

**Niche adaptations of the *Vibrionaceae*, from the coastal  
ocean to the laboratory**

by

Alison Francesca Takemura

B.S., Rice University (2008)

Submitted to the Microbiology Graduate Program  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

© Alison Francesca Takemura, MMXV. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly  
paper and electronic copies of this thesis document in whole or in part in any  
medium now known or hereafter created.

Author.....  
Microbiology Graduate Program  
August 31, 2015

Certified by.....  
Martin Polz  
Professor of Civil and Environmental Engineering  
Thesis Supervisor

Accepted by.....  
Kristala Prather  
Associate Professor of Chemical Engineering  
Director of the Microbiology Graduate Program

# Niche adaptations of the *Vibrionaceae*, from the coastal ocean to the laboratory

by

Alison Francesca Takemura

Submitted to the Microbiology Graduate Program  
on August 31, 2015, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## Abstract

Microorganisms play a significant role in biogeochemical cycling, thus their dynamics in the environment influence the biosphere. Yet how do features of the environment — such as abiotic conditions, resources, and predators — influence their activity and abundance, i.e. what constitutes their ecological niche? This study examines this question for members of a diverse marine heterotrophic family of bacteria, the *Vibrionaceae*. In chapter 2, I review the current knowledge of the environmental conditions and habitats in which *Vibrionaceae* populations are found. Through a meta-analysis of *Vibrio* abundance and bulk environmental variables, I show that temperature and salinity are strong correlates of *Vibrio*, but the patterns vary among species. By contrast, other commonly measured abiotic variables, like nitrogen and phosphate, are only weak correlates. Studies furthermore show that *Vibrio* engage in a diversity of lifestyles, from free-living to attached, in a wide range of habitats, though the patterns have largely not been characterized at a genetic or molecular scale. These observations motivate a finer-scale investigation of the microbial niche. In chapter 3, I explore how a single *Vibrio* strain is adapted to growth on different ecologically relevant resources, using nutrients extracted from habitat models — the copepod *Apocyclops royi*, and the brown alga *Fucus vesiculosus* — as well as the algal constituent, alginate. By selecting a transposon-mutant collection for growth on these resources, I find that *Apocyclops* is a replete resource, whereas *Fucus* is intermediate to *Apocyclops* and alginate in its anabolic requirements; that catabolic pathways have redundancy, which anabolic ones lack, that appears to mask fitness effects; and more generally, that these habitats contain complex resources that buffer fitness costs relative to growth on single carbohydrate resources. In appendix A, I determine how environmental phage isolates recognize the *Vibrio* strain: by its extracellular polysaccharide capsule. Losing the capsule enables the strain to resist infection from these bacteriophage; however, it suffers the tradeoff of becoming susceptible to others. By integrating environmental observations and genetic methods, this thesis provides an intimate view of the life of a marine microorganism.

Thesis Supervisor: Martin Polz

Title: Professor of Civil and Environmental Engineering



## Acknowledgments

I'm deeply grateful to my advisor Martin Polz, for his guidance, insight, and unfailing support; and to my committee members, Eric Alm, Jeff Gore, and Matt Johnson, for their infectious enthusiasm and eagerness to look at my science alongside me.

The members of the Polz lab, past and present, have contributed to me personally and intellectually — I count you not only as colleagues, but as friends: Michael Cutler, Chris Corzett (who was a partner in the mutant selections), Jan-Hendrik Hehemann, Sarah Preheim (who discovered the particle associations of *Vibrio* sp. F13), Frederique Le Roux (who created the mutagenesis plasmid), Yan Boucher, Hong Xue, Radhey Sharma, Sana Al Attas, Fatima Hussain, Kathryn Kauffman (who, with Fatima, enriched the phage investigation), Hans Wildschutte, Diana Chien, Phil Arevalo, Manoshi Datta, Dave VanInsberghe, Nate Cermak, and Joy Yang. I have learned so much from all of you, about microbes, programming, statistics, experimental design, dancing, persistence, and resilience. I'm keenly grateful for the wisdom and, I think, the unusually high concentration of humor and puns, you've brought to my journey. Fatima, Diana, and Kathryn have especially been a source of constant, unfailing cheer and encouragement. Undergraduates Matt Skalak, Yixin Li, and Mirim Yoo, were also not only partners in labor, but in thinking about my work.

The other members of the Parsons community, the Microbiology program, and beyond have made my time at MIT a joy, especially Jeff Rominger, Kyle Peet, Kelly Daumit, Ben Scandella, Mitul Luhar, Aarti Patel, Jen Nguyen, and Zach Berta. Jessie Berta-Thompson, with her microbial insight and quickness to refill my mug with tea, has also carved a special niche in my heart.

My friends at MIT have provided moral and intellectual support beyond reason, and opportunities for much-needed diversions, especially Kishor Nayar, Chris Smillie, Jonas Helfer, Jacob Colbert, Bonnielee Whang, Annelies Richmond, and Jaime Goldstein.

Beyond MIT, dear friends have also sustained and guided me with their visits, emails, texts, Skypes, and phone calls: Kathleen Hanley, her parents, Shivani Ghoshal, Karen Leu, Niki vonHedemann, Evan and King Chan, Yao Xu, Natt Supab, Norr Santz, Marcella Wong, Beatrice Viramontes, Hanna Hong, Robert Campbell, Jasmine Ng, Ron Matsuoka, Adam Wright, Cathy Hung, Sandi Chiu, Van Le, Brian Abe, and Uriel and Francesca Hernandez.

My family, always available, have eased the rough patches with their humor, perspective, and unconditional love, especially my parents Elaine Austria and Paul Takemura, and my brother Chris Takemura, who was for years the only published author in the family.

And my gratitude and love go to Patrick Brown, who has been an inspiration and a bedrock, even though he's more like a sunflower (for his optimism) or a redwood (for his fortitude).

THIS PAGE INTENTIONALLY LEFT BLANK

# Contents

<b>1</b>	<b>Introduction</b>	<b>17</b>
1.1	The dynamic microbial niche . . . . .	17
1.2	How does the environment shape genome content? . . . . .	19
1.3	Studying the niche with high-throughput tools that connect phenotype to genotype . . . . .	21
1.4	Aims of this thesis . . . . .	23
<b>2</b>	<b>Associations and dynamics of <i>Vibrionaceae</i> in the environment, from the genus to the population level</b>	<b>27</b>
2.1	Abstract . . . . .	27
2.2	Introduction . . . . .	28
2.3	Environmental correlates of <i>Vibrio</i> presence and abundance . . . . .	30
2.3.1	Total <i>Vibrio</i> . . . . .	30
2.3.2	<i>V. cholerae</i> , <i>V. parahaemolyticus</i> , and <i>V. vulnificus</i> . . . . .	33
2.3.3	Conclusions from meta-analysis . . . . .	39
2.4	Associations with complex and particulate marine growth substrates . . . . .	42
2.5	Biological niches for <i>Vibrio</i> . . . . .	43
2.5.1	Associations with plants . . . . .	43
2.5.2	Associations with microalgae and filamentous cyanobacteria . . . . .	48
2.5.3	Associations with macroalgae . . . . .	49
2.5.4	Associations with animals . . . . .	50
2.5.5	Population dynamics associated with macroinvertebrate hosts . . . . .	53
2.6	<i>Vibrio</i> proliferation in the water column . . . . .	54
2.6.1	<i>Vibrio</i> blooms . . . . .	55
2.6.2	The evidence for a planktonic, free-living lifestyle . . . . .	56
2.7	Using ecology to define cohesive populations . . . . .	57
2.8	Populations as ecological, genetic, and social units . . . . .	60
2.9	Conclusion . . . . .	60
2.10	Acknowledgments . . . . .	62

<b>3</b>	<b>Adaptations in <i>Vibrio</i> to ecological habitat resources, assayed by genome-wide fitness</b>	<b>63</b>
3.1	Abstract . . . . .	63
3.2	Introduction . . . . .	64
3.3	Results and discussion . . . . .	65
3.3.1	Construction of a transposon-mutant library . . . . .	65
3.3.2	Selection in five media conditions . . . . .	66
3.3.3	Validation of approach to test for significant changes in mutant abundance . . . . .	66
3.3.4	Exclusion of Extrachromosomal Element 1 . . . . .	70
3.3.5	Prediction of the 9CS106 essential genome . . . . .	70
3.3.6	Overview of fitness determinants for growth on ecologically relevant resources . . . . .	71
3.3.7	Selected mutant libraries more similar between <i>Fucus</i> and alginate, <i>Apocyclops</i> and 2216 . . . . .	71
3.3.8	Nutrients from habitats buffer fitness costs . . . . .	74
3.3.9	Anabolism and catabolism in the <i>Fucus</i> , alginate, and <i>Apocyclops</i> conditions . . . . .	77
3.3.10	Cofactor and ion metabolism . . . . .	79
3.3.11	Evidence of a costly colonization factor, MSHA, in the <i>Fucus</i> condition . . . . .	80
3.3.12	Intergenic regions . . . . .	80
3.4	Conclusion . . . . .	81
3.5	Materials and Methods . . . . .	81
3.5.1	Strains and mutagenesis plasmid . . . . .	81
3.5.2	Transposon mutagenesis . . . . .	81
3.5.3	Media for selective growth of 9CS106 mutant library . . . . .	93
3.5.4	Growth of <i>Vibrio</i> sp. F13 strain 9CS106 mutant library on selective media . . . . .	93
3.5.5	Sample preparation and Illumina sequencing . . . . .	94
3.5.6	Processing of sequencing reads and gene categorization . . . . .	97
3.5.7	Fitness calculation . . . . .	99
3.5.8	Genome annotation . . . . .	99
3.6	Supplementary Information . . . . .	100
3.6.1	Construction of mutagenesis plasmid . . . . .	100
<b>4</b>	<b>Conclusion</b>	<b>105</b>
<b>A</b>	<b>Identification of a <i>Vibrio</i> phage receptor</b>	<b>109</b>
A.1	Introduction . . . . .	109
A.2	Results and Discussion . . . . .	110
A.2.1	Isolation of phages to 9CS106 . . . . .	110

A.2.2	Selection for phage resistance . . . . .	110
A.2.3	Tradeoff in phage susceptibility in resistant mutants . . . . .	110
A.2.4	T1 and T6 phage have similar phenotypes . . . . .	112
A.2.5	Con-ARTIST to determine resistance factors . . . . .	113
A.2.6	Extracellular polysaccharide capsule is phage receptor . . . . .	113
A.2.7	Capsule dynamics and targeting by phage . . . . .	115
A.2.8	Capsule synthesis genes are diverse among the <i>Vibrio</i> . . . . .	116
A.3	Conclusion . . . . .	116
A.4	Materials and methods . . . . .	118
A.4.1	Phage isolation and purification . . . . .	118
A.4.2	Transposon mutagenesis . . . . .	119
A.4.3	Phage exposure and selection of resistant mutants . . . . .	119
A.4.4	Preparation and sequencing of resistant mutants . . . . .	119
A.4.5	Analysis of resistance-associated genes in phage-exposed mutant libraries . . . . .	119
A.4.6	Host range assay of isolates recovered from phage exposure . . . . .	119
A.4.7	Electron microscopy of host cells and isolated phages . . . . .	120
A.4.8	Annotation of 9CS106 Extrachromosomal Element 1 . . . . .	120
<b>B</b>	<b>Protocols</b> . . . . .	<b>121</b>
B.1	Tn-mutagenesis . . . . .	121
B.1.1	Materials . . . . .	121
B.1.2	Prep for cell growth . . . . .	121
B.1.3	Prep for mutant selection . . . . .	121
B.1.4	Grow cells . . . . .	122
B.1.5	Mate cells . . . . .	122
B.1.6	Check mutant yield . . . . .	123
B.1.7	Plate remainder of mating spot resuspension . . . . .	124
B.1.8	Collect mutants . . . . .	124
B.1.9	Mutant library resuspension . . . . .	125
B.2	Media preparation . . . . .	126
B.2.1	Marine Broth 2216 . . . . .	126
B.2.2	Artificial Sea Water (ASW) . . . . .	126
B.2.3	Minimal Medium (MM) Trace Elements and Vitamins . . . . .	126
B.2.4	Minimal Medium (MM) . . . . .	127
B.2.5	Freeze-dried, ground <i>Apocyclops</i> . . . . .	127
B.2.6	Powdered <i>Fucus</i> . . . . .	128
B.2.7	Alginate . . . . .	128
B.2.8	Glucose . . . . .	128
B.2.9	Phage isolation . . . . .	129
B.3	Phage selection . . . . .	130
B.3.1	Growth media . . . . .	130

B.3.2	Growth with phage . . . . .	130
B.3.3	Harvest resistant colonies. . . . .	131
B.4	DNA extraction using the MasterPure kit (epicentre) . . . . .	131
B.4.1	A. Cell Lysis . . . . .	131
B.4.2	B. Precipitation of Protein and DNA . . . . .	132
B.5	RNase A treatment . . . . .	133
B.6	Ethanol Precipitation of DNA . . . . .	133
B.7	Check DNA quantities and confirm amplifiable by qPCR . . . . .	134
B.8	Preparing DNA library for sequencing . . . . .	134
B.8.1	Main reagents and kits . . . . .	137
B.8.2	MmeI restriction digestion . . . . .	140
B.8.3	Blunting . . . . .	140
B.8.4	QIAquick Column Purification . . . . .	140
B.8.5	A-Tailing with Taq . . . . .	141
B.8.6	Anneal adapters . . . . .	142
B.8.7	Adapter ligation . . . . .	143
B.8.8	PCR 1. Amplification of transposon-associated DNA . . . . .	143
B.8.9	PCR 2. Addition of barcode, Illumina, and variability sequences . . . . .	144
B.8.10	Gel purify PCR product . . . . .	146
B.8.11	Submit samples for sequencing . . . . .	147
B.8.12	Gel-based quantification . . . . .	148
B.9	Processing raw sequencing reads to TA site tallies . . . . .	148
B.9.1	Downloading demultiplexed sequencing data . . . . .	148
B.9.2	Trimming reads using CLC . . . . .	149
B.9.3	Mapping reads to reference genome with Bowtie . . . . .	150
B.9.4	Creating a genome gene list and TA site maps for each chromosome . . . . .	150
B.9.5	Mapping genome-mapped reads to TA sites . . . . .	151
B.10	Checking data quality . . . . .	152
B.10.1	Calculating the overall library saturation . . . . .	152
B.10.2	Calculating average reads per TA site per gene . . . . .	153
B.10.3	Graphing a histogram of % of TA sites disrupted per gene . . . . .	153
B.10.4	Visualizing reads in Artemis . . . . .	154
B.11	Analyzing mutant strain fitness . . . . .	156
B.12	Running the ARTIST pipeline . . . . .	158
B.12.1	Generate input files for ARTIST . . . . .	158
B.12.2	Setting up the Matlab workspace to run ARTIST . . . . .	161
B.12.3	Run EL-ARTIST . . . . .	163
B.12.4	Run Con-ARTIST . . . . .	164
B.13	Analyze ARTIST results . . . . .	166
B.13.1	Annotate ARTIST ordered, non-overlapping loci . . . . .	166
B.13.2	Annotate loci.txt with RAST . . . . .	167

B.13.3	Save all Con-ARTIST output, separated by chromosome, to a single file	168
B.13.4	Separate Con-ARTIST results by each of the six MWUstats . . . . .	169
B.13.5	Pull out locus called significant in at least one experiment . . . . .	170
B.13.6	Excerpt other MWUstats based on locus significance in any experiment	171
B.13.7	Pull out genes ‘core’ (consistent) or ‘flexible’ between replicates . . .	173
B.13.8	Look for significant loci shared and not shared between conditions .	173
B.13.9	Pull out $P$ -values and enrichment direction for significant loci . . . .	174
B.13.10	Link direction of enrichment: de-enriched or enriched . . . . .	175
<b>C</b>	<b>Thesis defense presentation</b>	<b>177</b>

THIS PAGE INTENTIONALLY LEFT BLANK



# List of Figures

1-1	<b>Populations associated with zooplankton and vegetation particles,</b> from (Preheim, 2010). Maximum likelihood phylogenetic tree derived from concatenation of partial <i>adk</i> , <i>mdh</i> and <i>hsp60</i> sequences for all strains (clones represented by one concatenated sequence) isolated from detrital algal (previously described as plant-derived) and zooplankton particles with rings designating season (outer ring) and particle type (inner ring) of origin for each strain. Circles at the nodes of each population are colored according to the predicted habitat preference as determined by the AdaptML algorithm (Hunt et al., 2008a). Populations are alternatively shaded blue and gray and numbered corresponding to species as follows: 1, <i>Aliivibrio</i> sp.13; 2, <i>Aliivibriofischeri</i> ; 3, <i>Enterovibriocalviensis</i> -like; 4, <i>Enterovibrionorvegicus</i> ; 5, <i>V. breoganii</i> ; 6, <i>Vibrio</i> sp. F10; 7, <i>Vibrio</i> sp. F13; 8, <i>V. tasmaniensis</i> / <i>V. lentus</i> ; 9, <i>V. splendidus</i> . . . . .	25
1-2	<b><i>Vibrio</i> sp. F13 strain 9CS106.</b> Electron micrographs courtesy of Fatima Hussain and Kathryn Kauffman. . . . .	26
2-1	<b>An overview of regression analyses indicate that temperature and salinity explain most variation in bulk-water total <i>Vibrio</i> abundance.</b> The $R^2$ , or pseudo- $R^2$ , values associated with regression analyses are shown for selected environmental variables that are well-represented across studies. An individual study may perform multiple analyses because variables are considered for correlation independently (for ex. (Wetz et al., 2008)); because datasets are split (e.g. between seasons in Oberbeckmann et al. (2012)); or because different sets of variables are considered sequentially (e.g. two variables versus six variables in the two All Seasons models from Froelich et al. (2013)). Dots indicate bar heights, and where a dot occurs without a bar, $R^2$ was non-significant (i.e. $R^2 = 0$ ). Variables may have been log or exponentially transformed in references. . . . .	32

2-2	<p><b>Variation in <i>V. cholerae</i> abundance or percent positive samples is best explained by temperature, other organisms, and salinity.</b> <math>R^2</math>, or pseudo-<math>R^2</math>, values from analyses across studies are depicted grouped by variable, and then in rank order, with their associated reference. A reference may conduct multiple analyses for a given variable (e.g., on subsets of data or considering different variables combinations for data regression). Dots indicate bar heights, and where a dot occurs without a bar, <math>R^2</math> was non-significant (i.e. <math>R^2 = 0</math>).</p>	34
2-3	<p><b>Variation in <i>V. parahaemolyticus</i> abundance or percent positive samples is best explained by temperature and other organisms.</b> <math>R^2</math>, or pseudo-<math>R^2</math>, values from analyses across studies are depicted grouped by variable, and then in rank order, with their associated reference. A reference may conduct multiple analyses for a given variable (e.g., on subsets of data or considering different variables combinations for data regression). Dots indicate bar heights, and where a dot occurs without a bar, <math>R^2</math> was non-significant (i.e. <math>R^2 = 0</math>).</p>	36
2-4	<p><b>Variation in <i>V. vulnificus</i> abundance or percent positive samples is best explained by temperature, and other organisms, including <i>Vibrio</i>.</b> <math>R^2</math>, or pseudo-<math>R^2</math>, values from analyses across studies are depicted grouped by variable, and then in rank order, with their associated reference. A reference may conduct multiple analyses for a given variable (e.g., on subsets of data or considering different variables combinations for data regression). Dots indicate bar heights, and where a dot occurs without a bar, <math>R^2</math> was non-significant (i.e. <math>R^2 = 0</math>).</p>	38
2-5	<p><b><i>V. cholerae</i> favors lower salinity and occupies a broad temperature range.</b> <i>V. cholerae</i> concentrations, i.e. MPN-estimated CFU or molecular marker gene copies per 100 mL, reported in different studies are plotted against the temperature (<math>^{\circ}\text{C}</math>) and salinity values (ppt or psu) at which they were found. All studies report <i>V. cholerae</i>, including O1/O139 and non-O1/non-O139, except for Heidelberg et al. (2002a,b); DeLoney-Marino et al. (2003), whose genetic marker detected <i>V. cholerae/V. mimicus</i>. Circle (<math>\circ</math>) sizes correspond to concentrations, but note the breaks are scaled for clearer visualization, and not linearly. (<math>\times</math>) indicates no <i>V. cholerae</i> found in that sample.</p>	40
2-6	<p><b><i>V. parahaemolyticus</i> favors high temperatures but is relatively unconstrained by salinity.</b> Concentrations, i.e. MPN-estimated CFU or molecular marker gene copies per 100 mL, reported in different studies are plotted against the temperature (<math>^{\circ}\text{C}</math>) and salinity values (ppt or psu) at which they were found in bulk water samples. Circle (<math>\circ</math>) sizes correspond to concentrations, but note the breaks are scaled for clearer visualization, and not linearly. (<math>\times</math>) indicates no <i>V. parahaemolyticus</i> found in that sample.</p>	41

3-1	<b>Mutant library grown in selective media and harvested.</b> A transposon-mutant library of <i>Vibrio</i> sp. F13 strain 9CS106 was grown in five media: 2216, <i>Apocyclops</i> , <i>Fucus</i> , alginate, and glucose. The cell densities used to inoculate growth were such that harvesting coincided with the elapse of approximately nine to 10 generations, and prior to entry into stationary phase. The time points in the growth curve when samples were harvested are also highlighted. . . . .	67
3-2	<b>Genes and intergenic (IG) regions under condition-specific selection.</b> The ratio of output to input abundance per gene or IG region in the mutant library of <i>Vibrio</i> sp. F13 strain 9CS106 is compared between different nutrient conditions: (A) alginate vs. glucose media, and (B) <i>Fucus</i> vs. <i>Apocyclops</i> media. Genes and IG regions are highlighted by color indicating whether their abundance was unchanged, or significantly different from the input. Sixty-three IG regions a priori assumed to be neutral were also tested for significance as negative controls; only one locus in the glucose and <i>Fucus</i> conditions was significant. The grey areas indicate the $\log_2$ fold change threshold (0.585) for the negative binomial test used to categorize the genes and IG regions ( $P < 0.05$ ). To obtain abundance, raw read counts were normalized with DESeq2 and $\log_2$ -transformed. Note the differences in scale for plot axes. . . . .	69
3-3	<b>Samples cluster according to resource similarity.</b> Selected mutant libraries from individual replicates cluster by condition. Replicates are biological for media conditions, labelled a, b, c, and technical for the input mutant library, labelled 1, 2. Different conditions also cluster: alginate and <i>Fucus</i> , with glucose as an outgroup, and, separately, <i>Apocyclops</i> and 2216. Hierarchically clustered heatmap of the sample-to-sample Euclidean distances of their Illumina sequencing read counts, normalized using Tikhonov/ridge regression to reduce small count variability, implemented with DESeq2. . . . .	75
3-4	<b>Growth medium 2216 and habitat-derived nutrients buffer fitness costs.</b> The fitness, i.e. the growth rate of 9CS106 mutants relative to that of the wild type, is depicted for each gene and IG region, across conditions. The fitness was calculated from mean read counts after normalization with DESeq2 across replicates ( $n = 3$ ), as detailed in the Section 3.5. . . . .	76
3-5	<b>Pair-wise correlations between biological replicates by gene and intergenic region.</b> Two-dimensional histogram plots for every within-media type pair of samples, binned by total sequencing read counts per gene and IG region. Logarithm (base 10) counts per bin depicted in a heatmap. . . . .	97
3-6	<b>Sequencing results have greater variance than expected for a Poisson process.</b> Variance of the read counts per gene or intergenic region is larger than the mean, i.e falls above the 1:1 line. Reads counts given in $\log_2$ -transformed read counts (+1 to avoid taking a zero logarithm) . . . . .	98

3-7	<b>pSW25T::C9::ISCAT, the plasmid used for transposon-mutagenesis.</b> The <i>mariner</i> transposon modified inverted repeats (red triangles labeled “modified IR”) are labeled with the start of their respective nucleotide sequences: AGA (“aga”) and ACA (“aca”). The chloramphenicol acetyltransferase gene (“cat”) serves as the selection marker for mutagenized strains. Also labeled are the hypertransposase <i>himar1</i> C9 gene (“C9”), plasmid marker spectinomycin resistance cassette (“specR”), origin of transfer (“OriT”) and the <i>pir</i> protein-dependent region for plasmid replication (“oriR6K”). . . . .	103
A-1	<b>Phages T1 (A) and T6 (B) are podoviruses.</b> Electron micrographs courtesy of Kathryn Kauffman and Fatima Hussain. . . . .	111
B-1	<b>Preparation of transposon insertion libraries for Illumina sequencing</b> as adapted from van Opijnen (van Opijnen et al., 2014) and Waldor (unpublished) references, and carried out in this thesis. . . . .	136
B-2	<b>An example histogram of the percent of TA sites disrupted per gene</b> in a sample sequencing library: 2216 media, replicate c. . . . .	154
B-3	<b>An example Artemis plot, showing reads mapping to gene Chr1, gene 682, annotated as S-adenosylmethionine synthetase (EC 2.5.1.6).</b>	156
B-4	<b>Correspondence of fitness values for each mutant genotype, as opposed to gene, from replicates grown in filtered <i>Apocyclops</i> medium.</b>	157
B-5	<b>Correspondence of fitness values for each mutant genotype, as opposed to gene, from one replicates grown in filtered <i>Apocyclops</i> medium, and the other replicate grown in glucose.</b> . . . . .	158

# List of Tables

2.1	Plant and algae hosts for <i>Vibrio</i> , as demonstrated by numerical enrichment and biological mechanisms supporting association. . .	44
2.2	Animal hosts for <i>Vibrio</i> , as demonstrated by numerical enrichment and biological mechanisms supporting association. . . . .	46
3.1	Growth of <i>Vibrio</i> sp. F13 strain 9CS106 in selective media . . . .	68
3.2	Annotation category distributions for media-selected genes . . . .	72
3.3	Log <sub>2</sub> fold change for every identified fitness determinant . . . . .	82
3.4	Illumina sequencing statistics of media-selected 9CS106 mutants.	96
3.5	Primers for plasmid construction . . . . .	101
A.1	Isolated 9CS106 phage. . . . .	110
A.2	Phage host range assay on wild-type 9CS106 and phage-exposed transposon mutants . . . . .	112
A.3	Annotation of Extrachromosomal Element 1 (ECE-1) . . . . .	113
A.4	Putative phage-susceptibility genes in 9CS106 . . . . .	114
A.5	Capsule operon association genes enriched in phage-resistant mutants. . . . .	117
B.1	qPCR primer sequences. . . . .	134
B.2	Oligos for sequencing library . . . . .	138
B.2	Oligos for sequencing library . . . . .	139
B.3	Adapter sequences, as observed in sequencing reads. . . . .	145

THIS PAGE INTENTIONALLY LEFT BLANK

# Chapter 1

## Introduction

### 1.1 The dynamic microbial niche

Microbes occupy far-flung environments, from the tropical oceans to animal interior body cavities. They can occupy Arctic tundra, deep sea thermal vents, deserts, sea ice, acid mine drains, toxic Superfund sites, plant roots, and our elbows, illustrating a diversity in approaches to living that dwarfs the range of our own species. A central goal of microbial ecologists is to make sense of such boggling diversity, by structuring our thinking regarding when, where, and why microorganisms thrive, abide, and die in the wild.

Every organism has a niche, the sum of all its interactions with its environment. These interactions are both biotic, such as competition with neighbors for resources, and abiotic, such as an organism's requirement for light. Conceptualized as an N-dimensional hypervolume where each axis represents a resource or environmental condition (Hutchinson, 1957), the niche is an amorphous, complex entity, whose study is necessarily reduced to individual niche components, or axes.

In the study of microbial niches, a historical approach has been to monitor microbial abundance for a species of interest by culture or marker gene(s) in concert with the study of its environment. For instance, to develop predictive models of pathogen abundance in the coastal ocean, a potential health threat, researchers have correlated bulk environmental variables, such as water temperature and salinity, with microbial abundance (as reviewed in chapter 2). This approach has also been applied at a finer scale, by monitoring microbial species abundance alongside zooplankton abundance or microalgal blooms, both of which might provide habitats or dissolved organic matter to microbes of interest (Turner et al., 2009; Epstein, 1993). Assuming relationships are robust (environmental variables and abundance highly correlated), these approaches can give a rough picture of the niche space.

However, the diversity of microorganisms makes defining a niche in such a broad way problematic. Methods that focus on identifying phylogenetic and taxonomic groups can miss the ecologically significant functional variation within groups. For example, what is taken

to be pathogen abundance may in fact be the abundance of benign variants. The mistake is made when we conflate niches. Only interchangeable members of a group—which may be a species or a population—have the same niche; yet identifying microbial populations meaningfully defined by a shared niche is hampered by a “maddening diversity” (Polz et al., 2013) caused by horizontal gene transfer (HGT). In contrast to vertical descent, HGT mixes DNA across phylogenetic lineages (Doolittle and Papke, 2006); an analysis of over 600 bacterial genomes showed 20 % of genes seem to have been recently acquired (Popa et al., 2011). By incorporating foreign DNA, taken up from the environment, infecting bacteriophage, or microbial neighbors capable of conjugation, microorganisms can shift their niche. Genomic islands (regions of variable gene content where HGT occurs) were initially called pathogenicity islands, because of the discovery that they correlated with pathogen fitness—i.e. virulence or antibiotic resistance (Dobrindt et al., 2004). In the environment, where the selection pressures may not be as severe as those in hospitals, HGT still shuffles genomic content, variegating what once might have been considered one species (Shapiro et al., 2012). Therefore, studies that have used simple phenotype markers (such as growth on particular culture media) or single gene markers to infer a population have likely conflated ecologically distinct groups.

To address the problem of identifying ecological niches specific to populations, microbial ecology has embraced the rapid technological advances in massively parallel sequencing, in tandem with its decreasing costs, to access information across genomes and community metagenomes (Metzker, 2010). The analysis of whole genomes has enabled a much more finely resolved view of the potential niche breadth of a microbe or its population than was possible even a decade ago. In 2006, Johnson et al. (2006) published their marker-gene based biogeographic mapping of *Prochlorococcus* ecotypes, revealing distinct distributions of genetic clades along light and temperature gradients across the oligotrophic Atlantic ocean. In 2014, Kashtan et al. (2014) published a far finer patterning of diversity in the same system: through analysis of single-cell genomes, they found hundreds of genomic subgroups within *Prochlorococcus*. Incredibly, they inferred that if each grouping is a species (or at least an ecologically distinct group), then thousands of *Prochlorococcus* species may coexist globally. What defines these subgroups in *Prochlorococcus* is the distribution of two gene pools: distinct alleles of core genes that are shared among all members—a ‘core’ genomic backbone—and genes shared by members of a group but variable between groups—the ‘flexible’ gene set, which is changing due to horizontal gene transfer, a pattern that could not be observed without genome-wide data and deep population sampling.

Analysis of gene flow thus reveals that HGT both erodes species boundaries, and gives species cohesion. Recently, a study of two populations of *Vibrio cyclitrophicus*, revealed them to be nascently speciating due to the acquisition of genes for colonizing a new habitat. More similar core alleles and flexible gene content within the populations demonstrated less mixing between the population gene pools, indicating that they are diverging. Species can also be inferred by the dynamics of gene flow in the Archaea (Cadillo-Quiroz et al., 2012), illustrating that the biological species concept—that species are defined by reproductive



isolation (Mayr, 1982)—spans all three domains of life. HGT, as well as the other mechanisms of molecular evolution, mutation and chromosomal rearrangements, render the microbial genome a dynamic entity.

## 1.2 How does the environment shape genome content?

Various niche components exert different selection pressures on the evolution of microbial genomes. Habitats have been observed to select for endemic traits, and HGT of ecologically relevant genes occurs more frequently between microbes within the same habitat than between different habitats. For example, in a study of isolates of *V. cholerae* and a sister-species, *V. metecus*, from the Atlantic and Indian oceans, Boucher et al. (2011) found more frequent transfer of genes in integrons—rapidly evolving mobile genetic elements—between isolates from the same ocean, yet different species, than between members of the same species in different oceans. Enriched in secondary metabolism and cell surface modification, these genes could be adaptive toward local environmental conditions or predation pressures. A separate study generalized these findings: of 2,235 bacterial genomes, isolated from diverse environments, including the human body, soil, and the oceans, the authors demonstrated that ecological similarity of habitats, rather than geography or phylogeny, structures gene flow via HGT (Smillie et al., 2011). An isolate from a person in Japan shares a greater number of recent transfers with an isolate from a person in the US than it does with an isolate from the Sea of Japan.

Within habitats, predation—for example, by bacteriophage and protozoa—exerts negative-frequency, or kill-the-(reproductive)-winner, selection, in which the most abundant member of the population is all the more easily gleaned. This seems to be true for marine microbes, who have a high diversity in their exposed or extracellular components (Rodriguez-Valera et al., 2009), which their predators use to recognize them. The strength of viral predation in propelling rapid evolution is exemplified in regions of clustered regularly interspaced palindromic repeats (CRISPR), which contain short sequences that confer an adaptive immunity to viral infection in microorganisms; no two individuals sampled from a natural population appear to be identical in CRISPR loci (Tyson and Banfield, 2008). Phage can also exert positive, diversifying selection, as prey evolves to evade predator (Williams, 2013).

Resource utilization also appears to exert diversifying selection. Fractionating the water column, Hunt et al. (2008a) found *Vibrio* populations have distinct distributions across size fractions—proxies for habitat-associations, like colonizing zooplankton ( $\geq 63 \mu\text{m}$  size fraction) or remaining free-living ( $0.2 \mu\text{m}$  to  $1 \mu\text{m}$  size fraction)—indicative of niche-partitioning. Moreover, these distributions were largely consistent when sampled three years later, demonstrating the niche associations to be stable (Szabo et al., 2013). In the study of flexible genome content, which differentiates members of a clade, researchers have found genes related to metabolism that could facilitate niche-partitioning; for instance,

SAR11’s flexible genome contains genes involved in phosphorous metabolism, glycolysis, and C1 metabolism (Grote et al., 2012). As with Darwin’s finches, where beak size and morphology enable specialization for different types of fruits and seeds (Lack, 1947), the key to coexistence may be specialization for resources within a habitat where competition is high (Grant and Grant, 2006). Microbes that metabolize sugars in the human gut, for example, are thought to specialize because of the subtle variations in structure of available polysaccharides (Martens et al., 2014).

Yet in a dynamic environment like the coastal ocean, governed by gradients and rapid turnover of nutrients (Azam and Malfatti, 2007; Stocker, 2012), microbial populations (such as populations in the *Vibrio*) can encode a functionally diverse genome suggestive of a generalist strategy. How could this diversity be maintained? A proposed mechanism starts from the idea that members of a metapopulation colonize distinct types of resource patches. Despite potential selection for adaptations within a patch, the population can experience stochastic bottlenecks that limit the number of returning individuals to the seed pool, hindering fixation of fitter genes or alleles (Fraser et al., 2009). In addition, enrichment of these selected genes among returning members could contribute to, and thereby diversify, the metapopulation gene pool via HGT.

Bacterioplankton populations exhibit distinct patch colonization dynamics, however, that could influence their evolution in ways not captured by a metapopulation model. For instance, nascent species of *Vibrio* exhibit different propensities toward long-term colonization, with one able to form denser biofilms, and the other better able to disperse and seek new patches—a fugitive strategy (Yawata et al., 2014). Perhaps with more sustained habitat associations, a population might tend toward functional specialization or genome streamlining. For example, *Pelagibacter ubique*, which abides in an a relatively stable environment—the nutrient poor, oligotrophic ocean—has the smallest genome of a freeliving bacterium known (Giovannoni et al., 2005). However, predominantly abundant resources could also support “imperfect generalists,” who generalize in resources when they are readily available but specialize in times or areas of deprivation (Barrett et al., 2005; De León et al., 2014).

Social interactions also influence genome content. Cordero et al. (2012a) found that *Vibrio* strains exhibit a genetic pattern with a fitness advantage: loss of genes to produce extracellular iron-scavenging molecules, siderophores, but retention of genes for siderophore receptors that transport them back into the cell. *Vibrio* strains with this genetic pattern can “cheat” their neighbors by capturing the siderophore public good, without suffering the cost of producing it. Furthermore, cheating is more likely in larger particle habitats than smaller ones, postulated to be because cellular, and therefore siderophore, densities are higher in these environments. This finding illustrates that spatial structure influences social dynamics, which in turn influence genome content.

### 1.3 Studying the niche with high-throughput tools that connect phenotype to genotype

While genomics lends us insight into the functional potential of the organism, genomes provide but a static picture, as only a third to half of genes are expressed at any given time (Passalacqua et al., 2009). Other high-throughput techniques complement genomics, and enable us to further explore the microbial niche. These techniques fall into two broad categories: those which enable a view of the intact physiology of a microorganism in response to an environment: transcriptomics, proteomics, and metabolomics; and that which systematically perturbs its physiology: mutant selection. Whereas the genome represents an organism’s fundamental niche, the functional readouts that these techniques produce, are facets of its realized niche. Here I briefly overview each, with some examples of their application and discussion of limitations.

**Transcriptomics** looks at the level of gene expression, the total set of mRNA transcripts, historically using RNA hybridization (microarrays), or now more commonly, next-generation sequencing (RNA-seq) to determine abundance. The technology approaches complete coverage of expressed regions of the genome (González-Ballester et al., 2010).

By analyzing differential expression of genes between conditions, researchers can infer the different physiological responses of an organism or clade. For example, a microcosm perturbation experiment revealed that *Prochlorococcus* and *Pelagibacter* are poised to respond to organic nitrogen sources, reducing expression of genes involved in nitrogen stress rapidly (Sharma et al., 2014).

Researchers have also frequently used transcriptomics to get a holistic view of ecosystem functioning, an approach called metatranscriptomics (DeLong, 2009). In one of the early studies of community expression in natural sea water, Poretsky et al. (2005) discovered transcripts that could be linked to biogeochemical processes, including sulfur oxidation (*soxA*), assimilation of C1 compounds (*fdh1B*), and nitrogen assimilation via polyamine degradation (*aphA*).

Furthermore, next generation sequencing has made transcriptomics more technically feasible. Whereas microarrays required microgram quantities of RNA, RNA-seq requires only nanograms (Giannoukos et al., 2012)—an important feature, since cellular RNA is only 1 to 5% transcripts (Neidhardt and Umbarger, 1996). Automation has also allowed researchers to apply RNA-seq to hundreds to thousands of single cells, to look at the heterogeneity in physiological responses within an assemblage (Saliba et al., 2014).

Transcriptomics has its challenges, however. Transcripts are not a universal proxy of actual protein abundance, hindering the interpretation of the actual cellular response (Schwanhäusser et al., 2011; Waldbauer et al., 2012). Moreover, because changes in a gene’s transcripts do not necessarily correlate with changes in the organism’s fitness when the gene is mutated, the efficacy of using transcriptomic data to infer genes required for certain conditions is unreliable (Deutschbauer et al., 2011; Turner et al., 2014). The short half-life

of RNA, some as brief as 30 seconds (Belasco and Brawerman, 2012), can also bias results.

**Proteomics** characterizes the pool of actually synthesized proteins, predominantly using tandem mass-spectrometry (MS/MS). At present, it appears routine to characterize 1,000 to 2,000 proteins per sample (Yang et al., 2015), which is not a complete proteome for most organisms.

Like transcriptomics, proteomics has lent insights into the microbial world. In the oceans, metaproteomics revealed the unexpected importance of the TonB-dependent transporters, which dominated cell membrane associated proteins (19%) (Morris et al., 2010). In the acid mine drainage environment, proteomics, together with sequenced genomes of the most abundant community members, demonstrated the importance of proteins involved in responding to oxidative stress and protein refolding (Ram et al., 2005).

Proteomics also has the potential to reveal post-translational modifications and protein-protein interactions, further uncovering how cells not only express, but regulate function (Schneider and Riedel, 2010).

Drawbacks of the approach are the difficulty in protein and peptide identification (Nesvizhskii, 2010), and a lack of sampling depth that can create large biases in the readout, with typically only proteins present at greater than 1% abundance detected (Verberkmoes et al., 2009). However, this problem hinders community analysis more than analysis of a single organism, which could feasibly be sampled more deeply.

**Metabolomics** assesses intermediates and products of metabolism, predominantly using MS/MS.

Argued to be the 'omics technology that analyzes the level most reflective of the physiology of a cell ('the ultimate phenotype' (Saito and Matsuda, 2010)), this technique allows insight into the cell's chemical and nutritional environment. Currently, metabolomics has the ability to characterize hundreds of molecules per sample, with a typical size range of 50–1500 Da.

This technique has provided insights into the ecology of single strains. For example, to study the effect of inorganic carbon limitation—sometimes a constraint in aquatic environments for photosynthetic organisms—the cyanobacterium *Synechocystis* sp. PCC 6803 was shifted from ambient to low carbon dioxide levels. In response, glutamine levels decreased, indicative of slowed nitrogen assimilation (Eisenhut et al., 2008). Interestingly, this shift was much more pronounced at the metabolite level than the transcript or protein level characterized in previous studies (Eisenhut et al., 2008). A second study, this time of metabolites consumed and released by a distinct cyanobacterium, *Synechococcus* sp. PCC 7002, revealed unexpected behavior; the organism imports metabolites that it also produces, suggesting that it may take advantage of lysed members of its local population, by cannibalizing their constituents (Baran et al., 2011).

Metabolomics can also be used to study the metabolic fluxes of substrates with labelled isotopes,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$ . Recently, this technique showed how much of cellular nutrients viral infection redirects into virion particles:  $\sim 75\%$  in *Sulfitobacter* sp. 2047 (Ankrah et al., 2014).

Major challenges in the use of this technology parallel those of proteomics; it is difficult to evaluate the data (i.e. identify and quantify metabolites through MS spectra) (Schwarz et al., 2013) and it has low sampling depth relative to transcriptomics or even proteomics (Saito and Matsuda, 2010). However, as with proteomics, sampling depth is, at least, less prohibitive for studies of single organisms.

Used in combination, these high-throughput techniques can provide a detailed inspection of microbial physiological responses and adaptations. In a study combining transcriptomics, proteomics, and biochemical assays of carbohydrate utilization, for example, the authors found that, in a simplified model of the human gut (a gnotobiotic mouse), symbiotic bacteria facilitate niche creation, and adjust their niche in the presence of other phyla (Mahowald et al., 2009).

**Mutant selections**, which have often been used to determine what makes a pathogen virulent, have the ability to query the effect of single-gene disruptions on an organism's fitness. This approach has enabled us to learn which genes are important in particular processes or conditions, determined by the choice of selective environment. In turn, mutant selection has led to much of our understanding of gene function.

By combining unbiased transposon insertion with massively parallel sequencing of insertion sites, researchers can generate complex mutant libraries (with virtually every mutagenizeable site disrupted) and assay their fitness simultaneously (Barquist et al., 2013). This technique allows the determination of what genes are essential, i.e. those that cannot be mutagenized, and those that only affect the organism's fitness in certain conditions. By using the transposon to determine the sequence of the adjacent genomic DNA, each transposon mutant can be mapped and enumerated using next-generation sequencing (Tn-seq) (van Opijnen and Camilli, 2013). The technique has been applied to the elucidation of genes and pathways needed to colonize the mammalian gut to determine both virulence factors (Gawronski et al., 2009; Fu et al., 2013) as well as pathways non-pathogenic symbionts use for persistence (Goodman et al., 2009). Goodman et al. (2009) were the first not only to vary the host habitat (comparing wild-type and immunodeficient gnotobiotic mice) to study the effect on fitness, but also to test the influence of a mixed community of commensals that better reflects the native environment.

A limitation of the technique is that, for mutants disrupted in only one locus, multicopy genes or redundant pathways can buffer mutant fitness effects, making it difficult in these cases to infer if an organism still relies on the functions encoded in such genes. However, transcriptomics, proteomics, and metabolomics can complement this limitation.

## 1.4 Aims of this thesis

In this thesis, I explore the niche of members of the *Vibrionaceae*, asking (i) at the broad scale, can bulk environmental variables be used as predictors of *Vibrionaceae* species abundances (chapter 2), and (ii) at the fine-scale, to what extent do distinct habitat

resources strongly select for metabolic pathways in a single organism (chapter 3)? In a preliminary investigation, I also ask, (iii) what cell envelope features do environmental bacteriophage target for *Vibrio* infection (appendix A)?

The *Vibrionaceae* family is an ideal experimental model system for exploring the microbial niche concept, in that it is culturable, genetically tractable, and ecologically diverse. It encompasses potential human pathogens, including *Vibrio cholerae*, the causative agent of the diarrheal disease cholera, and *Vibrio parahaemolyticus*, whose infections cause a mortality rate of 25%, and *Allivibrio fischeri*, the organism that symbiotically colonizes the light organ of the bobtail squid, and bioluminesces. One of the most beautiful biological phenomena, bioluminescence led to the discovery of the microbial density phenomenon of quorum sensing. Diverse *Vibrio* species are also a broadly distributed component of coastal marine bacterioplankton communities, including variants with different trophic strategies and habitat preferences.

In an effort to build predictive models of *Vibrio* abundance, either of the entire genus or of its potential pathogens, numerous studies (as reviewed in chapter 2) have assayed cellular abundance and measured environmental variables, such as temperature, salinity, nitrogen, and phosphate. Aside from temperature and salinity, correlations tend to be weak, limiting our mechanistic understanding of the system and predictive capabilities. These observations underscore the need for fine-scale environmental and genetic experimental studies.

Over the past decade, the Polz lab has obtained thousands of *Vibrio* isolates, and sequenced hundreds of genomes, to better understand their population structure and how the environment shapes it. Using the statistical modeling algorithm, AdaptML (Hunt et al., 2008a), ecological populations have been predicted within the *Vibrio* that are unified by shared habitat distributions on distinct marine particle types collected from the coastal ocean: detrital algal particles and zooplankton, living and dead (Preheim, 2010) (Section 1.4). Of particular interest is one population which occupies both habitats: *Vibrio* sp. F13. An apparent generalist, skewed in its habitat distribution toward macroalgal-derived fragments, this population has also been isolated from live brown macroalga, *Fucus vesiculosus*. Yet, *Vibrio* sp. F13 can grow on substrates characteristic of a zooplankton environment, including several amino acids and chitin, the primary constituent of crustacean exoskeletons. What are the pathways this organism relies on for growth in the different habitats it may occupy?

To experimentally interrogate habitat adaptation, I use a mutant selection approach. Because of its high mutagenizability relative to other members of the population tested, I chose strain *Vibrio* sp. F13 strain 9CS106 (Figure 1-2) to construct a mutant library. This library was then selected on nutrients derived from contrasting model habitats, the copepod *Apocyclops royi* and *Fucus vesiculosus*, as well as the algal constituent polysaccharide, alginate.

Finally, to determine the markers phage recognize to infect a *Vibrio* host, the 9CS106 mutant library was selected against two environmentally isolated phage.

This study provides an experimental investigation of a microbial niche.

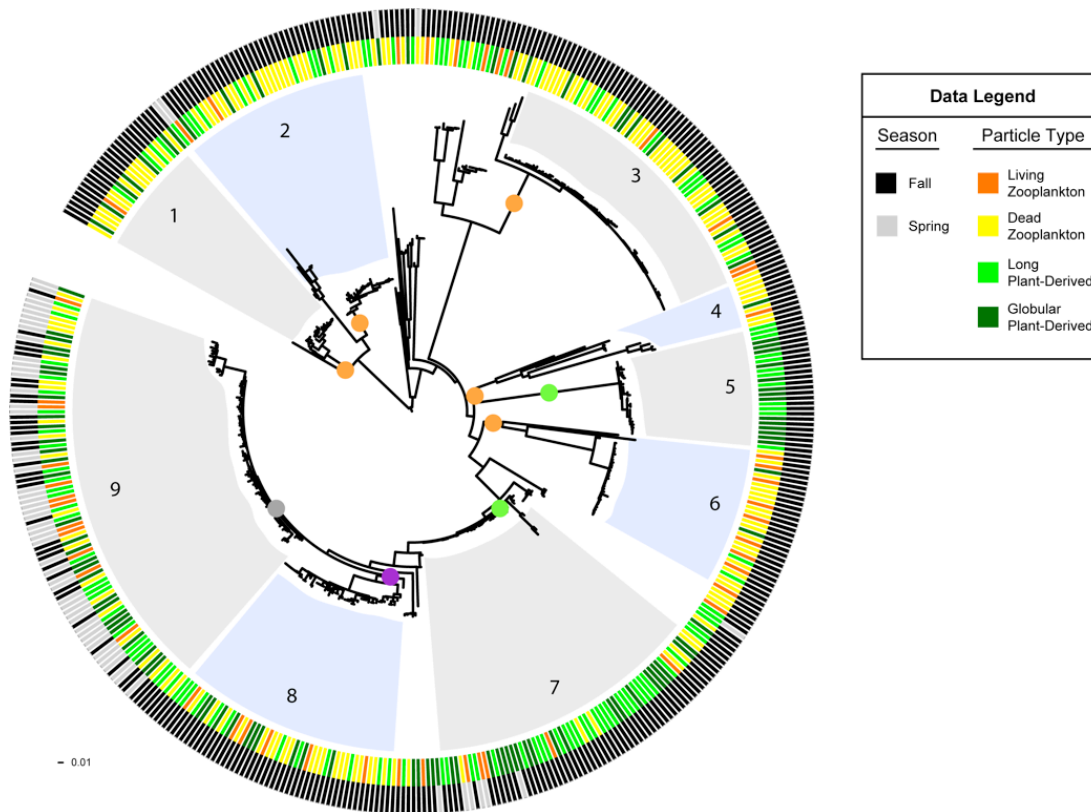


Figure 1-1: **Populations associated with zooplankton and vegetation particles**, from (Preheim, 2010). Maximum likelihood phylogenetic tree derived from concatenation of partial *adk*, *mdh* and *hsp60* sequences for all strains (clones represented by one concatenated sequence) isolated from detrital algal (previously described as plant-derived) and zooplankton particles with rings designating season (outer ring) and particle type (inner ring) of origin for each strain. Circles at the nodes of each population are colored according to the predicted habitat preference as determined by the AdaptML algorithm (Hunt et al., 2008a). Populations are alternatively shaded blue and gray and numbered corresponding to species as follows: 1, *Aliivibrio* sp.13; 2, *Aliivibriofischeri*; 3, *Enterovibriocalviensis*-like; 4, *Enterovibronorvegicus*; 5, *V. breoganii*; 6, *Vibrio* sp. F10; 7, *Vibrio* sp. F13; 8, *V. tasmaniensis*/*V. lentus*; 9, *V. splendidus*.

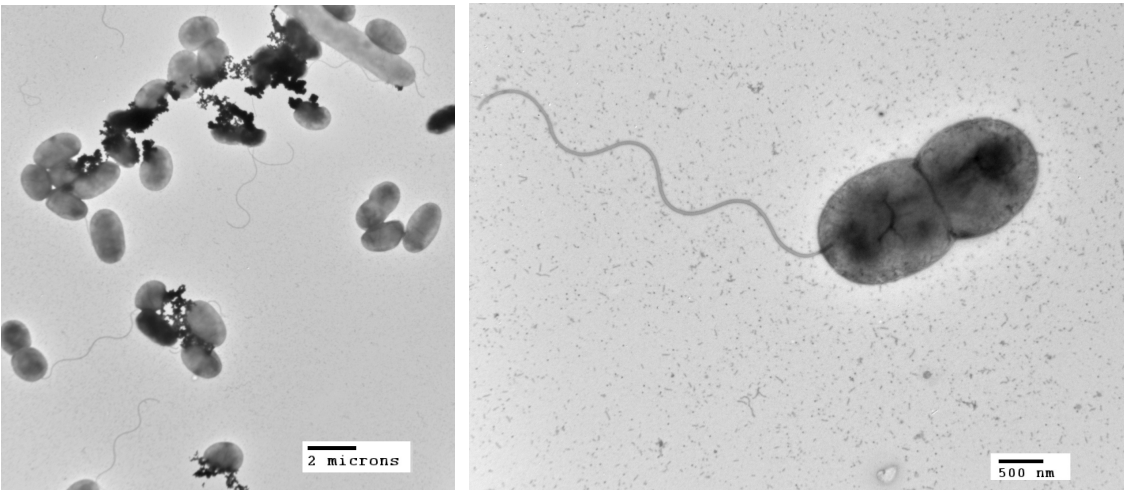


Figure 1-2: *Vibrio* sp. F13 strain 9CS106. Electron micrographs courtesy of Fatima Hussain and Kathryn Kauffman.



## Chapter 2

# Associations and dynamics of *Vibrionaceae* in the environment, from the genus to the population level

Alison F. Takemura, Diana M. Chien, and Martin F. Polz  
Published in *Front Microbiol.* 2014; 5: 38.

