v*Ah* (not their environmental counterparts) could use *myo*-inositol as a sole carbon source, as verified by growth on M9 minimal media [264]. Since its inception, M9 media has served as a consistent diagnostic for identifying v*Ah*, resulting in over 100 epidemic isolates with an affinity

for *myo*-inositol since 2012 (n=129) [222]. From these samples and subsequent research, sequence analyses of isolates taken from production ponds across Alabama and Mississippi also showed support for the consistent nature between this phenotypic trait and a possible linkage with the ability for v*Ah* to induce motile *Aeromonas* septicemia (MAS) [233,264]. Within the same study, this analysis was extended to include all phylogenetically confirmed members of *A. hydrophila*, which showed gene clusters associated with *myo*-inositol metabolism are unique to the v*Ah* pathotype within the species *A. hydrophila* (**Figure 1**) [1], thereby warranting investigation of the use of biological control agents that target this ubiquitous pathogen.

This research investigates two potential biological control approaches with the shared goal of breaking a cycle of nutrient accumulation linked with devastating outbreaks of MAS in channel catfish. This cycle begins when plant-based feeds (e.g. soy) that contain inositol hexakisphosphate/phytic acid (a carbon source that is indigestible by the catfish and by the normal microbiota within the catfish gut [265]) are introduced into aquaculture systems instead of fish meal-based feeds (e.g. forage fishes), resulting in approximately 80% of dietary phytic acid being released into the environment [124,266-268]. Once deposited into anoxic, low pH, sediment, the next step of this cycle begins as soil-associated microbes metabolize phytate/phytic acid, leading to the environmental accumulation of *myo*-inositol [257]; a carbon source that has been shown to abolish salicylic acid-dependent cell death and, as described above, enables intracellular proliferation in known pathogens [269,270]. For the recently classified v*Ah* pathotype associated with mortality in freshwater fishes [143], the abundance of *myo*-inositol is not only significant because this Gram-negative bacterium can use *myo*-inositol as a sole carbon source, thereby providing a selective advantage [263], but also because research shows that when the *myo*-inositol metabolic pathway is mutagenized, there are downstream regulatory effects, resulting in the

attenuation of v*Ah* virulence in an intraperitoneally injected channel catfish model [271]. In-line with previous research that links this molecule with pathogen invasion and intercellular replication, the underlying hypothesis is that v*Ah* evolved to upregulate virulence factor expression in concert with consumption of *myo*-inositol. After exposure to this environmental signal, v*Ah* is thought to switch from persistence within the environment towards the invasion, colonization, and degradation of host tissues, followed by rapid proliferation [1].

To silence this environmental signal, the enzyme phytase as well as the phytase-producing probiotic *Bacillus velezensis* strain AP193 were selected to be amended to catfish feed based on previous research based on their ability to hydrolyze phytate, releasing phosphorous for absorption by the catfish and its normal microbiota [265]. While phytase is already widely used in aquaculture to increase fish growth [112,272] and is poised to become one of the most pervasive methods used in v*Ah* biological control within the United States (Askelson, 2014 #7), this study is foundational in determining the system-wide effects of employing phytase and the phytase producing probiotic *B. velezensis* as a prophylactic agent. In addition to the capacity to phosphohydrolyze phytate, previous research has also shown *B. velezensis* AP193 as having antagonistic activity against v*Ah* as well as other freshwater pathogens *in laminam* (**Figure 2**) [4]; however, this research will be the first to quantify the effect of this probiotic within the catfish microbiome as well as with the probiotic's effect on increasing host survival.

To quantify these effects, multiple molecular diagnostics were used to produce results are both comparable with previous literature employs 16S rRNA gene sequencing methods and to produce results with greater taxonomic resolution and quantitative accuracy than previously possible by implementing new sequencing approaches, including the recently described

quantitative polybacterial PCR (qpPCR) method and whole metagenome sequencing with Oxford Nanopore's MinION platform (Oxford, United Kingdom).

This project also evaluates the disparity between challenges performed in production ponds with those performed in aquaria by designing two complementary experiments. The first experiment was designed as a disease challenge in aquaria to remove confounding variables associated with studies in natural systems and define the specific effect of the biocontrol agents on the abundance of v*Ah*. The second experiment aimed to evaluate these biocontrol agents in raceways by sampling production ponds with a history of v*Ah* outbreaks over the course of the channel catfish growing season. By coupling these methods, the biases of each experimental design are cross-checked by comparing results to address confounding variables that would otherwise remain as artifacts. In summary, this research tracks the persistence of v*Ah* within a complex bacterial community, provides insight into the interactions between these biocontrol agents and the metagenome of the catfish gut, evaluates reductions in mortality associated with amending phytase or the probiotic *B. velezensis* AP193 into catfish feed, and establishes new lines of evidence that refine our knowledge of the natural mode of infection for v*Ah* and direct the design of disease challenges that aim to emulate natural infection.

3. Methods.

3.1. Animal Welfare Statement.

All Channel catfish challenges were conducted under the approval of the Animal Care and Use Committee (IACUC) of Auburn University in compliance with U.S. regulatory standards for the humane care and use of laboratory animals. All field trials were conducted with

approval by the Alabama State veterinarian and USDA APHIS Center for Veterinary Biologics (Ames, IA).

3.2. Aquaria Challenge: Preparation of the Probiotic *B. velezensis* on Catfish Feed.

A modified sporulation protocol was performed on a culture of *B. velezensis* AP193 (GenBank: JX094285.1; partial genome is available upon request) that had been grown for 24 hours in tryptic soy broth (TSB) after inoculating the media with an isolated colony grown from a *B. velezensis* AP193 glycerol stock that was previously streaked for isolation on tryptic soy agar (TSA) at 28°C and taxonomically verified through PCR [4,273]. After cells were grown overnight on TSB, cells were transferred to spore preparation agar (**Table 1**) and incubated at 28°C for 7 days. Once spores were given time to form, 5 mL of sterile water used to dislodge bacteria (with a sterile loop), then cells were placed into a sterile test tube and held at 80°C for 15 min to destroy vegetative cells, leaving only spores. Subsequent concentrations were performed by serially diluted plate counts on TSA after an overnight incubation at 30°C. After verification of design integrity and confirmation of identity through PCR [4], spores were submitted to Osprey Biotechnics (Sarasota, FL) for large-scale sporulation. Resultant spore powder ($6.0x10^{11}$ CFU/gram) was resuspended in corn oil and spray-coated on commercial soy-based feed, based on the previously effective doses for *Bacillus* spp. at 0.5% of the total feed weight [4,126]. During the duration of the experiment, the integrity of spores was checked at random by plate counts on TSA after overnight growth.

3.3. Raceway Study: Preparation of the Probiotic *B. velezensis* on Catfish Feed.

An identical sporulation protocol and feed amendment was performed on a culture of *B. velezensis* AP193, as described above.

3.4. Aquaria Challenge: Preparation of the Phytase Enzyme on Catfish Feed.

The phytase enzyme was suspended in distilled water at a volute to weight ratio of 1% to achieve a 2,500 phytase units per kilogram of feed (FTU/kg) application rate. After dilution, an even coat was applied to feed while being actively rotated/tumbled in a mixer (batch size averaged 50 pounds). After drying, a 1% corn/fish oil topcoat was applied, then allowed to dry.

3.5. Raceway Study: Preparation of the Phytase Enzyme on Catfish Feed.

An identical phytase feed amendment protocol was performed for the raceway study, as described above for the challenge in aquaria.

3.6. Aquaria Challenge: Catfish Background.

Each aquarium used for this study (n = 12) contained 10 channel catfish fingerlings (Marion strain), with no history of disease, and were fed different diets (probiotic, phytase, or control [preparation described above]) prior to a standard fin clip (FC) challenge [274], while negative controls (n = 3; no v*Ah* inoculum) consisted of one tank with fish fed with phytase-coated pellets, one tank with fish fed with probiotic-coated pellets, and one tank with fish fed with untreated pellets, each with 10 catfish per tank. Water temperature was maintained at ~30°C for the duration of the study. Treatment and control groups were assigned using a randomized block design so that one replicate per group was ensured, totaling three replicates. Delivery of food was consistent with standard practices within catfish aquaculture, which translates to

supplying volumes of food ranging from 2–4% of fish weight per day. At the completion of the study, fish were exposed to tricaine mesylate (MS-222) for five minutes.

3.7. Raceway Study: Catfish Background.

Each raceway used for this study (n = 36) contained ~1,200 channel catfish fingerlings, with each pond containing a total of 16 raceways. Within each pond, three raceways were given food amended with phytase, three raceways were given food amended with the probiotic, three raceways were given food amended only with corn oil (serves as a blank), and the remaining three raceways were given a vaccine that is out of the scope of this study (results of the vaccine treatment are available in the dissertation "Disease Prevention in Channel Catfish (*Ictalurus punctatus*) Through the Use of an Attenuated *Aeromonas hydrophila* Vaccine or the Probiotic Effects of *Bacillus velezensis* AP193" by Charles Thurlow [2018]). Treatment and control groups were assigned using a randomized block design so that one replicate per group was allowed, totaling four replicates per pond, with a total of three ponds. Delivery of food was consistent with standard practices within catfish aquaculture, which again, translates to supplying volumes of food ranging from 2–4% of fish weight per day. Because of the nature of raceways, the potential for cross-contamination was evaluated by assessing water quality parameters and assaying water samples for the probiotic, as described in a parallel study [4]. As mortality occurred throughout the study, fish were removed. If fish exhibited characteristic symptoms of MAS [143], then samples were removed and evaluated for growth on M9 media. At the completion of the study, fish were counted to assess total mortality and sold at market.

3.8. Aquaria Challenge: Preparation of *A. hydrophila*.

To induce symptoms of the disease MAS, a pure culture of hypervirulent *Aeromonas hydrophila* ML10-51K was selected because of this strain's capacity to induce symptoms of disease by intraperitoneal injection as well as by immersion with fin clip [1,274]. After streaking ML10-51K for isolation on sheep's blood agar (SBA), five isolated colonies were used to inoculate 10 mL of tryptic soy broth (TSB) with 0.4 mM deferoxaminemesylate salt [119], which was then grown to log phase (optical density of 600 nm [OD_{600}] = 0.6) in a shaking incubator (120 RPM) at 30°C. Five milliliters of this culture was used to inoculate 500 mL of TSB with 400 µM of DFO, which was again grown to log phase (OD_{600} = 0.6) in a shaking incubator (120 RPM) at 30°C. Replicate plate counts yielded an average of 1.8×10^9 colony forming units (CFU)/mL. Challenge dose was 50 mL of v*Ah* culture per 10 L tank to lower overall mortality to emulate natural mortality rates [274].

3.9. Aquaria Challenge: Sample Collection.

Samples were collected from treatment and control groups at the start of the experiment, then after the disease challenge (5 days; n = 10 per treatment/control). Sample collection involved exposing catfish to MS-222 for five minutes, followed by surgical excision from behind the opercula to the stomach, then to the vent. After the body cavity was exposed, incisions were made below the stomach and at the anal pore. Connective tissue was lacerated with a scalpel. To maintain microbiome integrity after the intestine was separated from the body, the sample was transferred to a 50 mL conical tube and submerged in 500 µL RNA*later* (Ambion® ThermoFisher Scientific, Massachusetts, USA), then kept on ice during transit for approximately one hour. After transport, samples were maintained at -80°C until DNA extraction could be performed [275].

3.10. Raceway Study: Sample Collection.

Samples were collected from each treatment/control at the start of the experiment (referred to as time zero; n = 6 control group fish), then after one month and after three months (n = 6 per treatment/control group). Sample collection involved exposing catfish to MS-222 for five minutes, followed by surgical excision from behind the opercula to the stomach, then to the vent, as described above (**Figure 5**). After the body cavity was exposed, incisions were made below the stomach and at the anal pore. Connective tissue was lacerated with a scalpel. To maintain microbiome integrity after the intestine was separated from the body, the sample was transferred to a 50 mL conical tube and submerged in 500 µL RNA*later*, then kept on ice during transit for approximately three hours. After transport, samples were maintained at -80°C until DNA extraction could be performed [275] to assess shifts in microbial assemblage before, during, and after v*Ah* outbreaks, with data sampled synchronously from control groups unaffected by MAS or by treatment.

3.11. DNA Extraction.

After tissue homogenization of the entire gut from each sample collected from both the raceway study and the pond challenge with a Kinematica Polytron™ PT 10/35 GT homogenizer (Luzern, Switzerland), gDNA was isolated using a MoBio DNA fecal isolation kit (MoBio, California, USA), with bead-beating, with the aim of improving yield from both Gram-positive and Gram–negative bacteria. Extracted gDNA was quantified by Qubit™ fluorometric quantification using the dsDNA high sensitivity assay kit (ThermoFisher, USA) and preserved at -20°C for downstream analyses.

3.12. 16S rRNA Gene Sequencing.

Extracted DNA from the aquaria challenge and the raceway study with quantifiable concentrations (≥0.1 ng/μL; n = 60; 20 samples per group) was submitted to the University of Illinois at Chicago's Sequencing Core (UICSQC) for 16S rRNA gene sequencing using a standard 16S rRNA gene bacterial primer set, 515FB/806RB, which targets the V4 region of the 16S SSU and a high-fidelity polymerase to generate amplicons for sequencing using the Illumina MiSeq sequencing technology (CA, USA), with the 2 x 300 paired-end v2 sequencing kit and a 30% phiX control.

3.13. Raceway Study: qpPCR Microbiome Sequencing.

Numerous attempts were made to employ the quantitative polybacterial polymerase chain reaction (qpPCR) method, described in Chapter III, to this metagenomic data. Low concentrations of non-fragmented gDNA (a common trait of samples taken from gastrointestinal tracts) compounded by inefficient amplicon purification (discussed in depth in Chapter III), was hypothesized to lead to inconsistent amplification, which precluded the inclusion of this method from this study. If this method was included, there would be an inherent potential for not accurately capturing the microbiome (based on preliminary results with gDNA concentrations that would presumably be several orders of magnitude higher than those found within a complex microbiome). Therefore, while this limitation is also evident within 16S rRNA amplicon sequencing and, to a lesser extent, MinION 'whole metagenome' sequencing, the results of this approach were not included within the core findings of this project. All data from this approach is available upon request.