### 2.1 Abstract

The *Vibrionaceae*, which encompasses several potential pathogens, including *V. cholerae*, the causative agent of cholera, and *V. vulnificus*, the deadliest seafood-borne pathogen, are a well-studied family of marine bacteria that thrive in a diverse habitats. To elucidate the environmental conditions under which vibrios proliferate, numerous studies have examined correlations with bulk environmental variables—e.g., temperature, salinity, nitrogen and phosphate—and association with potential host organisms. However, how meaningful these environmental associations are remains unclear because data are fragmented across studies with variable sampling and analysis methods. Here, we synthesize findings about *Vibrio* correlations and physical associations using a framework of increasingly fine environmental and taxonomic scales, to better understand their dynamics in the wild. We first conduct a meta-analysis to determine trends with respect to bulk water environmental variables, and find that while temperature and salinity are generally strongly predictive correlates, other parameters are inconsistent and overall patterns depend on taxonomic resolution. Based on the hypothesis that dynamics may better correlate with more narrowly defined niches, we review evidence for specific association with plants, algae, zooplankton, and animals. We find that *Vibrio* are attached to many organisms, though evidence for enrichment compared

to the water column is often lacking. Additionally, contrary to the notion that they flourish predominantly while attached, *Vibrio* can have, at least temporarily, a free-living lifestyle and even engage in massive blooms. Fine-scale sampling from the water column has enabled identification of such lifestyle preferences for ecologically cohesive populations, and future efforts will benefit from similar analysis at fine genetic and environmental sampling scales to describe the conditions, habitats, and resources shaping *Vibrio* dynamics.

## 2.2 Introduction

The family *Vibrionaceae* (or vibrios for short) comprises a genetically and metabolically diverse group of heterotrophic bacteria that are routinely found in all ocean environments, ranging from coastal to open and surface to deep water (Thompson et al., 2004; Thompson and Polz, 2006). Moreover, a few *Vibrio* species have extended their range beyond the marine environment, occurring predominantly in brackish and even freshwater environments (Thompson et al., 2004). The study of the environmental distribution and dynamics of vibrios has a long history, largely because many species contain potential human and animal pathogens (Thompson et al., 2004, 2005). Hence there is considerable public health and economic interest in determining factors correlated to increased abundance of vibrios (Stewart et al., 2008). Moreover, vibrios are easily cultured on standard and selective media and thus were highly visible in the pre-molecular era of microbial ecology. In recent years, environmental dynamics have also been studied with culture-independent methods allowing for a more fine-scale assessment of environmental drivers of occurrence, and the vibrios have become a model for bacterial population biology and genomics. In fact, presently, the vibrios represent one of the best-studied models for the ecology and evolution of bacterial populations in the wild.

The early discovery that some fish species harbor high numbers of vibrios (e.g., Liston (1954, 1957); Aiso et al. (1968); Sera et al. (1972)) has led to the widespread notion that these bacteria are only transient members of microbial assemblages of the water column. Instead, vibrios were regarded as specifically associated with animals, and occurrence in water samples was thought to be primarily due to their excretion with fecal matter. This picture was enforced by the discovery that several luminescent *Vibrio* (*Allivibrio*) and related *Photobacterium* species form intimate symbioses with animals (e.g., fish, squid) (Ruby and Nealson, 1976; Stabb, 2006). More recent work has, however, revealed that the notion of vibrios being ‘enterics of the sea’ (Liston, 1954) represents an oversimplification. Many *Vibrio* species grow actively in ocean water either in the free-living phase or associated with various types of organic particles, many of which are of non-animal origin (Lyons et al., 2007; Froelich et al., 2012). Thus although association with animals can be an important part of the life cycle of many *Vibrio* species, there are others that only loosely associate with animals or not at all, an aspect we explore in detail in this review.

Another widely held belief about vibrios is that they play a relatively minor role in

chemical transformations in the ocean, despite the wide range of metabolisms [e.g., chitin degradation (Hunt et al., 2008a; Grimes et al., 2009)] of which they are capable. This belief is largely based on low to medium average relative abundance of *Vibrionaceae* in ocean water. Yet three considerations suggest that the role of vibrios has been underestimated. First, it has been pointed out that although vibrios' abundances are generally only around  $10^3$  to  $10^4$  cells per mL seawater (i.e., on the order of few percent of total bacteria), they have very high biomass (Yooseph et al., 2010). For example, an actively growing *Vibrio* can have  $100\times$  the biomass of *Pelagibacter*, which, at  $\sim 10^5$  cells per mL, is typically the most abundant heterotrophic member of bacterial assemblages in the ocean (Yooseph et al., 2010). Second, new time-series analysis shows that vibrios are capable of blooms in the water column during which they can even become the predominant members of the total bacterial assemblage (Gilbert et al., 2012). These blooms had been missed previously because they are of relatively short duration, yet they confirm that vibrios, which are capable of very rapid growth in laboratory media, can reach high doubling rates in the environment. Finally, vibrios might be disproportionately subject to predation by protozoa and viruses (Suttle, 2007; Worden et al., 2006), likely due to their comparatively large size. For example, cells were found in one study to measure more than three times the community average in volume, and, along with other similarly large genera, suffered especially high grazing mortality (Beardsley et al., 2003). Taken together, these considerations suggest that vibrios should be re-evaluated for their role in biogeochemical processes in the ocean since they have disproportionately high biomass that is subject to high turnover by rapid growth in concert with high predation.

The purpose of this review is to provide an overview of known environmental factors and ecological associations affecting *Vibrio* abundance and dynamics. We note that although we look at the dynamics of potentially pathogenic species, we purposefully exclude data on pathogenesis itself since this is outside the scope of this review. We first focus on total *Vibrio* (i.e., the assessment of occurrence of members of the genus or family), which have often been measured as a proxy for potential pathogen occurrence, asking whether they can be treated as an environmentally cohesive unit. To what extent do total vibrios correlate to specific environmental variables, and do these measures have predictive power for individual species? To address this question, we present meta-analyses of the dynamics of *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*, three species harboring genotypes potentially pathogenic to humans. The limitation to these three is necessary since public health interests have driven much of the research so that the literature is highly biased towards human pathogens. In this context, a further important question is to what extent easily measurable bulk parameters, such as temperature, salinity, nutrients, dissolved oxygen and/or chlorophyll a are good correlates for total vibrios or specific species, allowing easy and cost-effective risk assessment.

However, because our meta-analysis suggests poor or inconsistent performance of most bulk parameters, we researched alternative, frequently finer-scale environmental variables. These include associations with different animals, plants and algae, as well as organic

polymers, which may occur as suspended particulate matter in the water column and provide resources for attached bacteria. Although such attached lifestyles are common for vibrios, recent research also suggests that many species can occur free-living at least part of the time and be engaged in relative short-lived blooms.

Finally, we summarize recent research aimed at defining habitat characteristics and phylogenetic bounds of ecologically cohesive populations among co-existing vibrios, using the water column and macroinvertebrates as examples of adaptive landscapes. This research demonstrates that such populations, which may or may not correspond to named (taxonomic) species, represent eco-evolutionary units that allow testing of hypotheses of how populations are structured by environmental selection and gene flow.

## 2.3 Environmental correlates of *Vibrio* presence and abundance

To better understand under what conditions vibrios occur and proliferate, most studies have investigated environmental variables that can be measured from bulk seawater such as temperature, salinity, dissolved oxygen, nitrogen, phosphorus, and chlorophyll a concentrations. These are attractive since they are easily measured and many are observable remotely by buoy or satellite (e.g., Lobitz et al. (2000)) so that potential for presence of pathogenic vibrios might be easily assessed. In addition, several studies have extended measurements to more complex physicochemical and biotic variables, including dissolved organic carbon and zoo- and phyto-plankton taxa.

In the following, we first ask how informative these variables are by conducting a meta-analysis to compare correlations across studies, for both total *Vibrio* as well as the potential pathogens *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*, and, second, determine if the genus and species levels exhibit similar patterns. To determine the potential impact of environmental variables, we looked at how strong their correlations are by comparing coefficient of determination values,  $R^2$ , reported in the literature. A goodness of fit parameter,  $R^2$  varies from 0 (no explanation of variance in the dependent variable) to 1 (perfect explanation), giving us a means of assessing, for example, whether temperature better predicts abundance of total *Vibrio*, than salinity does. Studies included have regression analyses with associated  $R^2$  values, or Spearman or Pearson correlations, whose rho values were squared to obtain  $R^2$ . Additionally, we compare how their abundances trend along gradients in two particularly well-studied variables, salinity and temperature.

### 2.3.1 Total *Vibrio*

When correlations across studies are compared, we see that the strongest environmental correlates to total *Vibrio* are temperature and salinity. These two variables most often explain the greatest amount of variance in total *Vibrio* abundance in the water column

(**Figure 2-2**), whereas consideration of additional variables often makes only marginal improvements (e.g., in Heidelberg et al. (2002a,b); Oberbeckmann et al. (2012); Froelich et al. (2013)). However, a minority of analyses has found temperature and salinity to be non-significant toward explaining *Vibrio* abundance. This inconsistency might be a result of the ranges considered; for instance, temperature may be found non-significant due to a narrow range observed, such that *Vibrio* abundance varies little. In fact, evidence supports this hypothesis; the correlation strength of temperature to vibrios varies by season (Oberbeckmann et al., 2012; Froelich et al., 2013), suggesting the magnitude of the correlation may depend on the temperature range examined. For instance, Oberbeckmann et al. (2012) and Froelich et al. (2013) both observed the highest correlation of temperature and *Vibrio* during the seasons with the broadest temperature ranges, spring and fall, respectively. Additionally, it is possible that at lower temperatures vibrios exhibit less variation in abundance; two studies assessing total vibrios in the cooler waters of the Baltic Sea and North Sea found non-significant correlations (Eiler et al., 2006; Oberbeckmann et al., 2012).

Compared to salinity and temperature, other environmental measures usually explain less variance in total *Vibrio*. Dissolved oxygen has had little explanatory power; for instance, in Figure 2-1, its largest  $R^2$  was less than half that of temperature in the same analysis (Blackwell and Oliver, 2008). The same is true for nitrogen, whose highest  $R^2$  was still less than temperature's (Blackwell and Oliver, 2008). In the environments examined, phosphate, pH, and turbidity explain little variance, and dissolved organic carbon (DOC) explains none at all, albeit the number of studies used for DOC in this meta-analysis is limited. Of interest, though not depicted, potential host organisms, copepods, decapods, and cyanobacteria, have been found to explain relatively little variance in total vibrios when considered in a model that already incorporates temperature (Turner et al., 2009; Vezzulli et al., 2009), and similarly for dinoflagellates when salinity is first considered (Eiler et al., 2006). Turner et al. (2009) did observe that diatoms explained more variance than temperature. While this might imply a physical association, the correlation was negative, suggesting that total *Vibrio*, at least as a whole, do not associate with diatoms.

Chlorophyll a, on the other hand, has had noted importance in two datasets: the spring and summer of the study by (Oberbeckmann et al., 2012), with  $R^2$  values of 60 % and 26 %, respectively. These were in fact higher than correlations to temperature or salinity in these seasons. Perhaps during this period, as temperature warms, growth conditions favor phytoplankton blooms that impact *Vibrio* abundance (Oberbeckmann et al., 2012). However, Froelich et al. (2013) did not make these same observations in their seasonal datasets. This inconsistency may be a product of the fact that different *Vibrio* species likely affiliate with or feed on exudates of specific algal taxa only, rather than algae in general, a subject further discussed in Section 2.6.2.

Given the frequent strength of temperature and salinity as correlates, we asked, how do total vibrios distribute with respect to these variables when their combined effect is considered? A few studies have modeled the bivariate relationship, finding that total *Vibrio*

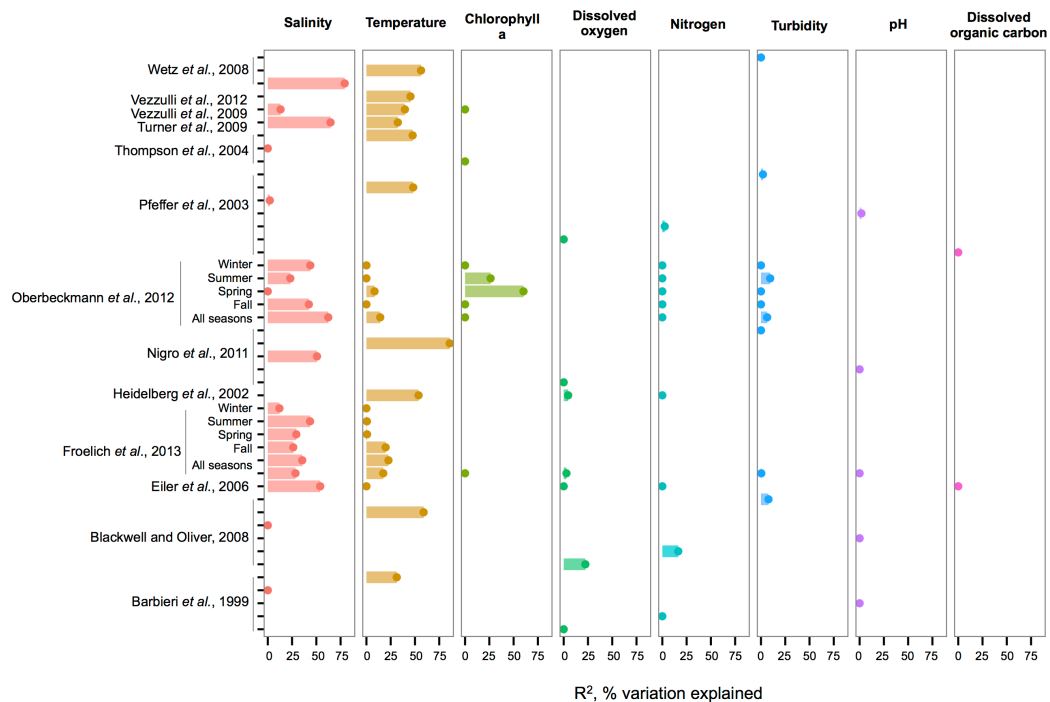


Figure 2-1: An overview of regression analyses indicate that temperature and salinity explain most variation in bulk-water total *Vibrio* abundance. The  $R^2$ , or pseudo- $R^2$ , values associated with regression analyses are shown for selected environmental variables that are well-represented across studies. An individual study may perform multiple analyses because variables are considered for correlation independently (for ex. (Wetz et al., 2008)); because datasets are split (e.g. between seasons in Oberbeckmann et al. (2012)); or because different sets of variables are considered sequentially (e.g. two variables versus six variables in the two All Seasons models from Froelich et al. (2013)). Dots indicate bar heights, and where a dot occurs without a bar,  $R^2$  was non-significant (i.e.  $R^2 = 0$ ). Variables may have been log or exponentially transformed in references.

abundance increases as temperature and salinity increase (Hsieh et al., 2008; Turner et al., 2009; Froelich et al., 2013). The ranges investigated were also broad, lending confidence that these results are general; for example, Hsieh et al. (2008) modeled from 2.5 to 32.5 °C and 0 to 27 ppt, respectively.

### 2.3.2 *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*

We compare environmental correlates and trends noted in total *Vibrio* to three species that have been well sampled across locales: *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*. While it would also be interesting to consider species beyond potential pathogens, their environmental data is much more limited.

In *V. cholerae*, we see an interesting shift from total *Vibrio* in the strength of correlating environmental variables: some biotic variables are as strong or, in fact, stronger than temperature or salinity (**Figure 2-2**). Total *Vibrio*, congeners *V. vulnificus* and *V. parahaemolyticus*, as well as a dinoflagellate genus (*Prorocentrum*) and cladoceran species (*Diaphanosoma mongolianum*) have all significantly correlated to *V. cholerae* abundance (Eiler et al., 2006; Blackwell and Oliver, 2008; Kirschner et al., 2011; Prasanthan et al., 2011). Moreover, *V. parahaemolyticus* abundance has explained more *V. cholerae* abundance variance than nitrogen, temperature, or salinity in (Prasanthan et al., 2011), and dinoflagellate abundance has explained more variance than phosphorus, salinity, or temperature (Eiler et al., 2006). While correlations to plankton may represent direct associations, such high correlation of vibrios to each other is likely not indicative of causal interactions, but rather stems from overlap in environmental ranges and/or habitats (Blackwell and Oliver, 2008). *E. coli* and total coliforms have also correlated to *V. cholerae* abundance, though both groups may simply be responding to anthropogenic nutrient influxes favoring growth of heterotrophs (Blackwell and Oliver, 2008).

Long thought to be a reservoir of toxigenic *V. cholerae*, zooplankton, and particularly copepods, are hypothesized to correlate to *V. cholerae* abundance. Surprisingly, however, when Magny et al. (2011) examined several zooplankton genera and species, including copepods *Cyclops* and *Diaptomus*, they did not find significant correlations to any zooplankton except the rotifer *Brachionus angularis* (not depicted in Figure 2-2, because Monte Carlo analysis did not yield  $R^2$  values). While the association between *V. cholerae* O1/O139 and the copepod *Acartia tonsa* has also been studied (Huq et al., 2005; Lizárraga-Partida et al., 2009), quantitatively significant correlation in the environment has remained elusive. For instance, Lizárraga-Partida et al. (2009) demonstrated only a qualitative link between *V. cholerae* O1 presence coincident with an increase in *A. tonsa*, even though laboratory studies have shown ready attachment (e.g., Huq et al. (1984); Rawlings et al. (2007)).

*V. cholerae* has also been hypothesized to correlate with chlorophyll a, a potential proxy of algal and zooplankton growth, and/or a eutrophic environment conducive to heterotroph growth, but chlorophyll a's general predictive value is unclear. While significant in Eiler et al. (2006), other studies have observed no correlation of chlorophyll a to *V. cholerae*

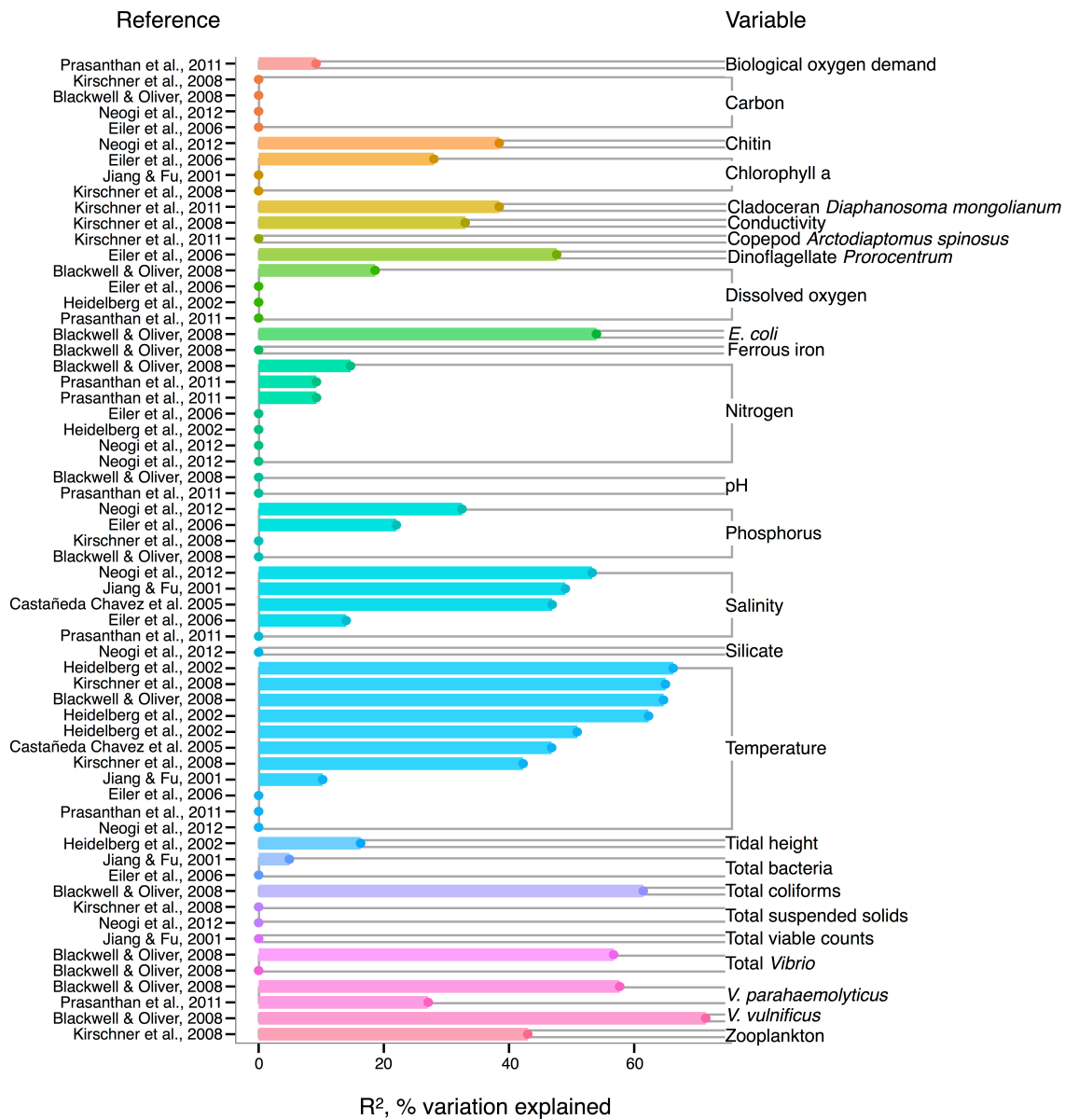


Figure 2-2: Variation in *V. cholerae* abundance or percent positive samples is best explained by temperature, other organisms, and salinity.  $R^2$ , or pseudo- $R^2$ , values from analyses across studies are depicted grouped by variable, and then in rank order, with their associated reference. A reference may conduct multiple analyses for a given variable (e.g., on subsets of data or considering different variables combinations for data regression). Dots indicate bar heights, and where a dot occurs without a bar,  $R^2$  was non-significant (i.e.  $R^2 = 0$ ).



abundance (Jiang and Fu, 2001; Kirschner et al., 2008; Mishra et al., 2012). Yet *V. cholerae* growth has been observed experimentally to depend on dissolved organic carbon, which could relate to phytoplankton abundance and thus chlorophyll a (Eiler et al., 2007). In microcosm experiments, Eiler et al. (2007) demonstrated that adding 2.1 mg carbon L<sup>-1</sup> of cyanobacterial-derived dissolved organic matter influenced bacterial growth more than a 12–25 °C change in temperature. The inconsistency of chlorophyll a, and, incidentally, bulk DOC (which showed no significant correlation) (Eiler et al., 2006; Blackwell and Oliver, 2008; Kirschner et al., 2008; Neogi et al., 2012) as correlates might be due to the quality of exudates; its composition of refractory humic substances (Kirschner et al., 2008) or derivation from different algal species, differentially stimulating *V. cholerae* growth ((Worden et al., 2006), see also Section 2.6.2). Interestingly, the lack of clear support for chlorophyll a's influence on *V. cholerae* environmental abundance is in contrast to the fact that chlorophyll a can correlate with cholera *disease* incidence (Magny et al., 2008), and has been used in predictive models for cholera in Bangladesh (Bertuzzo et al., 2012; Jutla et al., 2013).

Like *V. cholerae*, *V. parahaemolyticus* abundance in water samples is also strongly correlated to temperature, and was found significant in all but one analysis reviewed here (DePaola et al., 1990; Zimmerman et al., 2007; Blackwell and Oliver, 2008; Caburlotto et al., 2010; Julie et al., 2010; Johnson et al., 2010, 2012; Böer et al., 2013), with maximal  $R^2 = 50.6\%$  (Julie et al., 2010) (**Figure 2-3**). Blackwell and Oliver (2008) found that *V. parahaemolyticus* correlates both to total *Vibrio* and congeners, as well as coliforms and *E. coli*. These variables were only considered in a single study, however, so it is not known if the relationships hold across different sampling locations. The significance of salinity is variable for *V. parahaemolyticus* with only three of seven studies having non-zero  $R^2$  values (Figure 2-5) (Zimmerman et al., 2007; Caburlotto et al., 2010; Johnson et al., 2010), but this may be due to *V. parahaemolyticus* colonizing a large salinity range, as detailed below (Figure 2-4).

Correlation to environmental variables has also frequently been studied for *V. parahaemolyticus* occurring in sediment and shellfish, though trends remain unclear. In sediment, considered a potential reservoir (Vezzulli et al., 2009), individual regressions of *V. parahaemolyticus* abundance to temperature, salinity, and total organic carbon have yielded moderate  $R^2$  values, at times above 30% (Blackwell and Oliver, 2008; Julie et al., 2010; Johnson et al., 2012; Böer et al., 2013). However, some studies have found salinity or temperature to be a non-significant explanatory variable (Blackwell and Oliver, 2008; Julie et al., 2010; Johnson et al., 2010).

In shellfish, a common vehicle of virulent vibrios to humans, the incidence of temperature and salinity as correlates to *V. parahaemolyticus* is also inconsistent. Salinity has been found explanatory in some studies, with  $R^2$  as high as 42% (DePaola et al., 2003; Johnson et al., 2010, 2012) and non-significant in others (Deepanjali et al., 2005; Julie et al., 2010; Sobrinho et al., 2010). Temperature can explain moderate amounts of variance in *V. parahaemolyticus* abundance (DePaola et al., 1990, 2003; Cook et al., 2002; Johnson et al.,

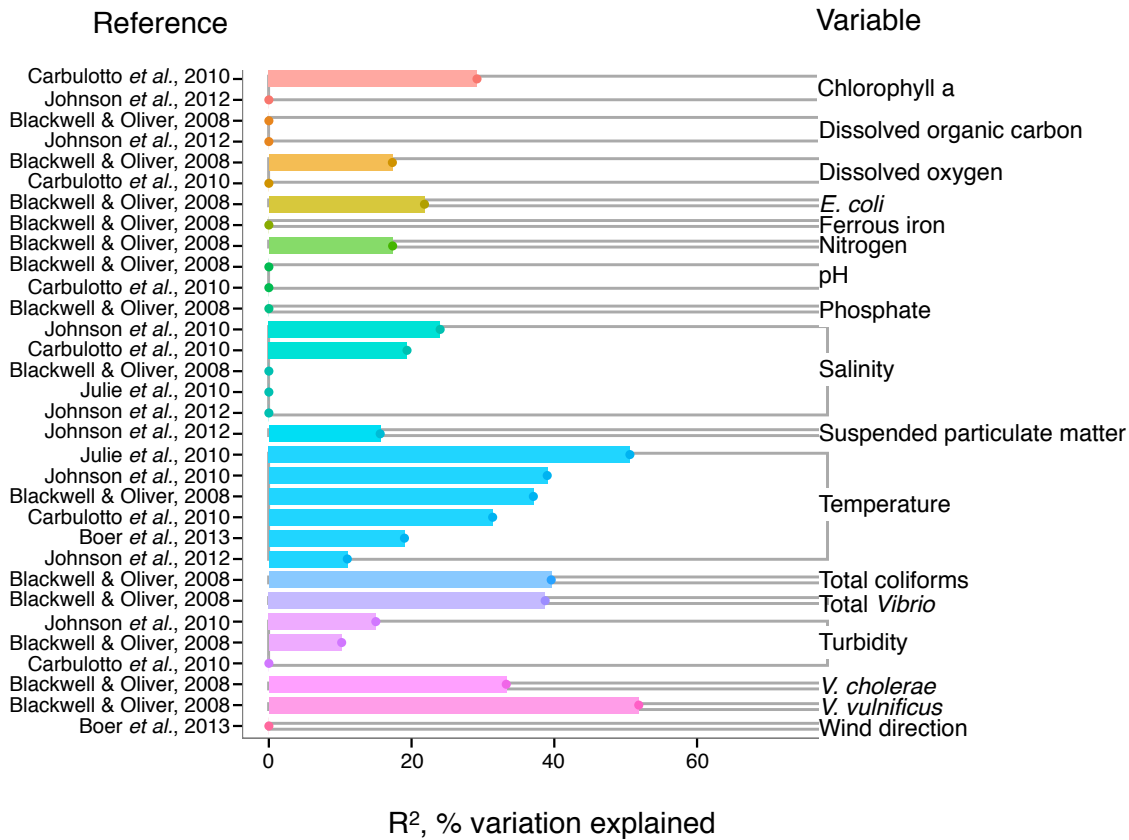


Figure 2-3: Variation in *V. parahaemolyticus* abundance or percent positive samples is best explained by temperature and other organisms.  $R^2$ , or pseudo- $R^2$ , values from analyses across studies are depicted grouped by variable, and then in rank order, with their associated reference. A reference may conduct multiple analyses for a given variable (e.g., on subsets of data or considering different variables combinations for data regression). Dots indicate bar heights, and where a dot occurs without a bar,  $R^2$  was non-significant (i.e.  $R^2 = 0$ ).

2010; Sobrinho et al., 2010; Johnson et al., 2012), with significant  $R^2$  as high as 44% (Cook et al., 2002), though other studies have found little or no correlation (Deepanjali et al., 2005; Duan and Su, 2005; Julie et al., 2010). The absence of correlation is surprising, given that temperature's effect is amplified by influencing shellfish's ability to concentrate *V. parahaemolyticus* from surrounding water. Oysters can enrich *V. parahaemolyticus* over 100-fold (DePaola et al., 1990; Shen et al., 2009), and the magnitude of concentration is temperature-dependent, with effects greatest at 32 °C and less, but still evident, in cooler waters (Shen et al., 2009).

For *V. vulnificus* isolated from the water column, temperature is the strongest correlate among measured environmental variables, and often explains more variance in *V. vulnificus* than for other species or total *Vibrio*; several analyses found temperature explained over 50% of the variance in *V. vulnificus* sampled from water (Motes et al., 1998; Randa et al., 2004; Blackwell and Oliver, 2008; Nigro et al., 2011) (**Figure 2-4**). Moreover, temperature has been a stronger correlate than chlorophyll a (Randa et al., 2004; Johnson et al., 2010, 2012), dissolved oxygen (Pfeffer et al., 2003; Blackwell and Oliver, 2008; Ramirez et al., 2009), and nitrogen (Pfeffer et al., 2003; Blackwell and Oliver, 2008). While DOC is an inconsistent correlate, it has been more explanatory than temperature in at least one study (Jones and Summer-Brason, 1998). The variable pH, however, is not a significant correlate (Lipp et al., 2001; Pfeffer et al., 2003; Blackwell and Oliver, 2008; Ramirez et al., 2009; Franco et al., 2012), nor is phosphorus (Pfeffer et al., 2003; Blackwell and Oliver, 2008). Turbidity has been found non-significant in several studies (Lipp et al., 2001; Pfeffer et al., 2003; Wetz et al., 2008; Ramirez et al., 2009), or not as explanatory as temperature (Blackwell and Oliver, 2008). While salinity, when significant, has generally been less informative than temperature (Motes et al., 1998; Randa et al., 2004; Warner and Oliver, 2008; Johnson et al., 2010), it has, in one analysis, been more (Lipp et al., 2001).

Biotic correlates have also been identified for *V. vulnificus*. Total bacteria (Pfeffer et al., 2003; Randa et al., 2004; Blackwell and Oliver, 2008), enterococcus (Wetz et al., 2008; Ramirez et al., 2009), coliforms (Pfeffer et al., 2003; Blackwell and Oliver, 2008) and *E. coli* (Pfeffer et al., 2003; Blackwell and Oliver, 2008; Wetz et al., 2008) have been studied only sporadically, but their correlation strength to *V. vulnificus* has usually been less than temperature's; one exception, however, is enterococcus in (Ramirez et al., 2009), potentially indicative of a surge in nutrients overtaking temperature's effect on growth. Interestingly, total *Vibrio* have explained substantial variance ( $R^2 = 43\text{--}54\%$ ) in *V. vulnificus* in more instances than for other *Vibrio* species (Pfeffer et al., 2003; Wetz et al., 2008; Blackwell and Oliver, 2008), suggesting they are responding similarly to their environments under the conditions studied. However, instances do occur where total *Vibrio* and *V. vulnificus* do not correlate (Høi et al., 1998; Wetz et al., 2008), underscoring that a species is not a constant component of a genus, and may respond to environmental conditions independently.

Isolations of the three potentially pathogenic species across salinity and temperature gradients were also looked at, and found to exhibit different patterns. *V. cholerae* has a wide temperature range ( $\sim 10\text{--}30$  °C) in brackish water (1–10 ppt), and generally decreases

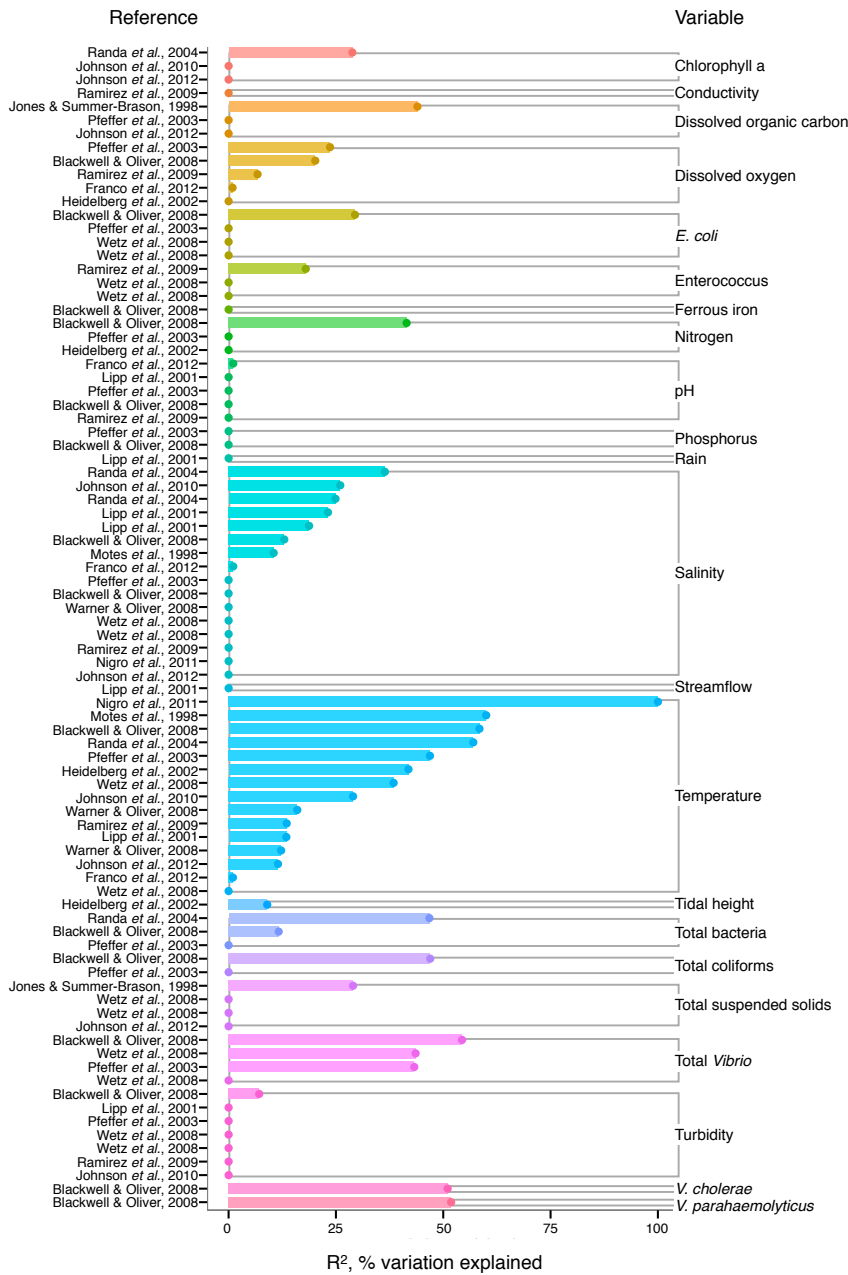


Figure 2-4: Variation in *V. vulnificus* abundance or percent positive samples is best explained by temperature, and other organisms, including *Vibrio*.  $R^2$ , or pseudo- $R^2$ , values from analyses across studies are depicted grouped by variable, and then in rank order, with their associated reference. A reference may conduct multiple analyses for a given variable (e.g., on subsets of data or considering different variables combinations for data regression). Dots indicate bar heights, and where a dot occurs without a bar,  $R^2$  was non-significant (i.e.  $R^2 = 0$ ). 38

with increasing salinity over the entire range examined (0–40 ppt) (**Figure 2-5**). Observed *V. cholerae* abundance is greatest around 20 °C and 0–10 ppt, on the order of 10<sup>3</sup> cells per mL. At less-favorable, higher salinities, *V. cholerae* has been found around this temperature, though in much lower abundances (on the order of 1 cell per mL). Interestingly, *V. cholerae*'s realized niche is much smaller than its fundamental one, as it has maximal temperature and salinity tolerances around 38 °C and 75 ppt (Materna et al., 2012), suggesting other controls on its abundance in the environment.

*V. parahaemolyticus* contrasts *V. cholerae* by having a more constant abundance that is broadly spread out over salinities of 3–35 ppt in a narrow, much warmer temperature range, centered roughly around 29 °C (**Figure 2-6**). Consistent with this finding, it has been noted that this species prefers warmer waters (>20 °C) (Martinez-Urtaza et al., 2012), and has been observed to grow best at 25 °C *in vitro* (Nishina et al., 2004). However, isolations from shellfish can exhibit different trends from those observed in the water column; Martinez-Urtaza et al. (2012) detected *V. parahaemolyticus* in mussels gathered in much cooler, 15 °C water, consistent with the potential for shellfish to concentrate *V. parahaemolyticus*.

A previous literature-based analysis showed *V. vulnificus* to have a more complicated relationship to temperature and salinity than either *V. cholerae* or *V. parahaemolyticus*. It has a narrow temperature range at higher salinities (>10 ppt) while at low salinities (between 5 and 10 ppt) its temperature range more than doubles—from 22–30 °C to 10–32 °C (Randa et al., 2004). This suggests that, in temperate climates, this species is found year-round in estuarine, low salinity environments but can expand into full strength seawater during warmer months. In the tropics, this species should be endemic to the ocean.

### 2.3.3 Conclusions from meta-analysis

From this meta-analysis, we find, first, that temperature and salinity often explain more variance than any other bulk water parameter, like phosphate, nitrogen, pH, or DOC. Yet some of the difficulty in making general statements regarding the relationship of vibrios to individual environmental variables likely stems from the fact that their strength can depend on the ranges examined, e.g. as for temperature, or in quality of the variable, such as DOC, which will encompass carbon derived from different sources that may impact *Vibrio* growth differentially. Second, we observe that trends that apply to the whole genus *Vibrio* do not necessarily reflect those of individual species. Total vibrios and the well-studied potential pathogens *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus* correlate with shared and distinct environmental variables. For *V. parahaemolyticus* and *V. vulnificus*, temperature often explains more variance than does salinity in the same analysis, and for *V. cholerae*, diverse biotic variables, including specific phyto- and zooplankton taxa, can be stronger correlates than abiotic variables. Unfortunately, biotic variables, particularly individual plankton taxa, have rarely been studied in more than one instance, making these observations difficult to generalize. But the correlations reviewed above hint that there may

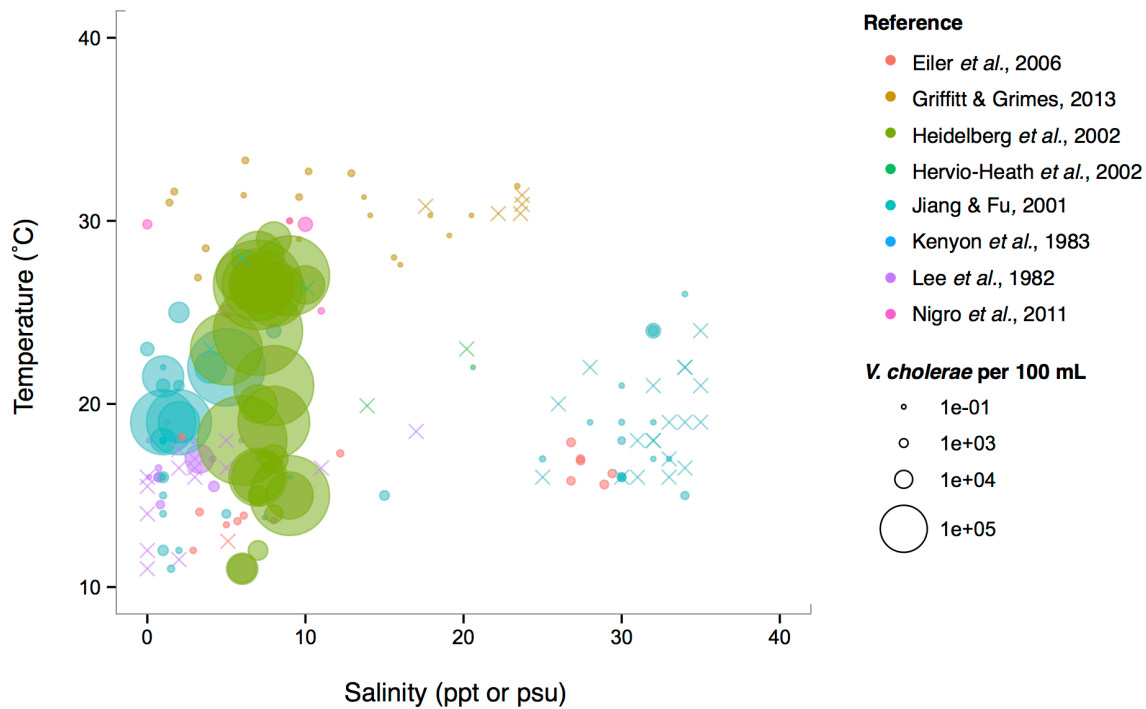


Figure 2-5: *V. cholerae* favors lower salinity and occupies a broad temperature range. *V. cholerae* concentrations, i.e. MPN-estimated CFU or molecular marker gene copies per 100 mL, reported in different studies are plotted against the temperature (°C) and salinity values (ppt or psu) at which they were found. All studies report *V. cholerae*, including O1/O139 and non-O1/non-O139, except for Heidelberg *et al.* (2002a,b); DeLoney-Marino *et al.* (2003), whose genetic marker detected *V. cholerae/V. mimicus*. Circle (○) sizes correspond to concentrations, but note the breaks are scaled for clearer visualization, and not linearly. (×) indicates no *V. cholerae* found in that sample.

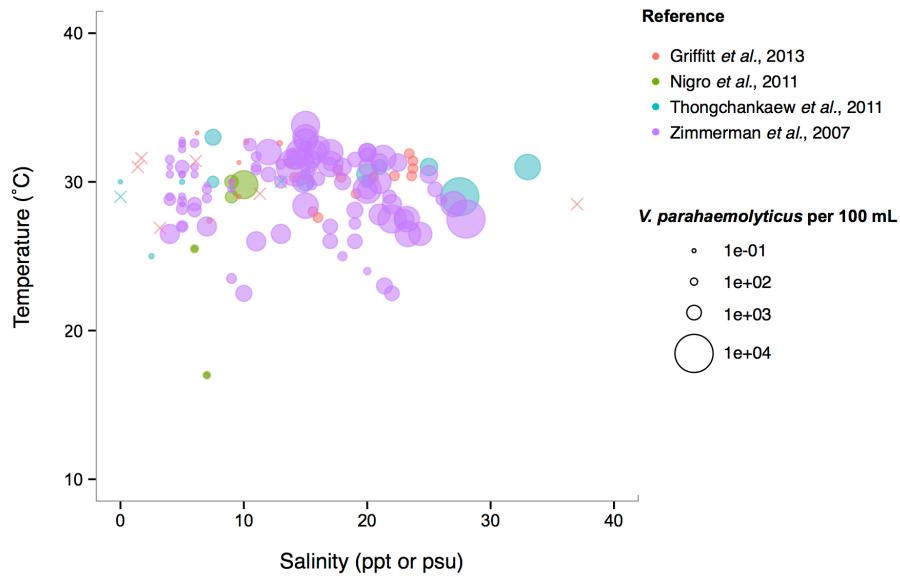


Figure 2-6: *V. parahaemolyticus* favors high temperatures but is relatively unconstrained by salinity. Concentrations, i.e. MPN-estimated CFU or molecular marker gene copies per 100 mL, reported in different studies are plotted against the temperature (°C) and salinity values (ppt or psu) at which they were found in bulk water samples. Circle (○) sizes correspond to concentrations, but note the breaks are scaled for clearer visualization, and not linearly. (×) indicates no *V. parahaemolyticus* found in that sample.

be ecological relationships between *Vibrio* and plankton that merit deeper investigation.

Across salinity and temperature gradients, the pattern also differs between total *Vibrio* and individual species, and species' patterns differ from each other. Indeed, differences may occur even within taxonomic species; *V. parahaemolyticus* pathogenic genotypes have been observed to be a variable fraction of total *V. parahaemolyticus* (Zimmerman et al., 2007). For example, at their Alabama site, total *V. parahaemolyticus*—detected via thermolabile hemolysin marker (tlh)—remained at a more constant concentration of between 1 and 10 cells per mL, while toxigenic genotypes—thermolabile hemolysin+ and thermostable direct hemolysin+ cells—fluctuated in a much wider range: between 0.0001 and 10 cells per mL. This result argues against using the total species to infer the potential pathogens. Taken together with the results from the meta-analysis, these findings suggest that finer-scale sampling—of both the environmental parameters and the *Vibrio* population of interest—is necessary to link ecological parameters to cellular abundances.

## 2.4 Associations with complex and particulate marine growth substrates

The previous sections demonstrate that, with the exception of temperature and salinity, parameters measured in bulk seawater have shown limited power in explaining the environmental dynamics of *Vibrio* species. This may, in part, be due to the narrow focus on only a few (potentially) pathogenic species, and frequently limited comparability of measured parameters across studies. It is also likely, however, that bulk measurements, such as dissolved oxygen, nitrogen and phosphate concentration in seawater, only poorly capture the ecological parameters that *Vibrio* populations are associated with or respond to. Vibrios are often presumed to primarily attach to biological surfaces, yet may also subsist on dissolved resources of biological origin while free-living. Taking these resource associations into account, their environmental dynamics may be somewhat decoupled from parameters measurable in bulk seawater, and may depend more on the concentration and properties of relevant solid or dissolved resources. We review in the following sections the ample evidence for surface-associated niches, as well as more recent evidence for environmental dynamics including free-living states and formation of blooms.

From the perspective of bacteria attaching to surfaces, these are either metabolically inert or can be degraded as a source of growth substrates. Vibrios have the ability to attach to and degrade a considerable number of polymeric substrates (Johnson, 2013), suggesting that specific association with surfaces is an important growth strategy. For example, nearly all vibrios can metabolize the abundant biopolymer chitin (present in both crustacean and diatom shells in the marine environment) (Hunt et al., 2008a; Grimes et al., 2009), and various representatives can metabolize an array of plant/algal polysaccharides: agar, alginate, fucoidan, mannan, cellulose, pectin, and laminarin (Goecke et al., 2010). In addition, vibrios may metabolize plastic wastes, as suggested by a recent study documenting



that vibrios make up the majority of bacteria attached to plastic wastes floating in the ocean, with electron microscopy showing individual cells residing at the bottom of pits (Zettler et al., 2013). Although this suggests that these plastics, which had been thought to be largely biologically inert, could be degraded by vibrios, such activity remains to be confirmed.

Evidence is also accumulating that vibrios may play a role in oil spill degradation: *Vibrio* representatives can metabolize oil-derived compounds (West et al., 1984; Moxley and Schmidt, 2010), and have been found to comprise a sizable fraction of oil-associated microbial communities from the Deepwater Horizon spill, both from sea-surface samples (>31 % in the molecular study of Hamdan and Fulmer (2011) and salt-marsh plants contaminated with oil mousse (57 % in the study of Liu and Liu (2013)). While a clear positive effect of crude oil on *Vibrio* growth has yet to be demonstrated *in vitro*, it appears that many vibrios can at least persist in the presence of oil (Stephens et al., 2013). *Vibrio* representatives furthermore show resistance to inhibition by the oil dispersant Corexit (Hamdan and Fulmer, 2011), which was widely used following the Deepwater Horizon spill; this resistance may additionally support an ability to persist after oil spills.

Most associations with specific surfaces have, however, been described for plants, algae, and animals, and the following section explores these organisms as potential biological niches for vibrios.

## 2.5 Biological niches for *Vibrio*

*Vibrio* have been detected on a plethora of aquatic biological surfaces, but which of these associations represent more than transient, incidental attachments? In the following sections we consider which aquatic plants Table 2.1 and animals Table 2.2 may represent sustained *Vibrio* niches, on the basis of (i) numerical enrichment compared to the surrounding medium, and (ii) knowledge of biological mechanisms, e.g., availability of nutrition and shelter, potentially supporting an association. In doing so, we also draw attention to the need for more quantitative and mechanistic approaches to understanding the ecological associations that allow vibrios to flourish—approaches that could underpin more powerful predictions of *Vibrio* dynamics arising from these diverse associations. We note also that many of the following observations are limited to *V. cholerae* because of its prominence as a pathogen, but the same niches may be available to other vibrios with similar biological activities.

### 2.5.1 Associations with plants

*Vibrio* survival is enhanced in association with certain freshwater and estuarine plants (Table 2.1). Plant hosts can provide nutrition (Andrews and Harris, 2000) and the opportunity to form predation-resistant biofilms (Matz et al., 2005), and have been postulated to modulate unfavorably cold temperatures as well (Criminger et al., 2007). Two freshwater

Table 2.1: Plant and algae hosts for *Vibrio*, as demonstrated by numerical enrichment and biological mechanisms supporting association.

Host	References, study site	Associated <i>Vibrios</i>	Enumeration method	Enrichment, survival advantage	Host site, mechanism of association
<b>PLANTS, FRESHWATER</b>					
<i>Eichhornia crassipes</i> (water hyacinth)	Spira et al., 1981: Bangladesh, freshwater bodies	<i>V. cholerae</i> O1 El Tor	Culture	<i>In situ</i> enrichment: 84% incidence on plants, 16% in water only. <i>In vitro</i> survival advantage: enriched by 10 <sup>2</sup> –10 <sup>3</sup> compared to surrounding water	Possible preference for root exudate
<i>Lemna minor</i> (duckweed)	Islam et al., 1990b: <i>in vitro</i>	<i>V. cholerae</i> O1: one clinical strain, one environmental (from Australian river water)	Culture	<i>In vitro</i> survival advantage: > 27 days survival of attached cells, vs. 15–21 days for cells in surrounding water	Whole plant; mechanism untested
<b>PLANTS, ESTUARINE</b>					
<i>Spartina alterniflora</i> , <i>Spartina patens</i> (marsh grass)	Bagwell et al., 1998; Lovell et al., 2008; Gamble et al., 2010: South Carolina estuary, USA	Spp. including <i>V. alginolyticus</i> , <i>anguillarum</i> , <i>diazotrophicus</i> , <i>parahaemolyticus</i>	Culture; molecular	<i>In situ</i> enrichment: > 50% of culturable diazotrophs; molecular evidence (Gamble et al., 2010) demonstrates stable abundance across seasons	Root association; anaerobic diazotrophy
<i>Juncus roemarianus</i> (marsh grass)	Larocque et al., 2004: South Carolina estuary, USA	<i>Vibrionaceae</i>	Culture	<i>In situ</i> enrichment: > 50% of culturable diazotrophs	Root association; anaerobic diazotrophy
<i>Salicornia virginica</i> (marsh herb)	Bergholz et al., 2001; Criminger et al., 2007: South Carolina estuary, USA	<i>Vibrionaceae</i>	Culture	<i>In situ</i> enrichment: > 50% of culturable diazotrophs	Root association; anaerobic diazotrophy
<b>MICROALGAE AND FILAMENTOUS CYANOBACTERIA, FRESHWATER</b>					
<i>Rhizoclonium fontanum</i> (filamentous green alga)	Islam et al., 1989: <i>in vitro</i>	<i>V. cholerae</i> O1 strains from Australian and Bangladeshi surface water; O1 Bangladeshi clinical isolates	Culture	<i>In vitro</i> survival advantage: 21 days survival of attached cells, compared to 3 days in surrounding water and in no-algae control	Mechanism untested
<i>Anabaena variabilis</i>	Islam et al., 1990b, 2002, 2006; Mizanur et al., 2002: <i>in vitro</i>	<i>V. cholerae</i> O1 Bangladeshi environmental isolates	Culture	<i>In vitro</i> survival advantage: up to 5 days survival of attached cells; > 6 survival in associated water. Persist as VBNC inside algal sheath up to 15 months	Mucilaginous sheath, with possible preference for heterocysts. Possible mechanism: benefiting from algal exudate while relieving oxygen inhibition of N <sub>2</sub> fixation and contributing CO <sub>2</sub> . Demonstrated mechanisms: chemotaxis to host mucus components; mucinase dependence of both chemotaxis and survival with host

(Continued)

Host	References, study site	Associated <i>Vibrios</i>	Enumeration method	Enrichment, survival advantage	Host site, mechanism of association
<b>MACROALGAE, MARINE</b>					
<i>Brown algae</i>					
<i>Ascophyllum nodosum</i>	Chan and McManus, 1969: Canada	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria; enriched by 10 <sup>2</sup> –10 <sup>4</sup> compared to water column	Algal polysaccharide metabolism
<i>Laminaria</i> spp.	Laycock, 1974: Nova Scotia, Canada; Wang et al., 2009	Spp. incl. <i>V. tasmaniensis</i>	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria	Algal polysaccharide metabolism; laminaranolytic activity in particular demonstrated
<i>Red algae</i>					
<i>Hypnea</i> spp.	Lakshmanaperumalsamy and Purushothaman, 1982: tropical estuary, Africa	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria	Algal polysaccharide metabolism
<i>Polysiphonia lanosa</i>	Chan and McManus, 1969: Canada. Islam et al., 1988: <i>in vitro</i> .; Wang et al., 2009	<i>Vibrio</i> spp., incl. <i>V. tasmaniensis</i> , <i>splendidus</i> ; <i>in vitro</i> experiments with <i>V. cholerae</i> O1	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria; enriched by 10 <sup>2</sup> –10 <sup>4</sup> compared to water column. <i>In vitro</i> survival advantage demonstrated	Algal polysaccharide metabolism
<i>Porphyra yezoensis</i>	Duan et al., 1995: China	<i>Vibrio</i> spp.	Culture, scanning electron microscopy	<i>In situ</i> enrichment: Dominant microscopically identifiable and culturable bacteria	Algal polysaccharide metabolism
<i>Green algae</i>					
<i>Chaetomorpha</i> spp.	Lakshmanaperumalsamy and Purushothaman, 1982: tropical estuary, Africa	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria	Algal polysaccharide metabolism
<i>Enteromorpha intestinalis</i> , <i>linza</i>	Lakshmanaperumalsamy and Purushothaman, 1982: tropical estuary, Africa. Islam et al., 1988: <i>in vitro</i>	<i>Vibrio</i> spp.; <i>in vitro</i> experiments with <i>V. cholerae</i> O1	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria. <i>In vitro</i> survival advantage demonstrated	Algal polysaccharide metabolism
<i>Ulva lactuca</i> , <i>pertusa</i>	Islam et al., 1988: <i>in vitro</i> ; Duan et al., 1995: China; Nakanishi et al., 1996; Patel et al., 2003; Tait et al., 2005	<i>Vibrio</i> spp.; <i>in vitro</i> experiments with <i>V. cholerae</i> O1	Culture, scanning electron microscopy	<i>In situ</i> enrichment: Dominant microscopically identifiable and culturable bacteria. <i>In vitro</i> survival advantage demonstrated	Algal polysaccharide metabolism; modulation of host processes: developmental morphogenic effects, spore germination stimulation

Table 2.2: Animal hosts for *Vibrio*, as demonstrated by numerical enrichment and biological mechanisms supporting association.