3.14. Aquaria Challenge: MinION Microbiome Sequencing.

Samples with extracted DNA, with a minimum total abundance of 1.5 μg, were sequenced with the Oxford Nanopore MinION technology (Oxford, United Kingdom). Equal numbers of samples from treatments and controls were selected that met this criteria (n = 6 per group; groups that exceeded this number had samples excluded at random to maintain equal sample sizes) and loaded sequentially on a separate flow cell per group (R9.4.1) after preparation with the ligation sequencing kit (SQK-LSK109). Washes were performed between samples with a flow cell wash kit (EXP-WSH002). Raw data produced by MinION sequencing was processed from FAST5 to FASTQ format with Guppy (v. 2.3.1; default parameters).

3.15. QIIME Processing of 16S rRNA Gene Sequencing Results.

Changes in relative abundance of bacterial taxa were assessed using the Python-based QIIME bioinformatics pipeline v.1.9.1 (http://qiime.org/) at the level of family and at the level of genus. The specific order of processing was as follows: library generation (files available upon request), barcode trimming with the script split_libraries.py, selection of operational taxonomic units (OTUs) with the script pick_otus.py, selection of a representative set with the script pick_rep_set.py, assignment of taxonomy with the script assign_taxonomy.py, generation of an OTU table with the script make_otu_table.py, summarizing taxa with the script summarize_taxa.py, testing evolutionary distance by building an alignment with the script align_seqs.py, hard sequence filtration of the alignment using a Lane mask with the script filter_alignment.py, generation of a phylogenetic tree with FastTree with the script make_phylogeny.py, rarefaction analyses with the script multiple_rarefactions.py, assessment of

alpha diversity with the script alpha_diversity.py, assessment of beta diversity with the script jackknifed_beta_diversity.py, and distance statistics were created with the script dissimilarity_mtx_stats.py. Identical parameters and scripts were used for the dataset generated from the pond study as well as the dataset generated from the aquaria challenge. All scripts referenced above are available on the QIIME website and are also available upon request.

3.16. Aquaria Challenge: Kraken 2 Processing of MinION Sequencing Results.

Changes in relative abundance of bacteria in the aquaria challenge were assessed using Johns Hopkins' Kraken 2 bioinformatics pipeline (v. 2.0.7-beta; https://ccb.jhu.edu/software/kraken2/) at the level of family and genus. In order to make these analyses inclusive, the standard Kraken 2 database was downloaded, which uses complete genomes from the National Center for Biotechnology Information's (NCBI) Reference Sequence database (RefSeq) for the bacterial, archaeal, and viral domains as well as entries for the human genome and a collection of known vectors. Default parameters were used for both database generation (data available upon request) and assignation of taxonomy were used as well as for k-mer length (35) and minimizer length (31), with the assignation threshold set to "-confidence 0.95" (assignations were only performed if 95% of k-mers mapped to the lowest common ancestor). Although k-mers are somewhat comparable to reads, the effect that k-mer size has on accuracy of mapping (and sequencing in general) cannot be understated [276,277]; however, no marked improvement was made after altering this parameter. Resultant analyses were transformed with a Python script for figure generation and statistical comparison in the software R (R Foundation for Statistical Computing, Vienna, Austria).

3.17. Statistical Analyses.

All statistical analyses were performed with the software R. Significance for differences in mortality was performed with a one-way analysis of variance [257] and Tukey's test for post-hoc analysis. Significance was set at the 95% confidence threshold; requiring a p-value ≤ 0.05. In addition to direct comparisons between microbial taxa that employed the previous method, differences in microbial abundance were analyzed on a larger scale by assessing model fit with survival as the response variable and the relative abundance of each bacterial/viral/fungal family or genus as the informative variables, then using the dredge function within the Multi-Model Inference (MuMIn) package to assess correlation using the statistical method Akaike information criterion; corrected for small sample sizes (AICc).

4. Results.

4.1. DNA Extraction.

Samples with DNA concentrations below 1 ng/μL were re-extracted up to three times. If the final DNA concentration did not meet this criterion, then the sample was excluded from all downstream analyses. Samples were excluded at random from other treatment/control groups if the total number of samples within a single treatment/control was uneven. The resulting samples within quantifiable ranges resulted in six per treatment/control for the aquaria challenge and six per treatment/control for the raceway study.

4.2. Aquaria Challenge.

Reductions in mortality were present for both treatment groups. Catfish that were fed a diet amended with the probiotic, *B. velezensis* strain AP193, showed reductions in mortality of

56.67% (n = 17; p-value ≥ 0.05), resulting in a total mortality of 3.33% after the four-day challenge. Similarly, catfish that had been fed a diet amended with the phytase enzyme, showed reductions in mortality of 23.34% (n = 7; p-value ≥ 0.05), resulting in a total mortality of 36.67% after the four-day challenge. In contrast, the control group had a total mortality of 60% after the conclusion of the study (**Figure 3**). The one-way ANOVA and Tukey Multiple Comparison Test showed that the differences between each group was significant (p-value ≥ 0.05).

4.2.1. QIIME Bioinformatics Pipeline Results; 16S rRNA Gene Sequencing.

The QIIME bioinformatics pipeline was performed after 16S rRNA gene sequencing for each treatment and for each control group. In the aquaria challenge, taxonomic assignation was limited to the levels of family to avoid spurious classification (**Figure 4**). While members of *Aeromonadaceae* and members of *Bacillaceae* were detected within each sample, there were no significant differences in relative abundance at the 95% confidence threshold between the control and either of the treatments. Beyond the stability of these two microbial families, notable shifts were present in the catfish gut microbiome predominantly consisting of of a shift from members of *Fusobacteriaceae* to members of *Pseudomonadaceae* and members of *Enterobacteriaceae* in both the treatment with phytase-amended feed and the treatment with the probiotic-amended feed. In contrast to the aquaria study, the shifts in microbial abundance of the genus *Bacillus* and of the genus *Aeromonas* were evident in raceways over the growing season when evaluated using read mapping (**Figure 5**). While these shifts appear to match expected trends, such as an increase in *Aeromonas* spp. in control fish during the growing season and increases in *Bacillus* spp. in fish fed the probiotic-amended feed, the variation between samples within the same treatment/control blurred potential effects. Interestingly, when the same dataset is analyzed by the QIIME pipeline,

the genus *Aeromonas* is not represented at any time point or group (**Figure 6**). Furthermore, the shifts in relative microbial abundance seen in the aquaria challenge were absent. Based on these disparities, future research should aim to increase sampling frequency and continue to cross-validate results, so the underlying interactions remain after factors that obscure these analyses are removed.

4.2.3. Kraken 2 Results from MinION Sequencing Data.

The microbial profile of each catfish gut was captured by raw sequence analysis of post-challenge fish (**Figure 7**). Considering the method to induce disease was an immersion challenge (with fin clip) and not induction through gavage or intraperitoneal injection, the lack of v*Ah* abundance within the catfish gut is a striking result. In effect, this result supports the hypothesis that the manifestation of disease is not precluded by gut colonization. Similar to the results processed with QIIME, species diversity remained virtually constant, irrespective of treatment/control.

4.3. Raceway Study.

Total mortality of each raceway was monitored through the duration of the study, with each measurement surveying and removing moribund fish daily. After the completion of the study, catfish that had been fed a diet amended with the probiotic, *B. velezensis* strain AP193, showed total reductions in mortality of 2.17% (p-value = 0.754) while fish that were fed a diet amended with the enzyme phytase showed an average reduction in mortality of 2.87% (p-value = 0.696) (**Figure 5**). As stated above, the one-way ANOVA and Tukey Multiple Comparison Test showed that the differences between the probiotic, phytase, and control groups were not

significant at the 95% confidence threshold, based on a one-way ANOVA and Tukey's Multiple Comparison Test (p-value > 0.05).

4.3.1. Raceway Study: Mortality in the Context of the Catfish Gut Microbiome.

The QIIME bioinformatics pipeline was performed on each treatment and control, with results being pooled into groups. Taxonomic assignation was limited to the levels of genus (**Figure 6**). As described above, taxa assigned to the genus *Bacillus* and to the genus *Aeromonas* were compared against mortality at each timepoint (**Figure 5**). While members of each of these genera are present within the environment, phylogenetic analyses at the level of species or subspecies are precluded by a lack of taxonomically informative nucleotides within the 16S rRNA gene. Combined with the disparity between read mapping results and the results from QIIME, before accurate descriptions of the microbiome state can be identified, deeper phylogenetic placement must be assessed, confounding variables must be addressed, and more frequent sampling methodologies must be implemented.

5. Discussion.

The core aims of this research were to assess the prophylactic efficacy of feed enriched with phytase as well as feed that was coated with the phytase-producing probiotic *B. amyloliquefaciens* strain AP193, to determine the specific effect that reducing *myo*-inositol concentrations within the production systems has on the catfish microbiome, to track the shift in microbial assemblage of ponds with culture-independent methods (e.g.16S rRNA gene-based sequencing, MinION sequencing, and the *gyrB* gene-based qpPCR technique), to identify if the gut composition of channel catfish was linked with susceptibility to v*Ah*, and to identify shifts in microbial assemblages within catfish guts prior to v*Ah* infection, with the aim of identifying

compositions of enteric bacteria that predispose fish to v*Ah* infection. Towards these aims, a challenge in aquaria and a raceway study in production ponds was performed. While the results of the raceway study show divergence from the results produced in aquaria challenges, these findings are discussed from several perspectives.

In the aquaria-based disease challenge, both the phytase enzyme (33% mortality; p-value > 0.05) as well as the *Bacillus* probiotic (3% mortality; p-value > 0.05) showed significant capacity as prophylactic agents against this devastating pathogen (60% mortality observed in the control) – an impressive feat considering v*Ah* was introduced in lethal concentrations (the ability for v*Ah* to proliferate through use of *myo*-inositol as a sole carbon source was not captured with this approach, but potential regulatory effects are) and a portal of entry is created when the catfish fin is clipped. With the exception of differences within the first 45 days of the study, the promising results observed in aquaria did not translate to production ponds. This could be attributable to factors capable of masking the effect of these agents, such outbreaks of the disease Columnaris (etiological agent *Flavobacterium columnare*) or environmental conditions (e.g. temperature-related kills). Furthermore, mortality associated with v*Ah* was masked by the addition of chemicals to prevent disease from pathogens including v*Ah* (e.g. formalin, dry/liquid copper, and chlorine). Therefore, although this approach may have contextualized the effect that phytase and probiotic treatment have in the presence of common therapeutic agents, the independent effects of these treatments were lost in translation, as evidenced by inconsistencies within treatments and overall disjointed results due to restocking after ~70 days.

Attributing the significant reduction in morality within aquaria to an inhibitory effect on v*Ah* presents an interesting finding, especially in the case of the enzyme phytase. Given the nature of the challenge (immersion with fin clip), the abundance of v*Ah* is not predicated by its

ability to use an otherwise unavailable carbon source. Therefore, the ~23% reduction in mortality in fish fed a diet with the enzyme phytase, may indicate metabolism of this anti-nutrient is in fact linked with the upregulation of virulence factors as previously postulated [1]. In effect, the standing hypothesis is that the genes associated with inositol catabolism (*iolA* - *iolR*) and with essential virulence factors, such as aerolysin, are governed under the same regulatory system [278,279]. While the reductions in mortality appear to support this hypothesis, future research exploring the molecular relationship between this carbon source and regulatory changes are necessary to more thoroughly challenge this possibility.

From a production perspective, the inhibitory selection observed in aquaria against v*Ah* in both the phytase and the probiotic treatments shows promise as a practical solution for use in production ponds. Even though these systems are artificial, understanding off-target effects of these introductions (e.g. shifts in microbial relative abundance) are a crucial step towards reducing the probability of off-target effects while also informing the development of future research aimed to improve production efficiency and safety. With that said, the phytate/phytic acid degrading activity of both the *Bacillus* probiotic and the enzyme phytase results in the release of nutrients already present within catfish feed. Therefore, the off-target effects from phytate/phytic acid degradation are conjectured to be minimal.

Considering the antagonistic activity of the probiotic against common aquaculture pathogens and the well-documented affinity for competitive exclusion within *Bacillus* spp.[127,232,280 4,126], it stands to reason these properties would result in a calculable shifts in mean species diversity within both the aquaria challenge and the raceway study. Based on QIIME evaluation of 16S rRNA gene sequencing, there were notable shifts were present in the catfish gut microbiome predominantly consisting of changes in relative abundance from members of

Fusobacteriaceae to members of *Pseudomonadaceae* and members of *Enterobacteriaceae* in both the treatment with phytase-amended feed and the treatment with the probiotic-amended feed (**Figure 8** and **Figure 9**). In contrast, Kraken2 results showed a nearly static microbiome, irrespective of treatment. Although the results from Kraken2 may be accurate, the prevalence of low abundance classifications that were then grouped under the 'Other' assignation accounted for ~70% of data, compared to ~7% of data assigned to the 'Other' assignation when using QIIME (**Table 2**). While each approach has clear methodological limitations, both findings were presented to increase the probability of cross-validation with future research.

Depending on the site of infection, the catfish gut microbiome may synchronously reflect trends in v*Ah* abundance as the pathogenic lifecycle of the bacterium progresses; however, this event was only identified within the raceway study when read mapping was performed and not when the QIIME pipeline was used. Contextualizing this information within the state of research aimed to understand the natural mode(s) of infection, such as correlation between feeding status with host susceptibility and the postulation that gastric vasculature may serve as an entry point (ligands/receptors via organ-specific ligands and/or receptors) [274,281,282], the lack of correlation between v*Ah* abundance in the gut with disease is a potential line of evidence that v*Ah*-induced disease in production systems manifests at another site. To build upon this research, future studies should use targeted analytical methods (e.g. pathotype- and strain-specific primers) to more accurately capture biological diversity at informative taxonomic levels.