Host	References, study site	Associated vibrios	Enumeration method	Enrichment, survival advantage	Host site, mechanism of association
<b>INVERTEBRATES</b>					
<i>Freshwater</i>					
<i>Acanthamoeba</i> protozoa	Abd et al., 2005, 2007, 2010; Sandström et al., 2010: <i>in vitro</i>	<i>V. cholerae</i> O1, O139; <i>V. mimicus</i>	Culture, microscopy	<i>In vitro</i> survival advantage: replicate intracellularly > 14 days	Cytoplasm, cysts; protected from antibiotics and predation
Chironomid midge egg masses	Broza and Halpern, 2001; Halpern et al., 2003, 2008: <i>in vitro</i>	<i>V. cholerae</i> isolates from Israeli rivers and waste-stabilization ponds	Culture	<i>In vitro</i> survival advantage: 10 <sup>5</sup> greater cell counts compared to growth in medium alone	Gelatinous egg matrix; can use gelatinous material as sole carbon source, degrading via secreted hemagglutinin/protease
Zooplankton: cladoceran <i>Diaphanosoma mongolianum</i> , from alkaline lake, Germany	Kirschner et al., 2011: <i>in vitro</i>	<i>V. cholerae</i> non-O1/non-O139 isolate from alkaline lake, Germany	Fluorescence <i>in situ</i> hybridization	<i>In vitro</i> survival advantage, but not enrichment: up to 6-fold increase in growth rate of cells in surrounding medium; 10 <sup>5</sup> –10 <sup>7</sup> cells attached compared to 10 <sup>6</sup> –10 <sup>7</sup> cells in surrounding medium	Probable use of host exudates
<i>Estuarine and marine</i>					
Zooplankton: Estuarine copepods, esp. <i>Acartia</i> and <i>Eurytemora</i>	Simidu et al., 1971: Japan; Sochard et al., 1979: Gulf of Mexico; Huq et al., 1983, 1984: <i>in vitro</i> ; Colwell, 1996: <i>in vitro</i> ; Mueller et al., 2007: <i>in vitro</i> ; Preheim et al., 2011a: Massachusetts estuary, USA	<i>Vibrio</i> spp., esp. <i>V. cholerae</i>	Culture	<i>In situ</i> and <i>in vitro</i> enrichment shown in some cases, with up to 10 <sup>5</sup> cells per host. Can dominate culturable surface- and gut-attached communities	Possible preference for oral region and egg sac, due to proximity to host exudates; preference for live versus dead hosts unclear
Corals, incl. <i>Acropora hyacinthus</i> , <i>Oculina patagonica</i> , <i>Mussimilia hispida</i> , <i>Stylophora pistillata</i>	Koren and Rosenberg, 2006: Israel; Kvennefors et al., 2010: Great Barrier Reef; Chimetto et al., 2008; Sharon and Rosenberg, 2008; Koenig et al., 2011; Krediet et al., 2013	Spp. incl. <i>V. alginolyticus</i> , <i>harveyi</i> , <i>splendidus</i>	Culture, molecular	<i>In situ</i> enrichment: can dominate mucus community, according to both culturing and molecular methods; can dominate culturable diazotrophs (found for <i>Mussimilia hispida</i> )	Mucus. Metabolize mucus; diazotrophs likely contribute nitrogen to hosts; may adapt to host antimicrobials via antibiotic-resistance gene acquisition; can inhibit pathogen colonization
Shellfish: blue crabs, <i>Callinectes sapidus</i>	Davis and Sizemore, 1982: Texas, USA	Spp. incl. <i>V. cholerae</i> , <i>vulnificus</i> , <i>parahaemolyticus</i>	Culture	<i>In situ</i> enrichment: Dominant culturable bacteria in hemolymph	Hemolymph; mechanism untested

(Continued)

Host	References, study site	Associated vibrios	Enumeration method	Enrichment, survival advantage	Host site, mechanism of association
Shellfish: oysters	Murphree and Tamplin, 1995; Froelich and Oliver, 2013	Spp. incl. <i>V. cholerae</i> , <i>parahaemolyticus</i> , <i>vulnificus</i>	Culture	<i>In situ</i> enrichment, via host filtration: can be concentrated by up to 10 <sup>4</sup> compared to surrounding water	Gut; unclear whether true gut microbionts, or transient occupants concentrated from food and water
Shellfish: abalone, <i>Haliotis</i>	Reviewed in Sawabe (2006)	<i>V. haliotis</i>	Culture	<i>In situ</i> enrichment: ~70% of culturable gut bacteria; reproducibly specific association	Gut; may contribute to host seaweed digestion via alginolytic activity
Squids: Sepiolid ( <i>Euprymna scolopes</i> ) and loligonoid	Reviewed in Ruby and Lee (1998); Stabb (2006)	<i>V. fischeri</i>	Culture, molecular	Exclusive light organ symbiotes	Bioluminescent symbiotes of nutrient-rich light organ. Colonize immature squid; in mature fish, are expelled and recolonize daily, outcompeting nonsymbiotes
<b>Vertebrates</b>					
Bluefish	Newman et al., 1972: New York, USA	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: can dominate gut bacteria	
Coral reef fishes, incl. surgeonfish <i>Acanthurus nigricans</i> , parrotfish <i>C. sordidus</i> , snapper <i>Lutjanus bohar</i>	Sutton and Clements, 1988; Smriga et al., 2010: Palmyra Atoll, northern Pacific	Spp. including <i>V. agarivorans</i> , <i>coralliilyticus</i> , <i>fortis</i> , <i>furnissii</i> , <i>ponticus</i> , <i>qinhuangdaora</i> , <i>nigripulchritudo</i> ; <i>Photobacterium</i> spp.	Culture, molecular	<i>In situ</i> enrichment: can dominate gut bacteria, according to both culturing and molecular methods. Molecular quantification: 10% of <i>A. nigricans</i> gut community, 71% of <i>C. sordidus</i> , 76% of <i>L. bohar</i>	Gut; unclear whether true gut microbionts, or transient occupants ingested from food (i.e., coral, for parrotfish) and water
Flashlight fishes (Anamalopidae) and anglerfishes (Ceratioidei)	Haygood and Distel, 1993	Novel <i>Vibrio</i> spp.	Molecular	Exclusive light organ symbiotes	Bioluminescent symbiotes of nutrient-rich light organ
Flatfishes incl. Rajidae skate, lemon sole <i>Microstomus kitt</i> , turbot <i>Scophthalmus maximus</i>	Liston, 1957: Scotland, UK; Xing et al., 2013: fish farm, China	Spp. incl. <i>V. cholerae</i> , <i>parahaemolyticus</i> , <i>cholerae</i> ; <i>Photobacterium</i> spp.	Culture, molecular	<i>In situ</i> enrichment: Can dominate gut bacteria, according to both culturing (35–74%, <i>M. kitt</i> ) and molecular (~80%, <i>S. maximus</i> ) methods	Gut; unclear whether true gut microbionts, or transient occupants ingested from food and water
Jackmackerel <i>Trachurus japonicus</i>	Aiso et al., 1968: Japan	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: 27% of stomach culturable bacteria, 100% of intestine	Gut; unclear whether true gut microbionts, or transient occupants ingested from food and water

(Continued)

Host	References, study site	Associated vibrios	Enumeration method	Enrichment, survival advantage	Host site, mechanism of association
Salmonidae, incl. pink salmon <i>Onchorhynchus gorbuscha</i> , chum salmon <i>O. keta</i> , sockeye salmon <i>O. nerka</i> , Chinook salmon <i>O. tshawytscha</i>	Yoshimizu and Kimura, 1976: Japanese coast, East Bering Sea	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: dominate gut bacteria of saltwater-dwelling (but not freshwater) salmonids; on average represent 69% of saltwater gut community	Gut; unclear whether true gut microbionts, or transient occupants ingested from food and water
Sea bream <i>Pagrus major</i> , <i>Acanthopagrus schlegelii</i>	Muroga et al., 1987: Japan	<i>Vibrio</i> spp.	Culture	<i>In situ</i> enrichment: ~45% of culturable gut bacteria	Gut; unclear whether true gut microbionts, or transient occupants ingested from food and water

aquatic plants have been observed to support both *in situ* enrichment (in freshwater bodies of Bangladesh) and *in vitro* survival advantage for *V. cholerae*: duckweed, *Lemna minor* (Islam et al., 1990a), and water hyacinth, *Eichhornia crassipes* (Spira et al., 1981), with preference for roots of the latter. Concentration on *E. crassipes* roots may indicate that root exudate is a particularly rich nutritional source, but may also be an artifact of the fact that the roots represent the greatest area exposed to water, and hence to inoculation by planktonic *Vibrio*. By contrast, duckweed’s minimal structure, lacking stem or developed leaves, means that almost the entire plant is in contact with the water and thus available for inoculation.

Among estuarine plants, nitrogen-fixing representatives of several *Vibrio* taxa—including *V. diazotrophicus*, *V. natriegens*, *V. cininnatiensis* (Urdaci et al., 1988), and *V. parahaemolyticus* (Criminger et al., 2007)—appear to be noteworthy members of the rhizosphere, given that they represent more than half of the culturable diazotrophs associated with the dominant marsh grasses *Spartina* sp. and *Juncus roemerianus* (Bagwell et al., 1998; Larocque et al., 2004), and the herb *Salicornia virginica* (Bergholz et al., 2001; Criminger et al., 2007). While this numerical dominance may reflect culturing bias, later molecular studies of the *S. alterniflora* rhizosphere confirmed that vibrios (not taxonomically resolved below the level of the family) are stable constituents of the community (Lovell et al., 2008), with little seasonal fluctuation (Gamble et al., 2010). Nitrogen fixation thus appears to be an effective strategy supporting *Vibrio* survival in the anaerobic rhizosphere, demonstrating the ecological breadth granted by vibrios’ facultatively anaerobic metabolism.

## 2.5.2 Associations with microalgae and filamentous cyanobacteria

While early culture-based studies have demonstrated numerical dominance of vibrios on phytoplankton surfaces compared to surrounding water, e.g., Simidu et al. (1971),

little is known about direct, physical associations with specific phytoplankton. Algal cells represent a nutritional opportunity in that they often excrete a high proportion of their photosynthetically fixed carbon, thereby creating a diffusive sphere (the phycosphere) around them, with elevated organic carbon concentration compared to the bulk (Paerl and Pinckney, 1996). However, *in vitro* survival advantage and persistence have been thus far been demonstrated only for *V. cholerae* in physical association with two microalgae: with the filamentous freshwater green alga *Rhizoclonium fontanum* (Islam et al., 1989), and inside the mucilaginous sheath of *Anabaena* sp. cyanobacteria under both freshwater (Islam et al., 1990b, 1999) and saline conditions (Ferdous, 2009) (Table 2.1).

Recent work has illuminated mechanistic details of the *V. cholerae* association with *Anabaena*, which may follow the canonical model of symbioses between heterotrophic bacteria and nitrogen-fixing freshwater cyanobacteria. In such associations, heterotrophs locate their hosts via chemotaxis and benefit from rich cyanobacterial exudate (Paerl and Gallucci, 1985). In return, their oxidative metabolism both relieves oxygen inhibition of nitrogen fixation (which would otherwise limit rapid algal growth), and generates carbon dioxide for photosynthetic assimilation (Paerl and Gallucci, 1985). For *V. cholerae*, chemotactic preference for components of the *Anabaena* mucilaginous sheath has been demonstrated (Mizanur et al., 2002). Furthermore, investigators have shown that both chemotaxis to and survival on *Anabaena* depend on *V. cholerae*'s expression of mucinase (Islam et al., 2002, 2006). The exact role of mucinase has yet to be defined, but activity of secreted mucinase might liberate from mucus the relevant chemotactic attractants, aid colonizing *Vibrio* in physical penetration of the mucilage, and/or convert mucilage to nutritive compounds supplementary to the cyanobacterial exudate.

### 2.5.3 Associations with macroalgae

Numerous studies have shown that vibrios are one of the most abundant culturable constituents of macroalgal communities (Table 2.1): a recent meta-analysis of 161, predominantly culture-dependent macroalgal-bacterial studies determined that vibrios on average comprised 10 % of these communities (Hollants et al., 2013), with 28 %, 28 % and 44 % of them found on brown, green, and red macroalgae, respectively. While no molecular studies have yet quantified *Vibrio* within macroalgal communities, numerical enrichment of culturable vibrios has been demonstrated for the brown algae *Ascophyllum nodosum* (Chan and McManus, 1969), and *Laminaria longicruris* (Laycock, 1974); the red algae *Hypnea* sp. (Lakshmanaperumalsamy and Purushothaman, 1982), *Polysiphonia lanosa* (Chan and McManus, 1969), and *Porphyra yezoensis* (Duan et al., 1995); and the green algae *Chaetomorpha* sp. (Lakshmanaperumalsamy and Purushothaman, 1982), *Enteromorpha* sp. (Lakshmanaperumalsamy and Purushothaman, 1982), and *Ulva pertusa* (Duan et al., 1995). For *V. cholerae*, *in vitro* survival advantage has been shown on the green algae *Ulva lactuca* and *Enteromorpha intestinalis* and the red alga *Polysiphonia lanosa* (Islam et al., 1988).

As mentioned above, vibrios can metabolize many algal polysaccharides; they have

furthermore been implicated in several other biological activities facilitating symbiosis with macroalgal hosts. These include antagonism directed towards potential bacterial or algal competitors for host surface area (Dobretsov and Qian, 2002; Kanagasabhapathy et al., 2008), developmental morphogenic effects on *Ulva pertusa* (Nakanishi et al., 1996), and stimulation of spore germination for *Ulva* sp. (Patel et al., 2003; Tait et al., 2005). Hence multiple lines of evidence point to significant *Vibrio* association with *Ulva* sp. (enrichment, survival, morphogenesis and spore modulation) and *Polysiphonia* sp. (enrichment, survival) in particular.

#### 2.5.4 Associations with animals

*Vibrio* interactions with animals include both specific, stable symbioses, and less well-defined associations (Table 2.2). Stable symbioses have been described for luminescent *V. fischeri* (*Aliivibrio*) with sepiolid squids (*Euprymna scolopes*) and loligonoid squids (Ruby and Lee, 1998), and for various luminescent *Vibrio* with flashlight fishes (Anamalopidae) and anglerfishes (Ceratioidei) (Haygood and Distel, 1993). The dynamics of the *V. fischeri*-*Euprymna symbiosis* have been particularly well explicated: *V. fischeri* from surrounding waters colonize the developing squid light organ, successfully outcompeting nonsymbionts in this process, which triggers a developmental program in the host. Once established, the symbionts undergo daily cycles of expulsion and regrowth (Ruby and Lee, 1998; Stabb, 2006). Thus the symbiosis regularly seeds the water column, such that luminous *V. fischeri* are enriched in the water surrounding *E. scolopes* (Ruby and Lee, 1998). This expedites continual recolonization of immature squid, which is likely further facilitated by *V. fischeri* chemotaxis toward squid mucus (DeLoney-Marino et al., 2003).

Some *Vibrio* have also been deemed facultative intracellular symbionts of *Acanthamoeba* protozoa: *Vibrio cholerae* O1 and O139, and *Vibrio mimicus* (Abd et al., 2005, 2007; Sandström et al., 2010; Abd et al., 2010). These vibrios can replicate intracellularly for at least 14 days without affecting host health, at least in nutrient-replete artificial medium, and have been observed in both cytoplasm and cysts of the protozoa. Like several other microbial taxa, then, most famously the pathogen *Legionella* (Rowbotham, 1980), vibrios appear capable of evading *Acanthamoeba* endocytosis to shelter intracellularly. Thus they gain protection from antibiotics (Abd et al., 2005, 2007, 2010), predation, and perhaps other adverse conditions, e.g. cold temperatures. Still to be investigated are the questions of why some *Acanthamoeba* cells encyst their *Vibrio* inhabitants while others do not; why the *Vibrio* do not appear to be detrimental to host survival; and how often *Vibrio* might be released following host lysis, or even actively ejected, thus returning to the water column. Moreover, all studies of the *Vibrio*-*Acanthamoeba* relationship have been experimental: *in situ* surveys are necessary to establish the environmental relevance of this potential symbiosis, and assess any effects on *Vibrio* population dynamics.

Vibrios may be neutral or benign inhabitants of coral hosts: they have been shown to comprise a significant portion of the mucus-dwelling bacterial community of healthy



corals (e.g., Koren and Rosenberg (2006); Kvennefors et al. (2010)), being able to subsist on coral mucus as their sole carbon and nitrogen source (Sharon and Rosenberg, 2008). *V. splendidus*, for example, constituted 50–68 % of clone libraries derived from *Oculina patagonica* coral mucus, but was scarce in the coral tissue itself (Koren and Rosenberg, 2006). Moreover, nitrogen-fixing *Vibrio* representatives, primarily *V. harveyi* and *V. alginolyticus*, have been found to dominate the culturable diazotrophs of the coral *Mussimilia hispida* (Chimetto et al., 2008), and likely share fixed nitrogen with either or both coral and zooxanthellae. Evidence also suggests immune interaction between *Vibrio* and coral hosts: adaptation of *Vibrio* commensals to coral antimicrobials has been suggested by significant antibiotic-resistance gene cassette content of their integrons (Koenig et al., 2011), while one *V. harveyi* coral isolate has been found to help defend its host by inhibiting colonization by a pathogen (Krediet et al., 2013).

In freshwater habitats, *V. cholerae* have been found to proliferate on egg masses of the abundant, widely distributed chironomid midges (Broza and Halpern, 2001; Halpern et al., 2008). These egg masses are embedded in thick, gelatinous material, which *V. cholerae* can use as a sole carbon source (Broza and Halpern, 2001); their degradation of the gelatinous matrix via secreted hemagglutinin/protease appears to be the primary cause of egg mass disintegration (Halpern et al., 2003). Accordingly, Halpern et al. (2006) were able to show correlations of chironomid egg mass with the abundance of attached *V. cholerae*, although they have not yet investigated any correlation of *V. cholerae* dynamics in the surrounding aquatic environment.

Zooplankton, primarily estuarine copepods such as *Acartia* and *Eurytemora*, have been investigated as a major reservoir of *V. cholerae* in particular, but while attachment has been demonstrated, it remains unclear whether the association is specific, and whether attached vibrios are consistently enriched compared to surrounding waters. Individual copepods have been shown to be able to host up to  $10^5$  *V. cholerae* cells (Colwell, 1996; Mueller et al., 2007), with preference often shown for attachment to the oral region and egg sac (next to the anal pore)—that is, regions offering close access to host exudates (Huq et al., 1983, 1984). Culture-based studies have detected enriched *Vibrio* occurrence on copepods compared to the surrounding water column (e.g., Simidu et al. (1971); Sochard et al. (1979)), and one culture-based study showed *Vibrio* dominance of wild copepods’ surface- and gut-attached bacterial communities (Sochard et al., 1979). However, other studies, both *in vitro* and *in situ*, have observed *V. cholerae* remaining predominantly free-living in the presence of copepods (Worden et al., 2006; Neogi et al., 2012) or attaching with greater preference to phytoplankton (Tamplin et al., 1990). Additionally, one culture-independent environmental study detected greater concentrations of *Vibrio*, including *V. cholerae*, in water compared to zooplankton (Heidelberg et al., 2002b,a). Perhaps such variability of association with copepods helps explain the difficulty in detecting correlated *Vibrio*-copepod dynamics, as mentioned above in Section 2.3.

Other uncertainties regarding *Vibrio* association with copepods exist. There is a lack of quantitative evidence demonstrating long-term proliferation of copepod-attached *Vibrio*:

existing studies assessing survival advantage of *Vibrio* cultured with copepods have only demonstrated increased abundance of *Vibrio* in surrounding water, without monitoring attached abundance (Huq et al., 1983, 1984). Finally, it is not clear whether vibrios prefer colonizing live or dead copepods. While several *in vitro* studies have noted *V. cholerae* attachment preference for dead or detrital copepods (Huq et al., 1990; Tamplin et al., 1990; Mueller et al., 2007), one study instead observed survival advantage only upon association with live copepods, and found little attachment to dead copepods (Huq et al., 1983). Perhaps this question could be resolved by investigating from which part(s) exactly of the copepod vibrios derive nutrition: from oral/anal exudates or gut contents of actively feeding copepods, from degradation of the chitinous exoskeleton (which for live copepods is protected by a waxy epicuticle that resists attachment (Tarsi and Pruzzo, 1999), or from degradation of other copepod detritus. In addition, variable host traits such as immune defenses, age, and time since molting or death (which likely affect epicuticle condition) should be taken into account. As of yet, evidence of association with live copepods as an ecological specialization has been demonstrated for only one *Vibrio* sp. nov. (F10) (Preheim et al., 2011a).

In addition, zooplankton other than copepods may represent potential *Vibrio* hosts as well. Kirschner et al. (2011) found cladoceran *Diaphanosoma mongolianum* to enhance growth more than the copepod *Arctodiaptomus spinosus* in microcosm experiments; when cladocerans were added, they enhanced the growth of *V. cholerae* strains in the surrounding medium relative to controls where cladocerans were excluded, while copepods did not. In addition, the number of cells attached to cladocerans per individual was on average 100 times higher than on copepods. When a back-of-the-envelope calculation is done to consider whether *V. cholerae* is enriched on zooplankton, however, we find that they are not, even on cladocerans; from six microcosms,  $10^5$ – $10^7$  cells were estimated attached and  $10^6$ – $10^7$  cells not attached, a result suggesting that cladocerans might enhance overall growth with frequent dispersal, rather than supporting exclusively attached growth.

For other animals in which *Vibrio* have been found to be abundant—fish, and shellfish—it has not yet been determined whether vibrios form specific, lasting associations as gut microbiota, or are merely transient occupants, temporarily proliferating on favorable nutrients until excreted or otherwise detached. In marine fish, numerous studies, both culture-dependent and -independent, have demonstrated that *Vibrio* are major gut inhabitants, often dominating the community, and hence are substantially enriched compared to surrounding seawater. Surveyed fish include flatfish (Liston, 1957; Xing et al., 2013), jack-mackerel (Aiso et al., 1968), bluefish (Newman Jr et al., 1972), salmonids (Yoshimizu and Kimura, 1976), sea bream (Muroga et al., 1987), and various coral reef fishes (Sutton and Clements, 1988; Smriga et al., 2010). Notably, *Vibrio* abundances often appear comparable between culture-based and -independent studies: e.g., 35–74% and 83.4%, respectively, of flatfish inhabitants (Liston, 1957; Xing et al., 2013). The ability of *Vibrio* representatives to resist low pH and bile supports their survival within the fish gut (Yoshimizu and Kimura, 1976). Whether food or water intake is the greater source of inoculation is an open question:

some studies have found a strong effect of food source on gut *Vibrio* composition (e.g., Grisez et al. (1997)), whereas others found a stronger influence of *Vibrio* representation in the water column (e.g., Blanch et al. (2009)). Conversely, *Vibrio* content of the fish gut has also been shown to be responsible for increasing *Vibrio* abundance in surrounding water when fish were introduced into a tank that did not otherwise support *Vibrio* growth, demonstrating significant excretion of viable cells from the fish gut (Sugita et al., 1985). Hence, regardless of length of association, the fish gut appears to represent a favorable refuge where *Vibrio* can rapidly proliferate, prior to being released again to the water column. Indeed, the bioluminescence of marine microbes, including many vibrios, has been suggested to be an adaptation encouraging fish ingestion: fish preferentially predate zooplankton that are glowing after having grazed bioluminescent *Photobacterium* (Zarubin et al., 2012).

Among shellfish, high *Vibrio* abundance has been reported on surfaces and in tissues of hosts including oysters (e.g., Murphree and Tamplin (1995); Froelich and Oliver (2013)), abalone (Sawabe, 2006), and blue crabs (Davis and Sizemore, 1982), with uptake and population dynamics particularly well documented for *V. vulnificus* in association with oysters ((Froelich and Oliver, 2013). *V. haliotis* has been suggested to stably associate with gut of the herbivorous *Haliotis* abalone on the basis of reproducibly specific occurrence: it has never been isolated from other seaweed-consuming invertebrates (reviewed in Sawabe (2006)). Being alginolytic, *V. haliotis* has also been suggested to aid its host's digestion of algal polysaccharides (Sawabe, 2006). Otherwise, it is not clear whether copious *Vibrio* representation might solely be the result of nonspecific uptake from food or water, particularly for filter-feeding shellfish, whose highly efficient filtration has been reported to increase *Vibrio* concentrations by up to 4 orders of magnitude in oysters compared to surrounding waters (Froelich and Oliver, 2013). Furthermore, filter feeders can produce copious amounts of mucus, which rapidly and efficiently removes associated microbes, so that their turnover may be high. Consequently, it is challenging to prove specific association on the basis of abundance. In the next section, we will review a metapopulation study that more explicitly addresses the problem of assessing *Vibrio* host specificity by analyzing population structure across and within macroinvertebrate hosts. Future application of the approach described could help to resolve the question of whether *Vibrio* colonization of animal hosts like fish and crabs is specific, or driven more by indiscriminate uptake from the water column.

### 2.5.5 Population dynamics associated with macroinvertebrate hosts

In a metapopulation study by Preheim et al. (2011a), relative abundances of *Vibrio* groups were compared across different shellfish and parts of shellfish. The study found that macroinvertebrates do not appear to be a strongly selective habitat for vibrios, when contrasted to preceding metapopulation studies of the water column, where differential associations of genotype clusters revealed ecologically distinct populations (described in detail in Section 2.7). When different body parts of mussels and crabs were sampled by

Preheim et al. (2011a), little host preference was evident, and the diversity and frequency of populations (identified by multi-locus sequence analysis) resembled that in water samples. For example, *V. splendidus* represented the dominant population in the water and on both animals. For mussels, which can retain particles when filter feeding (Vahl, 1972), the similarity between water column and animal-associated populations was particularly high, and there appeared to be relatively little difference when gills, stomach and gut walls and contents were compared. This was interpreted as population assembly being largely driven by filter-feeding activity, as was posited in the section above. In contrast to mussels' highly uniform population structure across individual hosts, crabs showed high variance in associated *Vibrio* populations, although composition across individuals' body parts was still similar to that in the water column. What causes the high variance among individual crabs is not known, although there was some evidence suggesting that they may be inoculated by food items, which could be of variable composition given their scavenging lifestyle.

The apparent lack of specificity for the animals was surprising considering that ecological theory predicts that habitats that are long-lived and stable compared to the colonizing species should be dominated by specialists (Kassen, 2002). Yet with regard to mussels and crabs as habitats, vibrios appear to be generalists whose population dynamics may be determined by direct inoculation from the water or via food items (Preheim et al., 2011a). A similar dynamic has recently been suggested to drive *V. vulnificus* accumulation in oysters (Froelich et al., 2010). These can only retain larger particles when filter feeding, and hence enrich pathogenic ecotypes of *V. vulnificus* that are particle-associated as compared to ecotypes that are predominantly free-living.

Overall, these studies demonstrate that colonization may be a complex process strongly influenced by dispersal. In contrast to water column populations, which showed varying degrees of specificity towards microhabitats (e.g., organic particles, zooplankton), *Vibrio* populations on larger invertebrates (mussels and crabs) showed little specificity either for host or host body parts. Whether similar patterns exist for other animals remains unknown; it will be valuable to test fish to determine whether their *Vibrio* inhabitants are true gut microflora. The above studies stress the importance of taking into account potential *Vibrio* sources, i.e. water and food, when assessing host association. For example, *V. splendidus* was the dominant population on both crabs and mussels, and on particles in the water column; had only mussels been sampled, *V. splendidus* may have appeared to have been a mussel specialist. Such erroneous conclusions can be avoided by “mass balancing” populations in a particular location by determining their frequency across different microhabitats or patches that are potentially connected by migration.

## 2.6 *Vibrio* proliferation in the water column

Ocean water is a heterogeneous landscape of varying ecological opportunities on small scales, with a highly patchy distribution of resources that may represent microhabitats

for vibrios. Some of these are hotspots of soluble organic material, which originates from exudates or excretions of larger organisms, while others are particulates of various origins. For example, as mentioned above, algal cells exude a zone of enriched organic material (Bell and Mitchell, 1972; Paerl and Pinckney, 1996). Several other processes can also generate ephemeral patches of dissolved nutrients, and it is likely that many bacteria, including vibrios, can chemotax towards these and take advantage of the elevated nutrient concentrations (e.g., for vibrios, Sjoblad and Mitchell (1979); Mizanur et al. (2001, 2002)). In addition, diverse processes are responsible for the formation of suspended particulate organic matter that can be colonized and degraded by bacteria. This includes dead biomass of small planktonic organisms, fecal pellets, and aggregates (marine snow) formed from polymers and other, smaller particles.

This section will address two main subjects, both seeking to situate *Vibrio* within the marine water column. Here, we will first review both experimental and environmental evidence that blooms of *Vibrio* can and do occur, despite their typically low representation in marine assemblages. Second, we will review the evidence for proliferation of *Vibrio* in the planktonic, free-living phase, expanding the view of their niche range beyond the longstanding proposition that their lifestyle is predominantly attached.

### 2.6.1 *Vibrio* blooms

Thompson and Polz (2006) summed up three key *Vibrio* traits supporting the ability to bloom on sporadic nutrient pulses: *Vibrio* can (i) survive long-term under resource-limited conditions, as indicated by continued respiratory activity in mesocosms (Ramaiah et al., 2002; Armada et al., 2003); (ii) recover from starvation and grow rapidly in response to substrate pulses, enabled by maintenance of high ribosome content (Hood et al., 1986; Flårdh et al., 1992; Kramer and Singleton, 1992; Eilers et al., 2000; Pernthaler et al., 2001); and (iii) actively seek out nutrient patches via chemotaxis (Bassler et al., 1991; Yu et al., 1993), including under starvation conditions (Gosink et al., 2002; Larsen et al., 2004).

*Vibrio* proliferation on natural dissolved resources alone has been experimentally demonstrated by rapid growth of inocula in mesocosms or microcosms of filtered water from algal blooms. *V. cholerae* strain N19691 grew at a rate of up to  $2.6 \text{ d}^{-1}$  in dinoflagellate (*Lingulodinium polyedrum*) bloom water (Mouriño-Pérez et al., 2003), and up to  $1.73 \text{ d}^{-1}$  in water from a dense picophytoeukaryote and dinoflagellate bloom, surpassing the  $0.76 \text{ d}^{-1}$  average growth rate of the separately incubated native bacterial assemblage (Worden et al., 2006).

Experiments have furthermore demonstrated conditions where algal resources were sufficient for *Vibrio* to overcome competition and/or grazing pressure. Taking competition into account, but in the absence of predation, strains of both *V. cholerae* and *V. vulnificus* have been shown capable of increasing in relative abundance when in direct competition with the total bacterial community for filtered homogenate of a cyanobacteria bloom (dominated by *Nodularia spumigena*) (Eiler et al., 2007). Meanwhile, *V. cholerae* N19691

has been shown to overcome substantial protozoan grazing when proliferating on filtrate of a particularly dense algal bloom (Worden et al., 2006). Ample algal dissolved organic material may have permitted this *V. cholerae* growth by relieving resource competition, as the *V. cholerae* inocula grew at the same rate with or without the whole bacterial community filtered out from their bloom-water amendments. Similarly, an analysis of *Vibrio* dynamics sampled from the Arabian Sea suggested that algal resource supply can be a more significant control on *Vibrio* abundance than predation, enabling rapid turnover (Asplund et al., 2011).

Reinforcing these experimental findings, Gilbert et al. (2012) observed an explosive *Vibrio* bloom in the environment, demonstrating that their potential for rapid growth is indeed relevant in the context of a full marine community. In one month, a single *Vibrio* sp., otherwise comprising only 0–2% of total rRNA genes, grew to constitute 54% of the community—the largest bloom of any bacterial group observed over the course of a six-year time series. Furthermore, there was a correlated bloom of the diatom *Chaetoceros compressus*, itself typically rare within the phytoplankton community. Hence, nutrients exuded by the unusually proliferating diatom taxon may have sparked the *Vibrio* bloom, whether by specifically appealing to the species’ metabolic palate, relieving resource competition, diluting protozoan grazing pressure by stimulating rapid growth of the surrounding bacterial community, or some combination of the three. Luminescent *Vibrio* blooming in association with algae have even been suggested to be responsible for the phenomenon dubbed “milky seas,” where significant stretches of surface water are rendered white with bioluminescence (Lapota et al., 1988; Nealson and Hastings, 2006); one recent case was expansive enough (>17,700 km<sup>2</sup>) to be detectable by satellite. Whether such bloom events are rare remains unknown due to currently infrequent sampling and lack of time series; however, the observations cited above provide evidence that *Vibrio* are capable of rapid growth in the environment.

### 2.6.2 The evidence for a planktonic, free-living lifestyle

The two mesocosm/microcosm studies discussed above (Mouriño-Pérez et al., 2003; Worden et al., 2006) both furnish evidence that vibrios can thrive while free-living. Mouriño-Pérez et al. (2003) demonstrate the ability of a *V. cholerae* strain to flourish purely on dissolved compounds derived from an algal bloom. Even more strikingly, Worden et al. (2006) observed *V. cholerae* N19691 remaining free-living in four out of their five seawater mesocosm experiments: one initiated from non-bloom seawater, and the other three initiated from seawater collected during distinctly different phytoplankton blooms. Notably, in two of these four experiments, *V. cholerae* attachment to cohabiting copepods was assessed and found to be insignificant (e.g., <1 *V. cholerae* cell found per copepod, averaged over a sampling of ten copepods, in one of the experiments). This stands in contrast to the theory that *V. cholerae* preferentially attach to copepods, as discussed above in Section 2.5.4. In the remaining experiment of Worden et al. (2006), in contrast to the mesocosms in which *Vibrio* remained free-living, the *V. cholerae* inoculum was initially almost entirely free-living,

but, as bloom decay progressed and algal detrital particles increased in size, the population became almost entirely particle-attached, presumably in response to nutrient limitation.

The factors determining whether *Vibrio* remain free-living versus particle-attached are still unknown, but both environmental and genetic determinants could come into play. Past studies have demonstrated effects of temperature, pH, ion concentration, and starvation state (Hood and Winter, 1997); salinity (Kumazawa et al., 1991; Hsieh et al., 2007); and growth-stage-dependent chitin content of diatom cell walls (Frischkorn et al., 2013) on *Vibrio* attachment. Perhaps encounters with relevant biological compounds, e.g., a specific algal cell wall component or polysaccharide, might also trigger lifestyle changes. Even less is known about the genetic mechanisms, diversity, and dynamics underlying *Vibrio* lifestyle association; this remains a rich field of inquiry. For example, Shapiro et al. (2012) recently discovered genomic patterns underlying the ongoing ecological differentiation of two *V. cyclitrophicus* populations: the population with preference for association with larger particles possessed genes for attachment and biofilm formation that were absent from the preferentially free-living population. Such evidence of genetic bases for habitat specificity will provide invaluable insights into selective pressures exerted by different marine microhabitats.

The findings described above suggest great flexibility in *Vibrio* lifestyle, permitting many lines of attack on marine substrates, with different ecological implications for vibrios' dynamics in the water column. For example, biofilm attachment on particulate resources can decrease susceptibility to protozoan predation (Matz et al., 2005), while association with larger particles might increase probability of ingestion by macrofaunal predators, which could in turn facilitate rapid proliferation and dispersal, as discussed above in the section on fish associations. Given vibrios' possibilities for rapid growth and association with diverse marine niches and resources, their impacts on marine nutrient cycling and trophic structure might be much greater than previously believed. Understanding their dynamics will help to elucidate these fundamental marine processes, as well as *Vibrio*-specific models of pathogen persistence and transmission.

## 2.7 Using ecology to define cohesive populations

The studies summarized above suggest potential for association of vibrios with plants, algae and animals as well as growth response to specific classes of particulate and dissolved organic matter; however, they have targeted primarily a single, taxonomically defined species, leaving several important questions unanswered. First, do such taxonomic species correspond to ecologically cohesive units, i.e., do they comprise several ecologically distinct populations or should they be merged with others to form one ecologically cohesive population? Second, if we can define such populations, do these partition resources or compete with each other? Finally, are vibrios primarily ecological generalists or specialists?

A series of studies explored to what extent ecologically coherent groups of vibrios could

be distinguished by determining the distribution patterns of genotypes among different potential microhabitats in the coastal ocean (Hunt et al., 2008a; Preheim et al., 2011a,b; Szabo et al., 2013). Initially, this was done by isolation of vibrios from four consecutive size fractionations of ocean water, collected in the spring and fall, to distinguish free-living and attached genotypes (Hunt et al., 2008a). The rationale of this sampling scheme was that different types of microhabitats (e.g., organic particles of various origin, zoo- and phytoplankton) have characteristic size spectra and hence will be enriched in a specific size fraction. Consequently, bacteria specifically associated with a microhabitat should be enriched in the same specific size fraction. Further, because ecological associations may evolve on faster time scales than rRNA genes, isolates were also characterized at higher genotypic resolution using several protein coding genes in a multilocus sequence analysis (MLSA) scheme, to better capture the eco-evolutionary dynamics of environmental populations. Because of the complexity of the data, a statistical clustering algorithm (AdaptML) was developed that allows identification of groups of related genotypes with distinct and characteristic distributions among the sampled parameters (size fractions and seasons) (Hunt et al., 2008a).

The analysis of >1000 isolates identified a large number of genotypic clusters with clear microenvironmental preferences, consistent with the notion of an ecological population (Hunt et al., 2008a). Seasonal differentiation was particularly strong, with little overlap between spring and fall samples, supporting the observed significant correlation of some species to temperature discussed in above sections. The study also revealed that several populations appear free-living or predominantly free-living, again supporting the notion that vibrios can pursue, at least temporarily (e.g., during a bloom), unattached lifestyles. Most populations, however, displayed various preferences for size fractions enriched in different types of organic particles or zoo- and phytoplankton. For example, *V. calviensis* appeared almost entirely free-living, while *V. alginolyticus* had significant representation in both the free-living and large-particle fractions, and *V. fischeri* occurred on small and large particle size fractions. Most strikingly, *V. splendidus* was broken up into several, very closely related populations with distinct distributions. Overall, 25 distinct populations could be identified in the two seasonal samplings. (Hunt et al., 2008a), demonstrating the fine-scale resource partitioning co-existing vibrios are engaged in.

To what extent does the commonly used rRNA marker gene resolve these populations? The *V. splendidus* example and several others demonstrate that at least some ecologically distinct genotypic clusters may not be resolved by rRNA analysis and do require high resolution protein-coding genes to identify genotypic clusters whose environmental distributions can be assessed (Preheim et al., 2011b; Shapiro et al., 2012). Most populations, however, were manifest as deeply divergent protein-coding gene clusters (Hunt et al., 2008a) that correspond to microdiverse rRNA gene clusters previously postulated to represent ecological populations (Acinas et al., 2004). Although overall reassuring for rRNA gene-based environmental surveys, variable performance of marker genes is expected since they are slowly evolving and may not capture populations at early stages of divergence (Shapiro



et al., 2012).

Additional studies carried out at the same coastal site refined the habitat resolution for several populations, allowed identification of ecological generalists and specialists, and also demonstrated reproducible associations (Preheim et al., 2011a,b; Szabo et al., 2013). The actual microhabitat of several attached populations was identified by hand-picking under the microscope visually identifiable types of particles and zooplankton (Preheim et al., 2011a,b). This revealed high habitat specificity for several populations while others occurred more broadly, indicating different levels of ecological specialization. For example, *V. breoganii* occurred on algal derived detritus while a not yet formally described species (*Vibrio* F10) was highly specific for living zooplankton. On the other hand, *V. crassostreae* was associated with both zooplankton and algal detritus. Metabolic potential in these species, measured by growth assays and comparative genomics, reflects these associations. Both *V. breoganii* and *V. crassostreae* are able to exploit alginate, a brown algal cell wall component, as the sole carbon source, yet the algae-associated *V. breoganii* has acquired the ability to grow on the algal storage polysaccharide laminarin but has lost the ability to grow on chitin, a trait ancestral to vibrios (Hunt et al., 2008b). Moreover, such high specificity for algal derived material was unexpected for vibrios, which are reputed to be animal associated, and supports the evidence provided above that vibrios encompass algal specialists.

A recent study that attempted to reproduce the original size fractionation of ocean water collected at a similar time point, but three years after the initial sampling, showed that population structure was preserved for many of the originally detected populations, but also revealed populations as dynamic and environmentally responsive entities (Szabo et al., 2013). For example, *V. breoganii*, *V. crassostreae* and *V. splendidus*, which range in ecological specialization from specialist to generalist, had highly reproducible distributions indicative of similar habitat associations. The study, however, also showed that several populations were nearly absent in either of the samplings, possibly due to the lower frequency of their habitat in the water samples. Moreover, some populations had shifted distributions among the size fractions. This was the case for a recently diverged population of *V. cyclitrophicus* that was associated with larger particles or organisms in the first study, but was highly represented in the free-living fraction in the second sampling. It was hypothesized that this shift represented a population expansion following a diatom bloom because the relative frequency of *V. cyclitrophicus* increased coincident with a shift from a copepod- to a diatom-dominated eukaryotic plankton community. Similarly, bloom dynamics, as have previously been observed for total vibrios in the water column, may cause the variable representation of several additional populations. Overall, the comparison of the two studies supports highly predictable population-habitat linkage but also provides additional support for the notion that vibrios may be subject to rapid population expansions or blooms in response to often overlooked or unknown environmental factors.

## 2.8 Populations as ecological, genetic, and social units

Populations as defined here are genotypic clusters (evident by MLSA) that act as ecologically cohesive units, i.e., their ecology is more similar within the cluster than between. Defining populations in this way has afforded the opportunity to test the hypothesis that, akin to sexual eukaryotes, gene flow boundaries across such clusters are strong enough for adaptive genes or alleles to spread in a population-specific manner. A population genomic analysis of two very recently diverged populations of *V. cyclitrophicus*, which are ecologically distinct but remain >99% similar in average nucleotide composition across their genomes, showed that specific genome regions have swept each of the two populations (recently reviewed in (Polz et al., 2013)). Moreover, annotation of these genome regions as well as behavioral and growth analysis suggest that these genome regions are adaptive for differential lifestyles (Shapiro et al., 2012; Yawata et al., 2014).

A second study showed that ecologically defined populations may also act as social units. This was evident in a test for potential of antagonistic interactions mediated by antibiotics between individuals from different ecological populations of vibrios (Cordero et al., 2012b). Because of higher niche overlap among close relatives, it was expected that antagonism be more advantageous if directed against members of the same population. In stark contrast, however, antagonism was primarily directed against members of other populations while members of the same population were resistant to antibiotics produced within their own populations. This suggests synergism on the population level, especially since multiple antibiotics were produced within each population but each only by relatively few members.

Overall, this research shows agreement between ecological, genetic and social population structure and suggests that, in many ways, populations can be regarded as species-like units in the wild. Importantly, these units are non-clonal, and their genetic exchange and social structure suggest that populations frequently coexist and re-assemble on small-scale habitats.

## 2.9 Conclusion

In this review, we examine what is known about *Vibrio* ecology at increasingly fine environmental and taxonomic scales, to reveal factors with potential for greater predictive and explanatory power for *Vibrio* dynamics.

We find that while bulk environmental variables are often inconsistent in their ability to explain variance in *Vibrio* abundances, at both the genus and species levels, temperature and salinity are usually the strongest abiotic correlates. Yet total *Vibrio* trends do not necessarily capture species-level trends, and thus it is necessary to monitor populations of interest directly to capture their dynamics. Correlations of species to specific plankton—like those of *V. cholerae* to dinoflagellate (Eiler et al., 2006), cladoceran (Kirschner et al., 2011) and rotifer (Magny et al., 2011) taxa—can provide the bases for hypotheses of biological associations, as was demonstrated by Kirschner et al. (2011) for the cladoceran *D.*

*mongolianum*. Further investigation is necessary to confirm reproducibility and biological significance of such correlations.

Indeed, the breadth of vibrios' metabolic and attachment abilities mean that they can appear quite generalist in their ecological associations, making it difficult to discern which relationships with other organisms are specific and stable, rather than simply the product of promiscuous attachment followed by proliferation. Among the diverse biological associations that we review, some may be true mutualisms, on the basis of vibrios exchanging benefits with their hosts. The symbioses of luminescent vibrios with certain squid and fish are well attested, while possible symbioses with other organisms are suggested by potentially mutual metabolic exchange (salt marsh plants, cyanobacteria, corals), or *Vibrio* modulation of host processes like development and reproduction (macroalgae), and response to infection (corals). Notably, diazotrophy may facilitate relationships with both marsh plants and corals. In numerous other cases, vibrios may simply be taking advantage of hosts as nutrient sources, and perhaps only temporarily and opportunistically be associated with microalgae, zooplankton, fish, shellfish, and chironomid egg masses, or as intracellular occupants of protozoa. Of these, we argue that evidence points towards a particularly significant ecological impact of *Vibrio* interactions with algae, given the abundant laboratory and environmental observations of vibrios' ability to live on algal exudates—including blooms as free-living cells, a historically underappreciated *Vibrio* lifestyle. Nonetheless, much work remains to be done in resolving more specific *Vibrio*-algae associations.

In light of these studies, we have several recommendations. Previous surveys of *Vibrio* abundance are predominantly culture-dependent; going forward, molecular methods, such as fluorescent *in situ* hybridization or quantitative PCR, can be used to gain less biased quantitative data. Such techniques also enable targeting of specific genotypic groupings, allowing better discrimination of pathogenic variants or ecologically meaningful populations than traditional taxonomic assays of species identity. Furthermore, to distinguish specialized association from incidental attachment, a “mass-balanced” approach is necessary: are *Vibrio* enriched on a given microhabitat (e.g., a specific organic particle type or zooplankton) compared to the surrounding water? Or, is the habitat enriched in *Vibrio* compared to other habitats? This approach has provided support for many of the potential symbioses noted above, and enabled identification of specialist *Vibrio* populations, e.g., *V. breoganii* for macroalgae-derived material and V. F10 for zooplankton (Hunt et al., 2008a; Preheim et al., 2011a,b; Szabo et al., 2013). It provides a strong basis from which to proceed to more detailed and, ideally, mechanistic elucidation of *Vibrio* associations: for example, identifying chemotactic preferences for or proliferation on host or host exudates, or taking advantage of vibrios' genetic tractability to demonstrate dependence of an association on particular metabolic pathways.

When considering the question of to what extent environmental affiliations may be shared among or within *Vibrio* taxa, we also explore the shifting perspective on the nature of microbial groupings: recent work has moved towards discerning ecologically cohesive *Vibrio* populations, rather than relying on named species as the unit of inquiry. Pursuing

this approach, whereby habitat associations are mapped onto genotypic clusters, has been successful in identifying ecological, genetic and social units among vibrios in the wild. We stress, however, that the initial identification of environment-genotypic cluster associations by the “mass-balanced” approach outlined above must be treated as a hypothesis of population structure to be further explored by more mechanistic investigation of, for example, dynamic habitat associations, biological interactions and gene flow boundaries. As demonstrated above, this approach has already helped to resolve apparently generalist *Vibrio* taxa into specialized populations and to identify mechanisms of how adaptive genes spread amongst nascent, ecologically differentiated populations. By sampling the environment at fine scales and molecularly characterizing associated *Vibrio*, we will gain a deeper understanding of the ways in which vibrios live in the environment. Such a population-based framework serves as a means of understanding the ecology of microorganisms in general.

## 2.10 Acknowledgments

This work was supported by grants from the National Science Foundation Evolutionary Ecology program (Evolutionary Processes), and the National Science Foundation and National Institutes of Health co-sponsored Woods Hole Center for Oceans and Human Health, the Moore Foundation, and the National Institutes of Health sponsored MIT Environmental Health Science Center (P30-ES002109).

## Chapter 3

# Adaptations in *Vibrio* to ecological habitat resources, assayed by genome-wide fitness

### 3.1 Abstract

Heterotrophic bacteria occupy diverse habitats in the ocean, from particles to transient gradients of dissolved organic matter. Though genomic, transcriptomic, and proteomic studies offer us a rich view of the genetic capabilities and physiological responses bacteria have to exploit different habitat-associated resources, we do not know either the extent to which distinct resources strongly select for metabolic pathways in a single organism, nor, moreover, if selection may be dominated by a single resource within a habitat. In this work, we bridge this knowledge gap by selecting a transposon mutant library of *Vibrio* sp. F13 strain 9CS106 on nutrients from model habitats—the copepod *Apocyclops royi* and the brown alga *Fucus vesiculosus*—and the *Fucus* constituent polysaccharide, alginate. We find that *Fucus* requires anabolic pathways for nucleosides and synthesis of some amino acids, but apparently provides some as well (proline and aromatic amino acid derivatives); that alginate does not dominate selection, as the catabolic requirements for growth on alginate (including the Entner-Doudoroff pathway) differ from those required for growth on *Fucus* (mannitol fermentation); and that *Apocyclops* exhibits neither strong selection for anabolic nor catabolic pathways, indicating that this habitat is a replete medium; unlike its growth on *Fucus*, 9CS106 growth on *Apocyclops* does not heavily rely on any single catabolic pathway, suggesting it instead uses several genes that have overlapping functions, such as permeases for peptides and free amino acids. Furthermore, the breadth and severity of selection for metabolic pathways in alginate is greater than in model habitats, indicating that complex resources buffer fitness costs—an effect that may influence evolution in natural environments. This work provides insight into heterotrophic adaptation to resources bacteria

may encounter in the wild.