In summary, the prophylactic efficacy of feed enriched by either the enzyme phytase or the *Bacillus* probiotic showed clear prophylactic activity when coated on feed, with the greatest reductions in mortality within treatments given the probiotic. Based on the differences in mortality associated with degradation of inositol-hexaphosphate, these findings provide another

line of evidence that metabolism/presence of this carbon source may result in the upregulation of virulence factors. With these new lines of evidence, future research should aim to challenge these inferences with targeted microbial quantification methods before, during, and after v*Ah* infection.

Table 1. Recipe for spore preparation media (agar).

Ingredient	Quantity (g/L)
Peptone	3.30
Beef extract powder	1.00
NaCl	5.00
K2HPO4	2.00
KCl	1.00
MgSO4 · 7H2O	0.25
MnSO4	0.01
Lactose	5.00
Agar	15.00

Table 2. Comparison of relative abundance for taxonomic mapping for the aquaria challenge. Summaries detailed below include data from 16S rRNA gene amplicons sequenced on the Illumina MiSeq (processed by QIIME) and whole metagenome sequencing with the Oxford Nanopore MinION (processed by Kraken2). Relative abundance thresholds below 0.1% were combined under the heading 'Other', with the exception of members within *Aeromonadaceae*.

	Control		**Phytase**		**Probiotic**	
Classification	**Illumina**	**MinION**	**Illumina**	**MinION**	**Illumina**	**MinION**
Aeromonadaceae	0.02%	0.00%	0.07%	0.00%	0.05%	0.00%
Bacillaceae	0.31%	4.44%	0.03%	6.30%	0.07%	5.10%
Bacteroidaceae	1.48%	0.00%	0.11%	0.00%	0.51%	0.00%
Barnesiellaceae	5.10%	0.00%	4.43%	0.00%	7.62%	0.00%
Burkholderiaceae	0.21%	0.00%	3.05%	0.00%	1.70%	0.00%
Campylobacteraceae	0.00%	0.92%	0.00%	1.26%	0.00%	1.03%
Clostridiaceae	15.46%	4.67%	0.99%	6.60%	2.65%	5.55%
Enterobacteriaceae	6.91%	0.00%	20.26%	0.00%	20.43%	0.00%
Erysipelotrichaceae	5.82%	0.00%	0.17%	0.00%	1.04%	0.00%
Flavobacteriaceae	0.00%	2.71%	0.00%	3.40%	0.00%	2.76%
Fusobacteriaceae	40.25%	0.00%	8.11%	0.00%	10.28%	0.00%
Lachnospiraceae	2.50%	0.00%	0.52%	0.00%	2.98%	0.00%
Lactobacillaceae	0.00%	1.85%	0.00%	2.48%	0.00%	2.27%
Marinifilaceae	0.06%	0.00%	0.15%	0.00%	0.43%	0.00%
Moraxellaceae	0.00%	1.22%	0.00%	1.74%	0.00%	1.39%
Mycoplasmataceae	0.00%	1.23%	0.00%	1.68%	0.00%	1.35%
Paenibacillaceae	0.00%	1.42%	0.00%	2.03%	0.00%	1.75%

Peptostreptococcaceae	18.38%	0.00%	0.61%	0.00%	1.91%	0.00%
Pseudomonadaceae	0.09%	2.93%	54.67%	3.81%	33.75%	3.43%
Rhizobiaceae	0.00%	1.81%	0.00%	2.59%	0.00%	2.06%
Ruminococcaceae	0.18%	0.00%	0.23%	0.00%	0.38%	0.00%
Staphylococcaceae	0.00%	1.67%	0.00%	2.47%	0.00%	1.99%
Streptococcaceae	0.00%	1.46%	0.00%	1.91%	0.00%	1.47%
Streptomycetaceae	0.00%	1.96%	0.00%	2.44%	0.00%	2.33%
Tannerellaceae	2.06%	0.00%	1.45%	0.00%	1.88%	0.00%
Vibrionaceae	0.00%	1.47%	0.00%	1.64%	0.00%	1.69%
Other	1.17%	70.26%	5.16%	59.65%	14.31%	65.85%

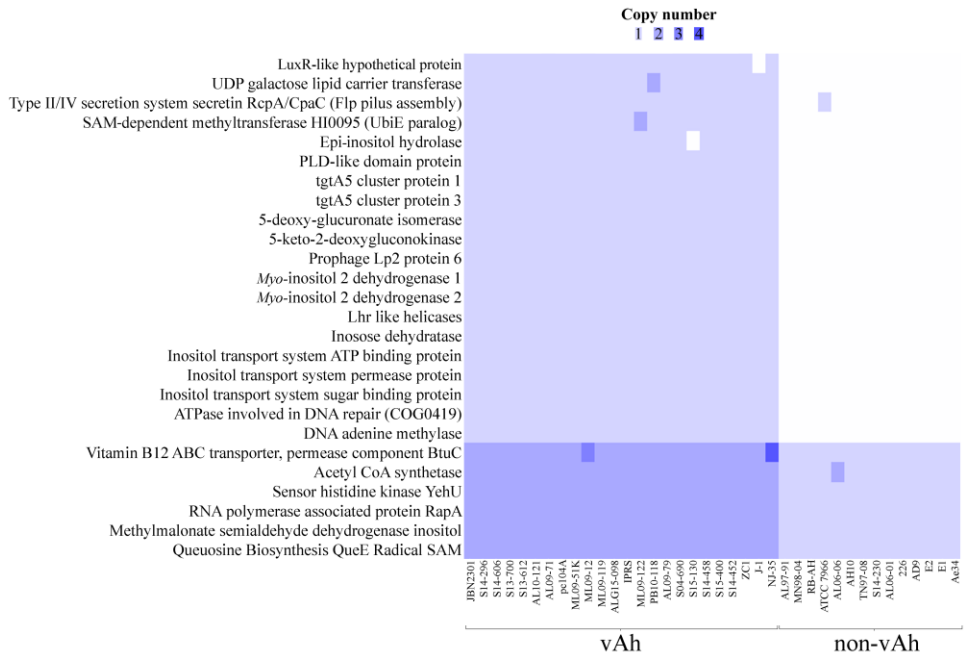


Figure 1. Comparative whole genome predicted gene-based analysis of all phylogenetically confirmed v*Ah* (n=26) and non-v*Ah* isolates (n=15) [1].

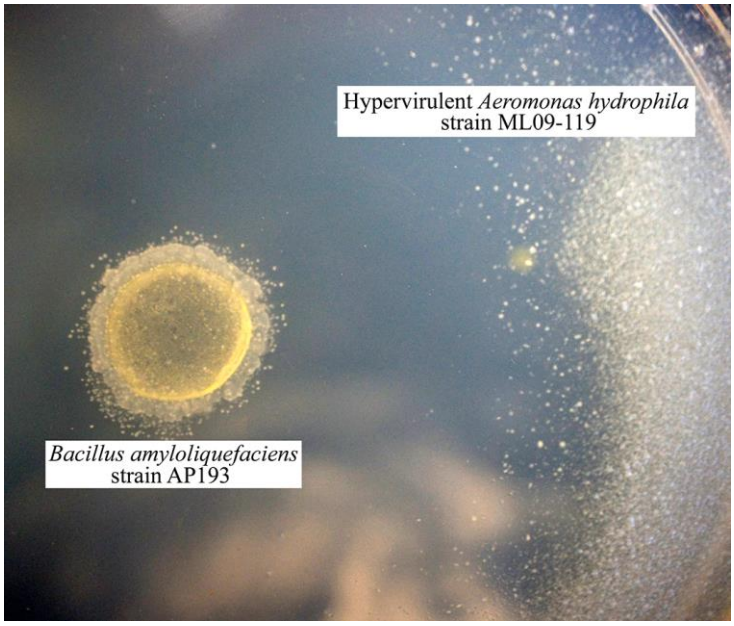


Figure 2. Soft agar overlay of probiotic *B. amyloliquefaciens* strain AP193 (left) demonstrating a clear zone of inhibition towards the growth of hypervirulent *A. hydrophila* strain ML09-119 [2 4].

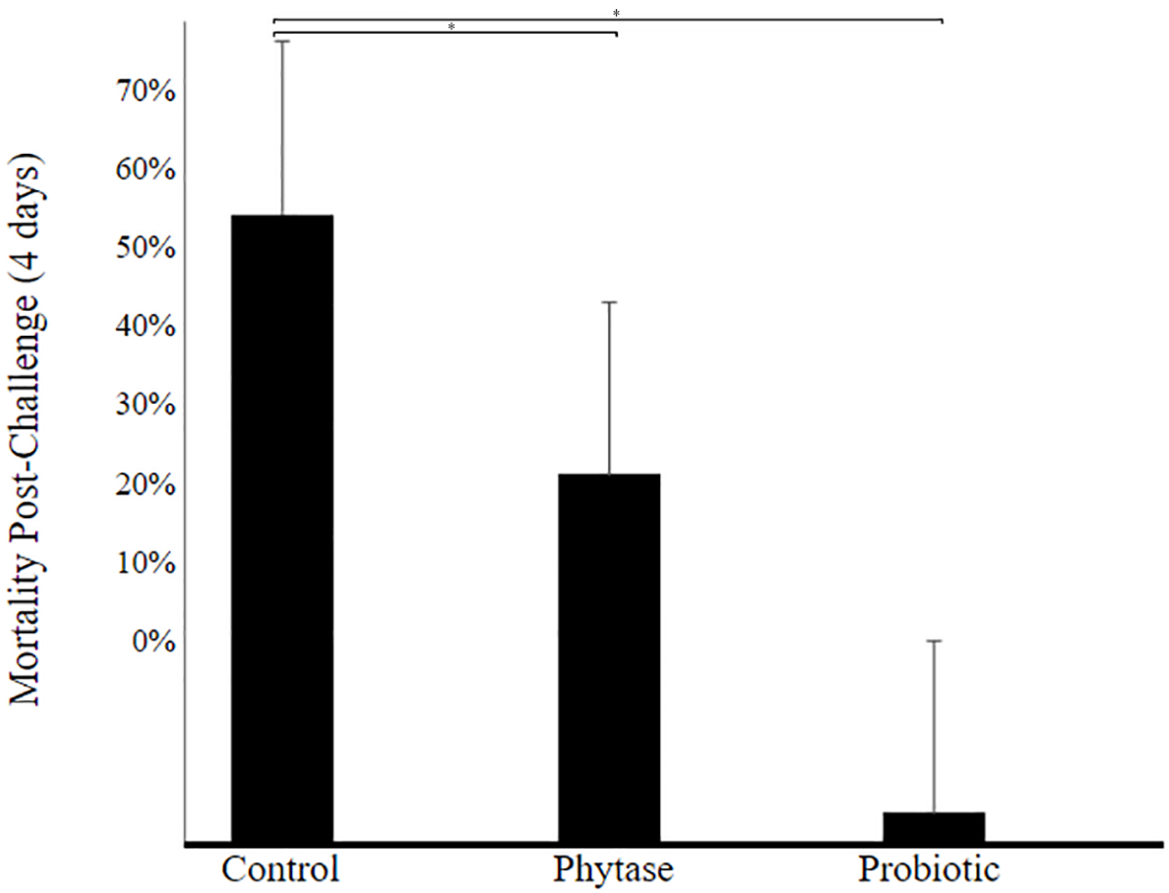


Figure 3. Mortality of channel catfish after a four-day disease challenge in aquaria (immersion with fin clip) with *A. hydrophila* ML10-51K. Treatments included feed coated with the enzyme phytase and the probiotic *B. velezensis* AP193. Control feed was coated with fish oil (the agent used to coat the treatments). Statistical significance (*) was set at the 95% confidence threshold.

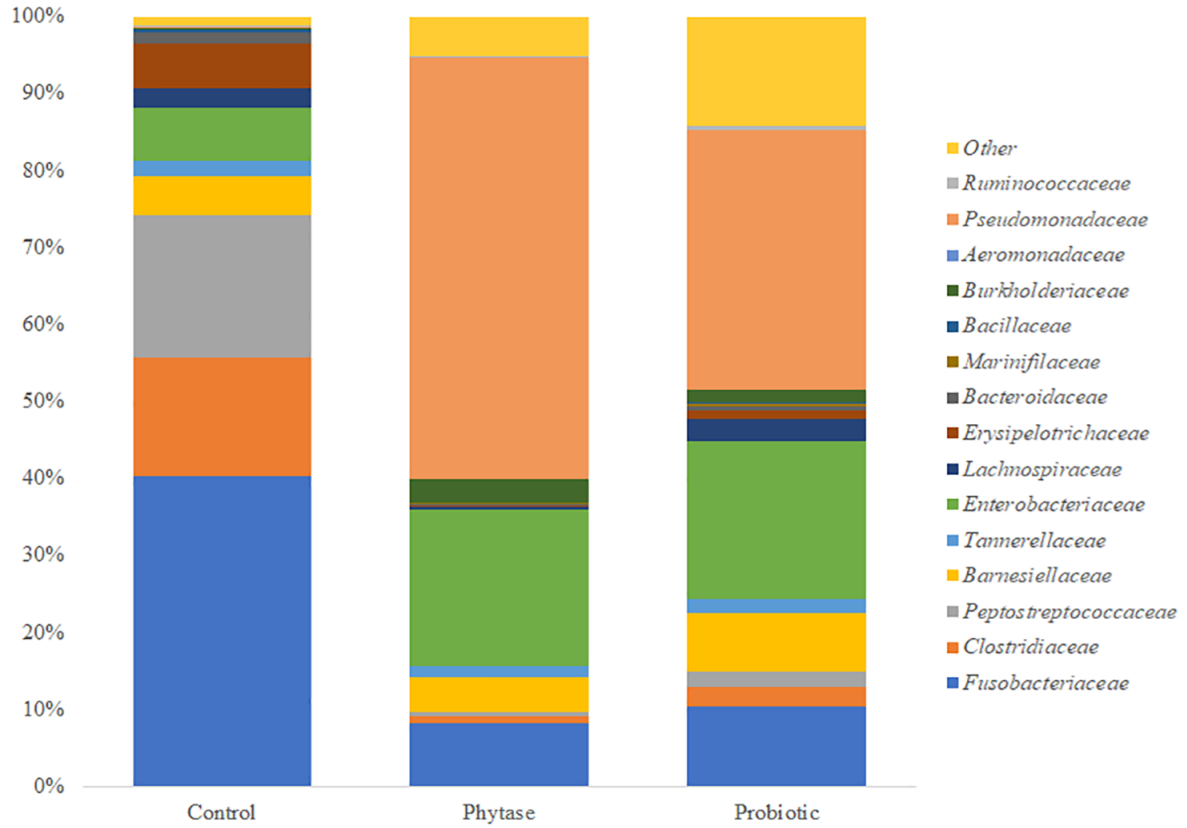


Figure 4. Microbial profile of the mid and hind gut of channel catfish after feed supplementation, but before the aquaria challenge. Results reflect 16S rRNA gene amplicon sequencing processed by QIIME at the taxonomic level of family.

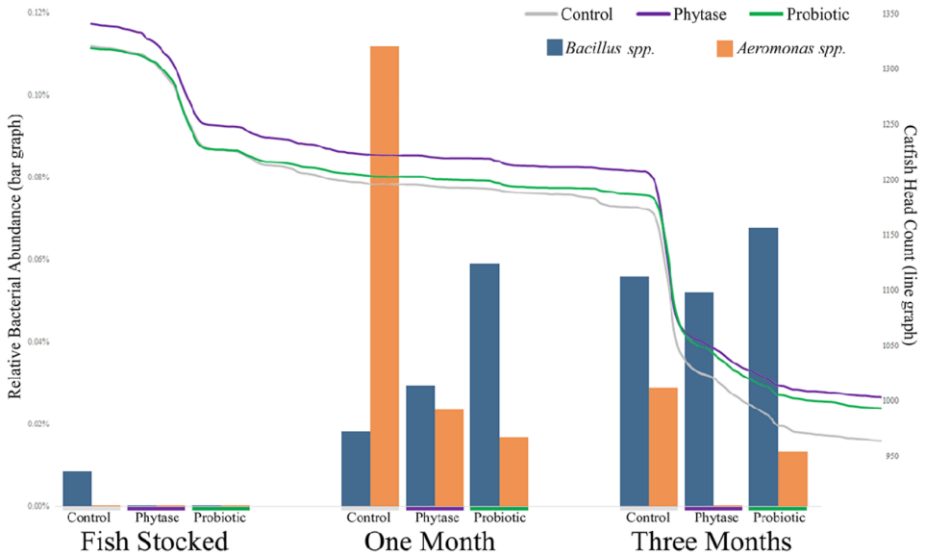


Figure 5. Microbiome composition based on 16S rRNA gene amplicon sequencing results with read mapping of the genus *Bacillus* and the genus *Aeromonas* raceway study after stocking of fish, then after one month and after four months of feeding in raceways with their respective treatment/control feed.

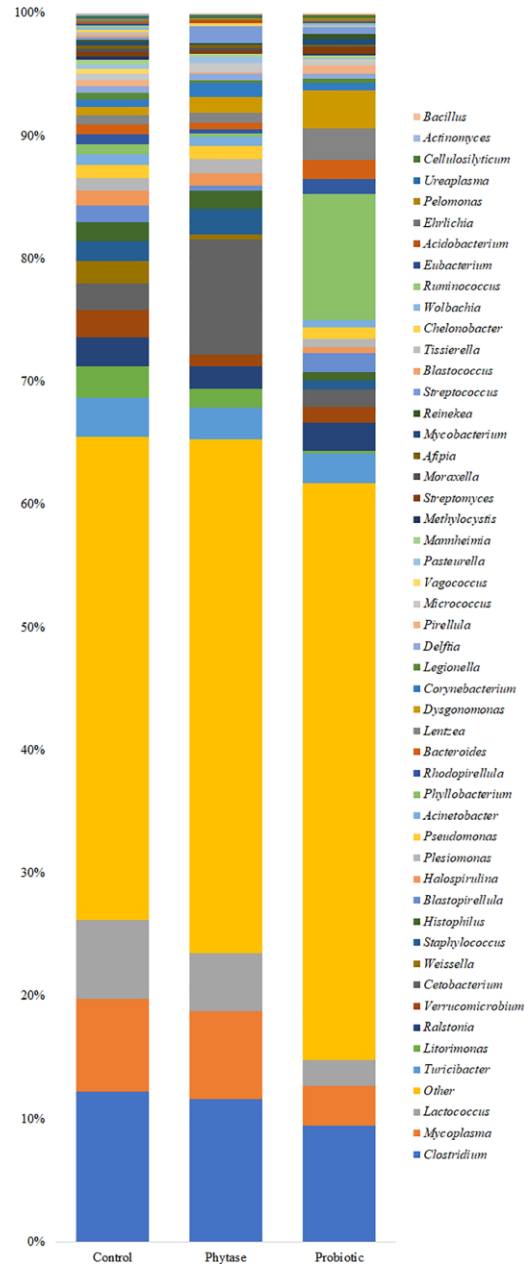


Figure 6. Catfish gut microbiome composition of after four months of feeding in raceways with their respective treatment/control feed based 16S rRNA gene amplicon sequencing (processed by QIIME).

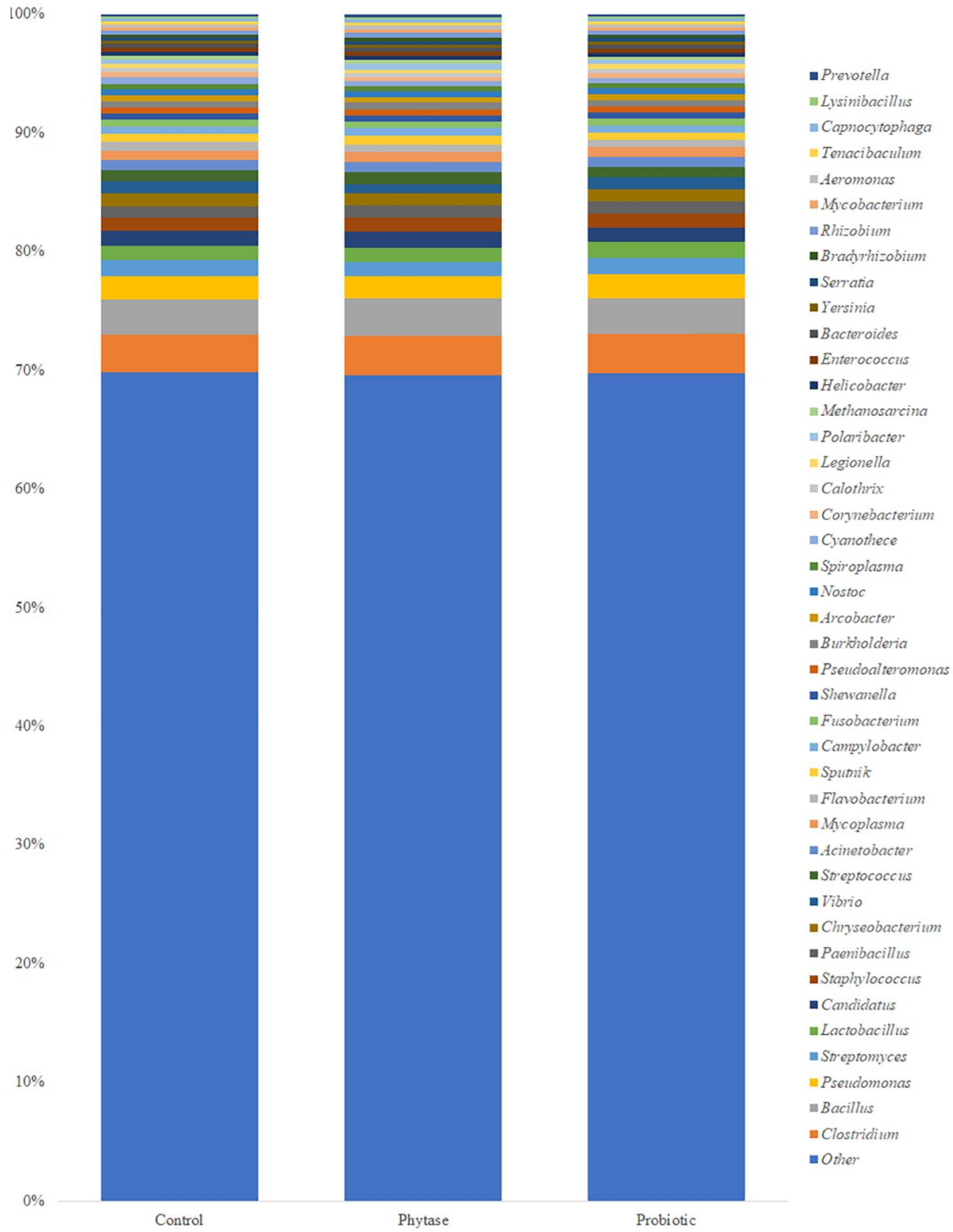


Figure 7. Kraken2-based microbiome composition results of the channel catfish gut after a four-day disease challenge in aquaria (immersion with fin clip) with *A. hydrophila* ML10-51K.

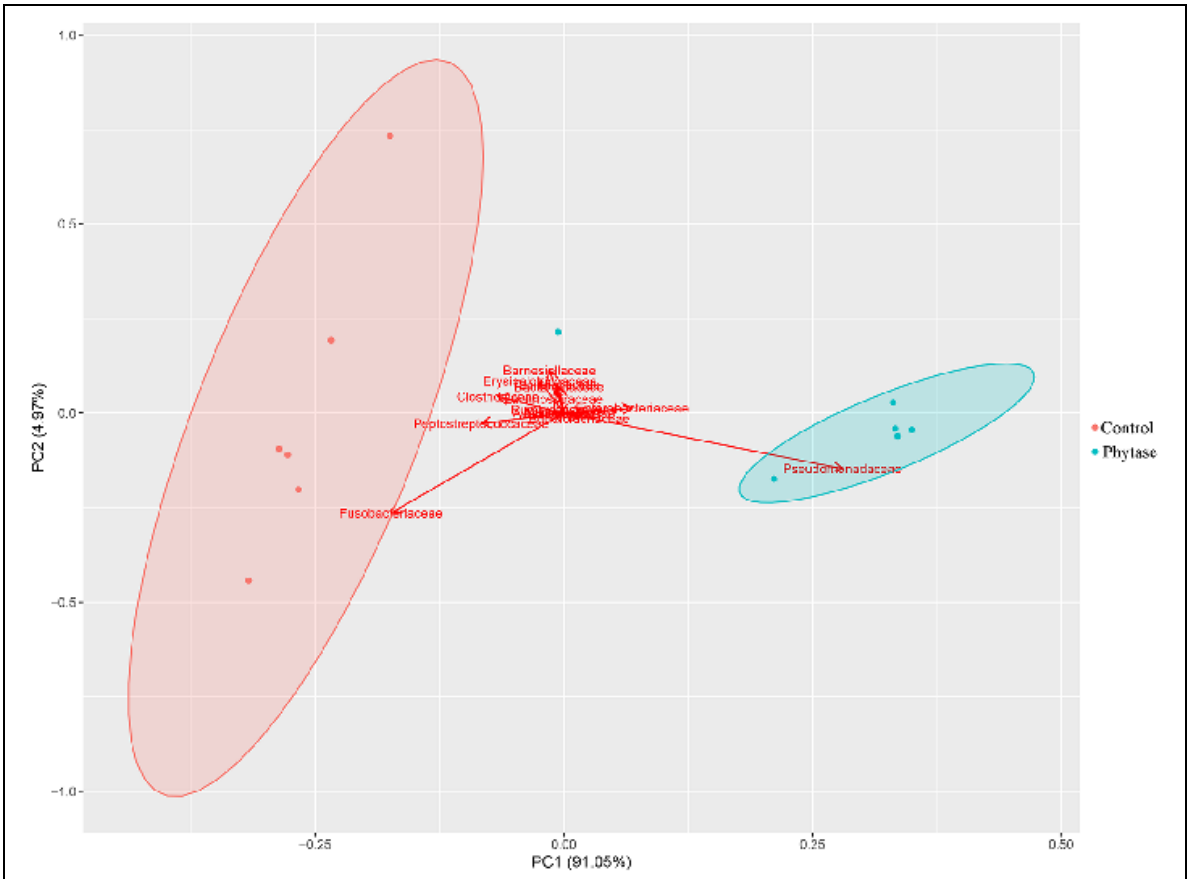


Figure 8. Principal component analysis of the control and phytase groups within the aquaria challenge based on QIIME phylogenetic classification of amplicons produced from the 16S rRNA gene (sequenced on the Illumina MiSeq).

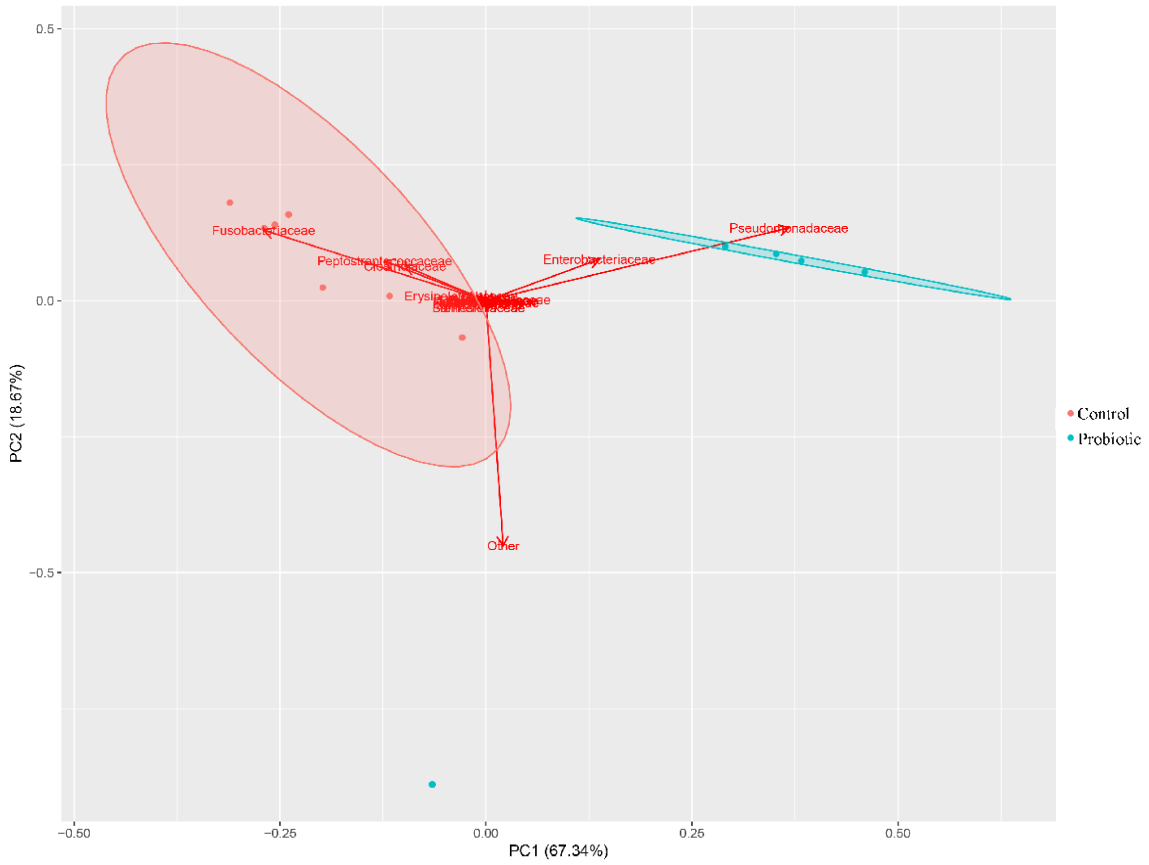


Figure 9. Principal component analysis of the control and probiotic groups within the aquaria challenge based on QIIME phylogenetic classification of amplicons produced from the 16S rRNA gene (sequenced on the Illumina MiSeq).