## 3.2 Introduction

Heterotrophic microorganisms contribute heavily to the flux of organic matter in the oceans (Azam, 1998). Dominating marine metabolic activity (Azam and Malfatti, 2007), microorganisms may actively colonize particles or degrade dissolved macromolecules (such as from algal exudates) to sustain growth (Arnosti, 2014). Thus, microbial habitats occupy a spectrum from surfaces to gradients, free-living to particulate (Stocker, 2012). While adaptations to these diverse niches have been increasingly inferred by exploration of individual genomes and metagenomes (Bengtsson-Palme et al., 2014; Hellweger et al., 2014; Fontanez et al., 2015; Berube et al., 2015; Xing et al., 2015), a complementary approach to link genotype to phenotype is experimental manipulation of culturable isolates.

Our recent characterization of differential microhabitat associations among closely related marine bacteria affords the opportunity to explore the selective landscape exerted by environmental resources (Hunt et al., 2008a; Preheim et al., 2011a; Szabo et al., 2013). We have recently identified an apparent generalist population: *Vibrio* sp. F13, which appears to specifically associate with different and divergent microhabitats in the water column. Isolates from this population have been cultured from both living and dead zooplankton, and detrital particles of algae. Using laboratory models of these habitats, and extracting constituent nutrients to sustain high cell densities that can then be interrogated genetically, we ask: what pathways do resources in different habitats most strongly select for?

To answer this question, we used an approach common to the investigation of virulence determinants and host colonization factors in pathogens: selection of a collection of mutants in an environment of interest. The approach allows the quantified fitness effects of single gene disruptions, enabling the simultaneous investigation of which genes are selected in a given condition and to what degree. Using a *Vibrio* sp. F13 strain that is genetically tractable and whose genome has recently been sequenced, 9CS106, we created a transposon-mutant library to analyze the effect of single gene disruptions during growth on nutrients from models of zooplankton and vegetation habitats. A high-throughput approach based on PCR amplification and sequencing of transposon junctions, Tn-seq, was used to tally mutant abundance pre- and post-selection.

The models of habitat used to select the mutant library were the copepod *Apocyclops royi* and the brown alga *Fucus vesiculosus*. *Apocyclops* and *Fucus* have distinct resource profiles; copepods are richer in protein and lipids than brown algae, and brown algae are richer in saccharides. *Apocyclops* contains an estimated 57% protein (information from provider) and 24% saccharide, whereas the dry mass of *Fucus*, though content varies depending on season and organism age, has an estimated 1.4% protein and 65% saccharide (Rioux et al., 2007).<sup>1</sup>

---

<sup>1</sup>Estimates of saccharide obtained by difference, as in Rioux et al. (2007), since no official method exists.

We further ask, to what extent does a single, dominant resource utilized by the organism select for the same pathways as total habitat resources? 9CS106 has been found to grow on *Fucus*' constituent polysaccharide, alginate (10–40 % of dry weight) (Percival and McDowell, 1967), as a sole carbon source, but whether it specializes for this carbon source during growth on *Fucus* is unknown. A high degree of overlap between pathways selected in these conditions would indicate specialization for alginate, whereas a low degree would suggest broader niche breadth and the ability to capitalize on the greater complexity of the *Fucus* habitat. In addition to alginate, *Fucus*, and *Apocyclops* media, two other media were used as points of reference: the rich culture medium marine broth 2216, and a single carbon source minimal medium with glucose.

To screen the large number of mutants in our collection, we used large amounts of habitat-derived nutrients distributed in a liquid medium. Though this condition negates that the habitats investigated here are solids to which the bacteria may attach, it nonetheless provides a realistic resource spectrum. By comparing fitness determinants across conditions, the selections provide insights into the ecology of this organism, as a pathway required in one condition but not another suggests a condition-specific difference. Though the *Vibrio* sp. F13 population has been isolated from particles, indicative of growth in attached biofilms, 9CS106 appears to grow predominantly planktonically under these nutrient-rich conditions.

We find that ecologically relevant resources indeed select for different metabolisms. *Fucus* selects for biosynthesis of certain amino acids and nucleosides, and catabolism of mannitol, but not detectably for catabolism of alginate. By contrast, *Apocyclops* does not select for any of these metabolisms, indicating that multicopy genes and redundant pathways may broadly buffer fitness effects, and that the metabolic pathways 9CS106 uses in this habitat are distinct from those in the *Fucus* habitat. Neither complex habitat condition selected for polysaccharide degradation of alginate and chitin in the *Fucus* and *Apocyclops* media, respectively. Moreover, selection is broader (i.e. for more pathways) and stronger (i.e. fitness costs are greater) for simple resource landscapes, like alginate, than for complex ones, like habitats *Fucus* and *Apocyclops*. Together, these results indicate that resource context exerts distinct selection pressures on 9CS106, contributing to distinct metabolic strategies.

## 3.3 Results and discussion

### 3.3.1 Construction of a transposon-mutant library

To select mutants in both ecologically relevant and reference media, we generated a mutant library, and characterized the initial mutant abundance. First, we created a transposon vector for *Vibrio* mutagenesis, pSW25T::C9::ISCAT. The vector contains a transposon— a chloramphenicol resistance cassette flanked by MmeI-modified *mariner* inverted repeats—and the hyperactive *mariner* transposase, *Himar1* C9 (Lampe et al., 1999), which targets TA dinucleotide sites for insertion. pSW25T::C9::ISCAT also contains the

machinery for conjugative transfer, and requires the *pir* protein for replication, rendering it a suicide plasmid in the *pir*-recipient. To create the mutant library, we conjugated *Vibrio* sp. F13 strain 9CS106 with an *E. coli* donor, EC3-4 (Section 3.5). Strain 9CS106 mutagenized with  $1.2 \times 10^{-6}$  efficiency (ratio of mutagenized cells to total).

Using a Tn-Seq approach to determine abundance of individual mutant strains (van Opijnen et al., 2009; Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009), we found that the original mutant pool contained  $\sim 100,000$  independent insertion sites (about 30 % saturation of TA sites). Grouping insertion sites by gene and IG regions,  $\sim 4400$  genes and  $\sim 3300$  IG regions were disrupted, of 4621 and 3776 total, respectively. That not all genes were disrupted was consistent with our expectation that a few hundred genes would be essential (see Section 3.3.5). These were distributed without apparent bias among both chromosomes (chromosome 1: 3.37 Mb; chromosome 2: 1.86 Mb) and both extrachromosomal elements (ECE) (ECE 1: 45.4 kb; ECE 2: 38.7 kb).

### 3.3.2 Selection in five media conditions

To identify genes that affect fitness in ecologically-relevant and reference conditions, we prepared media from dried and ground brown alga *Fucus* (yielding a turbid, brown particulate suspension) and freeze-dried copepod *Apocyclops* (a clear, light yellow filtrate), in addition to culture medium 2216 (a turbid, yellow particulate suspension), glucose (clear and colorless), and alginate (clear, with partially gelled particles). All media except 2216 were made with artificial seawater, and amended with dilute minimal medium (10 % v/v) as a source of buffering capacity and reduced nitrogen, phosphate, metals and cofactors. Mutants were grown in batch culture and harvested after approximately 9 to 10 generations at the end of exponential phase, to constrain selection to unlimited growth conditions.

The 9CS106 mutant library grew at comparable rates in the 2216 (doubling time,  $t_d$ , of  $(31 \pm 2)$  min), *Apocyclops* ( $t_d$  of  $(30 \pm 2)$  min), and *Fucus* ( $t_d$  of  $(32 \pm 4)$  min) media (Figure 3-1 and Table 3.1), suggesting the habitat-derived media to be relatively rich. By contrast, mutant library growth was slower in the single-carbon source media: in the glucose medium ( $t_d$  of  $(41 \pm 2)$  min), and slowest in the alginate medium ( $t_d$  of  $(55 \pm 3)$  min). We postulate that growth on alginate may have been slowed by the precipitation of the polysaccharide (unquantified). However, the degree of growth was sufficient to see changes in mutant fitness, as described below.

### 3.3.3 Validation of approach to test for significant changes in mutant abundance

To determine significant changes in mutant abundance after growth in the selective media, we first normalized the data, and validated a threshold for statistical significance.

After filtering for insertion sites with at least three reads, we corrected for sequencing depth, replicate variability, and small read counts, using the R package DESeq2 (Love et al.,



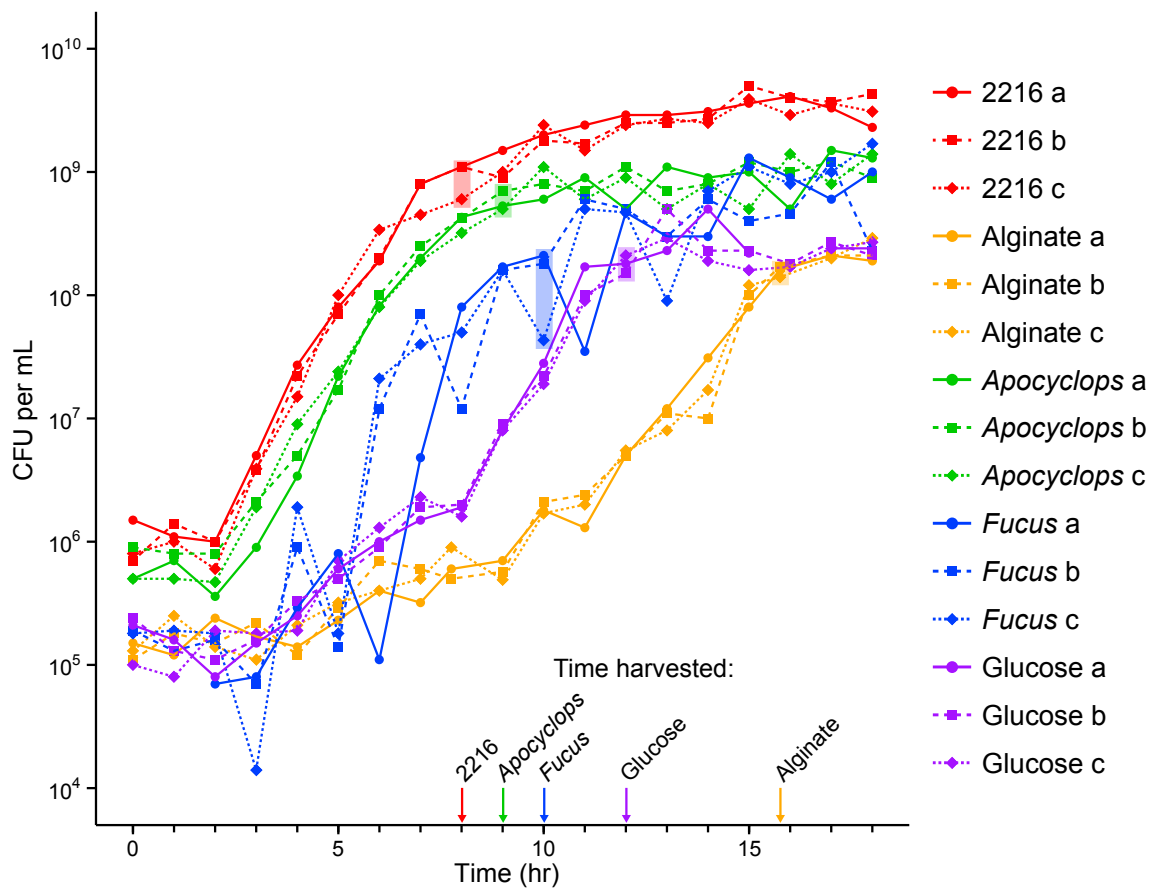


Figure 3-1: **Mutant library grown in selective media and harvested.** A transposon-mutant library of *Vibrio* sp. F13 strain 9CS106 was grown in five media: 2216, *Apocyclops*, *Fucus*, alginate, and glucose. The cell densities used to inoculate growth were such that harvesting coincided with the elapse of approximately nine to ten generations, and prior to entry into stationary phase. The time points in the growth curve when samples were harvested are also highlighted.

Table 3.1: Growth of *Vibrio* sp. F13 strain 9CS106 in selective media

Growth medium	Initial mean CFU per mL <sup>a</sup> , x 10 <sup>5</sup>	Time harvested (hr)	Harvest mean OD <sub>600</sub>	Harvested mean CFU per mL <sup>a</sup> , x 10 <sup>5</sup>	Mean Generations elapsed <sup>a</sup>	Growth rate <sup>b</sup> (per hour)	Doubling time <sup>c</sup> (min)	Carrying capacity <sup>b</sup> (CFU per mL)
2216 Marine Broth (autoclaved)	10.0 ± 4.36	8	0.38 ± 0.01	9.33 ± 2.89	9.9 ± 0.6	1.33	31 ± 2	7.0E+05
<i>Apocyclops</i> 1 g per L (pasteurized + 0.2- $\mu$ m filtered)	6.33 ± 2.31	9	0.19 ± 0.10	5.77 ± 4.61	9.8 ± 0.6	1.41	30 ± 2	5.0E+05
<i>Fucus</i> 10 per L (pasteurized)	1.90 ± 0.10	10	0.13 ± 0.10	1.44 ± 2.52	9.6 ± 0.2	1.33	32 ± 4	7.6E+04
Alginate 1 g per L (pasteurized)	2.53 ± 0.20	15.75	0.12 ± 0.034	1.57 ± 2.04	9.3 ± 0.9	0.76	55 ± 3	1.9E+05
Glucose 1 g per L (pasteurized)	1.83 ± 0.74	12	0.20 ± 0.002	1.8 ± 0.30	9.9 ± 0.9	1.02	41 ± 2	1.5E+05

a. Mean ± standard deviation.

b. From the fitted parameters of a non-linear least squares regression of the logistic growth curve, using the Gauss-Newton algorithm.

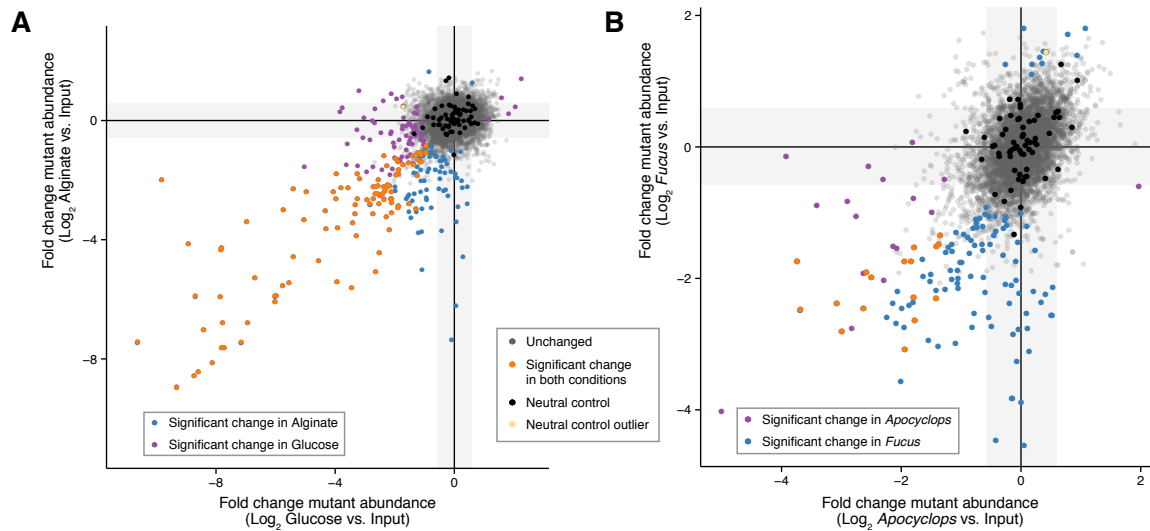
c. Doubling time based on growth rate ± lower (growth rate minus standard error) and upper (growth rate plus standard error) estimates.

2014).

To identify genes significantly over- or under-represented in the output library, we applied a negative binomial test for fold-changes greater than or equal to 1.5 (Benjamini-Hochberg adjusted  $P$  value,  $P < 0.05$ ). The approach was validated using a set of gene-sized (300 to 1300 bp) IG regions, which, because they are located downstream (3' end) of their flanking genes, are hypothesized a priori to be 'neutral' — an approach taken previously (Goodman et al., 2009). Of 63 IG regions in the genome that met this criteria, only one was under-represented in each of the 2216, *Fucus*, and glucose conditions (a false positive rate of 1.6%) whereas none were affected in the alginate and *Apocyclops* conditions (Figure 3-2).

Further confidence in this approach came from the detection of genes expected to have a fitness effect. The glucose condition required the glucose-specific IIB/IIC component of the phosphoenolpyruvate phosphotransferase system (PTS)—responsible for carbohydrate uptake—and the alginate condition required three genes known to be involved in metabolizing alginate: two copies of poly(beta-D-mannuronate) lyase and pectin degradation protein (KdgF) ( $P < 0.05$ , negative binomial test, 1.5 fold change). In both instances, mutants in these genes suffered no fitness effect in any other condition.

Analysis of the alginate-selected library underscored a second point: multicopy genes would likely be undetected in this analysis, even though they may be (and, sometimes, very likely are) relevant to a particular condition. Alginate metabolism is well-characterized, and the initial process involves extracellular degradation of polymers to importable oligomers using alginate lyases (Wargacki et al., 2012). Though 9CS106 has a battery of seven alginate lyases, disruption of any one does not have a measurable fitness effect, likely because the genes are sufficiently redundant, and the gene dosage high.



**Figure 3-2: Genes and intergenic (IG) regions under condition-specific selection.** The ratio of output to input abundance per gene or IG region in the mutant library of *Vibrio* sp. F13 strain 9CS106 is compared between different nutrient conditions: (A) alginate vs. glucose media, and (B) *Fucus* vs. *Apocyclops* media. Genes and IG regions are highlighted by color indicating whether their abundance was unchanged, or significantly different from the input. Sixty-three IG regions a priori assumed to be neutral were also tested for significance as negative controls; only one locus in the glucose and *Fucus* conditions was significant. The grey areas indicate the log<sub>2</sub> fold change threshold (0.585) for the negative binomial test used to categorize the genes and IG regions ( $P < 0.05$ ). To obtain abundance, raw read counts were normalized with DESeq2 and log<sub>2</sub>-transformed. Note the differences in scale for plot axes.

### 3.3.4 Exclusion of Extrachromosomal Element 1

One extrachromosomal element was excluded from the analysis. For 38 of ECE1’s 58 genes, and 15 of its IG regions, mutants were less abundant in the *Fucus* condition, and to a greater degree, in the glucose and alginate conditions (Table 3.3). Because of this large percentage, we postulate that the fitness effect of these mutations is not on the whole organism itself, but rather on the ability of ECE1 to replicate in more resource-limited conditions. ECE1 has few annotations, though one is a bacteriophage terminase, suggesting it may be a phage remnant.

### 3.3.5 Prediction of the 9CS106 essential genome

To identify essential genes, i.e. those with significantly fewer transposon insertions in the input mutant library, we used a Monte Carlo approach as previously described (Turner et al., 2015). This method randomizes the reads across insertions, creating 2000 ‘expected’ data sets from the pooled input library replicates. Against these input mutant abundances, the ‘observed’ pooled input mutant library abundances were tested for significant reduction relative to the expected, using a negative binomial test implemented with DESeq2. Genes were then clustered by their  $\log_2$  fold change (lfc) using the R package mclust (Fraley and Raftery, 2012) into two groups: one that appeared unchanged from the expected abundance, and the other exhibiting reduced abundance relative to that expected. For a locus to be considered essential, it had to meet two criteria: have a significant fold change (negative binomial test,  $P < 0.05$ ) and cluster into the reduced abundance group. From this analysis, we found 331 essential genes and 58 essential IG regions in the 9CS106 genome. The number of genes is similar to previous estimates of essential genome size (Juhas et al., 2011): for example, 303 in *E. coli* (Baba et al., 2006).

In the 9CS106 essential genome, as defined by colony growth on 2216, well-represented functional categories are cell division, DNA replication and repair, ribosome biogenesis, tRNA synthetases, translation factors, protein processing and secretion, capsule synthesis, lipopolysaccharide synthesis, cell envelope biogenesis, cofactor synthesis, and respiration. With respect to carbohydrate metabolism, fructose/sucrose utilization appears to be essential via fructokinase, and four of the 10 enzymes in the Embden-Meyerhof-Parnas (EMP) glycolysis pathway are essential: aldolase, triose phosphate isomerase, phosphoglycerate kinase (an ATP-generating step), and phosphoglycerate mutase. A few pathways involving amino acid metabolism are also essential: alanine biosynthesis via cysteine desulfurase, an arginine/ornithine antiporter, and synthesis of S-adenosyl methionine, involved in methyl group transfers from methionine. Interestingly, a “widespread colonization island,” which may have aided growth on the agar surface the mutant library was initially grown, and *toxR*, a transcriptional regulator, are also essential. *ToxR* regulates expression of cholera toxin (Miller and Mekalanos, 1984), but is also found in non-pathogenic strains, and regulates non-host associated phenotypes, such as dormancy (Almagro-Moreno et al., 2015).

### 3.3.6 Overview of fitness determinants for growth on ecologically relevant resources

Mutants libraries selected in the habitat conditions had different numbers of genes with significantly altered mutant abundance—or briefly, fitness determinants—which were distinctly similar in magnitude to those selected in the reference conditions. The *Fucus* condition had 98 fitness determinants; the alginate condition and glucose conditions, 133 and 129, respectively. The *Apocyclops* condition had 29; the 2216 condition, 38.

Pooling fitness determinants for all conditions, and excluding IG regions, which are discussed in greater detail below, the most represented category was amino acid and derivatives (52 genes). Other well-represented categories pertained to carbohydrates (24 genes), protein metabolism (18 genes), cell wall and capsule (13 genes), RNA metabolism (12 genes), respiration (11 genes), nucleosides (11 genes), membrane transport (seven genes), cofactors (six genes), DNA metabolism (six genes), stress response (six genes), and fatty acids and lipids (five genes). The categories are further delineated into subcategories and pathways (i.e. subsystems) for fitness determinants in each media in Table 3.2. Alginate and glucose conditions shared the greatest number of fitness determinants suggesting that single-carbohydrate media are more similar selective environments than ones that share a common resource; 77 genes (58% of alginate, and 60% of glucose) shared between the two Figure 3-2.

*Apocyclops* and *Fucus* appear to have only a varying degree of functional overlap; the two conditions shared 14 genes (48% of the number in *Apocyclops*, but only 14% of the number in *Fucus*) (Figure 3-2). *Fucus* and alginate, by contrast, have a greater degree of overlap suggestive of similar metabolisms 9CS106 employs on these carbohydrate-rich resources. The two conditions shared 46 genes (47% of *Fucus*, 35% of alginate).

In addition to the essential genome, only 0.6% of 9CS106 genes (excluding IGs) are fitness determinants in the *Apocyclops* condition, and 2% in the *Fucus* condition, out of the 4,519 genes combined in 9CS106's two chromosomes. Interestingly, these results contrast other whole-genome mutant studies, which typically observe hundreds for microorganisms selected in host habitats, with only one example of less than 150 genes having been reported—a study of a *Pseudomonas aeruginosa* strain grown in buffered sputum (Turner et al., 2015). We hypothesize that this difference is due to the harsher selective conditions found in a living organism.

### 3.3.7 Selected mutant libraries more similar between *Fucus* and alginate, *Apocyclops* and 2216

To see if the selected mutant libraries are more similar between ecologically similar conditions, we hierarchically clustered the output libraries by read abundance per gene / IG region, using a minimal normalization approach to reduce variability in small read counts.

Indeed, the output libraries cluster according to resource similarity, suggesting overlap

Table 3.2: Annotation category distributions for media-selected genes

Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	2216	Apocyclops	Fucus	Alginate	Glucose
Amino Acids and Derivatives	Alanine, serine, and glycine	Glycine and Serine Utilization			1		
Amino Acids and Derivatives	Alanine, serine, and glycine	Serine Biosynthesis			3	3	3
Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Biosynthesis extended			2	5	5
Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Deiminase Pathway	1			1	1
Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Urea decomposition				2	1
Amino Acids and Derivatives	Aromatic amino acids and derivatives	Aromatic amino acid biosynthesis				1	1
Amino Acids and Derivatives	Aromatic amino acids and derivatives	Chorismate Synthesis				1	1
Amino Acids and Derivatives	Aromatic amino acids and derivatives	Shikimate kinase containing cluster				1	1
Amino Acids and Derivatives	Aromatic amino acids and derivatives	Tryptophan synthesis				4	4
Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis			6	7	6
Amino Acids and Derivatives	Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis			1	2	2
Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis			2	5	6
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Cysteine Biosynthesis			1	3	5
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Lysine biosynthesis				1	1
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Methionine Biosynthesis			3	4	3
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Threonine and Homoserine Biosynthesis			1	4	3
Amino Acids and Derivatives	Proline and 4-hydroxyproline	Proline Synthesis				2	2
Carbohydrates	Aminosugars	Chitin and N-acetylglucosamine utilization				1	
Carbohydrates	Central carbohydrate metabolism	Dehydrogenase complexes				3	3
Carbohydrates	Central carbohydrate metabolism	Glycolate, glyoxylate interconversions	1	1		1	
Carbohydrates	Central carbohydrate metabolism	Glycolysis and Gluconeogenesis	1	1	1	2	3
Carbohydrates	Central carbohydrate metabolism	Glyoxylate bypass				1	
Carbohydrates	Central carbohydrate metabolism	Pyruvate metabolism I: anaplerotic reactions, PEP			2		
Carbohydrates	Central carbohydrate metabolism	TCA Cycle				1	
Carbohydrates	Di- and oligosaccharides	Maltose and Maltodextrin Utilization		1		1	
Carbohydrates	Fermentation	Fermentations: Mixed acid	2	1	2		
Carbohydrates	Monosaccharides	D-gluconate and ketogluconates metabolism				2	1
Carbohydrates	One-carbon Metabolism	One-carbon metabolism by tetrahydropterines				1	
Carbohydrates	Sugar alcohols	Glycerol and Glycerol-3-phosphate Uptake and Utilization					1
Carbohydrates	Sugar alcohols	Mannitol Utilization			1		
Carbohydrates		Glycogenolysis and glycogenesis				1	
Carbohydrates		Sugar utilization in Thermotogales			1		
Carbon sources		Eugenol utilization				1	
Cell Division and Cell Cycle		Macromolecular synthesis operon			1		
Cell Wall and Capsule	Capsular and extracellular polysaccharides	Alginate metabolism				3	
Cell Wall and Capsule	Gram-Negative cell wall components	KDO2-Lipid A biosynthesis		1	1		
Cell Wall and Capsule	Gram-Negative cell wall components	LOS core oligosaccharide biosynthesis	1	2	1	1	1
Cell Wall and Capsule	Gram-Negative cell wall components	Major Outer Membrane Proteins		1	1	1	2
Cell Wall and Capsule	Gram-Negative cell wall components	O-antigen biosynthesis		1	1		
Cell Wall and Capsule		CBSS-160492.1.peg.550			1		
Cell Wall and Capsule		Peptidoglycan Biosynthesis	1	2			1
Clustering-based subsystems		CBSS-342610.3.peg.1794			1	1	1
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate and pterines	Folate Biosynthesis					1
Cofactors, Vitamins, Prosthetic Groups, Pigments	Pyridoxine	Pyridoxin (Vitamin B6) Biosynthesis			1	1	1
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN, FAD	Flavodoxin	1	1	1		1
Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Cobalamin synthesis				1	
Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Coenzyme B12 biosynthesis				1	1
Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Heme and Siroheme Biosynthesis		1			1
DNA Metabolism	DNA repair	2-phosphoglycolate salvage			1		1

Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	2216	Apocyclops	Fucus	Alginate	Glucose
DNA Metabolism	DNA repair	DNA repair, bacterial				1	
DNA Metabolism	DNA repair	DNA repair, bacterial UmuCD system			1		
DNA Metabolism		DNA structural proteins, bacterial	1		1		
Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty acid biosynthesis			1	2	1
Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty acid degradation regulons	1				
Fatty Acids, Lipids, and Isoprenoids	Phospholipids	Glycerolipid and Glycerophospholipid Metabolism in Bacteria	1	1	1		
IG <sup>b</sup>	IG	IG	7	1	12	7	22
Ion Metabolism		Mg ion transport			1		
Ion Metabolism		Potassium ion uptake		1		1	
Iron acquisition and metabolism		Transport of Iron			1		2
Membrane Transport	ABC transporters	ABC transporter oligopeptide (TC 3.A.1.5.1)	1				1
Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Mannose-sensitive hemagglutinin type 4 pilus			2		
Membrane Transport	Protein secretion system, Type III	Type III secretion system orphans				1	1
Membrane Transport	Sugar Phosphotransferase Systems, PTS	Sucrose-specific PTS			1	1	1
Membrane Transport		Ton and Tol transport systems					1
Metabolism of Aromatic Compounds	Peripheral pathways for catabolism of aromatic compounds	Quinate degradation			1	1	1
Motility and Chemotaxis	Flagellar motility in Prokaryota	Flagellum				2	1
Nitrogen Metabolism		Allantoin Utilization				1	
Nitrogen Metabolism		Ammonia assimilation			1	2	2
No category		CBSS-364106.7.pcg.3204	1		1		
No category		Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins) subsys not found			1		
No category			4	1	4	4	6
Nucleosides and Nucleotides	Purines	De Novo Purine Biosynthesis			3	4	4
Nucleosides and Nucleotides	Purines	Purine conversions			1	2	2
Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis			2	5	5
Outer membrane transport		Translocation and assembly module TamB	1			1	
Phosphorus Metabolism		Phosphate metabolism		1	1	1	1
Protein Metabolism	Protein biosynthesis	NusA-TFII Cluster	1		1	1	
Protein Metabolism	Protein biosynthesis	Peptide processing	1		1		
Protein Metabolism	Protein biosynthesis	Ribosome biogenesis bacterial	1		2	1	1
Protein Metabolism	Protein biosynthesis	Translation elongation factors bacterial	1	1	2		
Protein Metabolism	Protein biosynthesis	Translation initiation factors bacterial	1				
Protein Metabolism	Protein biosynthesis	Translation termination factors bacterial		1			
Protein Metabolism	Protein biosynthesis	Universal GTPases	1		1		
Protein Metabolism	Protein degradation	Proteasome bacterial			2	1	
Protein Metabolism	Protein degradation	Putative TidE-TidD proteolytic complex	1	1			1
Protein Metabolism	Protein folding	Periplasmic disulfide interchange				2	
Protein Metabolism	Protein folding	Protein chaperones	1	1	1	2	
Protein Metabolism	Protein secretion	Protein transmembrane transport				1	
Regulation and Cell signaling	Quorum sensing and biofilm formation	Quorum-sensing in Vibrio				1	
Regulation and Cell signaling		cAMP signaling in bacteria				1	
Regulation and Cell signaling		Stringent Response, (p)ppGpp metabolism			2	1	
Regulation and Cell signaling		Zinc regulated enzymes	1			1	
Respiration	ATP synthases	F0F1-type ATP synthase		1		3	1
Respiration	Electron donating reactions	Succinate dehydrogenase				1	
Respiration	Sodium Ion-Coupled Energetics	Na+ translocating decarboxylases and related biotin-dependent enzymes					3
Respiration		Biogenesis of c-type cytochromes				1	1
Respiration		Ubiquinone biosynthesis -- gjo	1			1	1
RNA Metabolism	RNA processing and modification	ATP-dependent RNA helicases, bacterial	1	1	1	1	
RNA Metabolism	RNA processing and modification	RNA processing and degradation, bacterial		1	1		
RNA Metabolism	RNA processing and modification	tRNA modification Bacteria	1		4	3	
RNA Metabolism	Transcription	Rrf2 family transcriptional regulators			1	1	1
RNA Metabolism	Transcription	Transcription factors bacterial		1	2		1

Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	2216	Apocyclops	Fucus	Alginate	Glucose
Stress Response	Osmotic stress	Osmotic stress cluster	1	1		1	1
Stress Response	Oxidative stress	Glutaredoxins			1		
Stress Response	Oxidative stress	Glutathione: Redox cycle		1	1	1	
Stress Response	Periplasmic Stress	Periplasmic Stress Response				1	
Stress Response		Carbon Starvation			1		1
Stress Response		Hfl operon		1		1	1
Sulfur Metabolism		Thioredoxin-disulfide reductase			1		
Virulence, Disease and Defense	Resistance to antibiotics and toxic compounds	Copper homeostasis					1

a. Annotations manually curated from RAST, KEGG, Pfam, and Phyre2, as described in the Methods.

b. IG = Intergenic region.

in 9CS106's physiological response. Mutant libraries selected in the *Fucus* condition cluster with those selected with its polysaccharide constituent, alginate, and separately, libraries grown on *Apocyclops* with those on 2216 (Figure 3-3). Interestingly, the *Fucus*-selected libraries were also more similar to the *Apocyclops*-selected ones, than to the glucose-selected ones, suggesting *Fucus* and *Apocyclops* have a greater overlap in resources than *Fucus* and glucose. Because the mutant library was created on the 2216 medium, the input library samples clustering with 2216 was expected.

### 3.3.8 Nutrients from habitats buffer fitness costs

Because the habitats, being biological entities, may contain a variety of carbohydrates, amino acids, cofactors, and nucleosides that 9CS106 can utilize, we hypothesized that the severity of fitness costs, i.e. the magnitude to which mutants decreased in the selected libraries, might be less than in the single-carbon source conditions. For instance, loss of the ability to synthesize an amino acid may be benign when there are sufficient amino acids to scavenge.

To test this hypothesis, we calculated mutant fitness, and analyzed their distributions among the five conditions. Using the mean initial and final cell densities of the experiments, we converted the lfc metric used for hypothesis testing into the time-independent relative growth rate, i.e., the fitness ( $W$ ), per locus as previously described (van Opijnen et al., 2009).

Indeed, the *Fucus* and *Apocyclops* media, like 2216, buffered fitness costs; whereas mutants in these conditions had fitness values no less than approximately 0.4, alginate and glucose both showed long tails in their distributions of mutant fitness (decreasing to effectively zero)<sup>2</sup> (Figure 3-4). Because many of these severely deleterious mutations were

<sup>2</sup>A gene in the glucose condition has a negative fitness value, approximately -0.7. While no growth would give a fitness of 0, negative fitness indicates mutants in this gene are dying — reducing rather maintaining their absolute abundance.



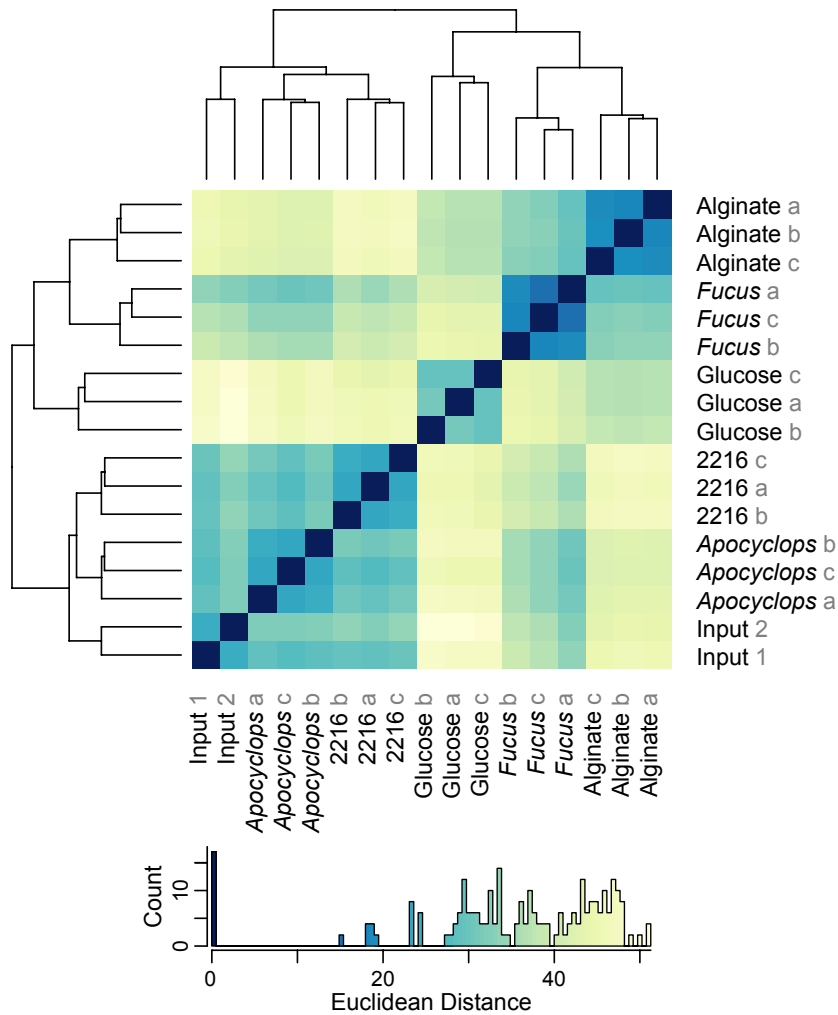


Figure 3-3: **Samples cluster according to resource similarity.** Selected mutant libraries from individual replicates cluster by condition. Replicates are biological for media conditions, labelled a, b, c, and technical for the input mutant library, labelled 1, 2. Different conditions also cluster: alginate and *Fucus*, with glucose as an outgroup, and, separately, *Apocyclops* and 2216. Hierarchically clustered heatmap of the sample-to-sample Euclidean distances of their Illumina sequencing read counts, normalized using Tikhonov/ridge regression to reduce small count variability, implemented with DESeq2.

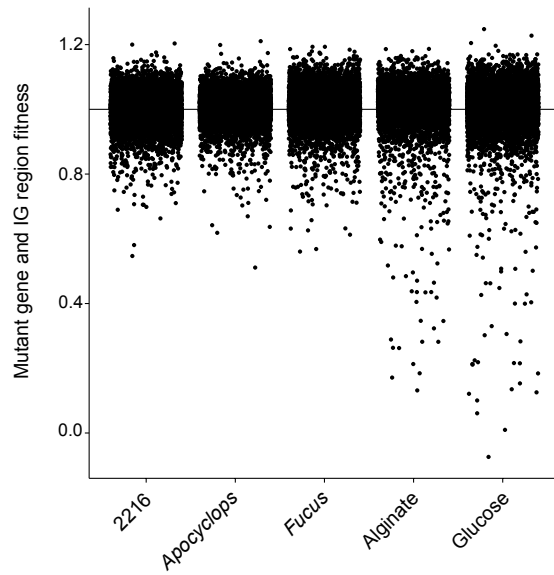


Figure 3-4: **Growth medium 2216 and habitat-derived nutrients buffer fitness costs.** The fitness, i.e. the growth rate of 9CS106 mutants relative to that of the wild type, is depicted for each gene and IG region, across conditions. The fitness was calculated from mean read counts after normalization with DESeq2 across replicates ( $n = 3$ ), as detailed in the Section 3.5.

in anabolic pathways (which are discussed in greater detail below), these results suggest that the *Fucus* and *Apocyclops* habitats are diverse in scavengeable biosynthetic resources. More broadly, they imply that organisms with less fit alleles may persist in nutrient-rich environments longer than in poorer ones.

For simplicity, we use the lfc metric instead of W when comparing the magnitude of fitness effects in the sections below.

### 3.3.9 Anabolism and catabolism in the *Fucus*, alginate, and *Apocyclops* conditions

The selected mutant libraries revealed that *Fucus* and alginate share anabolic requirements, but differ in their catabolic ones, and the *Apocyclops* condition, by contrast, does not exhibit strong requirements for either anabolic or catabolic pathways.

#### *Fucus*

The relatively protein poor *Fucus* condition required biosynthetic pathways for 11 amino acids (based on annotation) for competitive mutant growth: serine, glutamate, arginine, asparagine, lysine, threonine, methionine, histidine, and branched amino acids leucine, isoleucine, and valine. Notably, the inability to synthesize proline and aromatic amino acids did not have a fitness cost in the *Fucus* condition, whereas it was pronounced in the alginate and glucose media, with lfcs ranging from -2.2 to -8.9. To explain this difference, evidence suggests that *Fucus* is a source of proline and aromatic amino acid derivatives; proline is a known algal and plant osmolyte (Edwards et al., 1987; Yoshida et al., 1997), and phenolic compounds, such as tannins, which constitute up to 25 % of macroalgae by dry weight, derive from the same chorismate precursor as aromatic amino acids (Arnold and Targett, 2002).

*Fucus* also showed selection for de novo synthesis of nucleoside bases purine and pyrimidine, with reduced mutant fitness for six genes total in both pathways.

Because of *Fucus*' diverse polysaccharide composition, we hypothesized that genes involved in their metabolism might be fitness determinants. In addition to alginate, brown algae contain the sulfated polysaccharide fucoidan (5–20 % dry weight) (Black, 1953), in the cell walls and the overlying matrix (Davis et al., 2003); laminarin (2–34 % dry weight) (Bold and Wynne, 1985), a storage polysaccharide, and the sugar alcohol mannitol (5–30 % dry weight) (Lewis and Smith, 1967; Reed et al., 1985), both as a free monomer, and in branched chains in laminarin (which also contains branches of glucose).

In fact, genomic evidence suggests that pathways to metabolize fucoidan and laminarin are absent, and the mutant selection shows 9CS106 appears to be specializing to a greater degree on mannitol than alginate. The disruption of mannitol-1-phosphate 5-dehydrogenase, which converts mannitol-1-phosphate to fructose-6-phosphate, enabling its entry into the glycolytic pathway, incurred a large fitness cost (lfc = -3.8). Mutation in alginate genes,

poly(beta-D-mannuronate) lyases and pectin degradation protein, however, had no fitness cost in the *Fucus* condition, in contrast to their effect in the alginate condition as described in the Section 3.3.6.

Mannitol utilization could be characteristic in an algal habitat; a strain of *Vibrio* isolated from a kelp bed fermented mannitol (Davis and Robb, 1985), and recent findings by Ymele-Leki et al. (2013) suggest that mannitol and the mannitol-specific IIB component of the PTS activate biofilm formation, which could be advantageous in the colonization of macroalgae. Interestingly, the two genes required for mixed acid fermentation, phosphate acetyltransferase (PTA), and acetate kinase (ACK), were fitness determinants in the *Fucus* condition suggesting potential fermentative growth on mannitol. The other gene besides mannitol-1-phosphate 5-dehydrogenase in the mannitol utilization operon, a fusion that encodes the mannitol-specific IIA, IIB, and IIC, components, however, did not have a significant effect on fitness, but showed only a slight negative lfc (-0.5). We speculate that another transporter may allow for mannitol entry into the cell, and could be tested with targeted mutagenesis.

Furthering the evidence of a saccharide-based diet, 9CS106 required other genes involved in sugar metabolism for growth on *Fucus* resources, as well as on alginate: 6-phosphofructokinase, the first committed step of glycolysis, and phosphoenolpyruvate-protein phosphotransferase, the non-sugar specific PTS component.

## Alginate

Genes required for growth in the alginate condition were found in metabolic pathways distinct from those found in either the *Apocyclops* or *Fucus* conditions. Alginate metabolism relies on the Entner-Doudoroff (ED) pathway, and indeed, disruption of an enzyme in this pathway, and one generating the substrate that enters the pathway, incurred large fitness costs: 2-dehydro-3-deoxygluconokinase (lfc = -7.37), which phosphorylates the alginate metabolite 2-dehydro-3-deoxy-D-gluconate (KDG) into 2-dehydro-3-deoxy-D-gluconate 6-phosphate (KDPG), and KDPG aldolase (lfc = -4.57), which cleaves KDPG into glyceraldehyde-3-phosphate and pyruvate. These products are further metabolized in the classical glycolytic pathway (Embden-Meyerhof-Parnas) (Patra et al., 2012), and fed into the tricarboxylic acid (TCA) cycle, and, consistently, mutants in the TCA cycle's malate dehydrogenase, were less fit.

While amino acid and purine / pyrimidine biosynthesis were clearly required in the *Fucus* condition, the fitness costs in the alginate condition were significantly greater for genes required in both conditions (one-sided Mann-Whitney U test,  $P < 0.05$ ). For instance, mutants in the arginine synthesis enzyme acetylglutamate kinase had an lfc of -1.4 in the *Fucus* medium; yet, these mutants had an lfc of -8.7, a more than 64-fold greater effect, in the alginate medium. In the case of pyrimidine synthesis, mutants in carbamoyl-phosphate synthase small subunit exhibited a similar trend: an lfc of -2.2 in the *Fucus* medium, and -8.13 in the alginate medium.

## *Apocyclops*

In contrast to the *Fucus* condition, *Apocyclops* appears to be a replete medium, not requiring synthesis of any amino acid, nor nucleosides.

No individual catabolic pathways emerged as being unique to the *Apocyclops* medium, suggesting 9CS106 engages metabolisms on this habitat, that unlike growth on *Fucus*, insulate it from individual mutations. Though copepods are rich in protein, genes in protein catabolism (such as peptide transporters or permeases (Sussman and Gilvarg, 1971) were not observed, a fact that may be explained by functional overlap encoded in the genome; 9CS106 has a dipeptide ABC transporter (Dpp), and (what appears to be) two oligopeptide ABC transporters (Opp). Permeases for individual amino acids, lysine, thiamine, and polar amino acids could also immure fitness effects from being observed. Alternatively or in addition, 9CS106 may also be utilizing other carbon catabolism that similarly have no individual effects on fitness.

Growth on *Apocyclops*, like on *Fucus*, appears to elicit fermentative metabolism—but the selection is not as strong, perhaps because of a lower nutrient concentration. Whereas only one gene is required in the *Apocyclops* condition (PTA), the *Fucus* condition requires both genes (lfc = -2.1 and -2.6, respectively) and their fitness is more similar to that observed in 2216 (-1.7 and -2.8), than to that in *Apocyclops* (-1.0 and -1.7).

While copepods have chitinous exoskeletons, and 9CS106, like many vibrios, has extensive machinery to utilize this carbon source (Hunt et al., 2008b), chitin utilization was not expected for the time-scale of this study. Growth on chitin is expected to be a secondary metabolic tactic; in *V. harveyi* grown on culture broth with 2% colloidal chitin, chitinases were not expressed until after exponential phase (Rao et al., 2013).

### 3.3.10 Cofactor and ion metabolism

Cofactors were readily available in the *Apocyclops* condition, limited in the alginate condition, and varied in the *Fucus* condition. Whereas the vitamin B12 transporter, btuB, was required only in the single-carbon source, alginate and glucose, conditions, the vitamin B6 (pyridoxin) enzyme, D-erythrose-4-phosphate dehydrogenase, was required in the *Fucus*, alginate, and glucose conditions.

With respect to ions, *Fucus* exerts a stronger requirement for magnesium, with mutants in a transporter showing reduced fitness. In the alginate and *Apocyclops* conditions, however, a potassium transporter was required.

Ferric iron transport also appears to be influencing fitness in both of the habitat-derived media: OmpT, a porin that is hypothesized to be a portal for free iron (Craig et al., 2011) was a fitness determinant in all conditions except 2216, though the effect was stronger in the *Apocyclops* condition than in the *Fucus*. (The effect was strongest in glucose and alginate conditions.) Additionally, mutants in the iron-binding component of a ferric iron ABC transporter were selected against, in both *Fucus* and glucose conditions. Iron limitation

does not appear to be as severe in the *Fucus* condition as in the glucose one, however; mutants in the ATP-binding component of this transporter also had reduced fitness in the glucose condition. Why these mutants were not also selected against in the *Apocyclops* condition remains unclear.

### 3.3.11 Evidence of a costly colonization factor, MSHA, in the *Fucus* condition

Interestingly, mutants in the synthesis of the Type IV pilus, mannose-sensitive hemagglutinin (MSHA), genes *mshI* and *mshJ*, experienced a fitness benefit in the *Fucus* condition ( $lfc = 1.4$  and  $1.8$ , respectively). No effect was observed in any other condition for these genes.

The role of MSHA in colonization of biotic environments is an ongoing field of study; *Vibrio parahaemolyticus* has been found to use MSHA, in concert with another Type IV pilus, PilA, to attach to diatom-derived chitin (Frischkorn et al., 2013), and *Vibrio cholerae* has been found to rely on MSHA to adhere to cellulose, in addition to chitin (Chiavelli et al., 2001; Watnick et al., 1999), though it relies on the toxin co-regulated pilus to colonize the mammalian gut (Thelin and Taylor, 1996; Teschler et al., 2015). Intriguingly, another marine *Gammaproteobacteria*, *Pseudomonas tunicata*, uses the MSHA pilus to attach to the live green alga *Ulva australis* (Dalisay, 2006), suggesting MSHA might have a similar role in 9CS106.

Yet mutants in the gene have a measurable growth advantage, contrary to the expectation that MSHA may be beneficial to exploiting *Fucus* resources. We speculate that while the *Fucus* condition induces expression of MSHA, the experimental conditions are sufficiently rich in diffusible nutrients to allow for planktonic growth, and that a colonization factor is unnecessary, and in fact, detrimental. We hypothesize that the stimulated synthesis of MSHA is energetically costly, without the benefit of preferential access to resources an attached lifestyle might confer in the natural environment.

### 3.3.12 Intergenic regions

We hypothesized that IG regions might also confer fitness effects as regulatory regions (promoters or terminators), or non-coding RNA. To get a sense of the split between these types, we calculated for chromosomes 1 and 2 the number of fitness determining IG regions adjacent to a fitness determining gene (15; 11 of which were 5' of the gene, and four, 3' of the gene), and the number not adjacent to such a gene (25). The observation that 63% of IG regions appear to be independently affecting fitness suggests a potential for non-coding small RNAs, which have been implicated in niche adaptation (Shi et al., 2009), that will be investigated in future work.

## 3.4 Conclusion

Our results suggest that colonization of different habitats contributes to diverse metabolic pathways in a single organism. They also provide a quantitative context for why genomes of copiotrophs like the *Vibrio*, which take advantage of nutrient pulses to rapidly expand their populations, are large and varied, while oligotrophs, which abide in nutrient poor environments, are small and streamlined (Lauro et al., 2009).

This study provides a first step toward a genome-wide understanding of adaptation to marine habitats. In future work, looking at the dynamic expression of the genome with a complementary method, such as proteomics, would be informative to elaborate the insights and hypotheses described here, such as the nature of catabolic pathways utilized for growth on *Apocyclops* nutrients.

Finally, this study relies on gene annotations to infer the physiological meaning of a change in mutant abundance; however, it should be noted that the study itself furthers the annotations of these genes, which can be used in future work. For instance, exploring the conservation of genes identified here as adaptive within the *Vibrio* sp. F13 population — e.g., the partitioning of these genes between the core and flexible genomes — will shed light on the extent and tempo of this population’s evolution with respect to potential habitat adaptation.

## 3.5 Materials and Methods

### 3.5.1 Strains and mutagenesis plasmid

*Vibrio* sp. F13 13 strain 9CS106, isolated from the stomach of a male *Hemigrapsus sanguineus* crab collected from Plum Island Sound Estuary, Ipswich, MA on Oct. 3, 2007, was grown in marine broth 2216 (Difco) at room temperature prior to conjugation. *E. coli* strain, EC3-4, containing the *mariner* transposon plasmid pSW25T::C9::ISCAT, was grown in Luria-Bertani (LB) (Difco) with 12.5  $\mu\text{g}/\text{ml}$  chloramphenicol (Cm) and 0.3 mM diaminopimelate (DAP) at 37°C. The *mariner* transposon was modified to include MmeI restriction enzyme recognition sites for Tn-seq. For more detail on the construction of the plasmid, see Section 3.6.1.