References

1. Rasmussen-Ivey, C.R. *et al.* Classification of a Hypervirulent Aeromonas hydrophila Pathotype Responsible for Epidemic Outbreaks in Warm-Water Fishes. *Front Microbiol* **7**, 1615 (2016).
2. Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E. & Shay, J.W. Comparison of DNA quantification methods for next generation sequencing. *Scientific reports* **6**(2016).
3. Wheeler, D.L. *et al.* Database resources of the National Center for Biotechnology. *Nucleic Acids Research* **31**, 28-33 (2003).
4. Thurlow, C.M. *et al.* Bacillus velezensis AP193 exerts probiotic effects in channel catfish (Ictalurus punctatus) and reduces aquaculture pond eutrophication. *Aquaculture* **503**, 347-356 (2019).
5. Garrity, G. *et al.* Bergey's Manual® of Systematic Bacteriology: Volume Two: The Proteobacteria. *Springer Science & Business Media* (2006).
6. Cipriano, R.C., Bullock, G.L. & Pyle, S.W. Aeromonas hydrophila and motile aeromonad septicemias of fish. *U.S. Fish and Wildlife Publications* **Paper 134**(1984).
7. Figueras, M.J. *et al.* Aeromonas hemolytic uremic syndrome. A case and a review of the literature. *Diagn Microbiol Infect Dis* **58**, 231-4 (2007).
8. Monaghan, S.F., Anjaria, D., Mohr, A. & Livingston, D.H. Necrotizing fasciitis and sepsis caused by Aeromonas hydrophila after crush injury of the lower extremity. *Surg Infect (Larchmt)* **9**, 459-67 (2008).
9. Janda, J.M. & Abbott, S.L. The genus Aeromonas: taxonomy, pathogenicity, and infection. *Clin Microbiol Rev* **23**, 35-73 (2010).
10. Cahill, M.M. Virulence factors in motile Aeromonas species. *J Appl Bacteriol* **69**, 1-16 (1990).
11. Allan, B.J. & Stevenson, R.M. Extracellular virulence factors of Aeromonas hydrophila in fish infections. *Can J Microbiol* **27**, 1114-22 (1981).
12. Thornley, J.P., Shaw, J.G., Gryllos, I.A. & Eley, A. Virulence properties of clinically significant Aeromonas species: Evidence for pathogenicity. *Reviews in Medical Microbiology* **8**, 61-72 (1997).
13. Beaz-Hidalgo, R. & Figueras, M.J. Aeromonas spp. whole genomes and virulence factors implicated in fish disease. *J Fish Dis* **36**, 371-88 (2013).
14. Grim, C.J. *et al.* Functional genomic characterization of virulence factors from necrotizing fasciitis-causing strains of Aeromonas hydrophila. *Appl Environ Microbiol* **80**, 4162-83 (2014).

15. Colston, S.M. *et al.* Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using Aeromonas as a test case. *MBio* **5**, e02136 (2014).
16. Beaz-Hidalgo, R., Hossain, M.J., Liles, M.R. & Figueras, M.J. Strategies to Avoid Wrongly Labelled Genomes Using as Example the Detected Wrong Taxonomic Affiliation for Aeromonas Genomes in the GenBank Database. *Plos One* **10**(2015).
17. Huys, G. *et al.* Aeromonas hydrophila subsp. dhakensis subsp. nov., isolated from children with diarrhoea in Bangladesh, and extended description of Aeromonas hydrophila subsp. hydrophila (Chester 1901) Stanier 1943 (approved lists 1980). *Int J Syst Evol Microbiol* **52**, 705-12 (2002).
18. Beaz-Hidalgo, R., Martinez-Murcia, A. & Figueras, M.J. Reclassification of Aeromonas hydrophila subsp. dhakensis Huys et al. 2002 and Aeromonas aquariorum Martinez-Murcia et al. 2008 as Aeromonas dhakensis sp. nov. comb nov. and emendation of the species Aeromonas hydrophila. *Syst Appl Microbiol* **36**, 171-6 (2013).
19. Mateos, D., Anguita, J., Naharro, G. & Paniagua, C. Influence of growth temperature on the production of extracellular virulence factors and pathogenicity of environmental and human strains of Aeromonas hydrophila. *J Appl Bacteriol* **74**, 111-8 (1993).
20. Swann, L. & White, M. *Diagnosis and treatment of" Aeromonas Hydrophila" infection of fish*, (Aquaculture Extension, Illinois-Indiana Sea Grant Program, 1991).
21. Merino, S., Camprubi, S. & Tomas, J.M. Effect of growth temperature on outer membrane components and virulence of Aeromonas hydrophila strains of serotype O:34. *Infect Immun* **60**, 4343-9 (1992).
22. Gonzalez-Serrano, C.J., Santos, J.A., Garcia-Lopez, M.L. & Otero, A. Virulence markers in Aeromonas hydrophila and Aeromonas veronii biovar sobria isolates from freshwater fish and from a diarrhoea case. *J Appl Microbiol* **93**, 414-9 (2002).
23. Popoff, M. & Veron, M. A taxonomic study of the Aeromonas hydrophila-Aeromonas punctata group. *J Gen Microbiol* **94**, 11-22 (1976).
24. Yu, H.B., Kaur, R., Lim, S.M., Wang, X.H. & Leung, K.Y. Characterization of extracellular proteins produced by Aeromonas hydrophila AH-1. *Proteomics* **7**, 436-449 (2007).
25. Swift, S. *et al.* Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida: identification of the LuxRI homologs AhyRI and AsaRI and their cognate N-acylhomoserine lactone signal molecules. *J Bacteriol* **179**, 5271-81 (1997).
26. Lynch, M.J. *et al.* The regulation of biofilm development by quorum sensing in Aeromonas hydrophila. *Environ Microbiol* **4**, 18-28 (2002).
27. Wilhelms, M., Molero, R., Shaw, J.G., Tomas, J.M. & Merino, S. Transcriptional Hierarchy of Aeromonas hydrophila Polar-Flagellum Genes. *Journal of bacteriology* **193**, 5179-5190 (2011).
28. Wilhelms, M., Gonzalez, V., Tomas, J.M. & Merino, S. Aeromonas hydrophila Lateral Flagellar Gene Transcriptional Hierarchy. *Journal of bacteriology* **195**, 1436-1445 (2013).
29. Sandkvist, M. Type II secretion and pathogenesis. *Infect Immun* **69**, 3523-35 (2001).
30. Peabody, C.R. *et al.* Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* **149**, 3051-72 (2003).
31. Tseng, T.T., Tyler, B.M. & Setubal, J.C. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *Bmc Microbiology* **9**(2009).

32. Pang, M.B. *et al.* Novel insights into the pathogenicity of epidemic Aeromonas hydrophila ST251 clones from comparative genomics. *Scientific Reports* **5**(2015).
33. Galindo, C.L. *et al.* Aeromonas hydrophila cytotoxic enterotoxin activates mitogen-activated protein kinases and induces apoptosis in murine macrophages and human intestinal epithelial cells. *J Biol Chem* **279**, 37597-612 (2004).
34. Cianciotto, N.P. Type II secretion: a protein secretion system for all seasons. *Trends Microbiol* **13**, 581-8 (2005).
35. Aguilera-Arreola, M.G., Hernandez-Rodriguez, C., Zuniga, G., Figueras, M.J. & Castro-Escarpulli, G. Aeromonas hydrophila clinical and environmental ecotypes as revealed by genetic diversity and virulence genes. *FEMS Microbiol Lett* **242**, 231-40 (2005).
36. Galan, J.E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322-8 (1999).
37. Sha, J., Kozlova, E.V. & Chopra, A.K. Role of various enterotoxins in Aeromonas hydrophila-induced gastroenteritis: generation of enterotoxin gene-deficient mutants and evaluation of their enterotoxic activity. *Infect Immun* **70**, 1924-35 (2002).
38. Sierra, J.C. *et al.* Unraveling the mechanism of action of a new type III secretion system effector AexU from Aeromonas hydrophila. *Microb Pathog* **49**, 122-34 (2010).
39. Braun, M. *et al.* Characterization of an ADP-ribosyltransferase toxin (AexT) from Aeromonas salmonicida subsp. salmonicida. *J Bacteriol* **184**, 1851-8 (2002).
40. Burr, S.E., Stuber, K. & Frey, J. The ADP-ribosylating toxin, AexT, from Aeromonas salmonicida subsp. salmonicida is translocated via a type III secretion pathway. *J Bacteriol* **185**, 6583-91 (2003).
41. Ebanks, R.O. *et al.* Expression of and secretion through the Aeromonas salmonicida type III secretion system. *Microbiology* **152**, 1275-86 (2006).
42. Erova, T.E. *et al.* Mutations within the catalytic motif of DNA adenine methyltransferase (Dam) of Aeromonas hydrophila cause the virulence of the Dam-overproducing strain to revert to that of the wild-type phenotype. *Infect Immun* **74**, 5763-72 (2006).
43. Erova, T.E. *et al.* DNA adenine methyltransferase influences the virulence of Aeromonas hydrophila. *Infect Immun* **74**, 410-24 (2006).
44. Vilches, S., Jimenez, N., Tomas, J.M. & Merino, S. Aeromonas hydrophila AH-3 type III secretion system expression and regulatory network. *Appl Environ Microbiol* **75**, 6382-92 (2009).
45. Vilches, S. *et al.* Complete type III secretion system of a mesophilic Aeromonas hydrophila strain. *Appl Environ Microbiol* **70**, 6914-9 (2004).
46. Beaz-Hidalgo, R., Alperi, A., Figueras, M.J. & Romalde, J.L. Aeromonas piscicola sp. nov., isolated from diseased fish. *Syst Appl Microbiol* **32**, 471-9 (2009).
47. Sha, J. *et al.* The type III secretion system and cytotoxic enterotoxin alter the virulence of Aeromonas hydrophila. *Infect Immun* **73**, 6446-57 (2005).
48. Vilches, S. *et al.* Aeromonas hydrophila AH-3 AexT is an ADP-ribosylating toxin secreted through the type III secretion system. *Microb Pathog* **44**, 1-12 (2008).
49. Khajanchi, B.K. *et al.* Distribution of virulence factors and molecular fingerprinting of Aeromonas species isolates from water and clinical samples: suggestive evidence of water-to-human transmission. *Appl Environ Microbiol* **76**, 2313-25 (2010).
50. Yu, H.B. *et al.* A type III secretion system is required for Aeromonas hydrophila AH-1 pathogenesis. *Infect Immun* **72**, 1248-56 (2004).

51. Grim, C.J. *et al.* Characterization of Aeromonas hydrophila wound pathotypes by comparative genomic and functional analyses of virulence genes. *MBio* **4**, e00064-13 (2013).
52. Tan, Y.W., Yu, H.B., Sivaraman, J., Leung, K.Y. & Mok, Y.K. Mapping of the chaperone AcrH binding regions of translocators AopB and AopD and characterization of oligomeric and metastable AcrH-AopB-AopD complexes in the type III secretion system of Aeromonas hydrophila. *Protein Sci* **18**, 1724-34 (2009).
53. Carvalho-Castro, G.A. *et al.* Detection of type III secretion system genes in Aeromonas hydrophila and their relationship with virulence in Nile tilapia. *Vet Microbiol* **144**, 371-6 (2010).
54. Hossain, M.J. *et al.* Implication of Lateral Genetic Transfer in the Emergence of Aeromonas hydrophila Isolates of Epidemic Outbreaks in Channel Catfish. *PLoS One* **8**(2013).
55. Sierra, J.C. *et al.* Biological characterization of a new type III secretion system effector from a clinical isolate of Aeromonas hydrophila-part II. *Microb Pathog* **43**, 147-60 (2007).
56. Sha, J. *et al.* Further characterization of a type III secretion system (T3SS) and of a new effector protein from a clinical isolate of Aeromonas hydrophila--part I. *Microb Pathog* **43**, 127-46 (2007).
57. Bingle, L.E., Bailey, C.M. & Pallen, M.J. Type VI secretion: a beginner's guide. *Curr Opin Microbiol* **11**, 3-8 (2008).
58. Suarez, G. *et al.* Molecular characterization of a functional type VI secretion system from a clinical isolate of Aeromonas hydrophila. *Microbial Pathogenesis* **44**, 344-361 (2008).
59. Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R. & Lappin-Scott, H.M. Microbial biofilms. *Annu Rev Microbiol* **49**, 711-45 (1995).
60. Khajanchi, B.K. *et al.* N-acylhomoserine lactones involved in quorum sensing control the type VI secretion system, biofilm formation, protease production, and in vivo virulence in a clinical isolate of Aeromonas hydrophila. *Microbiology* **155**, 3518-31 (2009).
61. dos Reis Ponce-Rossi, A., Pinto, U.M., Ribon, A.D.O.B., Bazzolli, D.M.S. & Vanetti, M.C.D. Quorum sensing regulated phenotypes in Aeromonas hydrophila ATCC 7966 deficient in AHL production. *Ann. Microbiol*, 1-10 (2016).
62. Khajanchi, B.K., Kozlova, E.V., Sha, J., Popov, V.L. & Chopra, A.K. The two-component QseBC signalling system regulates in vitro and in vivo virulence of Aeromonas hydrophila. *Microbiology* **158**, 259-271 (2012).
63. Kozlova, E.V., Khajanchi, B.K., Popov, V.L., Wen, J. & Chopra, A.K. Impact of QseBC system in c-di-GMP-dependent quorum sensing regulatory network in a clinical isolate SSU of Aeromonas hydrophila. *Microb Pathog* **53**, 115-24 (2012).
64. Canals, R. *et al.* Analysis of the lateral flagellar gene system of Aeromonas hydrophila AH-3. *J Bacteriol* **188**, 852-62 (2006).
65. Fulton, K.M., Mendoza-Barbera, E., Twine, S.M., Tomas, J.M. & Merino, S. Polar Glycosylated and Lateral Non-Glycosylated Flagella from Aeromonas hydrophila Strain AH-1 (Serotype O11). *Int J Mol Sci* **16**, 28255-69 (2015).
66. Jiang, X. *et al.* FlgN plays important roles in the adhesion of Aeromonas hydrophila to host mucus. *Genet Mol Res* **14**, 6376-86 (2015).
67. Pepe, C.M., Eklund, M.W. & Strom, M.S. Cloning of an Aeromonas hydrophila type IV pilus biogenesis gene cluster: complementation of pilus assembly functions and

characterization of a type IV leader peptidase/N-methyltransferase required for extracellular protein secretion. *Mol Microbiol* **19**, 857-69 (1996).
68. Barnett, T.C., Kirov, S.M., Strom, M.S. & Sanderson, K. Aeromonas spp. possess at least two distinct type IV pilus families. *Microb Pathog* **23**, 241-7 (1997).
69. Kirov, S.M., Barnett, T.C., Pepe, C.M., Strom, M.S. & Albert, M.J. Investigation of the role of type IV Aeromonas pilus (Tap) in the pathogenesis of Aeromonas gastrointestinal infection. *Infect Immun* **68**, 4040-8 (2000).
70. Merino, S. *et al.* The role of the capsular polysaccharide of Aeromonas hydrophila serogroup O:34 in the adherence to and invasion of fish cell lines. *Res Microbiol* **148**, 625-31 (1997).
71. Martinez, M.J. *et al.* The presence of capsular polysaccharide in mesophilic Aeromonas hydrophila serotypes O:11 and O:34. *FEMS Microbiol Lett* **128**, 69-73 (1995).
72. Merino, S., Rubires, X., Aguillar, A., Guillot, J.F. & Tomas, J.M. The role of the O-antigen lipopolysaccharide on the colonization in vivo of the germfree chicken gut by Aeromonas hydrophila serogroup O:34. *Microb Pathog* **20**, 325-33 (1996).
73. Hossain, M.J. Molecular Interactions between phage and the catfish pathogen Edwardsiella ictaluri and Comparative Genomics of Epidemic strains of Aeromonas hydrophila. *Auburn University (doctoral dissertation)* (2012).
74. Thomas, S.R. & Trust, T.J. Tyrosine phosphorylation of the tetragonal paracrystalline array of Aeromonas hydrophila: molecular cloning and high-level expression of the S-layer protein gene. *J Mol Biol* **245**, 568-81 (1995).
75. Esteve, C. *et al.* Pathogenic Aeromonas hydrophila serogroup O:14 and O:81 strains with an S layer. *Appl Environ Microbiol* **70**, 5898-904 (2004).
76. Wadstrom, T., Ljungh, A. & Wretlind, B. Enterotoxin, haemolysin and cytotoxic protein in Aeromonas hydrophila from human infections. *Acta Pathol Microbiol Scand B* **84**, 112-4 (1976).
77. Asao, T., Kinoshita, Y., Kozaki, S., Uemura, T. & Sakaguchi, G. Purification and some properties of Aeromonas hydrophila hemolysin. *Infect Immun* **46**, 122-7 (1984).
78. Hirono, I. & Aoki, T. Nucleotide sequence and expression of an extracellular hemolysin gene of Aeromonas hydrophila. *Microb Pathog* **11**, 189-97 (1991).
79. Wang, G.H. *et al.* Detection and characterization of the hemolysin genes in Aeromonas hydrophila and Aeromonas sobria by multiplex PCR. *Journal of Clinical Microbiology* **41**, 1048-1054 (2003).
80. Wong, C.Y.F., Heuzenroeder, M.W. & Flower, R.L.P. Inactivation of two haemolytic toxin genes in Aeromonas hydrophila attenuates virulence in a suckling mouse model. *Microbiology-Uk* **144**, 291-298 (1998).
81. Erova, T.E., Kosykh, V.G., Sha, J. & Chopra, A.K. DNA adenine methyltransferase (Dam) controls the expression of the cytotoxic enterotoxin (act) gene of Aeromonas hydrophila via tRNA modifying enzyme-glucose-inhibited division protein (GidA). *Gene* **498**, 280-7 (2012).
82. Fadl, A.A. *et al.* Deletion of the genes encoding the type III secretion system and cytotoxic enterotoxin alters host responses to Aeromonas hydrophila infection. *Microb Pathog* **40**, 198-210 (2006).
83. Sha, J. *et al.* Molecular characterization of a glucose-inhibited division gene, gidA, that regulates cytotoxic enterotoxin of Aeromonas hydrophila. *Infect Immun* **72**, 1084-95 (2004).

84. Sha, J., Lu, M. & Chopra, A.K. Regulation of the cytotoxic enterotoxin gene in Aeromonas hydrophila: characterization of an iron uptake regulator. *Infect Immun* **69**, 6370-81 (2001).
85. Chopra, A.K. *et al.* The cytotoxic enterotoxin of Aeromonas hydrophila induces proinflammatory cytokine production and activates arachidonic acid metabolism in macrophages. *Infect Immun* **68**, 2808-18 (2000).
86. Duarte, A.S. *et al.* Aeromonas piscicola AH-3 expresses an extracellular collagenase with cytotoxic properties. *Lett Appl Microbiol* **60**, 288-97 (2015).
87. Rivero, O., Anguita, J., Mateos, D., Paniagua, C. & Naharro, G. Cloning and characterization of an extracellular temperature-labile serine protease gene from Aeromonas hydrophila. *FEMS Microbiol Lett* **65**, 1-7 (1991).
88. Cascon, A. *et al.* A major secreted elastase is essential for pathogenicity of Aeromonas hydrophila. *Infect Immun* **68**, 3233-41 (2000).
89. Li, J., Ni, X.D., Liu, Y.J. & Lu, C.P. Detection of three virulence genes alt, ahp and aerA in Aeromonas hydrophila and their relationship with actual virulence to zebrafish. *J Appl Microbiol* **110**, 823-30 (2011).
90. Chang, C.Y., Thompson, H., Rodman, N., Bylander, J. & Thomas, J. Pathogenic analysis of Aeromonas hydrophila septicemia. *Ann Clin Lab Sci* **27**, 254-9 (1997).
91. Hu, M., Wang, N., Pan, Z.H., Lu, C.P. & Liu, Y.J. Identity and virulence properties of Aeromonas isolates from diseased fish, healthy controls and water environment in China. *Lett Appl Microbiol* **55**, 224-33 (2012).
92. Rodriguez, L.A., Ellis, A.E. & Nieto, T.P. Purification and characterisation of an extracellular metalloprotease, serine protease and haemolysin of Aeromonas hydrophila strain B32: all are lethal for fish. *Microb Pathog* **13**, 17-24 (1992).
93. Esteve, C. & Birbeck, T.H. Secretion of haemolysins and proteases by Aeromonas hydrophila EO63: separation and characterization of the serine protease (caseinase) and the metalloprotease (elastase). *J Appl Microbiol* **96**, 994-1001 (2004).
94. Sha, J. *et al.* Surface-expressed enolase contributes to the pathogenesis of clinical isolate SSU of Aeromonas hydrophila. *J Bacteriol* **191**, 3095-107 (2009).
95. Stehr, F., Kretschmar, M., Kroger, C., Hube, B. & Schafer, W. Microbial lipases as virulence factors. *Journal of Molecular Catalysis B-Enzymatic* **22**, 347-355 (2003).
96. Anguita, J., Rodriguez Aparicio, L.B. & Naharro, G. Purification, gene cloning, amino acid sequence analysis, and expression of an extracellular lipase from an Aeromonas hydrophila human isolate. *Appl Environ Microbiol* **59**, 2411-7 (1993).
97. Merino, S. *et al.* Cloning, sequencing, and role in virulence of two phospholipases (A1 and C) from mesophilic Aeromonas sp. serogroup O:34. *Infect Immun* **67**, 4008-13 (1999).
98. Huang, L. *et al.* MinD plays an important role in Aeromonas hydrophila adherence to Anguilla japonica mucus. *Gene* **565**, 275-81 (2015).
99. Hernanz Moral, C. *et al.* Molecular characterization of the Aeromonas hydrophila aroA gene and potential use of an auxotrophic aroA mutant as a live attenuated vaccine. *Infect Immun* **66**, 1813-21 (1998).
100. Ji, Y. *et al.* Contribution of nuclease to the pathogenesis of Aeromonas hydrophila. *Virulence* **6**, 515-22 (2015).
101. Erova, T.E. *et al.* Cold shock exoribonuclease R (VacB) is involved in Aeromonas hydrophila pathogenesis. *J Bacteriol* **190**, 3467-74 (2008).

102. McCoy, A.J. *et al.* Cytotoxins of the human pathogen Aeromonas hydrophila trigger, via the NLRP3 inflammasome, caspase-1 activation in macrophages. *Eur J Immunol* **40**, 2797-803 (2010).
103. Suarez, G. *et al.* Actin cross-linking domain of Aeromonas hydrophila repeat in toxin A (RtxA) induces host cell rounding and apoptosis. *Gene* **506**, 369-76 (2012).
104. Pillai, L. *et al.* Molecular and functional characterization of a ToxR-regulated lipoprotein from a clinical isolate of Aeromonas hydrophila. *Infect Immun* **74**, 3742-55 (2006).
105. Del Castillo, C.S. *et al.* Comparative sequence analysis of a multidrug-resistant plasmid from Aeromonas hydrophila. *Antimicrobial agents and chemotherapy* **57**, 120-129 (2013).
106. Griffin, M.J. *et al.* Rapid quantitative detection of Aeromonas hydrophila strains associated with disease outbreaks in catfish aquaculture. *J Vet Diagn Invest* **25**, 473-81 (2013).
107. Zhang, X.J., Yang, W.M., Li, T.T. & Li, A.H. The genetic diversity and virulence characteristics of Aeromonas hydrophila isolated from fishponds with disease outbreaks in Hubei province. *Acta Hydrobiol. Sinica* **37**, 458-466.
108. Abbott, S.L., Cheung, W.K. & Janda, J.M. The genus Aeromonas: biochemical characteristics, atypical reactions, and phenotypic identification schemes. *J Clin Microbiol* **41**, 2348-57 (2003).
109. Martinez-Murcia, A.J. *et al.* Multilocus phylogenetic analysis of the genus Aeromonas. *Syst Appl Microbiol* **34**, 189-99 (2011).
110. Rodriguez, E., Han, Y. & Lei, X.G. Cloning, sequencing, and expression of anEscherichia coliacid phosphatase/phytase gene (appA2) isolated from pig colon. *Biochemical and Biophysical Research Communications* **257**, 117-123 (1999).
111. Nwanna, L. Effect of defatted and dephytinized soy proteins fortified with phytase on the growth, nutrient digestibility and phosphorus load of Rainbow trout. *African Journal of Fisheries and Aquatic Resources Management* **1**(2016).
112. Jackson, L.S., Li, M.H. & Robinson, E.H. Use of microbial phytase in channel catfish Ictalurus punctatus diets to improve utilization of phytate phosphorus1. *Journal of the World Aquaculture Society* **27**, 309-313 (1996).
113. Tewari, G., Ram, R. & Singh, A. Effect of plant base digestive enzyme 'Papain'on growth, survival and behavioural response of Cyprinus carpio. (2018).
114. Pontes, T., França, W., Dutra, F., Portz, L. & Ballester, E. Evaluation of the Phytase Enzyme in Granulated and Liquid Forms for Nile Tilapia (Oreochromis niloticus). *Archivos de zootecnia* **68**(2018).
115. Peatman, E. & Beck, B. Chapter 16 From floor sweepings to fish flesh–phytase superdosing in the US catfish industry. in *Phytate destruction-consequences for precision animal nutrition* 10-22 (Wageningen Academic Publishers, 2016).
116. Morales, G., Marquez, L., Hernández, A. & Moyano, F. Chapter 9 Phytase effects on protein and phosphorus bioavailability in fish diets. in *Phytate destruction-consequences for precision animal nutrition* 1-8 (Wageningen Academic Publishers, 2016).
117. Kies, A., Kemme, P., Sebek, L., Van Diepen, J.T.M. & Jongbloed, A. Effect of graded doses and a high dose of microbial phytase on the digestibility of various minerals in weaner pigs. *Journal of Animal Science* **84**, 1169-1175 (2006).