### 3.5.2 Transposon mutagenesis

To generate a single mutant library for all experiments, 9CS106 was mutagenized by conjugation with *E. coli* donor EC3-4. Conjugation was done between exponentially growing cultures. To inoculate, *E. coli* cells were taken from a freshly streaked plate, while 9CS106 cells were taken from a 30-day old streaked plate, grown at room temperature the first day, and stored at 4°C until used for inoculation. *E. coli* was grown in LB+DAP+Cm12.5 at 37°C, and 9CS106 was grown in 2216 at room temperature. After 7.5 hr, *E. coli* was

Table 3.3: Log<sub>2</sub> fold change for every identified fitness determinant

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consus_4	-		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Cysteine Biosynthesis	Sulfite reductase [NADPH] hemoprotein beta-component (EC 1.8.1.2)	1.12	-0.19	-2.38	-3.88	-3.29
9CS106_Chr1_consus_5	-		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Cysteine Biosynthesis	Sulfite reductase [NADPH] flavoprotein alpha-component (EC 1.8.1.2)	0.74	-0.11	-1.79	-2.21	-2.72
IG_9CS106_Chr1_consus_8			IG	IG	IG	IG_9CS106_Chr1_consus_8	-1.69	-0.50	-0.09	0.43	-3.76
9CS106_Chr1_consus_14	-		No category		No subsystem	Unannotated	0.08	0.18	1.10	0.34	0.15
9CS106_Chr1_consus_18	-		Carbohydrates	Central carbohydrate metabolism	Glycolysis and Gluconeogenesis	Glucose-6-phosphate isomerase (EC 5.3.1.9)	-2.02	-3.92	-0.15	-1.98	-9.82
IG_9CS106_Chr1_consus_25			IG	IG	IG	IG_9CS106_Chr1_consus_25	-2.84	-1.95	-1.74	0.28	-0.44
9CS106_Chr1_consus_28	-		RNA Metabolism	RNA processing and modification	RNA processing and degradation, bacterial	3'-to-5' exonuclease RNase R	-0.44	-0.58	-0.92	-0.56	-0.12
9CS106_Chr1_consus_36	-		Stress Response		Hfl operon	RNA-binding protein Hfq	-1.30	-1.80	-0.78	-2.22	-2.37
9CS106_Chr1_consus_37	-	from No subsystem to tRNA modification	RNA Metabolism	RNA processing and modification	tRNA modification	tRNA dimethylallyltransferase (EC 2.5.1.75)	-1.50	-1.26	-2.17	-1.76	-1.35
9CS106_Chr1_consus_54	-		Protein Metabolism	Protein biosynthesis	Translation elongation factors bacterial	Translation elongation factor P	-0.09	-0.99	-1.76	-0.53	-0.34
9CS106_Chr1_consus_69	+	from Methicillin resistance in Staphylococci to O-antigen biosynthesis	Cell Wall and Capsule	Gram-Negative cell wall components	O-antigen biosynthesis	Undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.-)	0.76	0.05	1.80	0.23	0.14
IG_9CS106_Chr1_consus_70			IG	IG	IG	IG_9CS106_Chr1_consus_70	-0.13	0.42	1.44	0.89	0.08
IG_9CS106_Chr1_consus_88			IG	IG	IG	IG_9CS106_Chr1_consus_88	1.90	1.63	0.54	1.40	2.25
9CS106_Chr1_consus_91	+		Cell Wall and Capsule	Gram-Negative cell wall components	LOS core oligosaccharide biosynthesis	ADP-L-glycero-D-mannoheptose-6-epimerase (EC 5.1.3.20)	-1.08	-3.41	-0.89	-2.59	-0.93
9CS106_Chr1_consus_97	-		Cell Wall and Capsule	Gram-Negative cell wall components	LOS core oligosaccharide biosynthesis	Putative two-domain glycosyltransferase	-1.54	-3.74	-1.73	-0.29	-1.74
9CS106_Chr1_consus_104	-	from No subsystem to O-antigen biosynthesis	Cell Wall and Capsule	Gram-Negative cell wall components	O-antigen biosynthesis	Unannotated	-0.24	-2.55	-0.29	0.72	0.05
9CS106_Chr1_consus_106	-		No category		No subsystem	FIG00920623: hypothetical protein	-0.03	-0.15	-0.11	-1.54	-0.43
9CS106_Chr1_consus_107	-	from Unannotated to RP-L33	Protein Metabolism	Protein biosynthesis	Ribosome biogenesis bacterial	RP-L33, large subunit ribosomal protein L33	-2.27	-2.09	-2.69	-2.93	-0.77
9CS106_Chr1_consus_113	-		Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis	Orotate phosphoribosyltransferase (EC 2.4.2.10)	-0.08	-1.05	-1.37	-8.42	-8.60
9CS106_Chr1_consus_146	+	from No subsystem to Stringent response	Regulation and Cell signaling		Stringent Response, (p)ppGpp metabolism	GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp synthetase II / Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2)	-1.21	-1.51	-1.95	-1.48	-1.17
9CS106_Chr1_consus_171	+	from No subsystem to Protein chaperones	Protein Metabolism	Protein folding	Protein chaperones	33 kDa chaperonin (Heat shock protein 33) (HSP33)	-2.99	-2.99	-2.81	-2.65	-0.93
9CS106_Chr1_consus_180	+	from Peptidoglycan Biosynthesis to Ammonia assimilation	Nitrogen Metabolism		Ammonia assimilation	Glutamine synthetase type I (EC 6.3.1.2)	0.24	-0.08	-1.74	-0.99	-0.75
9CS106_Chr1_consus_182	+	from No subsystem to Ammonia assimilation	Nitrogen Metabolism		Ammonia assimilation	Nitrogen regulation protein NtrB (EC 2.7.13.3)	-0.59	-0.22	-0.97	-2.21	-2.63
9CS106_Chr1_consus_183	+		Nitrogen Metabolism		Ammonia assimilation	Nitrogen regulation protein NtrI)	-1.08	0.05	-1.20	-2.20	-2.17



Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consens us_220	+		Stress Response	Oxidative stress	Glutathione: Redox cycle	Glutathione reductase (EC 1.8.1.7)	-0.64	-1.35	-1.35	-2.33	-0.29
9CS106_Chr1_consens us_237	-		Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	Ketol-acid reductoisomerase (EC 1.1.1.86)	-0.14	-0.42	-4.47	-7.44	-10.5
9CS106_Chr1_consens us_238	+	from Alanine Biosynthesis to Branched-chain Amino Acid Biosynthesis)	Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	HTH-type transcriptional regulator IlvY	-0.11	-0.26	-2.09	-2.61	-3.25
IG_9CS106_Chr1_consens us_247			IG	IG	IG	IG_9CS106_Chr1_consens us_247	-1.21	-0.35	-0.57	-0.40	-1.40
9CS106_Chr1_consens us_265	+	from yidC cluster to tRNA modification	RNA Metabolism	RNA processing and modification	tRNA modification Bacteria	GTPase and tRNA-U34 5-formylation enzyme TrmE	-1.28	-1.15	-1.59	-3.08	-1.27
9CS106_Chr1_consens us_266	+		Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN, FAD	Flavodoxin	Flavoprotein MioC	-2.25	-3.69	-2.48	-0.40	-2.46
9CS106_Chr1_consens us_267	+		RNA Metabolism	RNA processing and modification	tRNA modification Bacteria	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	-1.35	-1.52	-2.05	-3.72	-1.10
9CS106_Chr1_consens us_271	+	from Unannotated to ATP synthase atpI	Respiration	ATP synthases	F0F1-type ATP synthase	atpI; ATP synthase protein I	-0.91	-0.42	-0.17	-0.51	-1.11
9CS106_Chr1_consens us_272	+	from No subsystem to ATP synthase	Respiration	ATP synthases	F0F1-type ATP synthase	ATP synthase F0 sector subunit a	-1.15	-1.53	-1.40	-2.36	-1.98
9CS106_Chr1_consens us_276	+		Respiration	ATP synthases	F0F1-type ATP synthase	ATP synthase alpha chain (EC 3.6.3.14)	-0.85	-1.36	-1.42	-2.23	-1.99
9CS106_Chr1_consens us_278	+		Respiration	ATP synthases	F0F1-type ATP synthase	ATP synthase beta chain (EC 3.6.3.14)	-0.74	-2.13	-1.51	-2.70	-1.41
9CS106_Chr1_consens us_284	-		Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	Dihydroxy-acid dehydratase (EC 4.2.1.9)	0.35	0.14	-3.11	-5.54	-5.76
9CS106_Chr1_consens us_285	-	from alanine biosynthesis to Branched-chain Amino Acid Biosynthesis	Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	Branched-chain amino acid aminotransferase (EC 2.6.1.42)	-0.25	-0.07	-0.99	-1.26	-1.17
9CS106_Chr1_consens us_287	-		Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	Acetolactate synthase large subunit (EC 2.2.1.6)	0.63	-0.13	-2.84	-2.92	-1.89
9CS106_Chr1_consens us_290	-		Protein Metabolism	Protein folding	Periplasmic disulfide interchange	Periplasmic thiol:disulfide interchange protein DsbA	-1.01	0.31	0.85	-1.32	-0.32
IG_9CS106_Chr1_consens us_291			IG	IG	IG	IG_9CS106_Chr1_consens us_291	-0.43	0.38	1.45	-0.27	-0.45
9CS106_Chr1_consens us_296	+	from No subsystem to Transcription factor	RNA Metabolism	Transcription	Transcription factors bacterial	Hypothetical Transcriptional Regulator	-0.53	0.19	1.26	0.84	-0.77
9CS106_Chr1_consens us_300	-	from conserved gene cluster to Potassium uptake	Ion Metabolism		Potassium ion uptake	Trk system potassium uptake protein TrkA	-0.83	-2.29	-2.02	-2.95	-1.09
9CS106_Chr1_consens us_301	-	from clustering based subsystem to Ribosome biogenesis	Protein Metabolism	Protein biosynthesis	Ribosome biogenesis bacterial	Ribosomal RNA small subunit methyltransferase B (EC 2.1.1.-)	-0.26	-0.43	-1.05	-0.48	-0.17
9CS106_Chr1_consens us_303	-	from clustering based subsystem to Protein biosynthesis	Protein Metabolism	Protein biosynthesis	Peptide processing	Peptide deformylase (EC 3.5.1.88)	-1.74	-1.54	-2.95	0.05	-0.33
9CS106_Chr1_consens us_311	+		Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Heme and Siroheme Biosynthesis	Coproporphyrinogen III oxidase, aerobic (EC 1.3.3.3)	-1.39	-5.00	-4.02	-1.55	-5.03
IG_9CS106_Chr1_consens us_312			IG	IG	IG	IG_9CS106_Chr1_consens us_312	-0.49	-0.29	-0.33	-0.69	-1.50

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginata	Glucose
9CS106_Chr1_consensus_312	+	from clustering based subsystem to Aromatic amino acid biosynthesis	Amino Acids and Derivatives	Aromatic amino acids and derivatives	Aromatic amino acid biosynthesis	Shikimate 5-dehydrogenase I alpha (EC 1.1.1.25)	-0.90	-0.63	-1.28	-2.20	-2.25
IG_9CS106_Chr1_consensus_331			IG	IG	IG	IG_9CS106_Chr1_consensus_331	-1.30	-0.14	0.56	0.28	-1.00
9CS106_Chr1_consensus_335	-		No category		No subsystem	Putative signal peptide protein	-0.26	-0.52	-0.81	-1.46	-0.25
IG_9CS106_Chr1_consensus_348			IG	IG	IG	IG_9CS106_Chr1_consensus_348	-0.78	-0.20	-0.23	-0.17	-1.31
9CS106_Chr1_consensus_349	+		RNA Metabolism	Transcription	Transcription factors bacterial	Transcription termination factor Rho	-1.77	-2.10	-0.01	0.01	-2.75
9CS106_Chr1_consensus_351	+	from No category to Respiration	Respiration		Ubiquinone biosynthesis -- gjo	3-polyprenyl-4-hydroxybenzoate carboxylase (EC 4.1.1.-)	-1.79	-0.70	-0.96	-0.14	-1.36
9CS106_Chr1_consensus_360	+		Regulation and Cell signaling		cAMP signaling in bacteria	Adenylate cyclase (EC 4.6.1.1)	-0.79	-0.60	-0.15	-1.92	0.09
9CS106_Chr1_consensus_362	+	from clustering-based subsystem to Lysine biosynthesis	Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Lysine biosynthesis	Diaminopimelate decarboxylase (EC 4.1.1.20)	0.19	0.01	-1.04	-2.38	-2.12
9CS106_Chr1_consensus_392	+		Carbohydrates	Sugar alcohols	Glycerol and Glycerol-3-phosphate Uptake and Utilization	Thiosulfate sulfurtransferase GlpE (EC 2.8.1.1)	-0.71	-0.45	0.05	0.26	-1.36
9CS106_Chr1_consensus_395	+	from No category to Respiration	Respiration		Ubiquinone biosynthesis -- gjo	Chorismate--pyruvate lyase (EC 4.1.3.40)	-1.11	-0.64	-0.88	-2.38	-4.06
9CS106_Chr1_consensus_397	-		Fatty Acids, Lipids, and Isoprenoids	Phospholipids	Glycerolipid and Glycerophospholipid Metabolism in Bacteria	Glycerol-3-phosphate acyltransferase (EC 2.3.1.15)	-1.95	-1.94	-3.08	-0.52	0.07
IG_9CS106_Chr1_consensus_398			IG	IG	IG	IG_9CS106_Chr1_consensus_398	-0.86	-0.80	-1.23	-0.14	0.45
9CS106_Chr1_consensus_406	+		Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Coenzyme B12 biosynthesis	Outer membrane vitamin B12 receptor BtuB	-0.21	-0.11	-0.96	-4.45	-2.51
IG_9CS106_Chr1_consensus_423			IG	IG	IG	IG_9CS106_Chr1_consensus_423	-1.85	-1.48	-1.26	-1.11	-0.87
9CS106_Chr1_consensus_437	-	from Ligases that form carbon-nitrogen bonds to De Novo Purine Biosynthesis	Nucleosides and Nucleotides	Purines	De Novo Purine Biosynthesis	Phosphoribosylamine--glycine ligase (EC 6.3.4.13)	0.31	0.09	-2.54	-5.91	-7.85
IG_9CS106_Chr1_consensus_438			IG	IG	IG	IG_9CS106_Chr1_consensus_438	-0.17	-0.22	-0.23	-0.79	-1.42
IG_9CS106_Chr1_consensus_439			IG	IG	IG	IG_9CS106_Chr1_consensus_439	-0.83	-0.34	-0.67	-1.48	-1.72
9CS106_Chr1_consensus_440	-		DNA Metabolism		DNA structural proteins, bacterial	DNA-binding protein Fis	-0.42	-1.06	-2.08	-1.17	-0.36
9CS106_Chr1_consensus_445	-		Metabolism of Aromatic Compounds	Peripheral pathways for catabolism of aromatic compounds	Quinate degradation	3-dehydroquinate dehydratase II (EC 4.2.1.10)	-1.59	-1.63	-1.98	-7.63	-7.73
9CS106_Chr1_consensus_479	+		Carbohydrates	Central carbohydrate metabolism	Glycolysis and Gluconeogenesis	6-phosphofruktokinase (EC 2.7.1.11)	-1.50	-1.87	-2.41	-2.18	-2.56
IG_9CS106_Chr1_consensus_480			IG	IG	IG	IG_9CS106_Chr1_consensus_480	-0.07	0.28	0.49	0.48	-1.71
9CS106_Chr1_consensus_499	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Methionine Biosynthesis	Cystathionine gamma-synthase (EC 2.5.1.48)	0.37	0.11	-2.77	-5.90	-8.69
9CS106_Chr1_consensus_503	-		Carbohydrates	Central carbohydrate metabolism	Pyruvate metabolism I: anaerobic reactions, PEP	Phosphoenolpyruvate carboxylase (EC 4.1.1.31)	0.31	0.10	-1.20	-0.49	-0.73
IG_9CS106_Chr1_consensus_505			IG	IG	IG	IG_9CS106_Chr1_consensus_505	0.55	0.25	-0.80	-2.39	-4.97
9CS106_Chr1_consensus_505	+		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Biosynthesis extended	N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)	-0.01	0.04	-0.84	-2.79	-2.53

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginata	Glucose
9CS106_Chr1_consens us_506	+		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Biosynthesis extended	Acetylglutamate kinase (EC 2.7.2.8)	-0.02	-0.05	-1.44	-8.56	-8.73
9CS106_Chr1_consens us_507	+	from EC 6.3.4 to Arginine Biosynthesis	Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Biosynthesis extended	Argininosuccinate synthase (EC 6.3.4.5)	0.14	0.00	-0.26	-2.52	-2.49
9CS106_Chr1_consens us_508	+	from No subsystem to Arginine Biosynthesis	Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Biosynthesis extended	Argininosuccinate lyase (EC 4.3.2.1) / N-acetylglutamate synthase (EC 2.3.1.1)	-0.48	-0.38	-0.55	-1.79	-3.96
9CS106_Chr1_consens us_519	+	from No subcategory to Aromatic amino acid biosynthesis	Amino Acids and Derivatives	Aromatic amino acids and derivatives	Shikimate kinase containing cluster	Shikimate kinase I (EC 2.7.1.71)	-1.22	-1.36	-2.07	-2.63	-2.72
9CS106_Chr1_consens us_523	+		DNA Metabolism	DNA repair	2-phosphoglycolate salvage	Ribulose-phosphate 3-epimerase (EC 5.1.3.1)	-0.70	-1.05	-2.00	-0.41	-0.22
IG_9CS106_Chr1_consens us_524			IG	IG	IG	IG_9CS106_Chr1_consens us_524	-1.28	-0.76	-0.20	0.09	-1.70
9CS106_Chr1_consens us_524	+		DNA Metabolism	DNA repair	2-phosphoglycolate salvage	Phosphoglycolate phosphatase (EC 3.1.3.18)	0.03	-0.38	-1.11	-0.52	-1.61
9CS106_Chr1_consens us_527	+		Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate and pterines	Folate Biosynthesis	Para-aminobenzoate synthase, amidotransferase component (EC 2.6.1.85)	-0.40	-0.30	-0.38	-0.43	-2.14
9CS106_Chr1_consens us_528	+		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Biosynthesis extended	Acetylornithine aminotransferase (EC 2.6.1.11) / N-succinyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.17) / Succinylornithine transaminase (EC 2.6.1.81)	-0.30	-0.53	-2.59	-7.44	-7.15
IG_9CS106_Chr1_consens us_543			IG	IG	IG	IG_9CS106_Chr1_consens us_543	-0.03	-0.30	-1.64	-0.69	-0.13
IG_9CS106_Chr1_consens us_552			IG	IG	IG	IG_9CS106_Chr1_consens us_552	-1.14	-0.18	-0.48	-0.07	-1.51
IG_9CS106_Chr1_consens us_554			IG	IG	IG	IG_9CS106_Chr1_consens us_554	-0.91	-1.16	-1.48	-1.21	-0.61
9CS106_Chr1_consens us_595	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Cysteine Biosynthesis	Sulfate adenyltransferase subunit 2 (EC 2.7.7.4)	0.01	-0.01	0.05	-1.23	-1.76
9CS106_Chr1_consens us_596	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Cysteine Biosynthesis	Sulfate adenyltransferase subunit 1 (EC 2.7.7.4)	-0.17	-0.16	-0.86	-1.53	-2.27
9CS106_Chr1_consens us_598	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Cysteine Biosynthesis	Adenylylsulfate kinase (EC 2.7.1.25)	0.06	-0.37	-0.70	-1.53	-3.25
9CS106_Chr1_consens us_609	+	from uncharacterized protein ytlN to tamB, outer membrane transporter	Outer membrane transport		Translocation and assembly module TamB	tamB; translocation and assembly module TamB	-1.18	-0.30	0.41	-0.12	0.22
9CS106_Chr1_consens us_611	-	from putative ABC transporter to TamB	Outer membrane transport		Translocation and assembly module TamB	Fructose-1,6-bisphosphatase, type I (EC 3.1.3.11)	-0.09	-0.11	0.47	-6.20	0.06
9CS106_Chr1_consens us_617	-		Protein Metabolism	Protein degradation	Putative TldE-TldD proteolytic complex	FIG138315: Putative alpha helix protein	-2.60	-2.63	-1.92	-0.76	-2.04
IG_9CS106_Chr1_consens us_618			IG	IG	IG	IG_9CS106_Chr1_consens us_618	-1.36	-0.61	0.00	-0.27	-0.69
9CS106_Chr1_consens us_619	-	from No subsystem to Mg ion transport	Ion Metabolism		Mg ion transport	Magnesium transporter	-0.76	-1.18	-1.88	-1.06	0.04
9CS106_Chr1_consens us_621	-		No category		No subsystem	Hypothetical ATP-binding protein UPF0042, contains P-loop	-0.11	0.24	-1.58	-0.87	-0.31
9CS106_Chr1_consens us_624	-		Motility and Chemotaxis	Flagellar motility in Prokaryota	Flagellum	RNA polymerase sigma-54 factor RpoN	-0.97	-0.20	-1.36	-2.76	-2.11

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginata	Glucose
9CS106_Chr1_consens us_628	-		Cell Wall and Capsule	Gram-Negative cell wall components	KDO2-Lipid A biosynthesis	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)	-1.49	-2.50	-1.98	-0.24	-0.23
9CS106_Chr1_consens us_639	-		No category		No subsystem	Unannotated	-0.37	-0.22	-0.72	-3.77	-1.21
9CS106_Chr1_consens us_642	-		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Deiminase Pathway	Ornithine carbamoyltransferase (EC 2.1.3.3)	0.17	0.00	-1.22	-2.67	-2.93
9CS106_Chr1_consens us_644	+		RNA Metabolism	RNA processing and modification	RNA processing and degradation, bacterial	Ribonuclease E inhibitor RraB	-1.03	-1.28	-0.49	0.47	-0.22
9CS106_Chr1_consens us_670	-		Amino Acids and Derivatives	Proline and 4-hydroxyproline	Proline Synthesis	Pyroline-5-carboxylate reductase (EC 1.5.1.2)	0.16	0.42	-0.11	-2.63	-2.11
9CS106_Chr1_consens us_677	-		Stress Response	Oxidative stress	Glutaredoxins	Glutathione synthetase (EC 6.3.2.3)	-1.66	-2.01	-3.57	-1.10	-1.69
9CS106_Chr1_consens us_684	+		Cofactors, Vitamins, Prosthetic Groups, Pigments	Pyridoxine	Pyridoxin (Vitamin B6) Biosynthesis	D-erythrose-4-phosphate dehydrogenase (EC 1.2.1.72)	0.06	-0.29	-1.24	-3.68	-2.19
IG_9CS106_Chr1_consens us_686			IG	IG	IG	IG_9CS106_Chr1_consens us_686	-0.35	-0.84	-1.96	-0.87	-1.80
9CS106_Chr1_consens us_691	-		Amino Acids and Derivatives	Alanine, serine, and glycine	Serine Biosynthesis	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	0.36	-0.49	-2.74	-4.71	-4.55
9CS106_Chr1_consens us_692	-		Carbohydrates	Central carbohydrate metabolism	Glycolate, glyoxylate interconversions	Ribose 5-phosphate isomerase A (EC 5.3.1.6)	-2.83	-2.83	-2.77	-2.83	-2.83
9CS106_Chr1_consens us_693	-		Carbohydrates	One-carbon Metabolism	One-carbon metabolism by tetrahydropterines	5-formyltetrahydrofolate cyclo-ligase (EC 6.3.3.2)	0.39	0.06	-1.36	-2.64	-1.61
9CS106_Chr1_consens us_707	+		Protein Metabolism	Protein biosynthesis	Translation elongation factors bacterial	Translation elongation factor LepA	-1.69	-1.84	-1.73	-1.22	-0.64
9CS106_Chr1_consens us_714	-		Membrane Transport	Protein secretion system, Type III	Type III secretion system orphans	BarA sensory histidine kinase (= VarS = GacS)	-0.51	-1.25	-1.82	-3.44	-3.41
9CS106_Chr1_consens us_716	+		Regulation and Cell signaling		Stringent Response, (p)ppGpp metabolism	GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp synthetase I	0.08	-0.24	-1.08	-1.34	-0.43
9CS106_Chr1_consens us_730	+		DNA Metabolism	DNA repair	DNA repair, bacterial UmuCD system	RecA protein	-0.51	-0.78	-2.54	-1.08	-0.85
9CS106_Chr1_consens us_734	+		Respiration	Sodium Ion-Coupled Energetics	Na <sup>+</sup> translocating decarboxylases and related biotin-dependent enzymes	Oxaloacetate decarboxylase gamma chain (EC 4.1.1.3)	0.01	-0.41	1.01	-0.19	-1.74
9CS106_Chr1_consens us_735	+		Respiration	Sodium Ion-Coupled Energetics	Na <sup>+</sup> translocating decarboxylases and related biotin-dependent enzymes	Oxaloacetate decarboxylase alpha chain (EC 4.1.1.3)	0.24	0.14	0.12	-0.08	-1.07
9CS106_Chr1_consens us_736	+		Respiration	Sodium Ion-Coupled Energetics	Na <sup>+</sup> translocating decarboxylases and related biotin-dependent enzymes	Oxaloacetate decarboxylase beta chain (EC 4.1.1.3)	-0.01	0.23	-0.09	-0.16	-1.47
9CS106_Chr1_consens us_761	+		Carbohydrates	Central carbohydrate metabolism	Dehydrogenase complexes	Transcriptional repressor for pyruvate dehydrogenase complex	-0.23	0.02	0.19	-1.18	-1.05
9CS106_Chr1_consens us_763	+		Carbohydrates	Central carbohydrate metabolism	Dehydrogenase complexes	Dihydrolypoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	-1.62	-1.71	0.77	0.65	-2.39
9CS106_Chr1_consens us_767	-		Regulation and Cell signaling		Zinc regulated enzymes	Carbonic anhydrase (EC 4.2.1.1)	-4.35	-0.13	-0.10	-2.79	0.43
9CS106_Chr1_consens us_781	+		No category		No subsystem	Aconitate hydratase 2 (EC 4.2.1.3) @ 2-methylisocitrate dehydratase (EC 4.2.1.99)	-1.97	-1.65	-1.71	-1.65	-2.43
9CS106_Chr1_consens us_788	+		Iron acquisition and metabolism		Transport of Iron	Ferric iron ABC transporter, iron-binding protein	-0.33	-1.95	-2.74	-1.57	-2.73
9CS106_Chr1_consens us_790	+		Iron acquisition and metabolism		Transport of Iron	Ferric iron ABC transporter, ATP-binding protein	-0.58	-1.61	-0.37	-1.73	-2.94

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginata	Glucose
9CS106_Chr1_consens us_814	+		RNA Metabolism	Transcription	Transcription factors bacterial	Transcription elongation factor GreA	-0.64	-1.42	-2.31	-1.10	-0.72
9CS106_Chr1_consens us_820	+	from Murein hydrolase regulator of cell death to protein transmembrane transporter (GO term)	Protein Metabolism	Protein secretion	Protein transmembrane transport	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	-0.69	0.31	1.19	-2.02	0.24
IG_9CS106_Chr1_consens us_821			IG	IG	IG	IG_9CS106_Chr1_consens us_821	-1.47	-0.89	-0.11	0.02	-0.91
9CS106_Chr1_consens us_822	+	from clustering based subsystem to Protein biosynthesis	Protein Metabolism	Protein biosynthesis	NusA-TFII Cluster	Transcription termination protein NusA	-1.77	-1.25	-1.61	-2.04	-1.12
9CS106_Chr1_consens us_824	+		Protein Metabolism	Protein biosynthesis	Translation initiation factors bacterial	Ribosome-binding factor A	-2.55	-1.38	-0.99	-0.40	-1.27
IG_9CS106_Chr1_consens us_834			IG	IG	IG	IG_9CS106_Chr1_consens us_834	0.16	0.36	1.27	0.48	-0.53
9CS106_Chr1_consens us_838	+		Protein Metabolism	Protein biosynthesis	Translation termination factors bacterial	Peptide chain release factor 3	-0.59	-1.49	-0.99	0.14	-0.04
9CS106_Chr1_consens us_850	+		Amino Acids and Derivatives	Alanine, serine, and glycine	Serine Biosynthesis	Phosphoserine phosphatase (EC 3.1.3.3)	0.25	0.05	-4.54	-6.79	-6.93
IG_9CS106_Chr1_consens us_879			IG	IG	IG	IG_9CS106_Chr1_consens us_879	-1.42	-0.15	-0.02	-0.13	-0.94
9CS106_Chr1_consens us_903	+		No category		No subsystem	Glutamate 5-kinase (EC 2.7.2.11) / RNA-binding C-terminal domain PUA	0.06	0.04	-0.60	-4.13	-8.93
9CS106_Chr1_consens us_904	+		Amino Acids and Derivatives	Proline and 4-hydroxyproline	Proline Synthesis	Gamma-glutamyl phosphate reductase (EC 1.2.1.41)	0.42	0.26	-0.53	-3.33	-5.05
9CS106_Chr1_consens us_918	+		RNA Metabolism	RNA processing and modification	tRNA modification Bacteria	Thiamine biosynthesis protein thil	-2.46	-1.49	-1.31	-0.94	-1.84
9CS106_Chr1_consens us_939	+		Amino Acids and Derivatives	Alanine, serine, and glycine	Serine Biosynthesis	Serine hydroxymethyltransferase (EC 2.1.2.1)	-0.04	-0.48	-1.83	-2.36	-2.33
9CS106_Chr1_consens us_966	+		Stress Response	Periplasmic Stress	Periplasmic Stress Response	Membrane-associated zinc metalloprotease	-1.67	-1.04	-0.38	-2.45	-1.34
IG_9CS106_Chr1_consens us_967			IG	IG	IG	IG_9CS106_Chr1_consens us_967	-1.00	-0.71	-0.80	-2.01	-1.01
9CS106_Chr1_consens us_990	-		Nucleosides and Nucleotides	Purines	De Novo Purine Biosynthesis	Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2)	-0.14	-0.19	-0.49	-2.63	-2.14
9CS106_Chr1_consens us_1007	-	from No subsystem to Peptidoglycan biosynthesis	Cell Wall and Capsule		Peptidoglycan Biosynthesis	FIG009095: D,D-carboxypeptidase family protein	0.78	-1.81	0.07	0.65	0.57
9CS106_Chr1_consens us_1018	-		Membrane Transport	Sugar Phosphotransferase Systems, PTS	Sucrose-specific PTS	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	-0.83	-1.38	-3.03	-2.79	-1.52
IG_9CS106_Chr1_consens us_1019			IG	IG	IG	IG_9CS106_Chr1_consens us_1019	-0.57	-0.57	-1.89	-0.58	-3.06
9CS106_Chr1_consens us_1046	-		Amino Acids and Derivatives	Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	Asparagine synthetase [glutamine-hydrolyzing] (EC 6.3.5.4)	-0.27	-0.67	-0.07	-5.29	-6.70
9CS106_Chr1_consens us_1049	-		Carbohydrates	Aminosugars	Chitin and N-acetylglucosamine utilization	N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	-1.03	-0.91	-1.09	-5.01	-1.08
9CS106_Chr1_consens us_1061	+	from SeqA and Co-occurring genes to Glycogenolysis / glycogenesis	Carbohydrates		Glycogenolysis and glycogenesis	Phosphoglucomutase (EC 5.4.2.2)	-1.40	-1.23	-0.60	-1.66	-1.13
9CS106_Chr1_consens us_1064	-		Respiration		Biogenesis of c-type cytochromes	Citrate synthase (si) (EC 2.3.3.1)	1.37	-0.58	-0.05	-2.46	-2.40

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consens us_1067	+		Respiration	Electron donating reactions	Succinate dehydrogenase	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	0.55	-0.72	0.07	-2.21	0.44
9CS106_Chr1_consens us_1069	+		Carbohydrates	Central carbohydrate metabolism	Dehydrogenase complexes	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)	0.17	-0.10	-0.62	-1.96	-3.18
9CS106_Chr1_consens us_1070	+		Carbohydrates	Central carbohydrate metabolism	Dehydrogenase complexes	Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)	0.52	0.41	-0.83	-3.03	-4.33
9CS106_Chr1_consens us_1108	+		Amino Acids and Derivatives	Aromatic amino acids and derivatives	Chorismate Synthesis	Chorismate synthase (EC 4.2.3.5)	-0.99	-1.02	-1.91	-2.25	-2.29
9CS106_Chr1_consens us_1124	+	<i>from Colicin V and Bacteriocin Production Cluster to De Novo Purine Biosynthesis</i>	Nucleosides and Nucleotides	Purines	De Novo Purine Biosynthesis	Amidophosphoribosyltransferase (EC 2.4.2.14)	0.29	0.53	-2.13	-4.27	-7.82
9CS106_Chr1_consens us_1185	+		Regulation and Cell signaling	Quorum sensing and biofilm formation	Quorum-sensing in Vibrio	Regulatory protein LuxO	-0.25	-0.33	-0.68	-1.34	-0.48
9CS106_Chr1_consens us_1199	-		Carbohydrates	Fermentation	Fermentations: Mixed acid	Phosphate acetyltransferase (EC 2.3.1.8)	-2.85	-1.77	-2.64	-0.74	-0.31
9CS106_Chr1_consens us_1200	-		Carbohydrates	Fermentation	Fermentations: Mixed acid	Acetate kinase (EC 2.7.2.1)	-1.78	-1.05	-2.15	-0.85	-0.79
9CS106_Chr1_consens us_1206	-		Protein Metabolism	Protein folding	Periplasmic disulfide interchange	Periplasmic thiol:disulfide oxidoreductase DsbB, required for DsbA reoxidation	-1.11	0.50	0.38	-1.65	-0.63
9CS106_Chr1_consens us_1208	+		Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty acid degradation regulons	Transcriptional regulator for fatty acid degradation FadR, GntR family	-2.52	-1.68	0.15	-0.96	-1.08
9CS106_Chr1_consens us_1213	+		DNA Metabolism	DNA repair	DNA repair, bacterial	A/G-specific adenine glycosylase (EC 3.2.2.-)	-0.21	0.09	0.04	-1.95	-0.48
9CS106_Chr1_consens us_1214	-		Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Cobalamin synthesis	Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	0.39	0.03	-0.95	-2.48	-1.52
9CS106_Chr1_consens us_1220	+		No category		No subsystem	COG1399 protein, clustered with ribosomal protein L32p	-2.48	-3.08	-2.38	-0.81	-0.75
9CS106_Chr1_consens us_1224	+	<i>from mycolic acid synthesis to Fatty Acid Biosynthesis</i>	Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty acid biosynthesis	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)	-2.16	-1.99	-2.46	-2.40	-1.13
9CS106_Chr1_consens us_1227	+	<i>from No subsystem to Fatty Acid Biosynthesis</i>	Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty acid biosynthesis	3-oxoacyl-[acyl-carrier-protein] synthase, KASII (EC 2.3.1.179)	-1.18	-1.31	-1.02	-1.23	-1.58
9CS106_Chr1_consens us_1233	+	<i>from subsuby not found to Glycolysis</i>	Carbohydrates	Central carbohydrate metabolism	Glycolysis and Gluconeogenesis	PTS system, glucose-specific IIB component (EC 2.7.1.69) / PTS system, glucose-specific IIC component	0.05	0.28	-0.34	0.09	-3.04
IG_9CS106_Chr1_consens us_1271			IG	IG	IG	IG_9CS106_Chr1_consens us_1271	-0.62	-1.33	-0.42	-0.63	-1.85
9CS106_Chr1_consens us_1271	+		Cell Wall and Capsule	Gram-Negative cell wall components	Major Outer Membrane Proteins	Outer membrane protein OmpT	-0.98	-2.58	-1.90	-2.63	-3.97
9CS106_Chr1_consens us_1299	+		DNA Metabolism		DNA structural proteins, bacterial	Chromosome partition protein MukB	-1.87	-0.70	0.36	-0.09	0.15
IG_9CS106_Chr1_consens us_1305			IG	IG	IG	IG_9CS106_Chr1_consens us_1305	0.07	0.52	-2.56	-1.91	0.04
9CS106_Chr1_consens us_1320	+		Membrane Transport		Ton and Tol transport systems	tolB protein precursor, periplasmic protein involved in the tonB-independent uptake of group A colicins	-0.44	-1.92	-1.26	-1.34	-1.81

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consens us_1321	+	from No subsystem to Major Outer Membrane Proteins	Cell Wall and Capsule	Gram-Negative cell wall components	Major Outer Membrane Proteins	18K peptidoglycan-associated outer membrane lipoprotein; Peptidoglycan-associated lipoprotein precursor; Outer membrane protein P6; OmpA/MotB precursor	0.42	0.33	-0.17	0.29	-3.84
IG_9CS106_Chr1_consens us_1328			IG	IG	IG	IG_9CS106_Chr1_consens us_1328	-1.38	-1.06	-1.88	-1.37	-0.60
9CS106_Chr1_consens us_1378	-	from No subsystem to Tryptophan synthesis	Amino Acids and Derivatives	Aromatic amino acids and derivatives	Tryptophan synthesis	Indole-3-glycerol phosphate synthase (EC 4.1.1.48) / Phosphoribosylanthranilate isomerase (EC 5.3.1.24)	0.05	0.37	-0.60	-7.64	-7.81
9CS106_Chr1_consens us_1379	-	from Auxin biosynthesis to Tryptophan synthesis	Amino Acids and Derivatives	Aromatic amino acids and derivatives	Tryptophan synthesis	Anthranilate phosphoribosyltransferase (EC 2.4.2.18)	0.02	0.23	-0.73	-6.80	-7.78
9CS106_Chr1_consens us_1380	-		Amino Acids and Derivatives	Aromatic amino acids and derivatives	Tryptophan synthesis	Anthranilate synthase, amidotransferase component (EC 4.1.3.27)	-0.65	0.10	-0.97	-5.87	-5.98
9CS106_Chr1_consens us_1381	-		Amino Acids and Derivatives	Aromatic amino acids and derivatives	Tryptophan synthesis	Anthranilate synthase, aminase component (EC 4.1.3.27)	0.54	0.31	-0.46	-2.31	-5.40
9CS106_Chr1_consens us_1429	+		No category		No subsystem	FIG01199550: hypothetical protein	-2.05	-0.58	-0.44	-0.12	-0.71
9CS106_Chr1_consens us_1442	-		Sulfur Metabolism		Thioredoxin-disulfide reductase	Thioredoxin reductase (EC 1.8.1.9)	0.12	-0.16	-2.10	-0.49	-0.35
9CS106_Chr1_consens us_1480	+		No category		Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	LSU ribosomal protein L35p	-1.88	-1.80	-2.36	-1.47	-0.46
9CS106_Chr1_consens us_1487	-	from No subsystem to Eugenol utilization, Carbon sources	Carbon sources		Eugenol utilization	Aldehyde dehydrogenase (EC 1.2.1.3); Probable conferyl aldehyde dehydrogenase (EC 1.2.1.68)	0.51	0.64	0.84	1.27	0.61
9CS106_Chr1_consens us_1493	-	from No subsystem to D-gluconate etc	Carbohydrates	Monosaccharides	D-gluconate and ketogluconates metabolism	4-Hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) @ 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14)	-0.31	0.29	-0.38	-4.57	0.29
9CS106_Chr1_consens us_1494	-		Carbohydrates	Monosaccharides	D-gluconate and ketogluconates metabolism	2-dehydro-3-deoxygluconate kinase (EC 2.7.1.45)	0.03	-0.01	-0.70	-7.37	-0.08
9CS106_Chr1_consens us_1495	-		Cell Wall and Capsule	Capsular and extracellular polysaccharides	Alginate metabolism	Pectin degradation protein KdgF	0.14	0.00	-0.11	-2.24	0.27
9CS106_Chr1_consens us_1496	+		Cell Wall and Capsule	Capsular and extracellular polysaccharides	Alginate metabolism	poly(beta-D-mannuronate) lyase (EC 4.2.2.3)	-0.15	-0.18	-0.78	-1.73	-0.11
9CS106_Chr1_consens us_1497	+		Cell Wall and Capsule	Capsular and extracellular polysaccharides	Alginate metabolism	poly(beta-D-mannuronate) lyase (EC 4.2.2.3)	-0.35	-0.42	-1.01	-1.79	0.03
9CS106_Chr1_consens us_1508	+		Nitrogen Metabolism		Allantoin Utilization	2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60)	-0.79	-0.24	-1.74	-3.70	-0.43
9CS106_Chr1_consens us_1588	+		Carbohydrates	Monosaccharides	D-gluconate and ketogluconates metabolism	6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44)	-2.05	-2.09	-1.60	-1.49	-3.62
9CS106_Chr1_consens us_1643	+		Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis	Dihydroorotate dehydrogenase (EC 1.3.3.1)	1.00	-0.39	-1.51	-3.39	-6.97
9CS106_Chr1_consens us_1648	+		No category		No subsystem	FIG00920305: hypothetical protein	-2.42	-1.62	-0.49	-1.39	-0.79
9CS106_Chr1_consens us_1848	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Methionine Biosynthesis	Homoserine O-succinyltransferase (EC 2.3.1.46)	0.23	0.86	-1.60	-2.81	-2.83
9CS106_Chr1_consens us_1995	-		Carbohydrates	Central carbohydrate metabolism	TCA Cycle	Fumarate hydratase class I, aerobic (EC 4.2.1.2)	-1.32	0.36	-1.39	-3.27	-1.45
9CS106_Chr1_consens us_2047	+		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Arginine Deiminase Pathway	Arginine/ornithine antiporter ArcD	-2.38	-0.52	-1.20	-1.20	-1.40

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consens us_2114	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Methionine Biosynthesis	Cystathionine beta-lyase (EC 4.4.1.8)	-0.06	0.41	-2.24	-2.28	-0.52
9CS106_Chr1_consens us_2162	-		Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis	Imidazole glycerol phosphate synthase cyclase subunit (EC 4.1.3.-)	0.04	0.34	-2.36	-5.90	-6.02
9CS106_Chr1_consens us_2163	-		Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16)	0.32	0.17	-1.89	-5.45	-5.55
IG_9CS106_Chr1_consens us_2164			IG	IG	IG	IG_9CS106_Chr1_consens us_2164	-0.62	-0.05	-1.57	-2.25	-3.80
9CS106_Chr1_consens us_2164	-		Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis	Imidazole glycerol phosphate synthase amidotransferase subunit (EC 2.4.2.-)	0.30	1.04	-1.78	-6.07	-6.02
9CS106_Chr1_consens us_2165	-	from No subsystem to Histidine Biosynthesis	Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis	Histidinol-phosphatase (EC 3.1.3.15) / Imidazoglycerol-phosphate dehydratase (EC 4.2.1.19)	-0.32	0.53	0.07	-0.80	-3.59
9CS106_Chr1_consens us_2166	-		Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis	Histidinol-phosphate aminotransferase (EC 2.6.1.9)	0.59	-0.03	-2.23	-4.34	-7.84
9CS106_Chr1_consens us_2168	-		Amino Acids and Derivatives	Histidine Metabolism	Histidine Biosynthesis	ATP phosphoribosyltransferase (EC 2.4.2.17)	0.16	0.21	-1.02	-2.17	-2.24
9CS106_Chr1_consens us_2194	-		Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23)	-0.34	-0.54	-1.39	-2.48	-2.65
9CS106_Chr1_consens us_2195	-		Stress Response	Osmotic stress	Osmotic stress cluster	Heat shock (predicted periplasmic) protein YciM, precursor	-3.24	-2.90	-0.82	-2.26	-3.57
9CS106_Chr1_consens us_2257	-		No category		No subsystem	Mobile element protein	-0.03	-0.26	1.25	0.89	-0.34
9CS106_Chr1_consens us_2328	-		No category		No subsystem	Unannotated	-2.34	-1.52	0.47	0.72	-1.93
9CS106_Chr1_consens us_2354	-		Protein Metabolism	Protein degradation	Proteasome bacterial	ATP-dependent protease La (EC 3.4.21.53) Type I	-0.49	-0.27	-1.12	-0.38	-0.84
9CS106_Chr1_consens us_2356	-		Protein Metabolism	Protein degradation	Proteasome bacterial	ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92)	-0.06	-1.21	-1.54	-2.53	-0.21
IG_9CS106_Chr1_consens us_2357			IG	IG	IG	IG_9CS106_Chr1_consens us_2357	-0.76	-0.32	-0.23	-0.29	-1.27
IG_9CS106_Chr1_consens us_2432			IG	IG	IG	IG_9CS106_Chr1_consens us_2432	1.00	0.88	0.48	0.45	2.05
9CS106_Chr1_consens us_2442	-		Motility and Chemotaxis	Flagellar motility in Prokaryota	Flagellum	Flagellar basal-body rod protein FlgG	0.36	-0.17	-0.09	-1.72	-0.48
9CS106_Chr1_consens us_2513	+		Protein Metabolism	Protein biosynthesis	Universal GTPases	GTP-binding and nucleic acid-binding protein YchF	-1.75	-1.65	-1.71	-1.16	-1.29
9CS106_Chr1_consens us_2527	+		Cell Wall and Capsule		Peptidoglycan Biosynthesis	Penicillin-binding protein 2 (PBP-2)	-2.28	-2.30	-0.50	-0.38	-2.12
9CS106_Chr1_consens us_2578	-	from No subsystem to De Novo Purine Biosynthesis	Nucleosides and Nucleotides	Purines	De Novo Purine Biosynthesis	Phosphoribosylformylglycin amidine synthase, synthetase subunit (EC 6.3.5.3) / Phosphoribosylformylglycin amidine synthase, glutamine amidotransferase subunit (EC 6.3.5.3)	0.01	-0.07	-3.27	-7.03	-8.42
9CS106_Chr1_consens us_2590	-		Protein Metabolism	Protein folding	Protein chaperones	Chaperone protein DnaK	-1.12	-1.21	-0.80	-1.76	-0.31
9CS106_Chr1_consens us_2620	-		Nucleosides and Nucleotides	Purines	Purine conversions	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2)	-1.28	-0.81	-2.74	-4.56	-5.41
IG_9CS106_Chr1_consens us_2621			IG	IG	IG	IG_9CS106_Chr1_consens us_2621	-0.95	-0.09	0.17	0.21	-1.27
9CS106_Chr1_consens us_2621	-		Nucleosides and Nucleotides	Purines	Purine conversions	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	-0.03	0.37	-1.41	-2.83	-3.32
9CS106_Chr1_consens us_2633	-		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Urea decomposition	Urea ABC transporter, ATPase protein UrtD	-0.42	0.18	-0.25	-1.81	-1.95
9CS106_Chr1_consens us_2634	-		Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Urea decomposition	Urea ABC transporter, permease protein UrtC	0.34	-0.21	-0.87	-1.56	-1.25



Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consens us_2656	-		RNA Metabolism	Transcription	Rrf2 family transcriptional regulators	Iron-sulfur cluster regulator IscR	-1.03	-0.95	-1.32	-1.79	-1.89
9CS106_Chr1_consens us_2662	-		RNA Metabolism	RNA processing and modification	tRNA modification Bacteria	tRNA-guanine transglycosylase (EC 2.4.2.29)	-0.98	-0.46	-1.35	-0.55	-0.28
9CS106_Chr1_consens us_2688	-		Protein Metabolism	Protein biosynthesis	Ribosome biogenesis bacterial	Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)	-2.16	-2.04	-1.92	-1.64	-2.08
9CS106_Chr1_consens us_2724	-		No category		CBSS-364106.7.pcg.3204	Adenosine (5')-pentaphospho-(5')-adenosine pyrophosphohydrolase (EC 3.6.1.-)	-2.79	-1.09	-2.99	-1.83	-1.40
9CS106_Chr1_consens us_2737	+		RNA Metabolism	RNA processing and modification	ATP-dependent RNA helicases, bacterial	ATP-dependent RNA helicase SrmB	-1.88	-1.79	-2.30	-2.55	-1.57
9CS106_Chr1_consens us_2746	-		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Threonine and Homoserine Biosynthesis	Threonine synthase (EC 4.2.3.1)	0.18	0.62	-1.80	-3.63	-4.31
9CS106_Chr1_consens us_2747	-		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Threonine and Homoserine Biosynthesis	Homoserine kinase (EC 2.7.1.39)	0.00	0.58	-1.45	-2.19	-2.19
9CS106_Chr1_consens us_2748	-		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Threonine and Homoserine Biosynthesis	Aspartokinase (EC 2.7.2.4) / Homoserine dehydrogenase (EC 1.1.1.3)	0.00	0.50	-2.56	-5.40	-3.93
9CS106_Chr1_consens us_2757	+		Amino Acids and Derivatives	Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	0.24	0.00	-3.89	-8.95	-9.32
9CS106_Chr1_consens us_2768	-		Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis	Carbamoyl-phosphate synthase small chain (EC 6.3.5.5)	-0.24	-0.60	-2.22	-8.14	-8.12
9CS106_Chr1_consens us_2787	+	from clustering based subsystem to Cell Wall and Capsule	Cell Wall and Capsule		CBSS-160492.1.pcg.550	LppC putative lipoprotein	0.48	-0.08	-2.03	-1.56	-0.33
9CS106_Chr1_consens us_2792	-		Stress Response		Carbon Starvation	Stringent starvation protein A	-0.89	-2.06	-2.20	-1.81	-2.16
9CS106_Chr1_consens us_2805	-		Clustering-based subsystems		CBSS-342610.3.pcg.1794	3',5'-cyclic-nucleotide phosphodiesterase (EC 3.1.4.17)	-1.07	-1.15	-2.02	-1.64	-1.87
9CS106_Chr1_consens us_2815	+		Phosphorus Metabolism		Phosphate metabolism	Probable low-affinity inorganic phosphate transporter	-0.75	-1.37	-1.48	-1.27	-1.97
9CS106_Chr1_consens us_2827	+		Cell Division and Cell Cycle		Macromolecular synthesis operon	Transamidase GatB domain protein	-1.14	-1.42	-2.19	-0.60	-0.34
9CS106_Chr1_consens us_2833	+		Carbohydrates	Central carbohydrate metabolism	Pyruvate metabolism I: anaplerotic reactions, PEP	Pyruvate kinase (EC 2.7.1.40)	-1.83	-2.24	-2.59	-1.02	-1.32
IG_9CS106_Chr1_consens us_2845			IG	IG	IG	IG_9CS106_Chr1_consens us_2845	-0.15	-0.24	-0.66	-0.90	-2.01
9CS106_Chr1_consens us_2846	+		Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	3-isopropylmalate dehydrogenase (EC 1.1.1.85)	0.29	-0.15	-3.83	-5.61	-3.45
9CS106_Chr1_consens us_2848	+		Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis	3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)	-0.13	-0.04	-2.78	-3.23	-2.29
9CS106_Chr1_consens us_2866	-		Carbohydrates	Central carbohydrate metabolism	Glyoxylate bypass	Malate dehydrogenase (EC 1.1.1.37)	0.68	-0.33	0.63	-3.41	0.07
9CS106_Chr1_consens us_2898	-		Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Mannose-sensitive hemagglutinin type 4 pilus	MSHA biogenesis protein MshJ	0.44	1.08	1.80	1.32	-0.20
9CS106_Chr1_consens us_2899	-		Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Mannose-sensitive hemagglutinin type 4 pilus	MSHA biogenesis protein MshI	0.66	0.94	1.39	0.72	0.61
9CS106_Chr1_consens us_2907	+		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Threonine and Homoserine Biosynthesis	Aspartokinase (EC 2.7.2.4)	0.24	-0.33	-0.06	-1.06	0.21

Gene / Intergenic (IG) region	Gene orientation	Manual curation	Category <sup>a</sup>	Subcategory <sup>a</sup>	Subsystem <sup>a</sup>	Role <sup>a</sup>	Log <sub>2</sub> Fold Change <sup>b</sup>				
							2216	Apocyclops	Fucus	Alginate	Glucose
9CS106_Chr1_consens us_2908	-		Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Methionine Biosynthesis	5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	0.38	-0.01	-1.01	-5.07	-2.64
IG_9CS106_Chr1_consens us_2909			IG	IG	IG	IG_9CS106_Chr1_consens us_2909	-0.74	-0.02	0.71	0.47	-1.26
IG_9CS106_Chr2_consens us_26			IG	IG	IG	IG_9CS106_Chr2_consens us_26	0.05	0.21	-2.20	-2.02	0.61
9CS106_Chr2_consens us_38	+		Carbohydrates	Di- and oligosaccharides	Maltose and Maltodextrin Utilization	4-alpha-glucanotransferase (amylomaltase) (EC 2.4.1.25)	-0.50	-2.75	-1.05	-3.05	-1.52
9CS106_Chr2_consens us_125	-		Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis	Dihydroorotase (EC 3.5.2.3)	0.27	-0.83	-1.43	-3.00	-5.74
9CS106_Chr2_consens us_312	-		Carbohydrates		Sugar utilization in Thermotogales	Transaldolase (EC 2.2.1.2)	-0.27	-0.62	-1.98	-0.49	-1.02
9CS106_Chr2_consens us_396	+		No category		No subsystem	Putative threonine efflux protein	-0.72	-0.01	0.56	0.38	-2.66
9CS106_Chr2_consens us_430	-	from No subsystem to Fatty Acid Biosynthesis	Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty acid biosynthesis	Fatty acid desaturase (EC 1.14.19.1); Delta-9 fatty acid desaturase (EC 1.14.19.1)	0.70	0.54	-0.46	-2.88	0.12
IG_9CS106_Chr2_consens us_900			IG	IG	IG	IG_9CS106_Chr2_consens us_900	-0.41	0.69	1.48	0.90	-2.12
9CS106_Chr2_consens us_910	+		No category		No subsystem	Repressor protein	-0.29	-0.09	-0.96	-0.66	-2.83
IG_9CS106_Chr2_consens us_914			IG	IG	IG	IG_9CS106_Chr2_consens us_914	-0.17	0.40	0.19	1.02	-3.19
9CS106_Chr2_consens us_931	-		Amino Acids and Derivatives	Alanine, serine, and glycine	Glycine and Serine Utilization	2-amino-3-ketobutyrate coenzyme A ligase (EC 2.3.1.29)	-0.93	-1.08	-1.31	-1.02	-0.62
9CS106_Chr2_consens us_1096	+		No category		No subsystem	Mobile element protein	-1.27	-0.84	1.04	0.80	-1.75
9CS106_Chr2_consens us_1258	-		Membrane Transport	ABC transporters	ABC transporter oligopeptide (TC 3.A.1.5.1)	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)	1.19	0.64	0.19	0.06	1.16
9CS106_Chr2_consens us_1303	-		Carbohydrates	Sugar alcohols	Mannitol Utilization	Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17)	0.13	-0.14	-3.83	-0.23	-0.14
IG_9CS106_Chr2_consens us_1332			IG	IG	IG	IG_9CS106_Chr2_consens us_1332	1.42	1.31	1.22	0.76	1.59
9CS106_Chr2_consens us_1366	-		Amino Acids and Derivatives	Alanine, serine, and glycine	Serine Biosynthesis	Phosphoserine phosphatase (EC 3.1.3.3)	0.37	0.31	1.36	0.82	0.46
IG_9CS106_Chr2_consens us_1528			IG	IG	IG	IG_9CS106_Chr2_consens us_1528	0.00	0.79	1.71	1.65	-0.85
9CS106_Chr2_consens us_1580	-		Virulence, Disease and Defense	Resistance to antibiotics and toxic compounds	Copper homeostasis	FIG135464: Cytochrome c4	1.06	0.55	0.16	0.24	1.83

a. Annotations manually curated from RAST, KEGG, Pfam, and Phyre2, as described in the Methods.

b. Significant log<sub>2</sub> fold change in mutant abundance calculated with negative binomial test for log<sub>2</sub> fold change greater than 0.585,  $P < 0.05$ . Genes / IG regions with significant fold change are indicated by red text. Heatmap shows more negative fold change as red, unchanged as yellow, and positive fold change as green.

pelleted, rinsed of antibiotics, and mated with 9CS106 in a 1:1 ratio of 600  $\mu\text{L}$  of each culture. The cell pellet was resuspended in ASW and pipetted in twelve 20- $\mu\text{L}$  spots on 0.2- $\mu\text{m}$  polycarbonate filters (Whatman) on 2216+DAP agar plates. After incubation for 23 hr at 30°C, the cells from the spots were recovered, resuspended in ASW, and a 10  $\mu\text{L}$  aliquot taken to make dilutions on 2216+Cm to assay the concentration of mutants, while the remainder of the mating resuspension was stored in 25% (v/v) glycerol at -80°C. Once the concentration of mutants was determined, the mutants were rinsed and resuspended in ASW and grown on 148 100 x 15 mm 2216+Cm25 agar plates for two days at room temperature, to give approximately 200-600 colonies per plate. Colonies were scraped, resuspended in ASW, and mixed by pipetting and vortexing. The mutant library was mixed with glycerol (final 25% v/v), aliquoted in approximately 100 vials, and stored at -80°C until use.