118. Zeng, Z. *et al.* Effects of adding super dose phytase to the phosphorus-deficient diets of young pigs on growth performance, bone quality, minerals and amino acids digestibilities. *Asian-Australasian journal of animal sciences* **27**, 237 (2014).
119. Cowieson, A., Wilcock, P. & Bedford, M. Super-dosing effects of phytase in poultry and other monogastrics. *World's Poultry Science Journal* **67**, 225-236 (2011).
120. Guggenbuhl, P., Calvo, E.P. & Fru, F. Effect of high dietary doses of a bacterial 6-phytase in piglets fed a corn–soybean meal diet. *Journal of animal science* **94**, 307-309 (2016).
121. Manobhavan, M. *et al.* Effect of super dosing of phytase on growth performance, ileal digestibility and bone characteristics in broilers fed corn–soya-based diets. *Journal of animal physiology and animal nutrition* **100**, 93-100 (2016).
122. Li, M.H. & Robinson, E.H. Microbial Phytase Can Replace Inorganic Phosphorus Supplements in Channel Catfish Ictalurus punctatus Diets 1. *Journal of the World Aquaculture Society* **28**, 402-406 (1997).
123. Eya, J.C. & Lovell, R.T. Net absorption of dietary phosphorus from various inorganic sources and effect of fungal phytase on net absorption of plant phosphorus by channel catfish Ictalurus punctatus. *Journal of the World Aquaculture Society* **28**, 386-391 (1997).
124. Yan, W., Reigh, R.C. & Xu, Z. Effects of Fungal Phytase on Utilization of Dietary Protein and Minerals, and Dephosphorylation of Phytic Acid in the Alimentary Tract of Channel Catfish Ictalurus punctatus Fed an All-Plant-Protein Diet. *Journal of the World Aquaculture Society* **33**, 10-22 (2002).
125. Hossain, M.J. *et al.* Deciphering the conserved genetic loci implicated in plant disease control through comparative genomics of Bacillus amyloliquefaciens subsp. plantarum. *Frontiers in plant science* **6**, 631 (2015).
126. Ran, C. *et al.* Identification of Bacillus strains for biological control of catfish pathogens. *PLoS One* **7**, e45793 (2012).
127. Cha, J.-H., Rahimnejad, S., Yang, S.-Y., Kim, K.-W. & Lee, K.-J. Evaluations of Bacillus spp. as dietary additives on growth performance, innate immunity and disease resistance of olive flounder (Paralichthys olivaceus) against Streptococcus iniae and as water additives. *Aquaculture* **402**, 50-57 (2013).
128. Meidong, R. *et al.* Evaluation of probiotic Bacillus aerius B81e isolated from healthy hybrid catfish on growth, disease resistance and innate immunity of Pla-mong Pangasius bocourti. *Fish & shellfish immunology* **73**, 1-10 (2018).
129. Cerezuela, R., Guardiola, F.A., Meseguer, J. & Esteban, M.Á. Increases in immune parameters by inulin and Bacillus subtilis dietary administration to gilthead seabream (Sparus aurata L.) did not correlate with disease resistance to Photobacterium damselae. *Fish & shellfish immunology* **32**, 1032-1040 (2012).
130. Das, A., Nakhro, K., Chowdhury, S. & Kamilya, D. Effects of potential probiotic Bacillus amyloliquifaciens FPTB16 on systemic and cutaneous mucosal immune responses and disease resistance of catla (Catla catla). *Fish & shellfish immunology* **35**, 1547-1553 (2013).
131. Borriss, R. Bacteria in agrobiology: plant growth responses. *Use of plant-associated Bacillus-strains as biofertilizers and biocontrol agents in Agriculture*, 41-76 (2011).
132. de Simone, C. The unregulated probiotic market. *Clinical Gastroenterology and Hepatology* (2018).

133. Addo, S. *et al.* Effects of Bacillus subtilis strains on growth, immune parameters, and Streptococcus iniae susceptibility in Nile tilapia, Oreochromis niloticus. *Journal of the World Aquaculture Society* **48**, 257-267 (2017).
134. Nandi, A., Banerjee, G., Dan, S.K., Ghosh, K. & Ray, A.K. Probiotic efficiency of Bacillus sp. in Labeo rohita challenged by Aeromonas hydrophila: assessment of stress profile, haemato-biochemical parameters and immune responses. *Aquaculture research* **48**, 4334-4345 (2017).
135. Soltani, M. *et al.* Genus bacillus, promising probiotics in aquaculture: Aquatic animal origin, bio-active components, bioremediation and efficacy in fish and shellfish. *Reviews in Fisheries Science & Aquaculture* **27**, 331-379 (2019).
136. Krebs, B. *et al.* Use of Bacillus subtilis as biocontrol agent. I. Activities and characterization of Bacillus subtilis strains/Anwendung von Bacillus subtilis als Mittel für den biologischen Pflanzenschutz. I. Aktivitäten und Charakterisierung von Bacillus sibtills-Stämmen. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection*, 181-197 (1998).
137. Fan, B. *et al.* Bacillus velezensis FZB42 in 2018: the gram-positive model strain for plant growth promotion and biocontrol. *Frontiers in microbiology* **9**(2018).
138. Ventola, C.L. The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and therapeutics* **40**, 277 (2015).
139. Langendijk, P.S. *et al.* Quantitative fluorescence in situ hybridization of Bifidobacterium spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples. *Applied and environmental microbiology* **61**, 3069-3075 (1995).
140. Morgart, J.R., Filson, J.L., Peters, J.J. & Bhave, R.R. Bacteria removal by ceramic microfiltration. (Google Patents, 1993).
141. Glenn, S.D., Butchko, G., O'connell, E. & Smariga, P. Process and apparatus for removal of DNA, viruses and endotoxins. (Google Patents, 1993).
142. Anderson, R.L. *et al.* Factors associated with Pseudomonas pickettii intrinsic contamination of commercial respiratory therapy solutions marketed as sterile. *Applied and environmental microbiology* **50**, 1343-1348 (1985).
143. Rasmussen-Ivey, C.R. *et al.* Classification of a hypervirulent Aeromonas hydrophila pathotype responsible for epidemic outbreaks in warm-water fishes. *Frontiers in Microbiology* **7**(2016).
144. Andrews, J.H. & Harris, R.F. r-and K-selection and microbial ecology. in *Advances in microbial ecology* 99-147 (Springer, 1986).
145. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews* **68**, 669-685 (2004).
146. Staley, J.T. & Konopka, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology* **39**, 321-346 (1985).
147. Austin, B. The value of cultures to modern microbiology. *Antonie van Leeuwenhoek*, 1-10 (2017).
148. Gray, M.A., Pratte, Z.A. & Kellogg, C.A. Comparison of DNA preservation methods for environmental bacterial community samples. *FEMS microbiology ecology* **83**, 468-477 (2013).
149. Willerslev, E. *et al.* Long-term persistence of bacterial DNA. *Current Biology* **14**, R9-R10 (2004).

150. Mitchell, K.R. & Takacs-Vesbach, C.D. A comparison of methods for total community DNA preservation and extraction from various thermal environments. *Journal of industrial microbiology & biotechnology* **35**, 1139-1147 (2008).
151. O'neill, M., McPartlin, J., Arthure, K., Riedel, S. & McMillan, N. Comparison of the TLDA with the Nanodrop and the reference Qubit system. in *Journal of Physics: Conference series* Vol. 307 012047 (IOP Publishing, 2011).
152. Nakayama, Y., Yamaguchi, H., Einaga, N. & Esumi, M. Pitfalls of DNA quantification using DNA-binding fluorescent dyes and suggested solutions. *PloS one* **11**, e0150528 (2016).
153. Mamedov, T. *et al.* A fundamental study of the PCR amplification of GC-rich DNA templates. *Computational biology and chemistry* **32**, 452-457 (2008).
154. Henegariu, O., Heerema, N., Dlouhy, S., Vance, G. & Vogt, P. Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* **23**, 504-511 (1997).
155. Kang, J., Lee, M.S. & Gorenstein, D.G. The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: application to in vitro combinatorial selection of aptamers. *Journal of Biochemical and Biophysical Methods* **64**, 147-151 (2005).
156. Henke, W., Herdel, K., Jung, K., Schnorr, D. & Loening, S.A. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic acids research* **25**, 3957-3958 (1997).
157. Frackman, S., Kobs, G., Simpson, D. & Storts, D. Betaine and DMSO: enhancing agents for PCR. *Promega notes* **65**, 27-29 (1998).
158. Kanagawa, T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of bioscience and bioengineering* **96**, 317-323 (2003).
159. Speksnijder, A.G. *et al.* Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Applied and environmental microbiology* **67**, 469-472 (2001).
160. Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* **11**, 1026-1030 (1993).
161. Kurata, S. *et al.* Fluorescent quenching-based quantitative detection of specific DNA/RNA using a BODIPY® FL-labeled probe or primer. *Nucleic acids research* **29**, e34-e34 (2001).
162. Zhou, J., Bruns, M.A. & Tiedje, J.M. DNA recovery from soils of diverse composition. *Applied and environmental microbiology* **62**, 316-322 (1996).
163. Salonen, A. *et al.* Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *Journal of microbiological methods* **81**, 127-134 (2010).
164. Joseph, S. & David, W.R. Molecular cloning: a laboratory manual. *Gold Spring Harbor, New York* (2001).
165. Krsek, M. & Wellington, E. Comparison of different methods for the isolation and purification of total community DNA from soil. *Journal of Microbiological Methods* **39**, 1-16 (1999).
166. Cuív, P.Ó. *et al.* The effects from DNA extraction methods on the evaluation of microbial diversity associated with human colonic tissue. *Microbial ecology* **61**, 353-362 (2011).
167. Bergmann, I. *et al.* Influence of DNA isolation on Q-PCR-based quantification of methanogenic Archaea in biogas fermenters. *Systematic and applied microbiology* **33**, 78-84 (2010).

168. Ariefdjohan, M.W., Savaiano, D.A. & Nakatsu, C.H. Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. *Nutrition journal* **9**, 23 (2010).
169. Yu, Z. & Morrison, M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**, 808-813 (2004).
170. Nielsen, K.M., Johnsen, P.J., Bensasson, D. & Daffonchio, D. Release and persistence of extracellular DNA in the environment. *Environmental biosafety research* **6**, 37-53 (2007).
171. Rudi, K., Moen, B., Drømtorp, S.M. & Holck, A.L. Use of ethidium monoazide and PCR in combination for quantification of viable and dead cells in complex samples. *Applied and Environmental Microbiology* **71**, 1018-1024 (2005).
172. Wang, S. & Levin, R.E. Discrimination of viable Vibrio vulnificus cells from dead cells in real-time PCR. *Journal of Microbiological Methods* **64**, 1-8 (2006).
173. Hein, I., Schneeweiss, W., Stanek, C. & Wagner, M. Ethidium monoazide and propidium monoazide for elimination of unspecific DNA background in quantitative universal real-time PCR. *Journal of microbiological methods* **71**, 336-339 (2007).
174. Lee, J.-L. & Levin, R.E. Quantification of total viable bacteria on fish fillets by using ethidium bromide monoazide real-time polymerase chain reaction. *International journal of food microbiology* **118**, 312-317 (2007).
175. Zong, C., Lu, S., Chapman, A.R. & Xie, X.S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622-1626 (2012).
176. Nocker, A., Cheung, C.-Y. & Camper, A.K. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. *Journal of microbiological methods* **67**, 310-320 (2006).
177. Acinas, S.G., Marcelino, L.A., Klepac-Ceraj, V. & Polz, M.F. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *Journal of bacteriology* **186**, 2629-2635 (2004).
178. Ludwig, W. & Schleifer, K.-H. How quantitative is quantitative PCR with respect to cell counts? *Systematic and applied microbiology* **23**, 556-562 (2000).
179. Suzuki, M.T. & Giovannoni, S.J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and environmental microbiology* **62**, 625-630 (1996).
180. Angly, F.E. *et al.* CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* **2**, 11 (2014).
181. Rastogi, R., Wu, M., DasGupta, I. & Fox, G.E. Visualization of ribosomal RNA operon copy number distribution. *BMC microbiology* **9**, 208 (2009).
182. Kembel, S.W., Wu, M., Eisen, J.A. & Green, J.L. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol* **8**, e1002743 (2012).
183. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS one* **8**, e57923 (2013).
184. Felsenstein, J. Phylogenies and the comparative method. *The American Naturalist* **125**, 1-15 (1985).

185. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *nature* **473**, 174-180 (2011).
186. Schellenberg, J. *et al.* Pyrosequencing of chaperonin-60 (cpn60) amplicons as a means of determining microbial community composition. *High-Throughput Next Generation Sequencing: Methods and Applications*, 143-158 (2011).
187. Case, R.J. *et al.* Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Applied and environmental microbiology* **73**, 278-288 (2007).
188. Martinez-Murcia, A.J. *et al.* Multilocus phylogenetic analysis of the genus Aeromonas. *Systematic and applied microbiology* **34**, 189-199 (2011).
189. Green, S.J., Venkatramanan, R. & Naqib, A. Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. *PLoS One* **10**, e0128122 (2015).
190. Ross, M.G. *et al.* Characterizing and measuring bias in sequence data. *Genome biology* **14**, R51 (2013).
191. van Dijk, E.L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental cell research* **322**, 12-20 (2014).
192. Berry, D., Mahfoudh, K.B., Wagner, M. & Loy, A. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and environmental microbiology* **77**, 7846-7849 (2011).
193. Mokry, M. *et al.* Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles. *PloS one* **5**, e15092 (2010).
194. Quail, M.A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nature methods* **5**, 1005-1010 (2008).
195. Kennedy, K., Hall, M.W., Lynch, M.D., Moreno-Hagelsieb, G. & Neufeld, J.D. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Applied and environmental microbiology* **80**, 5717-5722 (2014).
196. Smith, D.P. & Peay, K.G. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS One* **9**, e90234 (2014).
197. Driskell, L.O., Tucker, A.M., Woodard, A., Wood, R.R. & Wood, D.O. Fluorescence activated cell sorting of Rickettsia prowazekii-infected host cells based on bacterial burden and early detection of fluorescent rickettsial transformants. *PloS one* **11**, e0152365 (2016).
198. Bonner, W., Hulett, H., Sweet, R. & Herzenberg, L. Fluorescence activated cell sorting. *Review of Scientific Instruments* **43**, 404-409 (1972).
199. Schmid, L., Weitz, D.A. & Franke, T. Sorting drops and cells with acoustics: acoustic microfluidic fluorescence-activated cell sorter. *Lab on a Chip* **14**, 3710-3718 (2014).
200. Ottesen, E.A., Hong, J.W., Quake, S.R. & Leadbetter, J.R. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *science* **314**, 1464-1467 (2006).
201. Espina, V. *et al.* Laser-capture microdissection. *Nature protocols* **1**, 586-603 (2006).
202. Arneson, N., Hughes, S., Houlston, R. & Done, S. Whole-genome amplification by degenerate oligonucleotide primed PCR (DOP-PCR). *Cold Spring Harbor Protocols* **2008**, pdb. prot4919 (2008).
203. Gawad, C., Koh, W. & Quake, S.R. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics* **17**, 175-188 (2016).

204. Blainey, P.C. The future is now: single-cell genomics of bacteria and archaea. *FEMS microbiology reviews* **37**, 407-427 (2013).
205. Dean, F.B., Nelson, J.R., Giesler, T.L. & Lasken, R.S. Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome research* **11**, 1095-1099 (2001).
206. Zhang, D.Y., Brandwein, M., Hsuih, T. & Li, H.B. Ramification amplification: a novel isothermal DNA amplification method. *Molecular Diagnosis* **6**, 141-150 (2001).
207. De Bourcy, C.F. *et al.* A quantitative comparison of single-cell whole genome amplification methods. *PloS one* **9**, e105585 (2014).
208. Quail, M.A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* **13**, 341 (2012).
209. Liu, L. *et al.* Comparison of next-generation sequencing systems. *BioMed Research International* **2012**(2012).
210. Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic acids research*, gku1341 (2015).
211. Nielsen, M.E. *et al.* Is Aeromonas hydrophila the dominant motile Aeromonas species that causes disease outbreaks in aquaculture production in the Zhejiang Province of China? *Diseases of Aquatic Organisms* **46**, 23-29 (2001).
212. Hemstreet, B. An update on Aeromonas hydrophila from a fish health specialist for summer 2010. *Catfish Journal* **24**, 4 (2010).
213. Xu, B.H., Yin, Z., Wu, Y.S. & Cai, T.Z. Studies on the taxonomy of pathogenic bacteria of the bacterial hemorrhagic septicemia in cultured fishes in freshwater. *Acta Hydrobiologica Sinica* **17**, 259-266 (1993).
214. Camus, A.C., Durborow, R.M., Hemstreet, W.G., Thune, R.L. & Hawke, J.P. Aeromonas Bacterial Infections: Motile Aeromonad Septicemia (No. 478). *Southern Regional Aquaculture Center* (1998).
215. da Silva, B.C. *et al.* Haemorrhagic septicaemia in the hybrid surubim (Pseudoplatystoma corruscans x Pseudoplatystoma fasciatum) caused by Aeromonas hydrophila. *Aquaculture Research* **43**, 908-916 (2012).
216. Chen, H.Q. & Lu, C.P. Study on the pathogen of epidemic septicemia occurred in cultured cyprinoid fishes in southern China. . *Journal of Nanjing Agricultural University*, 14, 87–91 (1991).
217. Zhang, Y.L., Arakawa, E. & Leung, K.Y. Novel Aeromonas hydrophila PPD134/91 genes involved in O-antigen and capsule biosynthesis. *Infect Immun* **70**, 2326-35 (2002).
218. Pang, M.D. *et al.* Tetrahymena: An Alternative Model Host for Evaluating Virulence of Aeromonas Strains. *Plos One* **7**(2012).
219. Deng, G.C. *et al.* Isolation, identification and characterization of Aeromonas hydrophila from hemorrhagic grass carp. *China* **36**(2009).
220. Hossain, M.J. *et al.* An Asian Origin of Virulent Aeromonas hydrophila Responsible for Disease Epidemics in United States-Farmed Catfish. *MBio* **5**(2014).
221. Hemstreet, W. Aeromonas summary. *Fish Farming News: Alabama Fish Farming Center* (2015).
222. Hanson, L., Liles, M.R., Hossain, M.J., Griffin, M. & Hemstreet, W. Motile Aeromonas Septicemia. *Fish Health Section Blue Book 2014 Edition* (2014).

223. Figueras, M.J., Beaz-Hidalgo, R., Hossain, M.J. & Liles, M.R. Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announcements* **2**, e00927-14 (2014).
224. Zhang, D., Xu, D.-H. & Shoemaker, C.A. Experimental induction of motile Aeromonas septicemia in channel catfish by water-borne challenge with virulent Aeromonas hydrophila. *Aquaculture Reports* **3**, 18-23 (2016).
225. Angiuoli, S.V. & Salzberg, S.L. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* **27**, 334-42 (2011).
226. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540-552 (2000).
227. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
228. Aziz, R.K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
229. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* **42**, D206-D214 (2014).
230. Richter, M. & Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19126-19131 (2009).
231. Rodriguez, L.M. & Konstantinidis, K. Kostas lab | ANI calculator. (2014).
232. Verschuere, L., Rombaut, G., Sorgeloos, P. & Verstraete, W. Probiotic bacteria as biological control agents in aquaculture. *Microbiol. Mol. Biol. Rev.* **64**, 655-671 (2000).
233. Rasmussen-Ivey, C., Figueras, M.J., McGarey, D. & Liles, M.R. Virulence Factors of Aeromonas hydrophila: in the Wake of Reclassification. *Frontiers in Microbiology* (2016).
234. Martinez-Garcia, P.M., Ramos, C. & Rodriguez-Palenzuela, P. T346Hunter: A Novel Web-Based Tool for the Prediction of Type III, Type IV and Type VI Secretion Systems in Bacterial Genomes. *Plos One* **10**(2015).
235. Leiman, P.G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4154-4159 (2009).
236. Pell, L.G., Kanelis, V., Donaldson, L.W., Howell, P.L. & Davidson, A.R. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4160-4165 (2009).
237. Shrivastava, S. & Mande, S.S. Identification and Functional Characterization of Gene Components of Type VI Secretion System in Bacterial Genomes. *Plos One* **3**(2008).
238. Kitaoka, M., Miyata, S.T., Brooks, T.M., Unterweger, D. & Pukatzki, S. VasH Is a Transcriptional Regulator of the Type VI Secretion System Functional in Endemic and Pandemic Vibrio cholerae. *Journal of Bacteriology* **193**, 6471-6482 (2011).
239. Pang, M.B. *et al.* Novel insights into the pathogenicity of epidemic Aeromonas hydrophila ST251 clones from comparative genomics. *Sci Rep* **5**(2015).
240. Beaz-Hidalgo, R. *et al.* Aeromonas aquatica sp nov., Aeromonas finlandiensis sp nov and Aeromonas lacus sp nov isolated from Finnish waters associated with cyanobacterial blooms. *Syst Appl Microbiol* **38**, 161-168 (2015).

241. Velayudhan, J., Jones, M.A., Barrow, P.A. & Kelly, D.J. L-serine catabolism via an oxygen-labile L-serine dehydratase is essential for colonization of the avian gut by Campylobacter jejuni. *Infection and Immunity* **72**, 260-268 (2004).
242. Severi, E., Hood, D.W. & Thomas, G.H. Sialic acid utilization by bacterial pathogens. *Microbiology-Sgm* **153**, 2817-2822 (2007).
243. Rohmer, L., Hocquet, D. & Miller, S.I. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends in microbiology* **19**, 341-348 (2011).
244. Martin, R.L. Comparison of Effects of Concentrations of Malachite Green and Acriflavine on Fungi Associated with Diseased Fish. *The Progressive Fish Culturist* **30**, 153-158 (1968).
245. Erova, T.E. *et al.* DNA adenine methyltransferase influences the virulence of Aeromonas hydrophila. *Infection and immunity* **74**, 410-424 (2006).
246. Kirke, D.F., Swift, S., Lynch, M.J. & Williams, P. The Aeromonas hydrophila LuxR homologue AhyR regulates the N-acyl homoserine lactone synthase, AhyI positively and negatively in a growth phase-dependent manner. *FEMS microbiology letters* **241**, 109-117 (2004).
247. Clock, S.A., Planet, P.J., Perez, B.A. & Figurski, D.H. Outer membrane components of the tad (tight adherence) secreton of Aggregatibacter actinomycetemcomitans. *Journal of bacteriology* **190**, 980-990 (2008).
248. Sandt, C.H., Hopper, J.E. & Hill, C.W. Activation of prophage eib genes for immunoglobulin-binding proteins by genes from the IbrAB genetic island of Escherichia coli ECOR-9. *Journal of Bacteriology* **184**, 3640-3648 (2002).
249. Wang, M. *et al.* Molecular characterization of a functional type VI secretion system in Salmonella enterica serovar Typhi. *Current Microbiology* **63**, 22-31 (2011).
250. Klappenbach, J.A., Saxman, P.R., Cole, J.R. & Schmidt, T.M. rrndb: the ribosomal RNA operon copy number database. *Nucleic acids research* **29**, 181-184 (2001).
251. Miñana-Galbis, D. *et al.* Reclassification of Aeromonas hydrophila subspecies anaerogenes. *Systematic and applied microbiology* **36**, 306-308 (2013).
252. Vos, M., Quince, C., Pijl, A.S., de Hollander, M. & Kowalchuk, G.A. A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS One* **7**, e30600 (2012).
253. Johnson, L.A., Chaban, B., Harding, J.C. & Hill, J.E. Optimizing a PCR protocol for cpn 60-based microbiome profiling of samples variously contaminated with host genomic DNA. *BMC research notes* **8**, 253 (2015).
254. Zhang, Y. *et al.* Bovine thrombin enhances the efficiency and specificity of polymerase chain reaction. *BioTechniques* **57**, 289-294 (2014).
255. Edgar, R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* **5**, 113 (2004).
256. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).
257. Mukhametzyanova, A., Akhmetova, A. & Sharipova, M. Microorganisms as phytase producers. *Microbiology* **81**, 267-275 (2012).
258. Breitwieser, F., Baker, D. & Salzberg, S.L. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome biology* **19**, 198 (2018).
259. O'Donnell, C.R., Wang, H. & Dunbar, W.B. Error analysis of idealized nanopore sequencing. *Electrophoresis* **34**, 2137-2144 (2013).

260. Rang, F.J., Kloosterman, W.P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome biology* **19**, 90 (2018).
261. Laver, T. *et al.* Assessing the performance of the oxford nanopore technologies minion. *Biomolecular detection and quantification* **3**, 1-8 (2015).
262. Kroger, C., Srikumar, S., Ellwart, J. & Fuchs, T.M. Bistability in myo-inositol utilization by Salmonella enterica serovar Typhimurium. *J Bacteriol* **193**, 1427-35 (2011).
263. Staib, L. & Fuchs, T.M. From food to cell: nutrient exploitation strategies of enteropathogens. *Microbiology* **160**, 1020-1039 (2014).
264. Hossain, M.J. *et al.* Implication of lateral genetic transfer in the emergence of Aeromonas hydrophila isolates of epidemic outbreaks in channel catfish. *PLoS One* **8**, e80943 (2013).
265. Kumar, V., Sinha, A., Makkar, H., De Boeck, G. & Becker, K. Phytate and phytase in fish nutrition. *Journal of animal physiology and animal nutrition* **96**, 335-364 (2012).
266. Baruah, K., Sahu, N., Pal, A. & Debnath, D. Dietary phytase: an ideal approach for a cost effective and low-polluting aquafeed. *NAGA, WorldFish Center Quarterly* **27**, 15-19 (2004).
267. Da, C.T., Lundh, T. & Lindberg, J.E. Digestibility of dietary components and amino acids in animal and plant protein feed ingredients in striped catfish (Pangasianodon hypophthalmus) fingerlings. *Aquaculture Nutrition* **19**, 741-750 (2013).
268. Naylor, R.L. *et al.* Feeding aquaculture in an era of finite resources (vol 106, 15103, 2009). *Proc Natl Acad Sci U S A* **106**, 18040-18040 (2009).
269. Spies, H.S. & Steenkamp, D.J. Thiols of intracellular pathogens. *European Journal of Biochemistry* **224**, 203-213 (1994).
270. Chaouch, S. & Noctor, G. Myo-inositol abolishes salicylic acid-dependent cell death and pathogen defence responses triggered by peroxisomal hydrogen peroxide. *New Phytologist* **188**, 711-718 (2010).
271. Sun, D. Auburn University (2014).
272. Cao, L. *et al.* Application of microbial phytase in fish feed. *Enzyme and Microbial Technology* **40**, 497-507 (2007).
273. Kenny, D. & Couch, T. Mass production of biological control agents for plant diseases, weeds and insects control. *Biological control in crop production. Papavizas, GC Ed. Allenheld, Totowa, NJ. 143p* (1981).
274. Peatman, E. *et al.* Mechanisms of pathogen virulence and host susceptibility in virulent Aeromonas hydrophila infections of channel catfish (Ictalurus punctatus). *Aquaculture* **482**, 1-8 (2018).
275. Larsen, A.M., Mohammed, H.H. & Arias, C.R. Comparison of DNA extraction protocols for the analysis of gut microbiota in fishes. *FEMS microbiology letters* **362**, fnu031 (2015).
276. Krishnakumar, R. *et al.* Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias. *Scientific reports* **8**, 3159 (2018).
277. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27**, 722-736 (2017).
278. Zhang, D., Pridgeon, J.W. & Klesius, P.H. Expression and activity of recombinant proaerolysin derived from Aeromonas hydrophila cultured from diseased channel catfish. *Veterinary microbiology* **165**, 478-482 (2013).

279. Liles, M.R. *et al.* Vaccines for control of epidemic Aeromonas hydrophila generated by markerless gene deletion. (Google Patents, 2016).
280. Lalloo, R., Moonsamy, G., Ramchuran, S., Görgens, J. & Gardiner, N. Competitive exclusion as a mode of action of a novel Bacillus cereus aquaculture biological agent. *Letters in applied microbiology* **50**, 563-570 (2010).
281. Baumgartner, W.A., Ford, L. & Hanson, L. Lesions caused by virulent Aeromonas hydrophila in farmed catfish (Ictalurus punctatus and I. punctatus× I. furcatus) in Mississippi. *Journal of Veterinary Diagnostic Investigation* **29**, 747-751 (2017).
282. Zhang, X.J. *et al.* Does the gastrointestinal tract serve as the infectious route of Aeromonas hydrophila in crucian carp (Carassius carassius)? *Aquaculture Research* **46**, 141-154 (2015).