### 3.5.3 Media for selective growth of 9CS106 mutant library

Both defined and undefined media were used for selective growth: rich complex media, marine broth 2216 (Difco), defined single carbon source D-glucose (Sigma) and alginate (Sigma), and undefined media of natural substrates, ground brown macroalgae *Fucus vesiculosus* (Starwest Botanicals, Canada) and freeze-dried copepod *Apocyclops royi* (Brine Shrimp Direct). All media except 2216 were prepared in artificial sea water (ASW) derived from sea salts (Sigma) and amended with minimal medium (MM) for trace nutrients (10% vol/vol) (for recipe, see appendix B.2.4). Glucose, alginate, and *Apocyclops* were used at 1 g per L, while *Fucus* was used at 10 g per L (amounts chosen for highest cell yields).

To sterilize the media, 2216 was autoclaved, glucose and alginate media were 0.2  $\mu\text{m}$  filter-sterilized, and macroalga and copepods were sterilized by a pasteurization-like method as follows: 1-L batches were placed in a stirred water bath that was brought from room temperature to 78-82°C at a rate  $>1^\circ\text{C}$  per minute, and held for 30 minutes. Media were cooled in an ice bath, left at room temperature for 2 days to give spores a chance to germinate, and the heating process repeated. Because we observed contaminating growth for the *Apocyclops* particles, this medium was additionally 0.2- $\mu\text{m}$  filtered for further sterilization. Thus, the final particulate state of each medium was as follows: 2216, particulates; alginate, some gel precipitation; glucose, no particulates; *Apocyclops*, no particulates; *Fucus*, particulates.

### 3.5.4 Growth of *Vibrio* sp. F13 strain 9CS106 mutant library on selective media

For selection in different media conditions, a frozen aliquot of the 9CS106 mutant library was rinsed and resuspended in ASW+MM, and kept at 4°C for 12 hours before use. Cells were then added to 150 mL of growth medium, in triplicate. Cells were input at concentrations such that expansion from their initial number corresponded to roughly

the end of the exponential phase (half the carrying capacity), and the elapse of 9 to 10 generations.

Growth, conducted in 300 mL flasks at room temperature with shaking at 240 rpm on a VWR orbital shaker, was monitored by colony counts on 2216, 1.5% Bacto Agar plates (without antibiotic selection) as well as by OD<sub>600</sub> using a BioTek Synergy 2 Multi-Mode Reader. Cell density was inferred by OD<sub>600</sub> values, calibrated for each medium. To titer CFU per mL from *Fucus* samples, they were sonicated in an Ultrasonic Cleaner water bath (Misonix) at 42 kHz for 1 min, which was observed to loosen cells from the *Fucus* surface without a reduction in viability.

Total cells harvested for sequencing was approximately 5e9 CFU to yield the approximately 5  $\mu$ g of DNA needed for sequencing. *Fucus* samples were sonicated for 1 min, centrifuged at 600 x g for 30 sec to pellet the particulates, and the supernatant saved. Cells were collected from this supernatant, as well as the other media, by centrifugation at 5,000 x g for 5 min, and frozen till DNA extraction.

### 3.5.5 Sample preparation and Illumina sequencing

Media-selected mutant library samples and one unselected mutant library were split six ways, and DNA extracted using the MasterPure Complete DNA purification kit (Epicentre), according to manufacturer instructions, including RNase A treatment. For better yield, 2216 samples were first rinsed twice with phosphate-buffered saline. Due to the presence of complex carbohydrates, alginate and *Fucus* samples were subsequently extracted using the PowerPlant Pro DNA isolation kit (MO BIO). To prevent coprecipitation of carbohydrates and DNA, 100  $\mu$ L of isopropyl alcohol was added to each sample with addition of 350  $\mu$ L of PD1 lysis buffer, and DNA extracted subsequently according to manufacturer instructions. After rehydration of DNA in 10 mM Tris, pH 8.0, the DNA was checked for RNA contamination by gel electrophoresis (0.7% agarose in 0.5x TAE buffer, and stained with ethidium bromide (0.5  $\mu$ g/ml) in both the buffer and the gel). An additional RNase A treatment was applied for further purification of DNA from total nucleic acids. Samples were then concentrated and purified via ethanol precipitation.

To prepare the samples for sequencing, DNA was fragmented, and transposon fragments amplified from a known adapter sequence. A schematic for library preparation is given in Figure B-1. To fragment the sample DNA, 5  $\mu$ g of DNA were digested with MmeI (NEB, Massachusetts, USA) for 2.5 h, as previously described (van Opijnen et al., 2014). Sequencing libraries were then prepared as published (Zhang et al., 2012). Briefly, fragment ends were repaired (Quick Blunt kit, NEB, Massachusetts, USA), a 3' A overhang was added using Taq (NEB, Massachusetts, USA), adapters ligated, and fragments subjected to two rounds of PCR amplification: the first round to enrich transposon-genome junctions, and the second round to incorporate Illumina P5 and P7 hybridization sequences, nucleotide variability for cluster detection on the flow cell, and barcodes for multiplexing. Barcode and primer oligo sequences are given in Table B.2. Samples from alginate and *Fucus* selections,

with one technical replicate of an unselected mutant pool, were sequenced on an Illumina HiSeq, Rapid mode, three times to achieve a sufficiently diluted concentration (runs had exhibited poor clustering due to low diversity); reads for these samples were pooled from all sequencing runs. Samples from 2216, glucose, and *Apocyclops* selections, and the second technical replicate of an unselected mutant pool were sequenced on an Illumina NextSeq. To estimate library saturation (for both insertion sites and loci), raw read counts from the unselected replicates were pooled, and negatively screened for genes and IG regions with less than 20 counts. Sequencing statistics for each sample are summarized in Table 3.4.

Table 3.4: Illumina sequencing statistics of media-selected 9CS106 mutants.

Sample <sup>a</sup>	Raw reads	Trimmed reads	Trimmed reads of length 14-17 bp	Reads aligned to genome	Total Reads that mapped to TA sites	% of Reads mapped to TA sites to Trimmed reads 14-17 bp	Total Reads (sites with at least 3 reads)	TA sites hit <sup>b</sup>	% of TA sites hit <sup>c</sup>	Avg reads per hit TA	Sequencing	Number of sequencing runs
2216.a	2.26E+07	2.21E+07	1.74E+07	1.71E+07	1.70E+07	97.41	1.70E+07	84,852	27	200	MIT, NextSeq	1
2216.b	2.48E+07	2.43E+07	1.92E+07	1.87E+07	1.86E+07	97.00	1.86E+07	83,379	27	223	MIT, NextSeq	1
2216.c	2.47E+07	2.41E+07	1.89E+07	1.84E+07	1.83E+07	97.02	1.83E+07	82,217	26	223	MIT, NextSeq	1
Alginate.a	2.27E+07	1.70E+07	1.70E+07	1.62E+07	1.60E+07	94.24	1.60E+07	98,522	31	162	Tufts, Hiseq Rapid	3
Alginate.b	2.27E+07	1.70E+07	1.70E+07	1.62E+07	1.60E+07	94.06	1.60E+07	97,488	31	164	Tufts, Hiseq Rapid	3
Alginate.c	2.16E+07	1.61E+07	1.61E+07	1.53E+07	1.51E+07	93.91	1.51E+07	98,929	31	152	Tufts, Hiseq Rapid	3
Apocyclops.a	2.43E+07	2.37E+07	2.07E+07	2.00E+07	2.00E+07	96.48	1.99E+07	86,495	28	230	MIT, NextSeq	1
Apocyclops.b	2.22E+07	2.17E+07	1.95E+07	1.88E+07	1.87E+07	96.06	1.87E+07	80,592	26	232	MIT, NextSeq	1
Apocyclops.c	2.37E+07	2.32E+07	2.04E+07	1.97E+07	1.96E+07	96.25	1.96E+07	87,248	28	224	MIT, NextSeq	1
<i>Fucus</i> .a	2.16E+07	1.53E+07	1.53E+07	1.44E+07	1.43E+07	93.45	1.42E+07	119,762	38	119	Tufts, Hiseq Rapid	3
<i>Fucus</i> .b	1.75E+07	9.37E+06	9.37E+06	8.80E+06	8.69E+06	92.72	8.66E+06	87,152	28	99	Tufts, Hiseq Rapid	3
<i>Fucus</i> .c	2.22E+07	1.38E+07	1.38E+07	1.30E+07	1.28E+07	92.84	1.28E+07	104,311	33	122	Tufts, Hiseq Rapid	3
Glucose.a	2.39E+07	2.33E+07	1.96E+07	1.92E+07	1.91E+07	97.49	1.91E+07	76,316	24	250	MIT, NextSeq	1
Glucose.b	2.56E+07	2.50E+07	2.09E+07	2.04E+07	2.03E+07	97.22	2.03E+07	76,148	24	266	MIT, NextSeq	1
Glucose.c	2.41E+07	2.36E+07	1.88E+07	1.82E+07	1.81E+07	96.66	1.81E+07	81,497	26	222	MIT, NextSeq	1
Input.1 <sup>d</sup>	2.20E+07	2.15E+07	1.90E+07	1.83E+07	1.82E+07	95.51	1.82E+07	94,245	30	193	MIT, NextSeq	1
Input.2	2.23E+07	1.65E+07	1.65E+07	1.55E+07	1.53E+07	92.89	1.53E+07	140,218	45	109	Tufts, Hiseq Rapid	3

a. Mutant libraries selected in media, or unselected (the input library). Letters represent biological replicates, and numbers, technical replicates.

b. Heatmap for ascending values, illustrates batch effects of sequencing centers/machines.

c. Total mutagenizable TA dinucleotide sites in the *Vibrio* sp. F13 strain 9CS106 is 314,091.

d. Reads from the unselected samples were pooled to estimate library saturation reported in the main body text. For further detail, see Methods.

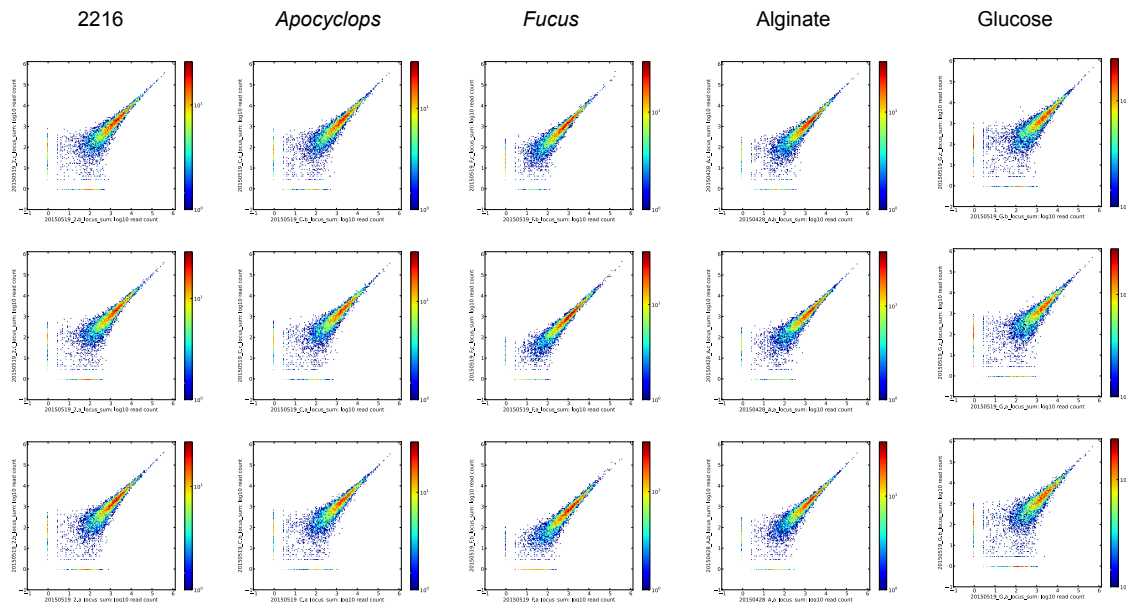


Figure 3-5: **Pair-wise correlations between biological replicates by gene and intergenic region.** Two-dimensional histogram plots for every within-media type pair of samples, binned by total sequencing read counts per gene and IG region. Logarithm (base 10) counts per bin depicted in a heatmap.

### 3.5.6 Processing of sequencing reads and gene categorization

Sequencing reads were demultiplexed by barcode and trimmed of adapter sequences, including the modified *mariner* inverted repeat (5'-GACTTATCATCCAACCTGT-3' and 5'-ATACCACGACCAGAT-3') in CLC Genomics Workbench. Trimmed reads, 14 to 17 bp in length, were then aligned to the 9CS106 genome with Bowtie v.1.1 (Langmead et al., 2009) with zero mismatches allowed and random allocation of reads that mapped to more than one site. Aligned reads were mapped to TA dinucleotide sites to filter for true *mariner* transposon insertions, whereas un-mapped reads were discarded.

Because sample noise reduced correlations between biological replicates when reads per insertion were evaluated, total raw read counts were summed per locus (for genes, reads within 90% of the 5' end to ignore insertions that may not affect gene function) (correlations depicted in Figure 3-5) and normalized using DESeq2 (Love et al., 2014) with default parameters in R. Briefly, DESeq2 performs a median-based normalization to account for sequencing depth across all samples. Rather than using the observed variance for each locus across replicates, DESeq2 empirically fits a relationship between mean counts and their variance, using the model value per gene. (Note that although read counts can be modeled, in principal, as a Poisson process, the observed variance is greater than the mean

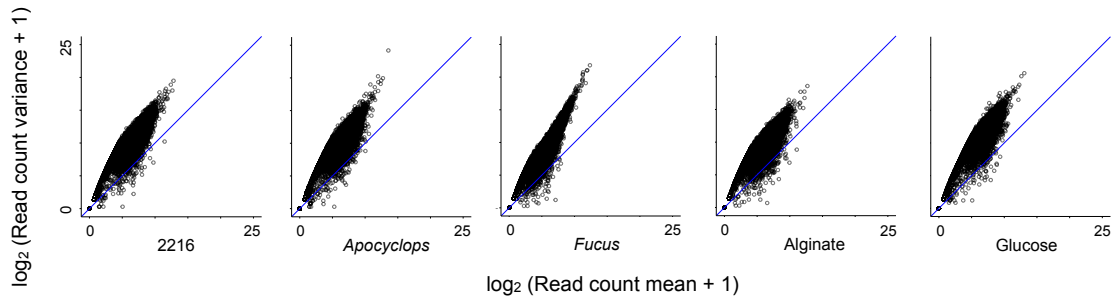


Figure 3-6: **Sequencing results have greater variance than expected for a Poisson process.** Variance of the read counts per gene or intergenic region is larger than the mean, i.e falls above the 1:1 line. Reads counts given in  $\log_2$ -transformed read counts (+1 to avoid taking a zero logarithm)

for this data (Figure 3-6.) After obtaining a maximum-likelihood estimate of the  $\log_2$ -fold change in counts from the output sequencing library to the input library, which can include genes with 0 read counts in a condition, DESeq2 shrinks estimates derived from small mean counts (which are more variable, thus tending toward false positives). Finally, significant change in gene abundance was calculated with DESeq2 using a negative binomial test, the  $P$ -value corrected for multiple hypothesis testing with the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

To identify essential genes (for which mutants were underrepresented in the initially constructed mutant library), a Monte-Carlo approach was taken as previously described (Turner et al., 2015). Briefly, using pooled reads from the technical replicates of the input library and removing sites with less than 20 reads, read abundance per insertion site was smoothed using locally weighted LOESS smoothing with parameter alpha 0.5 (for Chr1) or 1 (for Chr2, ECE1, and ECE2) to normalize for replication bias. Reads were then randomized, without replacement, to generate 2,000 simulated mutant pools. The mutant library was compared to the simulated pools using DESeq2, and underrepresented genes were identified using a negative binomial test and Benjamini-Hochberg adjusted  $P$  value (fold change  $>0$ ,  $P < 0.05$ ). The  $\log_2$ -fold change values per locus were clustered into two modes, “reduced” or “unchanged” by fitting the values to a parametrized bimodal Gaussian distribution using the R *mclust* package (Fraley and Raftery, 2012). Essential loci were those that were both underrepresented by the negative binomial test and clustered into the “reduced” group (uncertainty of mode fit  $< 0.05$ ). The process was done for each chromosome (and extrachromosomal element) separately.



### 3.5.7 Fitness calculation

Because growth was exponential in the selective conditions, mutant fitness ( $W_{mt}$ ) was calculated according to a rearrangement of the Malthusian, or exponential, growth equation, as previously described (van Opijnen and Camilli, 2013). The calculation depends on the input ( $t1$ ) and output ( $t2$ ) relative abundances of mutants per locus ( $N_{mt(t1)}$  and  $N_{mt(t2)}$ , respectively), estimated by sequencing read counts, together with knowledge of the expansion of the mutant library (number of all bacteria at  $t2$  / number of all bacteria at  $t1$ ) ( $d$ ).

$$W_{mt} = \frac{\ln \left[ N_{mt(t2)} \times \frac{d}{N_{mt(t1)}} \right]}{\ln \left[ (1 - N_{mt(t2)}) \times \frac{d}{1 - N_{mt(t1)}} \right]} \quad (3.1)$$

To estimate input and output relative abundance per locus, the DESeq2-normalized reads were used, and the expansion term per condition was simply the final CFU per mL density relative to the initial density. To also gain a sense of the confidence in each fitness value, a fitness range (minimum, maximum) was calculated for each locus based on the mean read abundance ( $\mu$ ) +/- the standard deviation of read abundance (estimated from the per locus dispersion parameter,  $\alpha$ , calculated in DESeq2), according to the equation (Love et al., 2014):

$$\sigma = \sqrt{\mu + \alpha\mu^2} \quad (3.2)$$

### 3.5.8 Genome annotation

The 9CS106 genome was first annotated using Rapid Annotations using Subsystems Technology (RAST) (Aziz et al., 2008). Genes with fold change greater than 1.5 ( $P < 0.05$ , negative binomial test) were additionally annotated in three ways: using the Kyoto Encyclopedia of Genes and Genomes (KEGG) tool blastKOALA (Kanehisa et al., 2014), Pfam (v. 28.0) (Finn et al., 2014), and Phyre2 (Kelley et al., 2015). Phyre2 annotations included only if the prediction was made with >95% confidence, sequence identity >40% and alignment coverage >90%. Where RAST prediction was unannotated, the KEGG, Pfam, or Phyre2 annotation was substituted. Subsystem, subcategory, and category were also elaborated based on role prediction when no or a misleading annotation was given by RAST. Forty-five changes were made, and are documented in Table 3.3.

## 3.6 Supplementary Information

### 3.6.1 Construction of mutagenesis plasmid

*E. coli* strains were grown in Luria-Bertani (LB) (Difco) at 37°C. Strains II 3813 and  $\beta$  3914 were used as a plasmid host for cloning and conjugation, respectively (Le Roux et al., 2007). Antibiotics were used at 25 or 5  $\mu\text{g}/\text{ml}$  for chloramphenicol (Cm). Diaminopimelate (DAP) was supplemented when necessary to a final concentration of 0.3 mM. *Vibrio* sp. F13 strain 9CS106 was grown in 2216 (Difco).

A 1023 bp fragment harboring the Cm resistance gene flanked by *mariner* transposon inverted repeats was amplified from pSW4426T (Le Roux et al., 2007) by PCR using primers Mar F and R (Table 3.5), which included single nucleotide substitutions to introduce MmeI restriction sites. Similarly, a 1332 bp fragment harboring the C9 transposase gene was amplified from pSC189 (Chiang and Rubin, 2002) using primers C9 F and R (Table Table 3.5). PCR reactions were done in 50  $\mu\text{L}$  volumes using the Herculase DNA polymerase (Agilent) following the manufacturer's instructions. Primers are listed in Table 3.5. Conditions for amplification were as follows: 95°C for 2 min, followed by 30 cycles of 95°C for 20 s, ( $T_m$  55°C) for 20 s, and 68°C for 60 s per kb.

Table 3.5: Primers for plasmid construction

Primer	Sequence (5' - 3')	Product size	Target gene
Mar F	GCCGGCacaggttggAtgataagtccccggtctCGCCGAATAAATACCTGTGACGG	1023 bp	<i>mariner cat</i>
Mar R	AGAcaggttggAtgataagtccccggtctGATATCGTCGCAGACCAAACG		
C9F	tcAGgaCCGCCAGTGTGATGGATGACA	1332 bp	C9 transposase
C9R	gcggccgcCTTGACGGGGAAAGCCCGCGCG		

Both amplicons were cloned in pCR blunt (Invitrogen) following the manufacturer's instructions. The resultant recombinant plasmids (pCR::ISCAT and pCR::C9) were sequenced by the Sanger method. After EcoR1 digestion of the plasmid pCR::C9, the generated fragment (C9) was cloned in the pSW25T suicide vector carrying the R6K origin of replication (oriVR6K), the RP4 origin of transfer (oriTRP4) and a spectinomycin selective marker (Demarre et al., 2005). The recombinant plasmid (pSW25T::C9) was verified by digestion with restriction enzymes. The plasmid pCR::ISCAT was digested by BamHI-XbaI and the generated fragment (ISCAT) was cloned in pSW25T::C9 leading to pSW25T::C9::ISCAT (Figure 3-7).

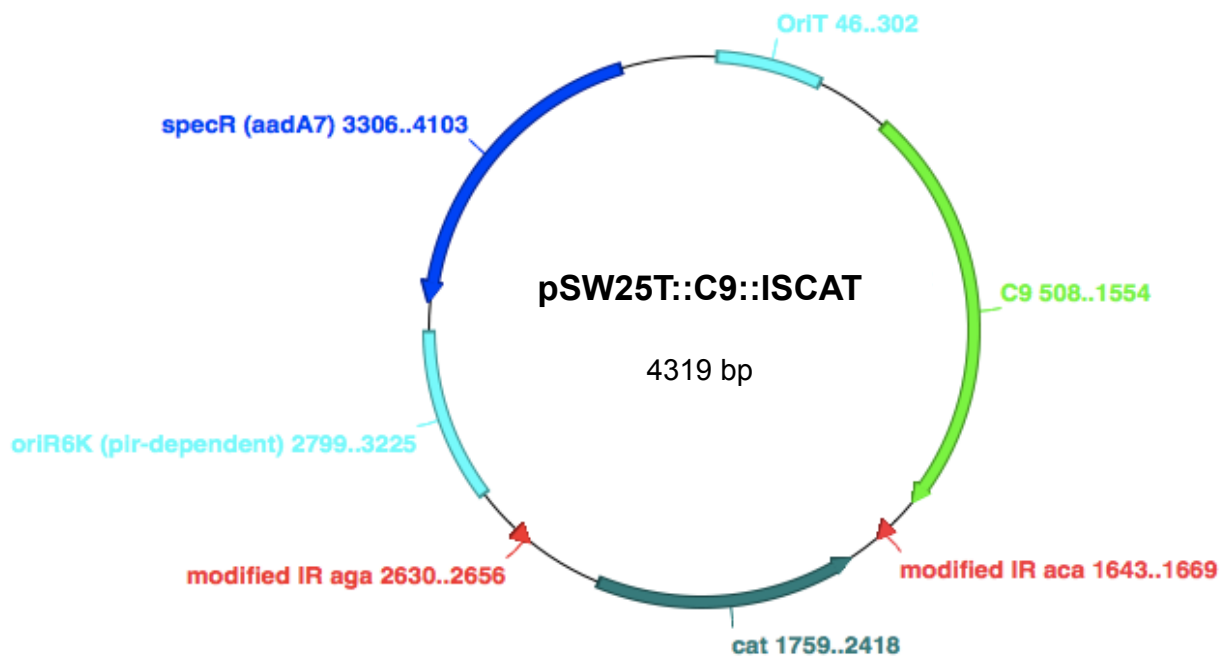


Figure 3-7: **pSW25T::C9::ISCAT**, the plasmid used for transposon-mutagenesis. The *mariner* transposon modified inverted repeats (red triangles labeled “modified IR”) are labeled with the start of their respective nucleotide sequences: AGA (“aga”) and ACA (“aca”). The chloramphenicol acetyltransferase gene (“cat”) serves as the selection marker for mutagenized strains. Also labeled are the hypertransposase *himar1* C9 gene (“C9”), plasmid marker spectinomycin resistance cassette (“specR”), origin of transfer (“OriT”) and the *pir* protein-dependent region for plasmid replication (“oriR6K”).

THIS PAGE INTENTIONALLY LEFT BLANK

## Chapter 4

# Conclusion

The significance of microorganisms both globally and clinically, motivates a deeper understanding of their ecology. Microorganisms are key drivers in the Earth's biogeochemical cycling; in the ocean, photosynthetic microorganisms are responsible for about half of total global productivity (Kirchman, 2012), and heterotrophic microorganisms dominate marine metabolic activity (Azam and Malfatti, 2007). Moreover, some marine bacteria are also human pathogens, which may become more prevalent in a shifting climate (Burge et al., 2014).

In this thesis, the chapters contrast different scales for analyzing the environmental niche of the *Gammaproteobacteria* family of marine heterotrophs, the *Vibrionaceae*. The first is coarse-grained: the environmental correlates in bulk sea water at the genus and population levels; and the second is fine-grained: habitat resource adaptation of a single strain.

In chapter 2, a meta-analysis of published observations demonstrated that most measured environmental variables, save temperature and salinity, are unreliable indicators of *Vibrio* abundance. Yet the scales at which the environment was sampled and the phylogenetic resolution determined could obscure meaningful environmental associations.

Microenvironments structure *Vibrio* populations; studies have shown cohesive populations defined by habitat distributions across size fractions of the water column (Hunt et al., 2008a; Szabo et al., 2013) or biotic particle (Preheim, 2010). Yet not every environment is a component of a microbial niche, and transient associations may also occur. *Vibrios* have stochastic associations with mussel and crab gastrointestinal tracts, for example (Preheim et al., 2011a). In this case, habitats may have been defined at a broader scale than that which bacteria deterministically assemble (such as on particles ingested by invertebrates).

To identify more robust environmental associations, future work in monitoring of microbial populations should leverage genetic tools—such as multilocus or whole-genome sequencing—and finer ecological sampling—of specific, non-overlapping ecological compartments. This approach, in fact, is exactly that which has been taken in the Polz lab to give

insight into environmental *Vibrio* associations. However, the approach has not currently been more widely adopted for the specific study of pathogens, whose dynamics many of the correlative studies sought to predict. Wider implementation of fine-scale ecological and phylogenetic sampling might improve models; for instance, it is still not known to what degree zooplankton are natural habitats of pathogenic variants of *V. cholerae*, an idea supported by attachment in laboratory settings, but with conflicting evidence in environmental ones, as reviewed in (Takemura et al., 2014). Even finer scale sampling, like different body sites within a zooplankton, may inform our understanding of a population's niche.

In chapter 3, selection of a single-strain mutant library on resources from different habitats revealed diverse genetic requirements, unexpected habitat resource complexity, and a diffuse reliance on catabolic pathways for growth on these resources. Brown alga *Fucus vesiculosus* resources selected for the ability to synthesize amino acids and nucleosides, and catabolism of mannitol via fermentation, whereas resources from the copepod *Apocyclops royi* did not select for biosynthetic pathways, and catabolism here appeared to rely on redundant pathways or genes, perhaps such as transporters of oligopeptides and amino acids — multicopy in the strain's genome.

Future work to elaborate these observations should manipulate the concentration and physical nature of habitat resources, and reduce the number of mutants screened at a time. High concentrations of nutrients were used to sustain sufficient culture densities to interrogate the large number of mutants (approximately 100,000). In these conditions, the *Vibrio* strain appeared to be predominantly planktonic; however, other isolates of the same population, *Vibrio* sp. F13, have been cultured from detrital algal and zooplankton particles, suggesting the attached lifestyle is a component of this population's niche. Therefore, is a predominantly attached lifestyle experimentally observed when nutrients are not homogeneously distributed, but rather concentrated in particles? This scenario is closer to what we would observe in the environment, where resources are patchily distributed (Azam and Malfatti, 2007; Stocker, 2012). Attachment of *Vibrio* sp. F13 may only be favored in a punctate resource landscape; if so, then targeted mutants could be constructed to test specific hypotheses suggested by this thesis, such as whether the Type IV pilus, MSHA, is a determinant for *Fucus* colonization. The fitness of targeted mutants could be assayed in competitive incubations with wild-type cells.

A second line of investigation is to explore the biotic dimension, in addition to the resource dimension, of natural habitats. Competitors, facilitators, mutualists, and predators also exert selective influence over bacteria in habitats, and the model system employed here should be extended to experimentally investigate the nature of social interactions. Another population, *V. breoganii*, can be used to study potential competition, as *V. breoganii* shares the *Fucus* habitat with *Vibrio* sp. F13. Using a strain of *V. breoganii* as a co-colonizer during selection on *Fucus*, does a *Vibrio* sp. F13 strain utilize different metabolic pathways not observed when a strain exploits habitat resources alone?

Finally, further research should investigate the distribution and natural selection of genes



identified as fitness determinants in ecologically relevant contexts. This approach would address, for example, whether particular habitats select for the same pathways in different populations. If they are conserved, do these genes show evidence of positive selection? Investigating the evolution of allelic variants for genes most needed in ecological settings will strengthen our understanding of how niche partitioning between populations occurs.

THIS PAGE INTENTIONALLY LEFT BLANK

## Appendix A

# Identification of a *Vibrio* phage receptor

### A.1 Introduction

Phage predation exerts strong selective pressure on microorganisms. With lytic phages estimated to daily kill approximately 20% of their number (Suttle, 2007), marine microorganisms must balance being recognized by their extracellular structures, with the adaptations those structures confer. Phage routinely target outer membrane nutrient transporters (for glucose, maltose, iron, etc.), flagella (used for motility), slime (which protects a cell from desiccation and xenobiotics), and structures for attaching to surfaces (pili and capsule) (as reviewed in (Vinga et al., 2006; Rakhuba et al., 2010)). Analysis of the distribution and evolution of phage receptors can lend insight into the balance of positive and negative selection for these features.

Here, we demonstrate a method to identify phage receptors using a genome-wide mutant library screen and massively parallel sequencing. We challenged a mutant library of a marine *Vibrio* with two environmental phage isolates, and sequenced resistant colonies. Resistant mutants were predominantly disrupted in the synthesis pathway of the capsule, a structure implicated in virulence and adhesion. A preliminary analysis suggests these genes are inconsistently represented and diverse in the *Vibrio*, which may indicate both negative frequency-dependent selection (gene presence/absence) and diversifying selection (allelic diversity) by phage — a hypothesis that will be examined in greater depth in future work.

Table A.1: **Isolated 9CS106 phage.**

Phage	Sample enriched from	Date sample collected
T1	IBYC <sup>a</sup> , 0.22 $\mu$ m filtered	09/13/2005
T6	Nahant, MA unsterilized <i>Fucus</i> in Artificial Sea Water + Minimal Medium (10% v/v)	07/16/2013

a. Ipswich Bay Yacht Club in Plum Island Sound, MA

## A.2 Results and Discussion

### A.2.1 Isolation of phages to 9CS106

Phages infecting *Vibrio* sp. F13 strain 9CS106 were successfully isolated from a water sample and a *Fucus* sample (Table A.1). These phages were obtained by enrichment culture, in which the sample was supplemented with the rich culture media 2216 marine broth, and incubated with 9CS106 cells, and left at room temperature, no shaking but twice daily by hand, for a week. A potential exists that this phage may be a lytic variant of a lysogen, which is discussed in greater detail below. Electron micrographs show both phage to belong to the *Podoviridae* (Figure A-1).

### A.2.2 Selection for phage resistance

To select for mutants resistant to phage infection, the mutant library was grown into exponential phase (OD<sub>600</sub> 0.5 to 0.9), and mixed separately with each phage to allow for adsorption before plating. After two to three days, the resistant colonies were harvested. Their DNA was extracted and prepared for transposon insertion sequencing (Tn-seq) as described in chapter 3.

### A.2.3 Tradeoff in phage susceptibility in resistant mutants

To assay any changes in susceptibility to environmental phage, we conducted a host-range assay using 16 other phages, in addition to T1 and T6 (Table A.2). In a preliminary selection of the mutant library, we isolated four resistant mutant strains per phage, named after the phage of exposure to which they were resistant (T1 or T6) followed by A, B, C, or D. The phages we used were from two enrichments conducted at the same time as T1 and T6: T2 and T3 (referred to as T phages), three from a previous enrichment in 2014 (B phages), and 11 from concentrated seawater (un-enriched) that infect *Vibrio* sp. F13 (names beginning with 1).

We found that the mutants have different phage sensitivity profiles, indicating tradeoffs in resistance. The wild type showed lytic infection by T phages and B phages, though B2

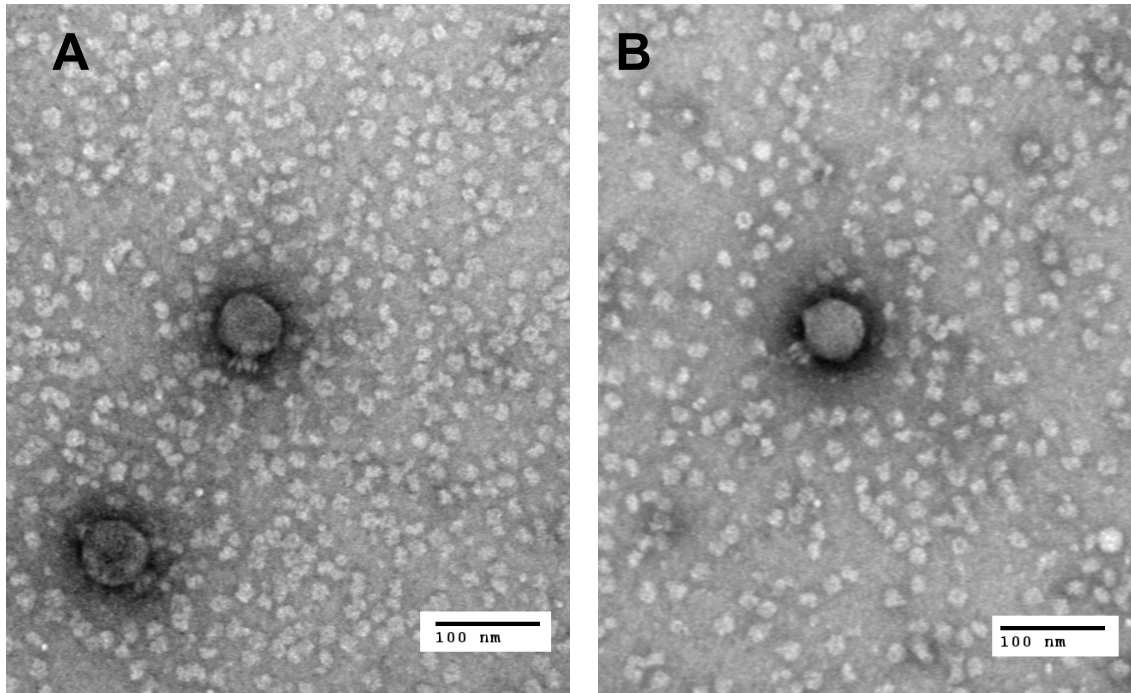


Figure A-1: **Phages T1 (A) and T6 (B) are podoviruses.** Electron micrographs courtesy of Kathryn Kauffman and Fatima Hussain.

Table A.2: Phage host range assay on wild-type 9CS106 and phage-exposed transposon mutants

Source	Phage	Morpho-type	Host Lawn <sup>a</sup>									
			WT-9CS106	T1-A	T1-B	T1-C	T1-D	T6-A	T6-B	T6-C	T6-D	
9CS106 Enrichment, this study	T1	Podo	xxx	xxx	0	0	0	xxx	0	0	0	
	T2	Myo	xxx	xxx	0	0	0	xxx	0	0	0	
	T3	Myo	xxx	xxx	0	0	0	xxx	0	0	0	
	T6	Podo	xxx	xxx	0	0	0	xxx	0	0	0	
Nahant Collection Phages infecting <i>Vibrio</i> sp. F13	1.030.O		0	0	0	0	0	0	0	0	0	
	1.135.O		0	0	0	0	0	0	0	0	0	
	1.155.O		0	0	0	0	0	0	0	0	0	
	1.164.O		0	0	0	0	0	0	0	0	0	
	1.210.O		0	0	0	0	0	0	0	0	0	
	1.213.O	Sipho	0	xxx	xxx	xxx	xx	xxx	xx	xx	xxx	
	1.215.A		0	0	0	0	0	0	0	0	0	
	1.216.O	Sipho	0	xx	xxx	x	x	0	xx	x	xx	
	1.238.A		0	0	0	0	0	0	0	0	0	
	1.238.B		0	0	0	0	0	0	0	0	0	
1.246.O		0	0	0	0	0	0	0	0	0		
9CS106 Enrichment, 2014	B1	Myo	xxx	xxx	0	0	0	xxx	0	0	0	
	B2		xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	
	B3	Myo	xxx	xxx	0	0	0	xxx	0	0	0	

a. Phage resistance susceptibility profile was determined for wild type 9CS106 (WT-9CS106) and eight phage-resistant transposon-mutants, four from each of the two phages, T1 and T6. Sensitivity is indicated by dark shading, and the number of x's indicates the number of replicates for which activity was discernible.

yielded a slightly turbid plaque. By contrast, T1-B is the only mutant that is strongly hit by phages 1.213 and 1.216 in all the replicates, and is also hit by B2. T1-A also shows a unique phenotype, being infected by all T phages and B phages, with turbid plaques for all except B2, which was strongly lytic (clear plaques). T6-A showed a third unique profile: it was infected lytically by all T phages and B phages, though infection by 1.213/216 phages resulted in turbid plaques. T1-C, T1-D, T6-B, T6-C, and T6-D, on the other, all had similar profiles; hit strongly only by B2 and weakly by both 1.213 and 1.216, except in one case where infection was absent, but we suspect this is an artifact of the low titer of this phage.

#### A.2.4 T1 and T6 phage have similar phenotypes

The host-range assay demonstrated that T1 and T6 have the same infection profile, though the plaque phenotype for one strain, suggested they may be unique. T6-A showed a turbid phenotype in the central clearing area of the plaque, with an additional sharp ring within, when infected with T1. By contrast, infection with T6 showed strong, even plaque clearing, with the grow back of resistant colonies.

Because of the similarity of T1 and T6, we wondered whether they might in fact be the lytic variant of an apparent lysogen residing in the 9CS106 genome. Annotation of the resident phage indicated it belongs to the *Siphoviridae*, however (Table A.3).

Table A.3: **Annotation of Extrachromosomal Element 1 (ECE-1)**

Locus Name	Annotation [Tool] <sup>a</sup>
9CS106_ECE_1_consensus_5	Probable sigma factor [PHYRE2]
9CS106_ECE_1_consensus_9	Probable Holliday junction resolvase [PHYRE2]
9CS106_ECE_1_consensus_13	Probable Muramoyl-pentapeptide carboxypeptidase [PHYRE2]
9CS106_ECE_1_consensus_14	Phage terminase, SSU [ACLAME]
9CS106_ECE_1_consensus_15	Phage terminase, LSU, in lambda-like prophages [RAST, ACLAME, PHYRE2]
9CS106_ECE_1_consensus_16	Possible head-to-tail joining protein [PHYRE2]
9CS106_ECE_1_consensus_17	Phage portal protein, in lambda-like prophages [ACLAME]
9CS106_ECE_1_consensus_18	Phage head protein, in lambda-like prophages [ACLAME, PHYRE2]
9CS106_ECE_1_consensus_19	Putative RecA/RadA recombinase, in lambda-like prophages [ACLAME]
9CS106_ECE_1_consensus_23	GroES-like [ACLAME, PHYRE2]
9CS106_ECE_1_consensus_25	Phage tail tape measure protein, in Siphoviridae prophages [ACLAME]
9CS106_ECE_1_consensus_38	Possible methyltransferase [PHYRE2]
9CS106_ECE_1_consensus_39	Plasmid partitioning protein A [ACLAME, PHYRE2]
9CS106_ECE_1_consensus_41	Exonuclease A [ACLAME]
9CS106_ECE_1_consensus_51	Phosphoadenosine phosphosulfate reductase [ACLAME, PHYRE2]
9CS106_ECE_1_consensus_53	Transcriptional regulator, PadR family [RAST, ACLAME, PHYRE2]
9CS106_ECE_1_consensus_56	MSHA pilin protein MshA [RAST, PHYRE2]

a. Initial whole-genome annotation [RAST]; database of mobile elements, evalule  $\leq 10^{-4}$  [ACLAME]; protein structure modeling [PHYRE2]. Annotations were described as "Probable" or "Possible" if Phyre2 annotations were at or above 90%/75% and 85%/65% for Confidence/Coverage, respectively.

### A.2.5 Con-ARTIST to determine resistance factors

To determine genes conferring phage resistance, we used the Tn-seq analysis pipeline, Con-ARTIST (Pritchard et al., 2014). Normalizing for replication bias and sequencing depth, and simulating the input library 100 times, Con-ARTIST uses a Mann-Whitney U test to predict genes with significantly different mutant abundances between the input and output libraries. We used this approach because it makes no assumptions on the complexity of the output library; another approach, normalization and hypothesis testing with R package DESeq2 (Love et al., 2014), assumes that mutants in most genes are equally abundant. This assumption does not hold crucial in a phage selection, where only very few genes are expected to confer resistance, and most mutants being killed. Con-ARTIST, on the other hand, constrains its hypothesis testing within genes; as long as the fitness effects are observed at the gene level (as opposed to within domains of a gene), Con-ARTIST can test whether it is significantly enriched or de-enriched in the output library. Here, we look at mutant enrichments, i.e. those with a fitness benefit under phage selection.

### A.2.6 Extracellular polysaccharide capsule is phage receptor

From the Con-ARTIST analysis, the results were strikingly clear; mutants in capsule synthesis were consistently more fit than other mutant genotypes, under selection by either phage T1 or T6, indicating the capsule is their receptor for host recognition Table A.4.

A surface-associated, thin layer of extrapolyssacharide, the capsule is highly diverse

Table A.4: Putative phage-susceptibility genes in 9CS106

Gene	Annotation <sup>a</sup>	Pathway <sup>b</sup>	Enrichment <sup>c</sup>				
			Max <sup>d</sup>	T1.a	T1.b	T6.a	T6.b
9CS106_Chr1_consensus_69	[RAST, KEGG] <i>wecA</i> ; Undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.-); [PHYRE2 100/83] transferase; phospho-n-acetylmuramoyl-pentapeptide-transferase	Capsule	955	955	939	831	814
9CS106_Chr1_consensus_141	[PHYRE2 20.1/26] NAD(P)-binding Rossmann-fold domains; Glyceraldehyde-3-phosphate dehydrogenase-like, N-terminal domain; [BLAST e-value 3e-84/identity 99%] exported hypothetical protein; [pfam e-value 8.53e-3]						
9CS106_Chr1_consensus_134	Prokaryotic membrane lipoprotein lipid attachment site	Capsule	296	147	91.1	296	103
9CS106_Chr1_consensus_139	[PHYRE2 99.9/90] transferase; udp-galactofuranosyl transferase <i>glft2</i>	Capsule	158	158	125	153	113
9CS106_Chr1_consensus_130	[RAST, KEGG] <i>kpsM</i> ; Polysialic acid transport protein	Capsule	141	136	108	141	92.7
9CS106_Chr1_consensus_127	[RAST] Glycosyltransferase	Capsule	124	116	82.8	124	81
9CS106_Chr1_consensus_135	[RAST] Glycosyl transferase, group 2 family protein	Capsule	110	110	80.6	105	83.5
9CS106_Chr1_consensus_140	[PHYRE2 100/89] transferase; the crystal structure of udp-galnac: polypeptide alpha-n-2 acetylglucosaminyltransferase-t1	Capsule	106	106	84.8	103	77.8
9CS106_Chr1_consensus_129	[RAST] <i>kpsD</i> ; Capsular polysaccharide export system periplasmic protein	Capsule	74.8	74.8	58.4	71.6	57.2
9CS106_Chr1_consensus_128	[PHYRE2 100/82] transferase, crystal structure of heparan sulfate 3-o-sulfotransferase2 isoform 1 in the presence of pap	Capsule	69.4	69.4	19.6	15.7	20.3
9CS106_Chr1_consensus_138	[PHYRE2 100/68] transport, transferase; crystal structure of heparan sulfate 2-o-sulfotransferase2 from gallus gallus as a maltose binding protein fusion	Capsule	59.9	59.9	51.8	55.1	54.7
9CS106_Chr1_consensus_281	[RAST, KEGG] <i>kpsE</i> ; Capsular polysaccharide export system inner membrane protein	Capsule	59.5	59.5	40.4	49.3	37.8
9CS106_Chr1_consensus_2894	[RAST] Multidrug resistance protein		49.8	49.8	45.1	28.4	48.2
9CS106_Chr1_consensus_599	[RAST, KEGG] <i>mshN</i> ; MSHA biogenesis protein MshN		48.6	42.6	39.3	48.6	43.6
9CS106_Chr1_consensus_2540	[PHYRE2 56.5/41] 2.15; PDB header:structural genomics, unknown function ; Chain: B: PDB Molecule:uncharacterized protein abo_0056		48	26	22.4	48	16.6
9CS106_Chr1_consensus_2765	[RAST, KEGG] DNA polymerase IV (EC 2.7.7.7)		47.9	34.2	44.3	43.3	47.9
9CS106_Chr1_consensus_136	[RAST, KEGG] <i>rbc</i> ; dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)	Capsule	44.8	36.3	44.8	30.9	43.4
9CS106_Chr1_consensus_1360	[PHYRE2 20.3/6] 2.3; PDB header:signaling protein/signaling protein ; Chain: E: PDB Molecule:sigma-e factor negative regulatory protein		37.5	37.5	29.6	28.6	31
9CS106_Chr1_consensus_126	[RAST, KEGG] <i>kpsT</i> ; Capsular polysaccharide ABC transporter, ATP-binding protein	Capsule	33.9	33.9	19.6	23.3	22.9
9CS106_Chr1_consensus_1147	[RAST] dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	Capsule	33	25.1	20.6	33	24.9
9CS106_Chr1_consensus_2062	[RAST, KEGG] <i>sppA</i> ; Signal peptide peptidase (EC 3.4.21.-)		31.9	31.9	30.6	29.2	25.6
9CS106_Chr1_consensus_1624	[RAST, KEGG] <i>malK</i> ; Various polyols ABC transporter, ATP-binding component (EC:3.6.3.-)		30.3	25.9	19.9	30.3	25.9
9CS106_Chr1_consensus_793	[PHYRE2 100/92] 1.75; PDB header:transport protein ; Chain: B: PDB Molecule:amino acid abc transporter, periplasmic amino acid-binding		26.1	0	24.8	0	26.1
9CS106_Chr1_consensus_20	[RAST, KEGG] Beta-N-acetylhexosaminidase, (GlcNAc)2 catabolism		25.9	13.2	18	6.68	25.9
9CS106_Chr1_consensus_2454	[PHYRE2 100/96] 3.2; PDB header:membrane protein ; Chain: A: PDB Molecule:outer membrane protein assembly factor bama		23.3	23.3	13.2	17.6	13.4
9CS106_Chr1_consensus_2846	[RAST] <i>flgO</i> ; Flagellar protein		21.1	20.9	20.4	0	21.1
9CS106_Chr1_consensus_2895	[RAST, KEGG] <i>leuB</i> ; 3-isopropylmalate dehydrogenase (EC:1.1.1.85)		21.1	9.65	21.1	6.21	10.6
9CS106_Chr1_consensus_379	[RAST, KEGG] <i>mshM</i> ; MSHA biogenesis protein MshM		19.8	4.36	4.24	19.8	4.19
9CS106_Chr1_consensus_2308	[RAST] Permease of the major facilitator superfamily		18.4	8.47	6.61	18.4	6.38
9CS106_Chr1_consensus_218	[KEGG] putative transposase		17.2	0	10	0	17.2
9CS106_Chr1_consensus_2704	[RAST, KEGG] <i>prlC</i> ; oligopeptidase A (EC:3.4.24.70)		16.9	13.9	16.9	13.4	15.1
9CS106_Chr1_consensus_643	[RAST, KEGG] <i>anmK</i> ; anhydro-N-acetylmuramic acid kinase (EC:2.7.1.170)		15.7	13.7	9.26	15.7	8.95
9CS106_Chr1_consensus_1233	[RAST, KEGG] <i>arginine deiminase</i> (EC:3.5.3.6)		14	14	11.2	9.19	12.4
9CS106_Chr1_consensus_2639	[RAST, KEGG] PTS-Glc-EIIB; PTS system, glucose-specific IIB component (EC 2.7.1.69) / PTS system, glucose-specific IIC component		13.4	11.5	12.4	13.4	11.1
9CS106_Chr1_consensus_2598	[KEGG] AARS; alanyl-tRNA synthetase (EC:6.1.1.7)		13.4	12.5	13.4	5.71	12.6
9CS106_Chr1_consensus_2801	[RAST] Mobile element protein		12.2	9.75	12.2	0	0
9CS106_Chr1_consensus_1708	[KEGG, RAST] <i>hhoB</i> ; serine protease, outer membrane stress sensor DegS (EC:3.4.21.-)		12	12	2.94	3.39	5.6
9CS106_Chr1_consensus_377	[PHYRE2 41.3/13] UNK ; PDB header:viral protein ; Chain: A: PDB Molecule:envelope small membrane protein		11.4	0	11.4	0	0
9CS106_Chr1_consensus_621	[RAST, KEGG] GAPDH; glyceraldehyde 3-phosphate dehydrogenase (EC:1.2.1.12)		10.7	10.7	7.02	8.28	7.26
	[RAST, KEGG] <i>yhbJ</i> ; UPF0042 nucleotide-binding protein		10.2	4.08	9.96	1.48	10.2

a. Annotations were made using RAST, KEGG, Phyre2, and, for 9CS106\_Chr1\_consensus\_141, BLAST. Phyre2 annotation includes scores: (Confidence)/(Coverage).

b. Pathways were designated for Capsule to indicate those genes analyzed in this study. Capsule genes were those either known to have a function involved in capsule synthesis, or within the same operon as these genes.

c. Enrichment is the average read count ratio of the output library to 100 simulations of the input library, carried out in Con-ARTIST as described in the Methods. Samples are given letters

d. The maximum average read count ratio for the gene seen in any of the phage selections.



(Wyres et al., 2015; Shu et al., 2009), with at least 70 capsular polysaccharides known in *Escherichia coli* (Sussman, 1997). In some gram-negative bacteria, including *Vibrio cholerae* O139, the capsule is formed by the polymerization of the lipopolysaccharide (LPS) O side chain (O antigen) (Waldor et al., 1994), and hence the capsule polysaccharide and O side chain are structurally identical (Knirel et al., 1995) (for a review of the capsule, see (Whitfield, 2006)). The capsule, like the LPS O side chain, is an adhesin that aids colonization of the mammalian gut (Waldor et al., 1994).

After selection with either phage, T1 or T6, the most abundant mutants were disrupted in gene *wecA* (average enrichment over simulated input abundance: 947 for T1, 877 for T2), undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.-). This gene encodes an enzyme known to be key in the synthesis of certain types of capsules, as it initiates the polymerization of the polysaccharide by transfer of the first sugar subunit: N-acetyl-glucosamine from uridine diphosphate (UDP) to undecaprenyl-phosphate (Und-P), a universal lipid carrier of glycan biosynthetic intermediates for extracellular polysaccharides (Cunneen et al., 2013).

Other genes involved in capsule formation were also significantly enriched, though the next most enriched, an exported protein of unknown function, was only about 10–30% as enriched as *wecA* (minimum and maximum enrichment of all selections: 91.1 and 296). Other genes required for phage infection were a gene with unknown function, but containing a glyceraldehyde-3-phosphate dehydrogenase-like, N-terminal domain; *udp-galactofuranosyl transferase glft2*; *kpsM*, polysialic acid transport protein; glycosyl transferase, group 2 family protein; *UDP-galNAc: polypeptide alpha-n-2 acetylgalactosaminyltransferase-t1*; *kpsD*, capsular polysaccharide export system periplasmic protein; a transferase; proteins with similarity to heparan sulfate 3-O-sulfotransferase isoforms 1 and 2; and *kpsE*, capsular polysaccharide export system inner membrane protein. Except for *wecA*, these genes are organized contiguously in an apparent operon on chromosome 1 for capsule synthesis (genes: *Chr1\_consensus\_122* through *Chr1\_consensus\_141*), indicated by the presence of the *kps* genes, suggesting that a single pathway is required for infection by the T1 and T6 phages. Of note, many genes in this pathway were unannotated by the initial annotation pipeline, Rapid Annotation Subsystem Technology (RAST), supporting the hypothesis that many unannotated genes in bacterial genomes are phage receptors (Rodriguez-Valera et al., 2009).

### A.2.7 Capsule dynamics and targeting by phage

Capsule presence/absence is known to affect ecological phenotype and disease association. In *Vibrio vulnificus*, for example, the capsule confers virulence (Wright et al., 1990), and is maintained at a frequency of approximately  $1 \times 10^{-4}$  by phase variation. Both opaque (capsuled) and clear (unencapsuled) colonies are frequently observed deriving from a single strain, and are associated with different phenotypes. Although both colony types can cause disease in eels (Biosca et al., 1993), only opaque colonies are isolated from diseased eels. Capsuled colonies have greater adherence to eel mucus and thus are more adapted to

transmission in water; in fact, the capsule is apparently essential for waterborne infectivity (Amaro et al., 1995).

Consistent with our findings on resistant mutant susceptibility to other phage, the capsule act as barrier for some phages (Scholl et al., 2005), but is a known target for others. In enterobacteria, “K-phages” carry multiple tail fibers that allow infection of hosts with different capsule types (Scholl et al., 2001). While phage adsorption to the capsule is a reversible process, it is often followed by binding to a surface protein receptor, an irreversible one. Some phages are known to then use hydrolitic tail fiber proteins to degrade the capsule and gain access to the cell surface (Tomlinson and Taylor, 1985).

Loss of the capsule increases the hydrophobicity of cell surfaces, which can result in cell flocculation (Wright et al., 1990). In this study, phage resistant mutants, in fact, were observed in several cases to be highly flocculating in liquid culture, unlike the wild type.

### A.2.8 Capsule synthesis genes are diverse among the *Vibrio*

To investigate how diverse the capsule may be within the *Vibrio*, we performed a nucleotide BLAST search using the significantly enriched gene products from the capsule synthesis operon (13 genes) against an internal *Vibrio* database of over 800 environmental isolates (Table A.5). Capsule associated genes were found to be abundant in the *Vibrio* (three genes present in over half of the collection), but can be highly diverse (four genes show less than 40% identity), suggesting that although capsule presence is likely important to an organism’s ecology, the specific capsular polysaccharide composition may not be.

The observation that changes in the capsule affect phage susceptibility profiles suggest that the capsule may be under diversifying selection specifically to outpace phage co-evolution.

## A.3 Conclusion

Using a high-throughput mutant screen, coupled to massively parallel sequencing, we identified a receptor recognized by two environmental phages: the extrapolysaccharide capsule.

For non-pathogenic strains, why might the capsule be retained? One hypothesis is that, despite the cost of phage susceptibility, the benefit of capsule formation is in the ability to colonize particulate organic matter, a known habitat type of the *Vibrio* sp. 13 population. Future work should test the effect of capsule loss on this phenotype using the resistant strains to identify if such a tradeoff exists.

Table A.5: Capsule operon association genes enriched in phage-resistant mutants.

Gene	Enrichment <sup>a</sup>	# of hits with e-value <0.001 <sup>b</sup>	% ID of best BLAST hit	Annotation <sup>c</sup>
Chr1_consensus_126	34	499	80	[RAST] Capsular polysaccharide ABC transporter, ATP-binding protein KpsT
Chr1_consensus_127	110	498	62	[RAST] Glycosyl transferase, group 2 family protein
Chr1_consensus_128	60	20	35	[PHYRE2 - 100/68] transport, transferase; crystal structure of heparan sulfate 2-o-sulfotransferase2 from gallus gallus as a maltose binding protein fusion
Chr1_consensus_129	69	4	28	[PHYRE2 - 100/82] transferase; crystal structure of heparan sulfate 3-o-sulfotransferase2 isoform 1 in the presence of pap
Chr1_consensus_130	124	103	58	[RAST] Glycosyltransferase
Chr1_consensus_131	33	498	86	[RAST] dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)
Chr1_consensus_134	158	149	40	[PHYRE2 - 99.9/90] transferase; udp-galactofuranosyl transferase glft2;
Chr1_consensus_135	106	135	47	[PHYRE2 - 100/89] transferase; the crystal structure of udp-galnac: polypeptide alpha-n-2 acetylgalactosaminyltransferase-t1
Chr1_consensus_136	45	493	83	[RAST] dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)
Chr1_consensus_138	60	22	61	[RAST] Capsular polysaccharide export system inner membrane protein KpsE
Chr1_consensus_139	141	39	71	[RAST] Polysialic acid transport protein KpsM
Chr1_consensus_140	75	526	96	[RAST] Capsular polysaccharide export system periplasmic protein KpsD
Chr1_consensus_141	296	13	98	[PHYRE2 - 20.1/26] NAD(P)-binding Rossmann-fold domains; Glyceraldehyde-3-phosphate dehydrogenase-like, N-terminal domain

a. Enrichment is the average read count ratio of the output library to 100 simulations of the input library, carried out in Con-ARTIST as described in the Materials and Methods. Samples are given letters. to represent replicates.

b. Protein-protein BLAST gainst the *Vibrio* internal database (>800 strains).

c. Annotated with RAST or Phyre2. Phyre2 annotation includes scores: (Confidence)/(Coverage).

## A.4 Materials and methods

### A.4.1 Phage isolation and purification

The bacteriophages used in this study, T1 and T6, are summarized in Table A.1. Bacteriophages that infect *Vibrio* sp. 13 strain 9CS106 were obtained by enrichment culture from water samples, by adding 1 mL of an overnight culture of wild-type 9CS106 to a solution of 500 mL sea water or 0.2  $\mu\text{m}$ -filtered ASW amended with Marine Broth 2216 (Difco) (37.4 g per L). Enrichment cultures were left on the bench, and briefly hand shaken twice a day for one week. Phages were purified and amplified into stock lysates from initial enrichments as follows. Source phage plaques were generated by spotting 0.22  $\mu\text{m}$  filtrate from each enrichment treatment onto a lawn of 9CS106 host formed by agar overlay. For all treatments where plaques were observed (1, 2, 3, and 6) a single plaque was selected for serial purification. Source plaques were touched with a pipette tip and streaked into molten agar overlays of host [100  $\mu\text{L}$  of overnight host culture in 2216 Marine Broth, 2.5 mL of top agar (2216, 0.4 % Bacto Agar, 5 % glycerol), on a bottom agar plate (2216 Marine Broth, 1 % Bacto Agar, 5 % glycerol)], allowed to form new plaques, and a single isolated plaques was then used to inoculate the next round of serial purification for a total of three re-streakings of the phage from each enrichment plating.

Lysates were prepared from purified phages by performing plaque soaks, primary lysates, and secondary lysates, as follows. Plaque soaks of isolated plaques from the third serial streak were used as source material for the primary lysate and were prepared by inserting a pipette tip into the agar at the location of a plaque, removing the agar plug containing the plaque, transferring this plug to 200  $\mu\text{L}$  of SS35+ media (Hussain, 2013) to allow phage particles to diffuse into the media, and incubating overnight at 4 °C. Primary lysates were generated from plaque soaks by transferring 100  $\mu\text{L}$  of overnight soak to a fresh tube, adding 100  $\mu\text{L}$  of fresh media, inoculating with 20  $\mu\text{L}$  of host overnight, holding at room temperature for 30 min without shaking, subsequently shaking for 90 min (11 hours for phage T2) and then centrifuging to pellet cells before transferring supernatant to fresh tube for temporary storage at 4 °C.

Primary lysate titers were determined by 3  $\mu\text{L}$  drop spot assays of serial dilutions onto host agar overlays and used to determine plating volumes for secondary lysates. Secondary lysates were plated on the same media as described above but using 150 mm petri dishes, 250  $\mu\text{L}$  of host overnight culture, and 7.5 mL of molten top agar. To harvest secondary late lysates 10 mL of SS35+ was added to the agar overlays with confluent lysis and the overlays immediately scraped and collected into a 50 mL tube, harvested overlays were soaked overnight at 4 °C, then centrifuged for 20 min at  $5000 \times g$  to remove agar, and finally 0.2  $\mu\text{m}$ -syringe-filtered using Sterivex barrel filters. Lysates were immediately titered using 5  $\mu\text{L}$  drop-spot plating of small-volume dilution series onto host agar overlays lawns and determined to be approximately as follows, given in PFU per mL: T1 -  $2 \times 10^6$ ; T2 -  $2 \times 10^4$ , T3 -  $2 \times 10^8$ ; T6 -  $4 \times 10^8$ .

#### **A.4.2 Transposon mutagenesis**

See chapter 3 for methods.

#### **A.4.3 Phage exposure and selection of resistant mutants**

To select resistant mutants, 9CS106 mutant library was mixed with phage to allow for adsorption, and the mixture plated. A frozen aliquot of the 9CS106 mutant library was rinsed and resuspended in ASW amended with 10% minimal medium (for composition, see appendix B.2.4), and kept at 4 °C for 12 hours before use. 900  $\mu$ L of the resuspended mutant library was grown in 15 mL of 2216 for approximately 3 hours until OD was within range OD<sub>600</sub> 0.5 to 0.9. Cells were centrifuged for 5 min at 5000  $\times g$ , and resuspended in 1 mL of 2216. To enable adsorption, 100  $\mu$ L each of phage concentrate and cells were gently mixed in a microfuge tube, and incubated for 20 min at room temperature. The mixture was then plated on bottom-agar plates (37.4 g 2216 dry media, 10 g Bacto Agar, 50 mL glycerol, and 950 mL H<sub>2</sub>O per L), and allowed to grow for 2 to 3 days. Resistant colonies were harvested from the plate by adding 2 mL of ASW and using a plate spreader to resuspend the cells into solution.

#### **A.4.4 Preparation and sequencing of resistant mutants**

See chapter 3 for methods.

#### **A.4.5 Analysis of resistance-associated genes in phage-exposed mutant libraries**

Resistance-associated genes were identified on the basis of overrepresentation of transposon insertions in libraries recovered following phage exposure. To determine significant overrepresentation, the pipeline Con-ARTIST (Analysis of high-Resolution Transposon-Insertion Sequences Technique for Conditionally essential loci) was used (Pritchard et al., 2014). Con-ARTIST normalizes for the difference in insertion site saturation between the control and experiment data by simulating equal saturation input libraries 100 times, and significant changes in gene abundance were assessed with the Mann-Whitney U test between the output library and each simulation.

#### **A.4.6 Host range assay of isolates recovered from phage exposure**

Resistance to phage after exposure was tested for a subset of colonies from each phage-selection experiment to both confirm resistant phenotype as well as assess changes in susceptibility to a panel of additional phages.

Four surviving colonies were picked and streaked from each of the phage-exposure plates, following outgrowth a single colony was picked from each of these streaks and inoculated

into 4 mL of 2216 Marine Broth and allowed to grow overnight at room temperature at 240 rpm (VWR orbital shaker) and then mixed 50:50 with 50 % glycerol to prepare a glycerol stock for storage at  $-80^{\circ}\text{C}$ .

Phage sensitivity was tested for all 16 mutant isolates and the wild type 9CS106 by application of 5  $\mu\text{L}$  drop spots of phage lysate onto agar overlays prepared with 100  $\mu\text{L}$  of overnight culture, 2.5 mL top agar (2216 Marine Broth, 0.3 % bacto agar, 5 % glycerol). Phage lysates used included four phages isolated in this study (T1, T2, T3, T6), three phages isolated on the same host using water samples collected in January of 2015 (B1, B2, B3; Hussain, unpublished), and 11 phages isolated on strains from the same host population in the summer of 2010 (1.030.O, 1.135.O, 1.155.O, 1.164.O, 1.210.O, 1.213.O, 1.215.A, 1.216.O, 1.238.A, 1.238.B, 1.246.O) (Kauffman, 2014).

#### **A.4.7 Electron microscopy of host cells and isolated phages**

Host and mutant cell samples were prepared for electron microscopy by inoculating 4 mL 2216 Marine Broth directly from glycerol stocks and growing for 9 hr on an orbital shaker at room temperature. Drops of 10  $\mu\text{L}$  of culture were then placed onto grids (Ted Pella 01803-F, Formvar/Carbon 200 mesh TH, Copper) that had been glow-discharged for 5 sec immediately prior to use and allowed to adsorb for 12 min before wicking and washing with 60  $\mu\text{L}$  of 1 % 0.02  $\mu\text{m}$ -filtered uranyl acetate stain. A residual drop of stain was allowed to remain on the grid after staining for an additional 30 sec before being wicked to dryness. Phage lysate samples were prepared for electron microscopy as described above with the exception that samples were held for 14 min and staining was by simple direct addition of 10  $\mu\text{L}$  of stain for 30 sec. All sample grids were viewed on a JEOL 1200 transmission electron microscope at 60 kV and images recorded using Advanced Microscopy Techniques XR41S side-mounted CCD camera and software.

#### **A.4.8 Annotation of 9CS106 Extrachromosomal Element 1**

Proteins of 9CS106 ECE1 were more extensively annotated due to detection of a phage-like gene in the initial Rapid Annotation Using Subsystem Technology (RAST)-based annotation applied to the entire genome (Aziz et al., 2008). A similarity search using BLAST was conducted against all mobile genetic element proteins available in the ACLAME v0.4 webserver database (Leplae et al., 2004). A protein homology search was conducted using the Protein Homology/analogy Recognition Engine v2.0, PHYRE2 (Kelley et al., 2015).

# Appendix B

## Protocols

### B.1 Tn-mutagenesis

#### B.1.1 Materials

**Strains.** *E. coli* donor, EC3-4 *Vibrio* strain

**Stocks.**

Dap 50 mM

Cm 34 mg / mL

Kan 10 mg / mL

#### B.1.2 Prep for cell growth

1. Make plates
  - 2216 plates for *Vibrio* (for. ex., 1 L, 1.5% agar)
  - LB plates for EC3-4 donor (for. ex., 1 L, 1.5% agar)
2. Add Dap, Cm, and Kan to individual plates as needed for streaking out EC3-4.

Final concn	Stock concn	Vol to add to 25-mL plate
DAP300 (mM)	50 mM	150 $\mu$ L
Cm12.5 ( $\mu$ g / mL)	34 mg / mL	9.19 $\mu$ L
Kan50 ( $\mu$ g / mL)	10 mg / mL	125 $\mu$ L

#### B.1.3 Prep for mutant selection

1. To select mutants, make 2216 Cm25 plates. To generate 36,000 mutants, which is a typical mutant library size, with 200 CFU per plate, you will need 180 plates.

- 6 L of 2216 for 200 plates of 2216 Cm25.
2. When partially cooled, add Cm (367.6  $\mu$ L per L of media).
  3. To facilitate plating of the mutants, let plates dry on bench 2 days. Suggested, though not strictly necessary.

#### B.1.4 Grow cells

1. Streak on plates for growth o/n.
  - *Vibrio* on 2216 and grown at room temperature.
  - EC3-4 on LB DAP300, Cm12.5, Kan50 at 37°C.
2. Note: For the 9CS106 mutant library, a 9CS106 colony used for the 5-mL culture was taken from a plate that had been streaked, left to grow for 1 day at room temperature, and stored at 4°C for 30 days. Further experiments had shown storing the recipient in 4°C worked consistently to give higher mutant yields for the 9CS106 strain, and so might be a consideration for applying this method to other strains.
3. The day before mating, inoculate 5 mL of the appropriate media with a single colony of each strain.
  - *Vibrio* in 2216
  - EC3-4 in LB with 30  $\mu$ L of DAP 50 mM and 1.84  $\mu$ L of Cm
4. Grow, shaking, for 7.5 hr.
  - *Vibrio* at room temperature.
  - EC3-4 at 37°C.

#### B.1.5 Mate cells

1. Set hybridization oven to 30°C.
2. For mating plates, make 2216 DAP300 plates: spread 150  $\mu$ L of DAP 50 mM on a 2216 plate.
3. Prepare plate for mating: place 4, 25 mm filters on the 2216 DAP300 plate. Note, although 25 mm 0.2- $\mu$ m Nuclepore Track-Etch membrane polycarbonate filters (Whatman, cat. no. WHA110606) were originally used to generate the 9CS106 mutant library, an alternative is to use MF Membrane Filters with nitrocellulose, filter type 0.45  $\mu$ m HA, ref HAWP02500. Fatima found papers using these to enable the mating spot to form a spreading biofilm.



4. Spin down 600  $\mu\text{L}$  of EC3-4, at 5000 x g for 5 min.
5. Rinse to remove the antibiotic with 1 mL artificial sea water (ASW).
6. Add 600  $\mu\text{L}$  of *Vibrio* recipient (ratio is 1:1).
7. Spin down.
8. Decant.
9. Resuspend in remaining liquid + extra ASW for final 240  $\mu\text{L}$  (20  $\mu\text{L}$  per 12 mating spots): e.g, add 220  $\mu\text{L}$  of ASW to ensure can get total cells plated.
10. Put four 20  $\mu\text{L}$  cell suspension spots on each filter.
11. Mate o/n (for 23 hr) at 30°C. Note start and end times.

### B.1.6 Check mutant yield

1. Get plates to check yield and mutation efficiency: 2216 Cm25 plate for mutants, and 2216 for total.
2. Resuspend the mating spots. If using the polycarbonate filters, scrape the mating spots off with a bent pipette tip and resuspend 1 spot per 100  $\mu\text{L}$  of salty media (ASW, 2216, or LB NaCl 0.5 M). To break up the spots, vortex for 20 sec or pipette. If using the nitrocellulose filters, place 6 filters in 2 mL of salty media and vortex on setting 6 until the filters clear. Transfer the supernatant to a clean tube.
3. Prepare a PCR strip tube with 90  $\mu\text{L}$  of ASW in each well.
4. Make dilutions (1e1-1e8x) in the strip tube by adding 10  $\mu\text{L}$  from the mating resuspension into the first dilution, and serially transferring 10  $\mu\text{L}$  to make each subsequent dilution.
5. Spot-plate dilutions in 10  $\mu\text{L}$  volumes on both the selective 2216 Cm25 and non-selective 2216 plates.
6. Let plates grow at room temperature.
7. Store the rest of the mating spot resuspension in 25% (v/v) glycerol. E.g., if 590  $\mu\text{L}$  of mating spot resuspension remains, add 590  $\mu\text{L}$  of 50% (v/v) glycerol to the resuspension.
8. Store at -80°C.

9. After a day, count mutant CFUs from each plate. A yield of >2,500 CFU per 100  $\mu\text{L}$  is great; with about 1.2 mL (volume prior to addition of glycerol) of mating spot resuspension stored, plating all of it would give 30,000 mutants. If much less, may have to do more mating spots to get a higher yield. And, in fact, a yield of over 100,000 mutant genotypes desirable to query about every mutagenizeable gene. You can calculate the efficiency of the mutagenesis by normalizing the mutant concentration by the total *Vibrio* concentration (estimated from growth on the 2216 plate).
10. If yield is good, continue on to plating the mutants.

### **B.1.7 Plate remainder of mating spot resuspension**

1. Rinse mutants of glycerol by centrifuging for 7 min at 5,000 x g.
2. Remove glycerol. Some glycerol can remain, since it is unlikely to affect colony growth.
3. Rinse with 1 mL ASW.
4. Centrifuge 3 min at 5,000 x g.
5. Decant.
6. Resuspend in 1.2 mL ASW.
7. Dilute the cells to give approximately 200 CFU per plate.
8. Have at least 2 friends working with you.
9. Each person takes 1 mL at a time from the stock of mutants. Keep cells on ice.
10. Either using glass beads or a plate spreader, plate 50  $\mu\text{L}$  of cells per 2216 Cm25 plate.
11. Can pipette cells onto 5 plates at a time, with no need to flame sterilize spreader in between spreading, since samples are the same between plates.
12. Wait a day or two for CFUs to come up.

### **B.1.8 Collect mutants**

1. For each plate of mutant colonies, pipette 2 mL of ASW.
2. Use a plate spreader to scrape off mutant colonies, resuspending them in the ASW.
3. Pipette the resuspension. Approximately 1 mL is what's not soaked into the plate.
4. Spin down the approximately 1 mL collected for 1 min at 5,000 x g. (Centrifuge longer if the pellet is too loose.)

5. Remove 700  $\mu\text{L}$  of liquid.
6. Resuspend the pellet, and transfer to a 50-mL conical tube kept on ice.
7. Once all colonies are harvested, try to get the cells well distributed by:
  - vortexing at setting 5 for 30 sec, and
  - sonicating for 30 sec
  - filtering with a 5  $\mu\text{m}$  filter (Whatman)
8. Aliquot mutants into eppendorf tubes: 250  $\mu\text{L}$  of 50% glycerol + 250  $\mu\text{L}$  of resuspended mutant cells.
9. Store at  $-80^{\circ}\text{C}$ .

### **B.1.9 Mutant library resuspension**

1. Get ice.
2. Get an aliquot of mutant library from downstairs  $-80^{\circ}\text{C}$  freezer.
3. Thaw the aliquot on ice.
4. Spin down 5,000 x g for 7 min.
5. Remove glycerol.
6. Resuspend in 1 mL ASW for rinse 1.
7. Spin down 5,000 x g for 3 min.
8. Pipette off the supernatant.
9. Resuspend in 1 mL ASW for rinse 2.
10. Spin down 5,000 x g for 3 min.
11. Pipette off the supernatant.
12. Resuspend in 1 mL ASW + Minimal Medium.
13. Store at  $4^{\circ}\text{C}$  for 12 hr before use.

## B.2 Media preparation

### B.2.1 Marine Broth 2216

1. In clean bottle, mix 37.4 g per L of MQ H<sub>2</sub>O.
2. Autoclave (121°C) on liquid run for 50 min.

### B.2.2 Artificial Sea Water (ASW)

1. In clean bottle with stir bar, mix 40 g Sigma Sea Salts per L of MQ H<sub>2</sub>O.
2. Mix for 5-10 min.
3. Filter sterilize with Corning 0.2 μm filter system.
4. Transfer to autoclaved bottles if needed.

### B.2.3 Minimal Medium (MM) Trace Elements and Vitamins

Filter through 0.2 μm pore size, and store at 4°C in the dark (i.e., wrapped in foil).

1. Trace Element Solution (1000x) after (Tibbles and Rawlings, 1994).

	1 x 1 L
H <sub>3</sub> BO <sub>3</sub>	2.86 g
MnCl <sub>2</sub> ·4 H <sub>2</sub> O	1.81 g
ZnSO <sub>4</sub> ·7 H <sub>2</sub> O	0.24 g
CuSO <sub>4</sub> ·5 H <sub>2</sub> O	0.075 g
CaCl <sub>2</sub> ·6 H <sub>2</sub> O	0.0494 g
NiCl · 6 H <sub>2</sub> O	0.005 g

2. Vitamin Mix (1000x)

	1 x 1 L
aminobenzoate	40 g
D(+) Biotin	10 g
Folate	30 g
Lipoate	10 g
Nicotinate	100 g
Ca-D-(+) pantothenate	50 g
Pyridoxaminedehydrochloride	100 g
Thiamine hydrochloride	100 g
Vitamin B12	50 g

### B.2.4 Minimal Medium (MM)

1. In a clean, autoclaved beaker with a stir bar, make Part 1.

1 x 500 mL	
NaCl	25.95 g
MgSO <sub>4</sub> ·7H <sub>2</sub> O	3 g
MgCl <sub>2</sub> ·6H <sub>2</sub> O	2 g
CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.12 g
Tris 1 M pH 8.0	25 mL
Na <sub>2</sub> EDTA 0.5 M	2.7 mL
NH <sub>4</sub> Cl 1 M	10 mL
MilliQ H <sub>2</sub> O (MQ H <sub>2</sub> O )	400 mL

2. Adjust pH 7.8 (typically from a pH of 8.1).
3. Bring volume to 500 mL with MQ H<sub>2</sub>O .
4. In another clean, autoclaved beaker with a stir bar, make Part 2.

1 x 500 mL	
K <sub>2</sub> HPO <sub>4</sub>	0.8 g
K H <sub>2</sub> PO <sub>4</sub>	0.2 g
MQ H <sub>2</sub> O	400 mL

5. Adjust to pH 8.0.
6. Bring volume to 500 mL with MQ H<sub>2</sub>O.
7. Autoclave both Part 1 and 2, keeping in separate bottles.
8. Once cooled, when ready to use, add Part 1 and 2 in equal proportions in a sterile bottle.
9. Add 1000x Vitamins, 1000x Trace metals for NDG, 1000x Na<sub>2</sub>MoO<sub>4</sub> (0.03 g per mL), 1000x FeSO<sub>4</sub> \* 7 H<sub>2</sub>O (0.015 g per mL) to final concentration 1x, i.e. 1 mL per L. (Stocks are 0.2- $\mu$ m filter-sterilized. 1000x refers to vol/vol.)

### B.2.5 Freeze-dried, ground *Apocyclops*

1. Measure out approximately 10 g of freeze-dried *Apocyclops roi* (Brine Shrimp Direct), kept in drawer.
2. Blend in Osterizer blender at Liquefy setting for 30 sec.

3. In clean bottle with stir bar, mix 1 g into 900 mL of sterile ASW.
4. Pasteurize.
5. Filter sterilize with Corning 0.2  $\mu\text{m}$  filter system.
6. Mix in 100 mL MM.

### **B.2.6 Powdered *Fucus***

1. Take *Fucus* (Starwest Botanicals) from  $-20^{\circ}\text{C}$ .
2. Measure 10 g of *Fucus* per L ASW.
3. In clean bottle with stir bar, mix 10 g into 900 mL of sterile ASW.
4. Pasteurize.
5. Mix in 100 mL MM.

### **B.2.7 Alginate**

1. In an acid-washed glass bottle with stir bar, mix 1 g alginate (Sigma A2158-100G) with 50 mL Part II Minimal Medium (MM).
2. Stir for 2 hr to dissolve.
3. Sterile filter by hand with reloadable filter cartridge (filter: 25 mm 0.2- $\mu\text{m}$  Nuclepore Track-Etch membrane polycarbonate filters (Whatman, cat. no. WHA110606).
4. Add 50 mL of Part I MM.
5. Add 900 mL of sterile ASW.
6. Add 100  $\mu\text{L}$  of the additives given in appendix B.2.4.

### **B.2.8 Glucose**

1. In a 50 mL falcon tube, add 1 g glucose to 50 mL Part II Minimal Medium (MM).
2. Invert approximately 50x to dissolve.
3. Sterile filter by hand with reloadable filter cartridge (as for alginate).
4. Add 50 mL of Part I MM.
5. Add 900 mL of sterile ASW.
6. Add 100  $\mu\text{L}$  of the additives given in appendix B.2.4.

## B.2.9 Phage isolation

### Phage enrichment.

1. Mix 5x 2216 with MQ H<sub>2</sub>O.
2. Autoclave 5x 2216 and 1-L flasks (1 per sample enriching phage from).
3. Dilute 5x 2216 to 1x with sea water samples to volume of 500 mL + o/n culture 1 mL.
4. For a *Fucus* sample, dilute 5x 2216 to 1x with Artificial Sea Water amended with 10% v/v Minimal Medium (ASW+MM).
5. If doing from a potentially toxic/anoxic sample, use 150 mL sample + 100 mL 5x 2216 + ASW+MM to 500 mL.
6. Leave on bench, shake 2x day, for about 1 week. In the meantime, prepare for the phage isolation.

### Preparation for phage isolation.

1. Make agar plates and top agar.
  - (a) 37.4 g 2216 dry media
  - (b) 10 g Bacto Agar (1%)
  - (c) 50 mL Glycerol
  - (d) 950 mL H<sub>2</sub>O
2. The day before isolating phage, grow a 5 mL culture of the strain of interest.

### Phage isolation.

1. Warm the top agar to 50°C with stirring.
2. Filter 10 mL of the phage enrichment culture through a 0.2  $\mu$ m Sterivex filter.
3. On a large bottom agar plate, add 7 mL of top agar, and 250  $\mu$ L of o/n cell culture. Mix the two vigorously on the plat. Be careful to keep the pipette sterile, returning it to its plastic casing between aliquoting 7 mL of top agar.
4. For a streak plate, add 10  $\mu$ L of the phage filtrate.
5. Make a first streak, gently pulling a 1-mL pipette tip from the liquid spot of phage into the top agar. Switch tips and make a second streak.
6. For a drop spot plate, make a numbered grid on the back of the plate, and add 5  $\mu$ L of the phage filtrate in its corresponding grid cell. Include a negative control with media, e.g. 2216, and positive controls (previously isolated phage) if any are available.

## B.3 Phage selection

### B.3.1 Growth media

1. Make bottom agar plates, and leave to dry for two days to prevent smeared clearings.
2. Decide on ratios and volumes of phage : mutant cells. For example, 100  $\mu\text{L}$  phage : 100  $\mu\text{L}$  cells, and 10  $\mu\text{L}$  phage : 100  $\mu\text{L}$  cells. Each mixture of phage : mutant cells will be plated on one plate. For a large plate, the maximum volume approximately 250  $\mu\text{L}$ . For a small plate, the maximum volume approximately 100  $\mu\text{L}$ . The volumes do have some play in them, but ideally, you want a volume that will allow for distinct colonies of resistant mutants.
3. On the day before selection, rinse and resuspend cells in ASW, as described in appendix B.1.9.
4. Keep o/n at 4°C, at least 10-12 hours.
5. If doing a wt control, grow o/n of wt cells.

### B.3.2 Growth with phage

1. Jumpstart cell growth.
2. Take 900  $\mu\text{L}$  of resuspended mutant library.
3. Make 2216 15 mL blank, along with tubes needed to jumpstart growth.
4. Grow wt in 15 mL of 2216 at room temp for approximately 3 hours to jumpstart growth (OD<sub>600</sub> approximately 0.5 to 0.9).
5. Spin down to pellet cells - 5 min at 5,000 x g.
6. Resuspend in 1 mL 2216 if ready to use; else ASW + min medium (10% by volume).
7. Adsorb and plate phage with mutants.
8. In microfuge tubes, add phage to cells (e.g. 100  $\mu\text{L}$  phage : 100  $\mu\text{L}$  cells, and 10  $\mu\text{L}$  phage : 100  $\mu\text{L}$  cells).
9. Be gentle, don't vortex.
10. Incubate 20 min at room temp.
11. Plate the phage+cells mixture.
12. Plate controls:



- (a) a positive control that cells grow (plate cells only)
  - (b) a drop spot assay to confirm clearing by phage: Plate cells with beads, wait to set, and add 5  $\mu\text{L}$  of phage onto lawn.
13. Check cell concentration by plating a serial dilution series on 2216 bottom agar.

### **B.3.3 Harvest resistant colonies.**

1. After allowing 1-2 days for resistant colonies to grow, add 2 mL ASW to plate.
2. Resuspend colonies using spreader.
3. Collect into eppendorf tube.

## **B.4 DNA extraction using the MasterPure kit (epicentre)**

Timing: 3-4 hr per Part A (appendix B.4.1) and Part B (appendix B.4.2), separated by a break.

Number of samples: Can do 48 samples at a time (24 fit in the centrifuge).

Number of cells per sample kit designed for: 0.5-1e6 mammalian cells; 0.1-0.5 ml of an overnight culture of *E. coli*, which is likely approximately 1e8-1e9 CFUs. If pellet is larger than kit can accommodate, split sample into multiple tubes for DNA extraction.

### **B.4.1 A. Cell Lysis**

1. Set up tube holder with water in hybridization oven at 65°C.
2. Dilute 1  $\mu\text{L}$  of Proteinase K into 300  $\mu\text{L}$  of Tissue and Cell Lysis Solution for each sample.
3. Pellet cells by centrifugation 13,000 x g for 1 min and discard the supernatant, leaving approximately 25  $\mu\text{L}$  of liquid.
4. Pipette mix to resuspend the cell pellet.
5. Add 300  $\mu\text{L}$  of Tissue and Cell Lysis Solution containing the Proteinase K and mix thoroughly by vortexing, e.g. setting 6, 10 sec.
6. Incubate at 65°C for 15 minutes; vortex mix every 5 minutes.
7. Cool the samples by keeping them at room temperature for 5 min, and add 1  $\mu\text{L}$  of 5  $\mu\text{g}/\mu\text{L}$  RNase A to the sample; mix thoroughly by briefly vortexing, e.g. setting 6, 3 sec.
8. Incubate at 37°C for 30 minutes. Briefly vortex every 10 min.

9. Place the samples on ice for 3-5 minutes, spin down briefly — i.e, run the samples in the centrifuge to 5,000 x g and let the centrifuge come back down — and then proceed with DNA precipitation in Part B.
10. Turn on refrigerated centrifuge in the Chisholm lab: set 4°C, "Fast Temp" to cool to 4°C.

#### **B.4.2 B. Precipitation of Protein and DNA**

1. Make one tube with 500  $\mu\text{L}$  isopropanol per sample to precipitate.
2. Add 175  $\mu\text{L}$  of MPC Protein Precipitation Reagent to 300  $\mu\text{L}$  of lysed sample and vortex vigorously to mix, e.g. setting 6, for 10 seconds.
3. Pellet the debris by centrifugation at 4°C for 10 minutes at 13,000 x g, then an additional 3 min at room temp, 13,000 x g in Polz Lab microcentrifuge.
4. Transfer the supernatant to the isopropanol tube and discard the pellet.
5. Invert the tube several (30-40) times.
6. Incubate at -20°C for 30 min to a day or more. If continuing on to precipitate DNA, make fresh, ice cold 75% ethanol. Turn on centrifuge, hit "Fast Temp" to cool to 4°C.
7. Optional: break.
8. Pellet the DNA by centrifugation at 4°C for 10 minutes at 16,000 x g.
9. Carefully pour off the isopropanol without dislodging the DNA pellet.
10. Rinse twice with fresh, ice cold 75% ethanol, being careful not to dislodge the pellet. Centrifuge briefly if the pellet is dislodged.
11. Centrifuge 13,000 x g 1 min to pool remaining ethanol. Remove all of the residual ethanol with a pipet.
12. Let dry in bench drawer 5 to 30 min.
13. Resuspend the DNA in 35  $\mu\text{L}$  of 0.2  $\mu\text{m}$ -filtered Qiagen EB (Tris 10 mM, pH 8.0). Leave at room temp for approximately 30 min, and leave at 4°C o/n to rehydrate. Vortex to resuspend well (e.g. setting 4 for 2 sec). Can also put in 60°C water bath for 1 hr; periodically mix the solution by gently vortexing (e.g. setting 4 for 2 sec).

## B.5 RNase A treatment

1. Dilute sample to 500  $\mu\text{L}$  and add 1  $\mu\text{L}$  of MasterPure RNase A (5  $\mu\text{g}$  per  $\mu\text{L}$ ).
2. Incubate at 37°C for 1 hr.
3. Follow with Ethanol Precipitation.

## B.6 Ethanol Precipitation of DNA

1. Put 100% ethanol on ice - in -20°C. Make fresh 75% ethanol.
2. Add salt (3 M sodium acetate, pH 5.2) to a final concentration of 10% (vol/vol) to sample.
3. Mix well, e.g. vortex at setting 4-5 for 3 sec, and spin down.
4. Add to 2x volume of ice-cold 100% ethanol, vortex, and spin down.
5. Keep on ice or put at -20°C for 30 min.
6. Turn on centrifuge for 0°C.
7. Centrifuge samples for 10 min at 16,000 x g, at 0°C.
8. Retrieve immediately.
9. Remove supernatant with a pipette.
10. Add 500  $\mu\text{L}$  ice cold 75% ethanol.
11. Spin 10 min at 16,000 x g at 4°C.
12. Remove supernatant with a pipette.
13. Spin briefly to approximately 16,000 x g at room temp.
14. Remove supernatant with a 200  $\mu\text{L}$  pipette.
15. Leave to dry, caps open, in bench for 30 min.
16. Resuspend in 35  $\mu\text{L}$ , or desired volume, of EB (10 mM Tris, pH 8.5).
17. Put at 4°C o/n, and let rehydrate in 60°C oven for 30 min to 1 hr the next day.

Table B.1: **qPCR primer sequences.**

Primers	Sequence (5' - 3')	Product size	Target gene
cat tn-IR 1127F	CGTTTGTGATGGCTTCCATG	200 bp	<i>cat</i> -tn IR
cat tn-IR 1326R	CAGGTTGGATGATAAGTCCC		
cras hsp60 1F	CGACGACAGCAACAGTTCTTGCT	111 bp	<i>hsp60</i>
cras hsp60 110R	GATAACTGCTTTGTCGATACCGCGCT		

## B.7 Check DNA quantities and confirm amplifiable by qPCR

On genomic DNA extracted from mutant library selections, use primers to amplify the antibiotic-transposon junction, as well as the single-copy housekeeping gene in *Vibrio* sp. F13, formerly thought *V. crassostreae*, *hsp60*.

All primers (Table B.1) used at 10  $\mu$ M concentration.

### 1. Reaction conditions:

1 x 25 $\mu$ L reaction volume	
Quantitect	12.5 $\mu$ g
primer F	1.25 $\mu$ L
primer R	1.25 $\mu$ L
DNA	10 $\mu$ L

### 2. Cycling conditions:

- (a) 95°C for 15 min
- (b) 50 cycles of
  - i. 95°C for 45 sec
  - ii. 61°C for 30 sec
  - iii. 72°C for 10 sec
  - iv. Plate read (for SYBR)
- (c) 72°C for 5 min
- (d) Melt Curve 50°C to 95°C: Increment 0.5°C for 5 sec.

## B.8 Preparing DNA library for sequencing

Based on protocols from Timothy van Opijnen's lab and Matthew Waldor's lab, with some modifications. The following protocol takes the MmeI digest steps from the van Opijnen protocol (van Opijnen et al., 2014), and the subsequent steps — blunting, A-tailing,

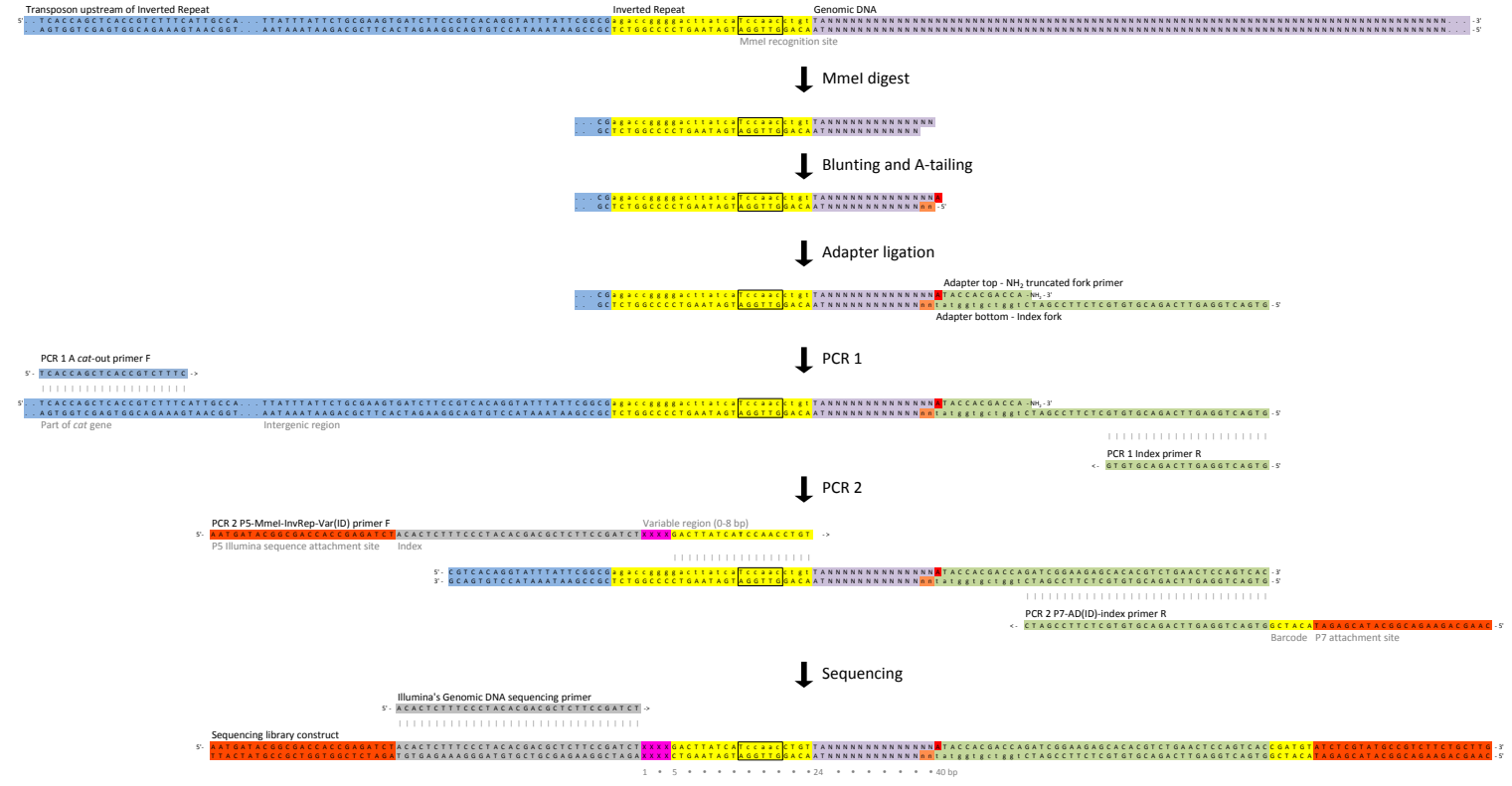
ligation, and PCRs — from the Waldor protocol. A schematic for the particular construct and primers developed in the thesis is shown in Figure B-1.

When preparing a sequencing library, an appropriate negative control is to prepare DNA from un-mutagenized wild-type cells of the same strain alongside mutant DNA. I observed product formation with wild-type cells using the van Opijnen protocol, thus convincing me to use the Waldor protocol despite its greater complexity.

Note that the protocol here is, for the most part, paraphrased from a protocol written by Michael Chao and Yoshiharu Yamaichi in the Waldor lab.

On timing, the MmeI digestion up to the ligation reaction can be done in a single day (approximately 10 hr) for the preparation of around 10 samples. Just block off the whole day, and enjoy doing some molecular biology!

Figure B-1: Preparation of transposon insertion libraries for Illumina sequencing as adapted from van Opijnen (van Opijnen et al., 2014) and Waldor (unpublished) references, and carried out in this thesis.



### **B.8.1 Main reagents and kits**

1. MmeI restriction enzyme (2,000 U / mL) (NEB, cat. no. R0637L) - comes with SAM and CutSmart buffer
2. Quick Blunting Kit (New England Biolabs, cat. no. E1201L)
3. Taq Polymerase (NEB, cat. no. MO273L)
4. dATP solution (NEB, cat. no. N0440)
5. T4 Ligase (400,000 U / mL) (NEB, cat. no. M0202L)
6. QIAquick PCR Purification Kit (Qiagen, cat. no. 28104)
7. Phusion polymerase (NEB, cat. no. M0530L)
8. Oligos (Table B.2)

Table B.2: Oligos for sequencing library

Primer	Sequence <sup>a, b</sup> (5' - 3')
Adapter top - NH2 truncated fork primer	TACCACGACCA-NH2
Adapter bottom - Index fork	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTCTGGTAT
PCR 1 A <i>cat</i> -out primer F	TCACCAGCTCACCGTCTTTC
PCR 1 index primer R	GTGACTGGAGTTCAGACGTGTG
P5-MmeI-InvRep-Var1F <sup>a, b</sup>	AATGATACGGGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT GACTTATCATCCAACCTGT
P5-MmeI-InvRep-Var2F	AATGATACGGGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT CGACTTATCATCCAACCTGT
P5-MmeI-InvRep-Var3F	AATGATACGGGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT ATGACTTATCATCCAACCTGT
P5-MmeI-InvRep-Var4F	AATGATACGGGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT TGTGACTTATCATCCAACCTGT
P5-MmeI-InvRep-Var5F	AATGATACGGGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT TCGACGACTTATCATCCAACCTGT
P5-MmeI-InvRep-Var6F	AATGATACGGGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT GCAGCGACGACTTATCATCCAACCTGT
P7-AD002-index-R	CAAGCAGAAAGACGGCATAACGAGAT ACATCG GTGACTGGAGTTCAGACGCTGT- GCTCTTCCGATC
P7-AD004-index-R	CAAGCAGAAAGACGGCATAACGAGAT TGGTCA GTGACTGGAGTTCAGACGCTGT- GCTCTTCCGATC
P7-AD005-index-R	CAAGCAGAAAGACGGCATAACGAGAT CACTGT GTGACTGGAGTTCAGACGCTGT- GCTCTTCCGATC
P7-AD006-index-R	CAAGCAGAAAGACGGCATAACGAGAT ATTGGC GTGACTGGAGTTCAGACGCTGT- GCTCTTCCGATC
P7-AD007-index-R	CAAGCAGAAAGACGGCATAACGAGAT GATCTG GTGACTGGAGTTCAGACGCTGT- GCTCTTCCGATC



Table B.2: **Oligos for sequencing library**

Primer	Sequence <sup>a, b</sup> (5' - 3')
P7-AD012-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>TACAAG</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC
P7-AD013-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>TTGACT</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC
P7-AD014-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>GGAAct</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC
P7-AD015-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>TGACAT</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC
P7-AD016-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>GGACGG</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC
P7-AD018-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>GCGGAC</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC
P7-AD019-index-R	CAAGCAGAAGACGGCATAACGAGAT <i>TTTCAC</i> GTGACTGGAGTTCAGACGTGT-GCTCTTCCGATC

a. P5 Sequences were modified to reflect the substitution of G to T to generate an MmeI recognition site in the inverted repeat of the mariner transposon.

b. Sequences in italics are variants within the P5 oligos, and different barcodes in the P7 oligos.

### B.8.2 MmeI restriction digestion

I give the essentials below, but I also wanted to include just a note about how I do these reactions. I estimate DNA concentration on a gel, using an *E. coli* DNA standard (see appendix B.8.12), since I found the NanoDrop to overestimate DNA prepared from the epicentre kit. Once DNA concentrations are determined, I set up reactions of DNA + DNase-free water to 176.56  $\mu\text{L}$ . Then, I make a mastermix containing the other reagents, and aliquot 23.44  $\mu\text{L}$  of this mastermix to the DNA+water samples. This approach of preparing the sample and a mastermix is used throughout this protocol.

1. Mix reagents. For one 200- $\mu\text{L}$  reaction (per sample):

<hr/>	
1 x 200 $\mu\text{L}$	
<hr/>	
DNA	5 $\mu\text{g}$
10x CutSmart Buffer	20 $\mu\text{L}$
S-adenosyl methionine	0.44 $\mu\text{L}$
MmeI	3 $\mu\text{L}$
Water	to 200 $\mu\text{L}$

2. Incubate for 2.5 hr at 37°C.
3. Proceed immediately to Ethanol precipitation (appendix B.6), and then Blunting (appendix B.8.3).

### B.8.3 Blunting

Quick Blunting Kit (New England Biolabs, cat. no. E1201L)

1. Mix reagents. For one 55- $\mu\text{L}$  reaction (per sample):

<hr/>	
1 x 55 $\mu\text{L}$	
<hr/>	
digested DNA	up to 5 $\mu\text{g}$
10x Buffer	5.5 $\mu\text{L}$
1 mM dNTP (from kit, not 10 mM dNTPs for PCR)	4 $\mu\text{L}$
Blunting enzyme mix	2 $\mu\text{L}$
Water	to 55 $\mu\text{L}$

2. Incubate the reaction at room temp for 30 min.
3. Proceed immediately to Column purification (appendix B.8.4).

### B.8.4 QIAquick Column Purification

QIAquick PCR Purification Kit (Qiagen, cat. no. 28104). Adapted from p. 19 of the instruction manual.

1. Warm elution buffer (EB) at 37°C.
2. Get 2 sets of tubes ready: 1 set with sides labelled, caps cut, to elute in, and 1 set well-labelled for sample storage. A well-labelled tube, for example:

- Top: “1. Mut 3  $\mu\text{g}$ ”
  - Side: “1. Mut 3  $\mu\text{g}$ , End-filled, 20150407, Clean (Qiagen)”
3. Add 5 volumes of PB to the sample: 275  $\mu\text{L}$  of PB to 55  $\mu\text{L}$  of Quickblunt reaction.
  4. Vortex 2-3 sec on setting 5 and spin down.
  5. Add total volume to column.
  6. Spin down at 17,900 x g for 30 sec.
  7. To wash, open up caps, add 750  $\mu\text{L}$  of PE to columns.
  8. Spin down at 17,900 x g for 30 sec.
  9. Turn the tubes 180°, do a second spin.
  10. Transfer to the clean, side-labelled eppendorf tubes.
  11. Keep caps open for 5 min on the bench to dry.
  12. Elute with 35  $\mu\text{L}$  by adding EB to the filter membrane.
  13. Let stand for 10 min at 37°C before spinning down.
  14. Spin down 1 min at 17,900 x g.
  15. Turn the tubes 180°, do a second spin.
  16. Transfer to well-labelled tubes.
  17. Can pause before continuing on to the next step.

### B.8.5 A-Tailing with Taq

Use the total eluate (approximately 32  $\mu\text{L}$ ) from the previous Column Purification (appendix B.8.4) step.

1. Start 72°C water bath in hybridization oven.
2. Mix reagents. For one 50- $\mu\text{L}$  reaction (per sample):

	1 x 50 $\mu\text{L}$
blunted DNA	32 $\mu\text{L}$
10x PCR standard Buffer	5 $\mu\text{L}$
10 mM dATP <sup>a</sup> .	10 $\mu\text{L}$
Taq Polymerase	3 $\mu\text{L}$

a. Make fresh, or make aliquots and freeze, but do not freeze/thaw.

3. Incubate reaction at 72°C for 45 min.
4. Proceed immediately Column Purification (appendix B.8.4), eluting in 52  $\mu\text{L}$  of EB.

### B.8.6 Anneal adapters

Anneal oligos to form the adapters that will be ligated onto the digested DNA. Using adapters enables subsequent PCR amplification of adapter-transposon containing DNA fragments, and the ability to sequence flanking genomic DNA lying downstream of the transposon insertion.

You will need to make the adapter fresh every time, as storage leads to random loss of basepairs at the ends, negatively affecting ligations. Typical yields from the A-tailing reaction require 2-3 ligation reactions.

The volumes are scaled up from those originally given in the Waldor lab protocol, to make it unnecessary to quantify the DNA (a process that can be unreliable, by NanoDrop, or take time, if done with an agarose gel image). Therefore, the amounts specified per reaction here are assuming a 5  $\mu\text{g}$  amount of DNA per sample, the maximum yield.

Adapter sequences:

**Adapter top - NH<sub>2</sub> truncated fork primer**

5'-TACCACGACCA-NH<sub>2</sub>-3' (fork truncated NH<sub>2</sub>)

**Adapter bottom - Index fork**

5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCtggtcgtggtat-3'

1. Mix reagents. For one ligation, 3.33  $\mu\text{L}$ , scaled from the original reaction mixture for ligating 1.2  $\mu\text{g}$ , to the theoretical maximum 5  $\mu\text{g}$  of DNA:

	1 x 5 $\mu\text{L}$
100 $\mu\text{M}$ Adapter top - NH <sub>2</sub> truncated fork primer	1.60 $\mu\text{L}$
100 $\mu\text{M}$ Adatper bottom - index fork	1.60 $\mu\text{L}$
2 mM MgCl <sub>2</sub>	0.13 $\mu\text{L}$

2. Run on the Chisholm lab thermocycler (settings for 'Anneal', saved in Folder 'F') <sup>1</sup>:

Timing: 1 hr 21 min to run

- (a) 95°C for 4 min
- (b) 95°C for 1min
- (c) Go to 2, 75×
- (d) -1°C per cycle
- (e) End

---

<sup>1</sup>This program incubates the sample for 5 min at 95°C, and slowly ramps down to 20°C.

### B.8.7 Adapter ligation

Modified from the original Waldor protocol, so that the ligation is scaled to the theoretical maximum of 5  $\mu\text{g}$ .

1. Mix reagents. For five ligations, 5  $\mu\text{L}$ , using 1  $\mu\text{L}$  per reaction (scale up as necessary):

<hr/>	
1 x 62.56 $\mu\text{L}$	
<hr/>	
DNA sample (up to 5 $\mu\text{g}$ )	48.8 $\mu\text{L}$
Annealed adapter	3.33 $\mu\text{L}$
10x T4 DNA ligase buffer	6.26 $\mu\text{L}$
T4 DNA ligase (400,000 U per mL)	4.17 $\mu\text{L}$

2. Aliquot 13.76  $\mu\text{L}$  to one PCR strip tube per sample.
3. Transfer 48.8  $\mu\text{L}$  of the DNA sample to the tube.
4. Incubate at 16°C o/n in the thermocycler.
5. The following day, prepare the following to spike into the ligation reaction:

<hr/>	
1 x 41.7 $\mu\text{L}$	
<hr/>	
Water	33.36 $\mu\text{L}$
10x T4 DNA ligase buffer	4.17 $\mu\text{L}$
T4 DNA ligase (400,000 U per mL)	4.17 $\mu\text{L}$

6. Incubate at 16°C for 2 hr.
7. Proceed immediately to Column Purification (appendix B.8.4), but when washing with buffer PE, let PE sit on column for 10 min, then spin down. Wash a second time with PE buffer. Elute in 52  $\mu\text{L}$  of 37°C EB.

### B.8.8 PCR 1. Amplification of transposon-associated DNA

After the adapter ligation, the sample contains genomic DNA fragments (some with transposons and mostly without) with forked adapters on the ends. The following PCR amplifies out from the transposon, into the downstream genomic sequence, to the ligated adapter. Amplification will only occur using the primers, and not from shorter strand of the forked adapter, because of its  $-\text{NH}_2$  block. For the primer and template orientation, see the schematic in Figure B-1.

1. Dilute 500 ng of the sample ligated DNA with PCR water to 190  $\mu\text{L}$ .

2. Make a mastermix of the following reagents:

	1 x 60 $\mu\text{L}$
5x Phusion Buffer	50
100 $\mu\text{M}$ PCR 1 A primer F	1.25
100 $\mu\text{M}$ PCR 1 Index primer R	1.25
10 mM dNTP mix	6.25
Phusion DNA polymerase	1.25

3. Aliquot 60  $\mu\text{L}$  of the mastermix to the 500 ng of ligated DNA.
4. Split the 250  $\mu\text{L}$  volume into five 50- $\mu\text{L}$  reactions in a thin-walled PCR plate.
5. Carry out the following cycling parameters:
  - (a) 98°C for 1 min
  - (b) 98°C for 10
  - (c) 53°C for 30 sec
  - (d) 72°C for 30 sec
  - (e) Go to 2, 29 times
  - (f) 72°C for 10 min
  - (g) 15°C forever
6. Pool the five PCR reactions, and then proceed immediately to Column Purification (appendix B.8.4). Elute in 52  $\mu\text{L}$  of 37°C EB.
7. NanoDrop the sample for the concentration. (NanoDrop works sufficiently well at this stage.)

### B.8.9 PCR 2. Addition of barcode, Illumina, and variability sequences

The sample DNA is now enriched in transposon-genomic DNA junctions: 5' end of the *cat* gene + modified Tn inverted repeat + genomic DNA fragment + linker sequence + index sequence 3'. For the primer and template orientation, see the schematic in Figure B-1.

1. Mix equal amounts (e.g. 10  $\mu\text{L}$ ) of the six P5 Forward primers with variable sequences to create 'P5 MmeI InvRep Var mix (1-6).'
2. Dilute 500 ng of the sample ligated DNA with PCR water to 190  $\mu\text{L}$ .

Table B.3: **Adapter sequences, as observed in sequencing reads.**

Adapter	Index sequence observed
AD02	CGATGT
AD04	TGACCA
AD05	ACAGTG
AD06	GCCAAT
AD07	CAGATC
AD012	CTTGTA
AD013	AGTCAA
AD014	AGTTCC
AD015	ATGTCA
AD016	CCGTCC
AD018	GTCCGC
AD019	GTGAAA

3. Add 1.25  $\mu\text{L}$  of a P7 barcode primer (note which used for which sample, information that will be needed when assigning reads to samples).

Adapters were synthesized with a variable 6-bp sequence. However, this sequence is not what is used to demultiplex the samples, because the sequencing read matches the forward strand, while the adapter sequence was used to synthesize the reverse strand. Therefore, the reverse complement is the observed sequence (given in Table B.3) in the Illumina data.

4. Mix reagents as a mastermix. For one sample:

	1 x 58.75 $\mu\text{L}$
5x Phusion Buffer	50 $\mu\text{L}$
100 $\mu\text{M}$ PCR 2 P5 MmeI InvRep Var Forward primer mix (1-6)	1.25 $\mu\text{L}$
10 mM dNTP mix	6.25 $\mu\text{L}$
Phusion DNA polymerase	1.25 $\mu\text{L}$

5. Add mastermix (58.75  $\mu\text{L}$ ) to the sample.
6. Split the 250  $\mu\text{L}$  volume into five 50- $\mu\text{L}$  reactions in a thin-walled PCR plate.
7. Carry out the following cycling parameters:
  - (a) 98°C for 1 min
  - (b) 98°C for 10

- (c) 55°C for 30 sec \*note the annealing temperature change
  - (d) 72°C for 30 sec
  - (e) Go to 2, 17 times \*note the shorter number of cycles
  - (f) 72°C for 10 min
  - (g) 15°C forever
8. Pool the five PCR reactions, and then proceed immediately to Column Purification (appendix B.8.4). Elute in 32  $\mu\text{L}$  of 37°C EB.
  9. Run a diagnostic gel (2% agarose dissolved in 0.5x TAE with EtBr, run with NEB QuickLoad 100 bp ladder) to check for the 168- to 174-bp product (dependent on the variable region length).
    - (a) Load 2  $\mu\text{L}$  of sample + 8  $\mu\text{L}$  of nuclease-free water + 3  $\mu\text{L}$  of 6x Loading dye.
    - (b) Run for 40 min at a constant 70 V.

### B.8.10 Gel purify PCR product

1. Have a clean, EtBr-free gel box, gel cast, and medium-sized combs (15 teeth) by rinsing them with water.
2. Melt 2 g of agarose in 100 mL of 1x TAE for a 2% gel.
3. While molten and in the flask, add 30  $\mu\text{L}$  of Gel Green to stain nucleic acids.
4. Prepare samples: 4  $\mu\text{L}$  of sample + 6  $\mu\text{L}$  of nuclease-free water + 3  $\mu\text{L}$  of 6x bromophenol blue.
5. Run gel for 55 min at a constant 70 V.
6. Prepare an eppendorf tube for the excised gel piece. Note the weight of the empty tube (generally 0.9 to 0.95 g).
7. Visualize the gel on the DeLong lab small transilluminator.
8. Excise the band using a SafeXtractor gel cutter, and put it in the eppendorf tube.
9. Use the Macherey-Nagel NucleoSpin Gel and PCR Clean-up kit to purify the DNA from the gel.
  - (a) Get 2 sets of tubes ready: 1 set with sides labelled, caps cut, to elute in, and 1 set well-labelled for sample storage. A well-labelled tube, for ex.
    - Top: “1. 2.a”



- Side: “1. 2.a sequencing library 20150529”
- (b) Turn on hybridization oven to 55°C. NTI can be preheated to extract 2% or higher gels.
- (c) Add NTI to the gel sample. Use 200  $\mu\text{L}$  per 100 mg gel for <2%. For 2% or higher gels, use 400-500  $\mu\text{L}$  per 100 mg gel. For example, add 600  $\mu\text{L}$  to 150 mg gel.
- (d) Incubate the sample at 55°C for 15 min, vortexing for 5 sec at setting 6, every 2-3 min. Make sure the gel slice is completely dissolved.
- (e) Add the sample to the column.
- (f) Spin down (maximum volume of 750  $\mu\text{L}$  at a time) for 30 sec at 11,000 x g.
- (g) Wash the column with 650  $\mu\text{L}$  of NT3.
- (h) Spin down for 30 sec at 11,000 x g.
- (i) Repeat wash with 650  $\mu\text{L}$  of NT3.
- (j) Spin down for 30 sec at 11,000 x g.
- (k) To dry the silica membrane, spin down for 1 min at 11,000 x g. Transfer immediately to clean eppendorf tubes.
- (l) Dry on the columns on the bench, with caps open, 5 min.
- (m) Elute DNA with 35  $\mu\text{L}$  of EB. Let sit for 5 min on bench.
- (n) Centrifuge at 11,000 x g for 1 min
- (o) Re-check concentration on a diagnostic gel (2% agarose dissolved in 0.5x TAE with EtBr, run with NEB QuickLoad 100 bp ladder) to check for the 168- to 176-bp product (dependent on the length variable region).
- (p) Store -20°C.

### B.8.11 Submit samples for sequencing

Samples can be submitted to the BioMicro Center (headed by Stuart Levine) in Bldg. 68.

1. Ensure samples have a minimum concentration of 10 ng per  $\mu\text{L}$ .
2. Aliquot 10  $\mu\text{L}$  for the BioMicro Center.
3. Fill out the sequencing submission form. See <http://openwetware.org/wiki/BioMicroCenter:Forms>. For example, on 20150601, I submitted 10 samples for the NextSeq, to be QC'd and pooled by the BioMicro Center. The Fwd read length I specified was 50 bp, and the index (the barcode) 6 bp. The sample type was TnSeq.
4. Ensure that the given barcode (index) sequence is correct per sample. See barcodes (Table B.3).

### B.8.12 Gel-based quantification

1. Make an 0.8% gel. For example, a medium-sized gel is 0.8 g of agarose with 100 mL 0.5x TBE with EtBr.
2. Prepare samples (e.g., 1  $\mu$ L) to run alongside *E. coli* DNA standards: 600 ng, 530 ng, 300 ng, 220 ng, 100 ng, and 30 ng.
3. Load samples and standards.
4. Run gel at constant 70 V for 40 min.
5. Image gel, invert coloration, and export without overlays, at original scan resolution, as .tiff file.
6. Quantify DNA amounts in samples using ImageJ software to quantify pixel area.

## B.9 Processing raw sequencing reads to TA site tallies

2

Once the raw sequencing data is obtained, it has to be processed: the Illumina adapter sequences for making the library are trimmed away to leave the putative stretch of genomic sequence, this candidate genomic sequence is filtered on length, with only those sequences between 14 and 17 bp taken, and then these fragments are aligned to the genome. If they map to TA dinucleotide sites, insertion sites for the *mariner* transposon, the sequences are considered to have derived from actual transposon-insertion mutant genotypes. The read abundance of the genotypes can then be used to infer fitness effects of gene disruption in different conditions.

### B.9.1 Downloading demultiplexed sequencing data

1. In Terminal (on Macs), logon to the server space to save the data to, e.g.:  

```
1 $ ssh Admin@arraypolz.mit.edu
2 $ password: t***h**
```
2. Secure copy (scp) the data (.fastq files), recursively (-r) from the Biomicro center url (polz\_ill@bmc-150.mit.edu:./your data path/\*.fastq) to your local directory (.). To finish the download should take an hour or two. Note that the data should already have been demultiplexed by the BMC, because the construct that the Waldor protocol

---

<sup>2</sup>I wanted to include some details in the thesis about data processing and bioinformatic analysis, since this kind of workflow was mysterious to me at the outset. As well as being for future reference, this documentation is intended to help demystify computational work for beginners.

yields is typical for an Illumina run (unlike the van Opijnen construct), and thus they can easily parse which sequence is the barcode.

```
1 $ scp -r polz_ill@bmc-150.mit.edu:./your data path/*.fastq .
2 password: polz#9Lk
```

3. A full example:

```
1 $ ssh Admin@arraypolz.mit.edu
2 $ password: t***h**
3 $ cd /cygdrive/d/alisontakemura/150601_sequencing/150601_raw_demultiplexed/
4 $ scp polz_ill@bmc-150.mit.edu:~/150601Pol/D15-**-1421L/*.fastq .
5 password: polz#9Lk
```

4. You can check that the quality scores look good by going to the url of the data, and looking at the .fastqc html file.

## B.9.2 Trimming reads using CLC

To obtain the genomic portions of the sequencing reads, we trim away the transposon (5') and adapter sequence (3'). (See Figure B-1.)

1. In CLC Genomics Workbench, in your working directory, import the .fastq file.
2. Click on the input (.fastq) file, and go to the menubar: Toolbox > Workflows > trim\_adapters\_and\_on\_length\_workflow.

Configurable parameters (with default values) include Quality limit (0.005), which filters reads based on quality score, and Maximum number of nucleotides in reads (17) and Minimum number of nucleotides in reads (14), which filters trimmed reads based on length.

The file containing the adapter sequences is saved in the workflow (Alison\_TN\_Adapter\_List-1). The sequences that will be trimmed:

End of the read	Trimmed sequence
5'	GACTTATCATCCAACCTGT
3'	ATACCACGACCAGAT

Note, while I expected 16-17 bp of genomic DNA (see Figure B-1), I have observed 14-15 bp (20150609).

3. Note, you should have somewhere in the range of 100 million reads if you sequenced on the HiSeq (160 to 220 million is the reputed yield), or 200 million (500 million reputed maximum) on the NextSeq. I've observed 21 million reads per sample of 10 multiplexed samples in a single NextSeq lane, and after trimming, observed 17 million reads.

### B.9.3 Mapping reads to reference genome with Bowtie

1. Export the trimmed reads file as a .fastq file. Change filenames so no spaces (e.g. replace spaces with '.').
2. On the CLC machine, execute the Bowtie command, specifying: .sam output (`--sam`), no mismatches (`-v 0`), report all alignments (`--all`), random assignment of reads that map to more than 1 location in the reference genome (`-M1`), guarantee alignment is best in terms of stratum (`--best --strata`), and input is .fastq (`-q`).

Note, Bowtie for Windows can only be used through the command prompt (and not cygwin, the linux-like interface).

```
1 $ python C:\Users\Admin\Downloads\bowtie-1.1.1\bowtie --sam -v 0 -a -M1
2 --best --strata -q < path to the .fastq file >
3 < path and filename for the output .sam file >
```

For ex.

```
1 $ python C:\Users\Admin\Downloads\bowtie-1.1.1\bowtie --sam -v 0 -a -M1
2 --best --strata -q D:\alisonakemura\reference_genomes\indexes\Vpop13_9CS106
3 D:\alisonakemura\150601_sequencing\150601_fasta\20150519_2.a.trimmed.
   trimmed
4 .fastq1
5 D:\alisonakemura\150601_sequencing\20150519_2.a.trimmed.trimmed.sam
```

### B.9.4 Creating a genome gene list and TA site maps for each chromosome

1. Generate a gene list from the genome, using `make_genelist.py`. Essential fields: chromosome = [0], gene = [1], start = [5], end = [6]. For example:

```
1 file: 9CS106_reseq_consensus_solexa_contigs_prod_genelist.txt
2
3 head:
4
5 9CS106_Ch1_consensus 9CS106_Ch1_consensus_1      259 556  +  ""gene_id
   Mobile element protein""
6 9CS106_Ch1_consensus 9CS106_Ch1_consensus_2      603 867  +  ""gene_id
   Unannotated""
```

2. Generate the bp position map of TA sites for each chromosome, using `make_TA_sitelist.py`. Note, the bp count restarts at 1 for each chromosome. All files for 9CS106 as of 20150620:

```
1 9CS106_Ch1_consensus_TA.txt
2 9CS106_Ch2_consensus_TA.txt
```

```
3 9CS106_ECE1_consensus_TA.txt
4 9CS106_ECE2_consensus_TA.txt
```

For example:

```
1 file: 9CS106_Chr1_consensus_TA.txt
2
3 head:
4
5 misc_feature 14
6 misc_feature 59
7 misc_feature 89
```

### B.9.5 Mapping genome-mapped reads to TA sites

Both as input for the ARTIST pipeline, as well as to check that the reads are coming from TA dinucleotide sites, where the transposase inserts the *mariner* transposon. The reads are mapped to TA sites from each chromosome/contig of the organism genome. As of 20150612, V. Pop. 13 has 4 'chromosomes': Chr. 1, Chr. 2, Extrachromosomal element (ECE) 1, and ECE 2.

1. Navigate to folder with .sam genome-mapped reads file(s).

```
1 file: .sam
2
3 head (after header lines):
4
5 No 0 9CS106_ECE1_consensus 160 255 15M * 0 0 TACATTCCGATCATA
   EEEEEEEEEEEEEEE   XA:i:0 MD:Z:15 NM:i:0
6
7 No 16 9CS106_ECE1_consensus 35703 255 15M * 0 0 GTTGAAGACTGGCTA
   EEEEEEEEEEEEEEE   XA:i:0 MD:Z:15 NM:i:0
8
9 No 0 9CS106_Chr1_consensus 1533590 255 15M * 0 0 TAACAGAAGACGTTTC
   EEEEEEEEEEEEEEE   XA:i:0 MD:Z:15 NM:i:0
```

2. Ensure have the following reference input files:

- The \_genelist.txt file (generated above).
- The \_TA.txt file for each chromosome (generated above).

3. The command structure is

```
1 $ python <python script to map reads to TA sites>
2
```

```

3 <path to the files with TAsites per chromosome>
4
5 <path to the the organism gene list>

```

From the folder with .sam file(s), execute the command. For example:

```

1 script: sam_to_TAmap.py
2
3 command:
4
5 $ python /cygdrive/d/alisontakemura/150601_sequencing/
6 sam_to_TAmap.py
7
8 /cygdrive/d/alisontakemura/reference_genomes/
9
10 /cygdrive/d/alisontakemura/reference_genomes/
11 9CS106_reseq_consensus_solexa_contigs_prod_genelist.txt

```

4. The output is a .TAmap file for each .sam file, with which chromosome the TA site is on, the position of the TA site (start and end), the locus it is in, and the tallies of reads that mapped to each TA site, separated by whether it mapped to the forward or reverse strand.

```

1 file: .TAmap (tab-delimited)
2
3 head:
4
5 Chr TA_start  TA_end  Locus Reads_forward Reads_reverse
6 9CS106_Chr2_consensus 2 3 9CS106_Chr2_consensus_1 0 0
7 9CS106_Chr2_consensus 52 53 9CS106_Chr2_consensus_1 0 0
8 9CS106_Chr2_consensus 74 75 9CS106_Chr2_consensus_1 2 0
9 9CS106_Chr2_consensus 80 81 9CS106_Chr2_consensus_1 10 0
10 9CS106_Chr2_consensus 86 87 9CS106_Chr2_consensus_1 0 0

```

## B.10 Checking data quality

### B.10.1 Calculating the overall library saturation

1. In the terminal, navigate to the folder with the .TAmap file(s).
2. Run Tapercent\_AFT\_read\_ct\_per\_hit\_TA.py from this folder. For example,

```

1 $ python Tapercent_AFT_read_ct_per_hit_TA.py

```
3. The output is directed to the screen, and shows counts how many TA sites in the organism, how many reads for the library, how many TA sites were hit, what percent

of TA sites were hit to the total number of TA sites, and the average read count per TA site. For example,

```
1 20150519_mutlib.trimmed.trimmed.sam.TAmap
2 TA = 314091
3 Reads = 18196070
4 Sites Hit = 94245
5 Percent TAs hit = 30
6 Average Read Count per hit TA = 57
7 -----
```

4. Values for each sample can be copied and stored into a tabular format for comparison.

### B.10.2 Calculating average reads per TA site per gene

1. Calculate the percent saturation of TA sites per gene, by navigating to the directory with the .TAmap file, and running the TAccounts\_gene\_AFT\_v2.py script. Note, it will generate an output file (.genePercentTAhit) for every .TAmap file in the directory.

Command structure:

```
1 $ python <script.py>
```

For example,

```
1 $ python /Users/alimura/Dropbox/Public/ARTIST/scripts_and_files_Mike_Chao/
2 TAccounts_gene_AFT_v2.py
```

The output file is shows the percent of TA sites disrupted per gene, as well as other statistics. For example:

```
1
2 file: .genePercentTAhit
3
4 head:
5
6 Locus # of hit TAs  Total TAs Fraction TAs hit  Total reads Avg_Reads per TA
7 9CS106_Chrl_consensus_1 11  14  0.7857142857142857  307 21.928571428571427
8 9CS106_Chrl_consensus_10 80  144 0.5555555555555556  5611
   38.965277777777778
9 9CS106_Chrl_consensus_100 21  72  0.2916666666666667  1294
   17.972222222222222
```

### B.10.3 Graphing a histogram of % of TA sites disrupted per gene

To visualize the library saturation as the distribution of the percentage of TA sites disrupted per gene, use the script Generate\_hist\_geneTApercenthit.py .

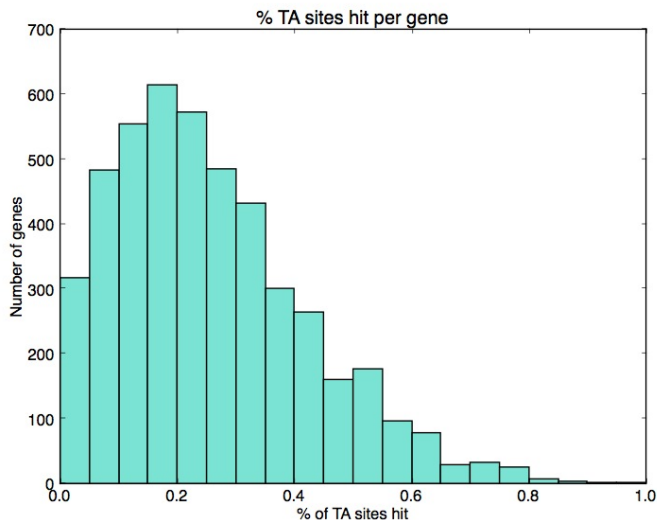


Figure B-2: **An example histogram of the percent of TA sites disrupted per gene** in a sample sequencing library: 2216 media, replicate c.

1. In the terminal, navigate to the folder with the .genePercentTAhit file(s).
2. Run `Generate_hist_geneTApercenthit.py` from this folder.
 

```
1 $python Generate_hist_geneTApercenthit.py
```
3. The output is a histogram. For an example, see Figure B-2.

#### B.10.4 Visualizing reads in Artemis

The read tallies can be graphed along the genome, with demarcated identified orfs, using the freely available program Artemis. To use Artemis, reads from the .TAmmap file are converted into a text file format Artemis can read using the script `Artemize_FR_AFT.py`.

1. Navigate to the folder with .TAmmap file(s).
2. Execute the command
 

```
1 $ python Artemize_FR_AFT.py
```
3. The output extension is .artemis, and contains, for each TA site, forward and reverse read counts. For example,



```

1 file: 20150519_2.a.trimmed.trimmed.sam.TAmap.artemis
2
3 head:
4 #BASE Finsert Rinsert
5 colour 5:150:55 225:0:0
6 0 0
7 63 0
8 1456 0
9 0 0
10 0 0
11 192 0
12 551 0
13 0 0

```

4. To map reads to the genome with orf predictions, you will need .gbk files for each chromosome:

```

1 9CS106_reseq_consensus_solexa_contigs_prod_Chr1.gbk
2 9CS106_reseq_consensus_solexa_contigs_prod_Chr2.gbk
3 9CS106_reseq_consensus_solexa_contigs_prod_ECE1.gbk
4 9CS106_reseq_consensus_solexa_contigs_prod_ECE2.gbk

```

5. Start the program Artemis (I'm using 16.0.0).
6. Import the .gbk file of the genome by going to File >Open File Manager and navigating to and selecting the file. You may get an error, which can be ignored (click No).
7. Import the .artemis data by going to Graph >Add user plot >and navigating to and selecting the file.
8. When the graph appears, right click on it to adjust the visualization parameters:
  - Left click menu to Set the Window Size... and change it to 1. This enables single nucleotide resolution.
  - Left click menu to Set Min/Max Values... and change them to fit your data (e.g., 0 and 100). Disable Scaling.
  - Left click menu to Configure... to adjust the color / weight / fill of the lines.
9. Zoom in / out on the display by adjusting the vertical sliders on the right side of the window.
10. Scroll through the genome by adjusting the horizontal sliders throughout the window.

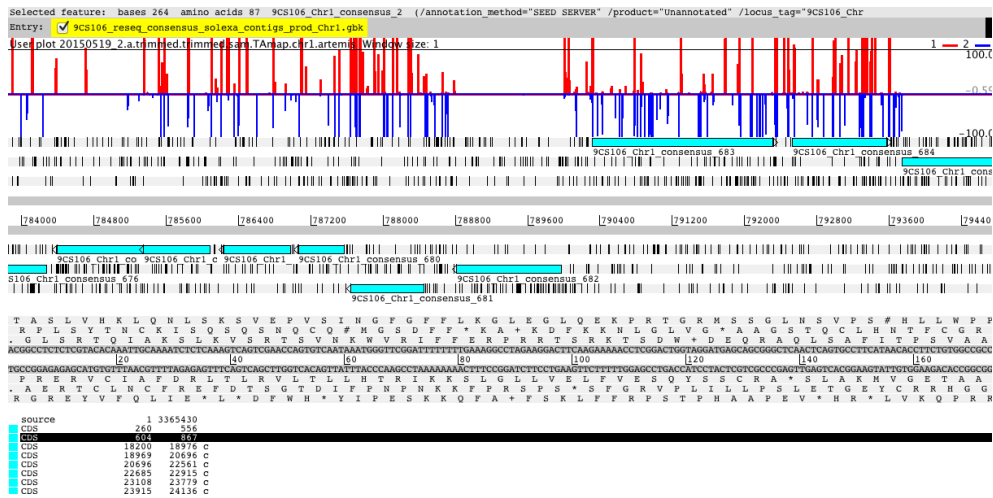


Figure B-3: An example Artemis plot, showing reads mapping to gene Chr1, gene 682, annotated as S-adenosylmethionine synthetase (EC 2.5.1.6).

## B.11 Analyzing mutant strain fitness

Calculates the fitness associated with disruption of each TA site, as long as it was observed in the input control library. (There has to be an initial frequency the mutant genotype was observed at.)

1. Navigate to .TAmap files.
2. Have file with initial and final cell concentrations, e.g. 20150617\_expt\_NO\_N\_generations.txt.
3. Run script:
 

```

1 sys.argv[1] = expt mapping file to initial cell density N0 and final cell
   density N.
2 sys.argv[2] = control .TAmap file
3
4 $ python TAmap_to_fitness.py
5 20150617_expt_NO_N_generations.txt
6 20150519_mutlib.trimmed.trimmed.sam.TAmap
      
```
4. Output is .fitness. For example,
 

```

1 file: 20150519_C.a.trimmed.trimmed.sam.TAmap.fitness
2
3 head:
4 Chrom Locus TAsite_bp Initial_proportion Final_proportion Fitness
5 9CS106_Chrl2_consensus 9CS106_Chrl2_consensus_1 74 1.6504372751e-07 0.0 0.0
      
```

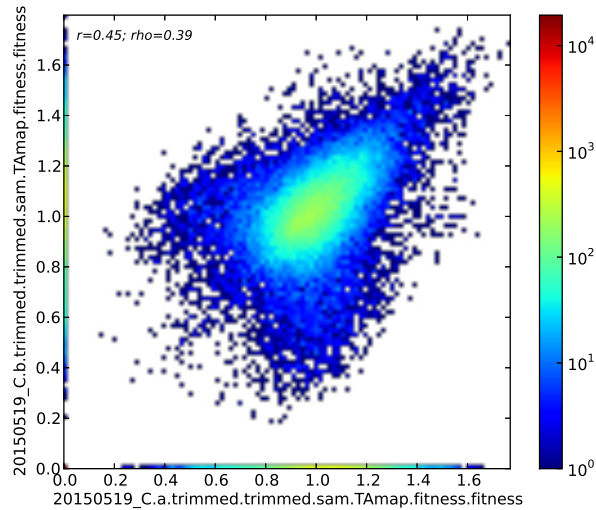


Figure B-4: Correspondence of fitness values for each mutant genotype, as opposed to gene, from replicates grown in filtered *Apocyclops* medium.

```

6 9CS106_Chr2_consensus 9CS106_Chr2_consensus_1 90 2.25009615172e-05
   3.85223894836e-05 1.07718876924
7 9CS106_Chr2_consensus 9CS106_Chr2_consensus_2 289 4.1811077636e-06
   5.76832655028e-06 1.04619671677

```

5. Graph the correspondence of replicates using `two_files_to_scatter_plot_file2_vs_file1.py`. The script generates a scatterplot or 2D-histogram (represented as a heatmap).

- (a) Open the script.
- (b) Change the user-defined portion.
- (c) Run the script, e.g.:

```

1 $ python two_files_to_scatter_plot_file2_vs_file1.py
2 20150519_C.a.trimmed.trimmed.sam.TAmap.fitness
3 20150519_C.b.trimmed.trimmed.sam.TAmap.fitness

```

- (d) The output is an .eps image. For example, see Figure B-4.

Note that the correlation between replicates of the same treatment type appears higher than between replicates of different treatments: see Figure B-5.

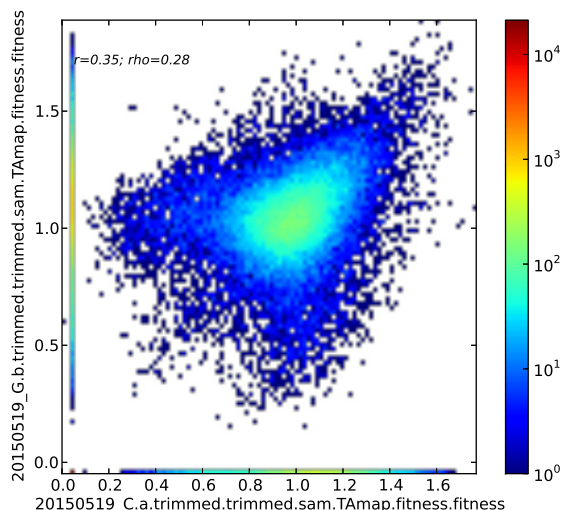


Figure B-5: Correspondence of fitness values for each mutant genotype, as opposed to gene, from one replicates grown in filtered *Apocyclops* medium, and the other replicate grown in glucose.

## B.12 Running the ARTIST pipeline

ARTIST is a pipeline for analyzing transposon insertion site (TIS) data (Pritchard et al., 2014). Both the original paper as well as the TIS nanocourse the Waldor lab ran in January 2015 (click here to access the nanocourse materials) describe the pipeline in detail, and walk the user through steps to implement it. For the most part, I have adopted only those commands necessary to the analysis of my data (with HMM analyses skipped due to the lower TA-site saturation of the mutant library relative to those analyzed by Pritchard and Chao. I have wrapped the commands into three scripts, which can be run successively, without necessitating the user type each command in the ARTIST pipeline by hand. If you would like to implement different functionalities of the ARTIST pipeline, or export different data variables generated, than the ones currently specified, simply open the script, and add the desired commands.

Analyses are conducted separately for each chromosome in the organism.

### B.12.1 Generate input files for ARTIST

1. **.totreads.** The ARTIST pipeline requires summed forward and reverse reads per TA site. To generate this kind of file, `.totreads`, run the following script, `extract_total_TAtally_totalreads_to_column_for_ARTIST.py`.

Navigate to the directory with the .TAmap file that will be passed as one argument to the script. For example,

```
1 $ python extract_total_TAtally_totalreads_to_column_for_ARTIST.py
2 20150519_G.a.trimmed.trimmed.sam.TAmap
```

The output is a .totreads file for each chromosome. E.g.,

```
1 20150519_G.a.trimmed.trimmed.sam.TAmap.9CS106_Chr1_consensus.totreads
2 20150519_G.a.trimmed.trimmed.sam.TAmap.9CS106_Chr2_consensus.totreads
3 20150519_G.a.trimmed.trimmed.sam.TAmap.9CS106_ECE1_consensus.totreads
4 20150519_G.a.trimmed.trimmed.sam.TAmap.9CS106_ECE2_consensus.totreads
```

And the contents of a .totreads file:

```
1
2 file:
3 20150519_G.a.trimmed.trimmed.sam.TAmap.9CS106_Chr2_consensus.totreads
4
5 head:
6 0
7 0
8 0
9 0
10 4
11 624
```

This script can also be run on multiple files by using a Unix command. For example, given that all the files to be processed have the structure 2\*.TAmap:

```
1 for file in 2*.TAmap;
2 do python extract_total_TAtally_totalreads_to_column_for_ARTIST.py "$file";
3 done
```

2. **.TAsites.txt.** ARTIST also requires a TAsites file per chromosome, which gives the nucleotide position of the T in each TA dinucleotide (counter restarts at 1 between chromosomes). In ARTIST, these are assigned to the variable 'TAsites'. Generate \_TAsites.txt files from the TA.txt files Phil Arevalo made using the Unix command:

```
1
2 For 9CS106, all the TAsites files:
3
4 \begin{lstlisting}
5 9CS106_Chr1_consensus_TAsites.txt
6 9CS106_Chr2_consensus_TAsites.txt
7 9CS106_ECE1_consensus_TAsites.txt
8 9CS106_ECE2_consensus_TAsites.txt
```

And for example,

```
1 file: 9CS106_Ch1_consensus_TAsites.txt
2
3 head:
4 14
5 59
6 89
7 120
8 154
9 187
```

3. **.TA\_Locusid.txt.** TA\_Locusid.txt files have locus IDs corresponding to each TA site per chromosome. In ARTIST, these are assigned to the variable 'TAid'. Generate these files using the script `sam_to_TAmapping.py`.

- (a) Open the script.
- (b) Set the user-defined flag `write_TAid_files = 1`. If you do not need to generate .TAmapping files or gene annotation files, set `write_chr_annotations_files = 0` and `write_TAmapping_files = 0`.
- (c) Run the script (as before in Mapping genome-mapped reads to TA sites (appendix B.9.5)). The command structure is

```
1 $ python <python script to map reads to TA sites>
2
3 <path to the files with TAsites per chromosome>
4
5 <path to the the organism gene list>
```

For example,

```
1 $ python /cygdrive/d/alisontakemura/150601_sequencing/
2 sam_to_TAmapping.py
3
4 /cygdrive/d/alisontakemura/reference_genomes/
5
6 /cygdrive/d/alisontakemura/reference_genomes/
7 9CS106_reseq_consensus_solexa_contigs_prod_genelist.txt
```

The output is a TA\_Locusid.txt file for each chromosome. E.g.,

```
1 9CS106_Ch1_consensus_TA_Locusid.txt
2 9CS106_Ch2_consensus_TA_Locusid.txt
3 9CS106_ECE1_consensus_TA_Locusid.txt
4 9CS106_ECE2_consensus_TA_Locusid.txt
```

And the contents of a TA\_Locusid.txt file:

```

1 head of 9CS106_Ch1_consensus_TA_Locusid.txt:
2
3 IG_9CS106_Ch1_consensus_1
4 IG_9CS106_Ch1_consensus_1
5 IG_9CS106_Ch1_consensus_1
6 IG_9CS106_Ch1_consensus_1
7 IG_9CS106_Ch1_consensus_1
8 IG_9CS106_Ch1_consensus_1
9 IG_9CS106_Ch1_consensus_1
10 IG_9CS106_Ch1_consensus_1
11 IG_9CS106_Ch1_consensus_1
12 IG_9CS106_Ch1_consensus_1
13 IG_9CS106_Ch1_consensus_1
14 IG_9CS106_Ch1_consensus_1
15 IG_9CS106_Ch1_consensus_1
16 9CS106_Ch1_consensus_1

```

4. **Chromosome size.** Finally, determine the size of the chromosome to be analyzed.

Look up the chromosome size in the genbank (.gbk) file. The .gbk file for 9CS106, (9CS106\_reseq\_consensus\_solexa\_contigs\_prod.gbk), was manually split by chromosome into the individual .gbk files:

```

1 9CS106_reseq_consensus_solexa_contigs_prod_Ch1.gbk
2 9CS106_reseq_consensus_solexa_contigs_prod_Ch2.gbk
3 9CS106_reseq_consensus_solexa_contigs_prod_ECE1.gbk
4 9CS106_reseq_consensus_solexa_contigs_prod_ECE2.gbk

```

And the contents of the file, for example, show the size of the chromosome:

```

1 file: 9CS106_reseq_consensus_solexa_contigs_prod_Ch2.gbk
2
3 head:
4
5 LOCUS 9CS106_Ch2 1859082 bp DNA UNK 01-JAN-1980

```

### B.12.2 Setting up the Matlab workspace to run ARTIST

Utilize a powerful machine, like the CLC, to run the ARTIST pipeline.

1. Open Matlab 2014b.
2. Navigate to the folder with the matlab (.m) scripts needed to run ARTIST. The scripts are available at the TIS nanocourse google drive folder.

The ARTIST scripts:

```

1 Bottleneck_analysis.m
2 callconditional.m
3 checkseqdepth.m
4 discretize5.m
5 discretizeconditional_3.m
6 genome_parser_TA.m
7 genome_parser.m
8 getuniquetanames.m
9 hmmconverge.m
10 hmmtrainmwu.m
11 MWUbytransposon.m
12 MWUsummarystats.m
13 output_essential.m
14 runmwuallboots.m
15 SAMreader_TA.m
16 SAMreader_Tn5.m
17 simulateequalsaturation.m
18 Slidingwindow.m
19 window_average.m

```

And my own scripts:

```

1 Preload_ARTIST_vars.m
2 ELARTIST_loop_20150621.m
3 CONARTIST_loop_20150621.m

```

Note, I edited MWUsummarystats.m so it can accept small chromosomes (<100 TA sites) as well as large ones. My copy of the script is on the CLC machine at /cygdrive/d/alisontakemura/ARTIST.

### 3. Load chromosome-specific variables into Matlab.

For the ARTIST pipeline, I have written scripts to analyze all the experimental data pertaining to one chromosome at a time; for example, all .totreads data files that correspond to Chr1 will be analyzed en batch. Variables corresponding the chromosome being analyzed have to be added as variables to the Matlab workspace.

- (a) Copy variables from files Preload\_ARTIST\_vars\_<chromosome\_name>.m into Preload\_ARTIST\_vars.m.

Chromosome-specific variables files for 9CS106:

```

1 Preload_ARTIST_vars_Chr1.m
2 Preload_ARTIST_vars_Chr2.m
3 Preload_ARTIST_vars_ECE1.m
4 Preload_ARTIST_vars_ECE2.m

```

- (b) Into the Matlab command line, type:



```
1 >> Preload_ARTIST_vars
```

### B.12.3 Run EL-ARTIST

EL-ARTIST looks for de-enriched regions by a sliding window analysis. As given in the nanocourse documentation:

The EL-ARTIST pipeline identifies regions that are required for optimal growth under a given condition (typically an in vitro grown library on rich medium). .. [S]liding window analysis is used to define regions that appear underrepresented in reads...

I highly recommend reading the TIS nanocourse documentation for further understanding of what the ELARTIST pipeline does (simulation of datasets, P-value threshold, etc.).

To begin implementing EL-ARTIST for a data set particular to a single chromosome, I created a Matlab script to automatically load the chromosome-specific variables (in `Preload_ARTIST_vars.m`) and iteratively run through all experiment files (`.totreads`).

1. Move all files `.totreads` files to run the into the directory with the ARTIST scripts. (Later, move them into an appropriately named directory.)
2. Open the script `ELARTIST_loop_20150621.m`.
3. Edit the User-defined portion:

```
1 file_name_structure = '2*.9CS106_Chr1_consensus.totreads'
```

4. Run the script by typing its name in the Matlab command line:

```
1 >> ELARTIST_loop_20150621
```

5. The script will output files with extension `.window_size_<bp>.normalized_ELARTIST.txt`, which have a single column of data (1, 0) indicating whether the locus was called as essential. The ordered list of loci is created in ARTIST and must be exported separately.
6. Once the script has iterated over all `.totreads` files, clear variables by typing in the Matlab command line:

```
1 >> clearvars
```

#### B.12.4 Run Con-ARTIST

Con-ARTIST will conduct a comparison between an Input, control, sample and an Output, or experimental, sample, to determine genes selected for (enriched) and against (de-enriched). Note that Con-ARTIST can be run independently of EL-ARTIST.

The Mann-Whitney U (MWU) test is a non-parametric statistical test, which compares whether reads at a gene in one library and significantly more or less than the reads in another library. However, as it does not assume normal distribution (which TIS data is not), it is robust to noise. The MWU takes the reads at all TA sites for every gene in both libraries and ranks the reads. Then it asks whether the ranks of the one library is significantly different than another. MWU tests are limited to annotated loci and will not define domains within genes that may be required for growth in different conditions.

Note, if you wish, you can change the encoded parameters (P-value cutoff, fraction of tests with P-values below cutoff) under which a locus is called significantly different in the Output to the Input.

1. On the CLC machine, open Matlab 2014b.
2. Load the baseline variables, as described in Setting up the Matlab workspace (appendix B.12.2).
3. Open the script CONARTIST\_loop\_20150621.m.
4. Edit ther User-defined portion:

```
1
2 % Specify the Input .totreads file that every Expt file will be compared
3 against
4 Input_file = 'D:\alisonakemura\ARTIST\20150519_mutlib.trimmed.trimmed.sam.
5 TAmap.9CS106_Chr1_consensus.totreads';
6 Input = dlmread(Input_file);
7
8 % Specify the Input sample name so it can be appended to the outfile name
9 Input_name = '20150519_mutlib';
10
11 % Specify the file name structure of the .totreads files to process
12 file_name_structure = '2*.9CS106_Chr1_consensus.totreads';
```

5. Run the script by typing its name in the Matlab command line:

```
1 CONARTIST_loop_20150621
```

6. The script will output `_vs<control sample>normalized_CONARTIST.txt` files, which will have six columns of information. As described by the authors:
  - (a) Column 1 is the proportion of MWU tests in which the locus was called significant (i.e., P-value was less than that specified by the user).
  - (b) Column 2 is whether this locus was reproducibly significant in more tests than the user cutoff (1 = yes; 0 = no).
  - (c) Column 3 is the average P-value for each locus across all MWU simulations.
  - (d) Column 4 is the standard deviation in P-values for each locus across all MWU simulations.
  - (e) Column 5 is the average read count ratio from all insertions within each locus between the experimental and control dataset across 100 simulations. The ratio is effectively (Experiment reads / Control simulation reads); large ratios denote enriched loci, while low ratios indicate conditional essentiality.
  - (f) Column 6 is the standard deviation of read counts for each locus across the 100 different simulations.
7. Once the script has iterated over all `.totreads` files, make a file of the `uniquenames` variable from the Matlab run, which are the loci, by row, that the output correspond to.

For example, for 9CS106 chromosome 1,

```

1 file: Chr1_loci.txt
2
3 head:
4
5 IG_9CS106_Chr1_consensus_1
6 gene_id Mobile element protein
7 IG_9CS106_Chr1_consensus_2
8 gene_id Unannotated
9 IG_9CS106_Chr1_consensus_3
10 gene_id Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8)

```

8. Clear variables by typing in the Matlab command line:

```
1 >> clearvars
```

9. Once the ARTIST analyses are complete, move the input (`.totreads`) and output files (`_vs<control sample>normalized_CONARTIST.txt`) into a directory with a descriptive name, e.g. `20151216_best_data_ever`. Or `20151216_Fucus_selection_ARTIST_results`.

For reference, `CONARTIST_loop_20150621.m` is given below:

## B.13 Analyze ARTIST results

Given below is one way to begin to analyze the ARTIST outputs.

### B.13.1 Annotate ARTIST ordered, non-overlapping loci

In order to link function information to the genes in the Con-ARTIST output, for each chromosome, the loci need to be copied out of Matlab, and matched to available annotations, such as from RAST.

1. In Matlab, navigate to the ARTIST working directory.
2. For each `Preload_ARTIST_vars.m` script (which are split by chromosome),
  - (a) Run the script by typing the name at the Matlab command line.
  - (b) Open the `uniquenames` variable, which has an ordered list of the unique, non-overlapping loci that the TA sites were split into.
  - (c) Copy the `uniquenames` values into a text file. Remove the quote marks by Find (') and Replace All (with nothing).
  - (d) Save the file with `<chromosome name >_nonoverlappingloci.txt`.

3. For 9CS106, these files are:

- 1 `Chr1_nonoverlappingloci.txt`
- 2 `Chr2_nonoverlappingloci.txt`
- 3 `ECE1_nonoverlappingloci.txt`
- 4 `ECE2_nonoverlappingloci.txt`

4. Using the organism `genelist.txt` file, and each `_nonoverlappingloci.txt` annotate the ordered loci with their RAST / SEED role annotation.

```
1 Script: match_ARTIST_loci_names_to_gene_annotations.py
2
3 sys.argv[1] = database file (genelist)
4 sys.argv[2] = queries file (_nonoverlappingloci.txt)
5 > output
```

Example command to run the script:

```
1 $ python match_ARTIST_loci_names_to_gene_annotations.py
2
3 /cydrive/d/alisontakemura/reference_genomes/
4 9CS106_reseq_consensus_solexa_contigs_prod_genelist.txt
5
6 Chr1_nonoverlappingloci.txt > Chr1_loci.txt
```

5. The output contains the gene annotation given in the genelist. For example,

```
1 file: Chr1_loci.txt
2
3 head:
4
5 IG_9CS106_Chr1_consensus_1 IG_9CS106_Chr1_consensus_1
6 9CS106_Chr1_consensus_1 "gene_id Mobile element protein"
7 IG_9CS106_Chr1_consensus_2 IG_9CS106_Chr1_consensus_2
8 9CS106_Chr1_consensus_2 "gene_id Unannotated"
9 IG_9CS106_Chr1_consensus_3 IG_9CS106_Chr1_consensus_3
10 9CS106_Chr1_consensus_3 "gene_id Phosphoadenylyl-sulfate
11 reductase [thioredoxin] (EC 1.8.4.8)"
12 9CS106_Chr1_consensus_4 "gene_id Sulfite reductase [NADPH]
13 hemoprotein beta-component (EC 1.8.1.2)"
14 9CS106_Chr1_consensus_5 "gene_id Sulfite reductase [NADPH]
15 flavoprotein alpha-component (EC 1.8.1.2)"
16 IG_9CS106_Chr1_consensus_6 IG_9CS106_Chr1_consensus_6
17 9CS106_Chr1_consensus_6 "gene_id Thymidylate kinase"
```

6. For 9CS106, these output files are:

```
1 Chr1_loci.txt
2 Chr2_loci.txt
3 ECE1_loci.txt
4 ECE2_loci.txt
```

## B.13.2 Annotate loci.txt with RAST

1. Download the SEED annotations file from 've referred to the file as SEED\_subsystems2role.txt.

2. From the directory with \_loci.txt files, run the script loci\_to\_RASTannotations.py.  
Pass two variables, the .ssa file and the SEED\_subsystems2role.txt file. E.g.,

```
1 $ python loci_to_RASTannotations.py /Users/alimura/Dropbox/Public/ARTIST/
2 genome_9CS106_reseq_consensus_solexa/9CS106_reseq_consensus_solexa_cds_prod.
   ssa
3 /Users/alimura/Dropbox/Public/Annotations/annotations_databases/
4 SEED_subsystems2role.txt
```

3. The output extension is .RAST, and contains four columns:

- (a) Column 1 is Role
- (b) Column 2 is Subsystem
- (c) Column 3 is Subcategory

(d) Column 4 is Category

For example,

```
1 file: Chr1_loci.txt.RAST
2
3 head:
4
5 9CS106_Chr1_consensus_16 Zinc uptake regulation protein ZUR Zinc regulated
   enzymes Regulation and Cell signaling
6 IG_9CS106_Chr1_consensus_17 IG_9CS106_Chr1_consensus_17 IG_no_subsys
   IG_no_subcat IG_no_cat
7 9CS106_Chr1_consensus_17 Chemotaxis protein CheX Bacterial Chemotaxis
8 Motility and Chemotaxis
9 IG_9CS106_Chr1_consensus_18 IG_9CS106_Chr1_consensus_18 IG_no_subsys
   IG_no_subcat IG_no_cat
10 9CS106_Chr1_consensus_18 Glucose-6-phosphate isomerase (EC 5.3.1.9)
    subsys_not_found subcat_not_found cat_not_found
```

Note, I also have a .conservative version of these annotations, where I only used the .ssa annotations for subsystem. If no subsystem was called, then it (as well as category and subcategory) are not included in the annotations.

```
1 files:
2
3 Chr1_loci.txt.RAST.conservative
4 Chr2_loci.txt.RAST.conservative
5 ECE1_loci.txt.RAST.conservative
6 ECE2_loci.txt.RAST.conservative
```

### B.13.3 Save all Con-ARTIST output, separated by chromosome, to a single file

(a) Navigate to your saved results, e.g. .normalized\_CONARTIST.txt output files for each chromosome per sample. E.g.,

```
1 For sample 2.a:
2
3 20150519_2.a.trimmed.trimmed.sam.TAmap.
4 9CS106_Chr1_consensus.totreads.vs20150519_mutlib.normalized_CONARTIST.
   txt
5
6 20150519_2.a.trimmed.trimmed.sam.TAmap.
7 9CS106_Chr2_consensus.totreads.vs20150519_mutlib.normalized_CONARTIST.
   txt
8
```

```

9 20150519_2.a.trimmed.trimmed.sam.TAmap.
10 9CS106_ECE1_consensus.totreads.vs20150519_mutlib.normalized_CONARTIST.
    txt
11
12 20150519_2.a.trimmed.trimmed.sam.TAmap.
13 9CS106_ECE2_consensus.totreads.vs20150519_mutlib.normalized_CONARTIST.
    txt

```

- (b) To combine the results per chromosome into one file, use script `combine_chr_MWUstats_to_all_chrom_MWUstats.py`. Navigate to the folder with the CONARTIST output separated by chromosome, then run the script:

```
1 $ python combine_chr_MWUstats_to_all_chrom_MWUstats.py
```

- (c) Move the output to a separate folder, e.g., `CONARTIST_mwustats_all_chrom_combined/`.

### B.13.4 Separate Con-ARTIST results by each of the six MWUstats

When analyzing the significant genes, you may want to look at the underlying average P-values over the 100 simulations, or whether the gene was de-enriched or enriched. To do so, first aggregate the MWUstats by type across samples, into individual files. These tables can later be subsetted, to look at the data for just the loci of interest.

1. Navigate to folder with Con-ARTIST data for all chromosomes combined, e.g., `CONARTIST_mwustats_all_chrom_combined/`.
2. Open the script `mwustats_to_separate_files.py`, and edit the user-defined portion.
3. Run the script:

```
1 $ python mwustats_to_separate_files.py
```

4. Output is tables of one MWUstat, for each locus across all samples.

```

1 20150519_CONARTIST_MWUstat.1.txt
2 20150519_CONARTIST_MWUstat.2.txt
3 20150519_CONARTIST_MWUstat.3.txt
4 20150519_CONARTIST_MWUstat.4.txt
5 20150519_CONARTIST_MWUstat.5.txt
6 20150519_CONARTIST_MWUstat.6.txt

```

5. If necessary, remove some columns (some experiments) with scrip `remove_columns.py`. Open the script, edit User-defined, and run it:

```

1 sys.argv[1] = file to remove columns from
2
3 python remove_columns.py
4 20150705_both_mutant_libraries_CONARTIST_MWUstat.2.consensus.txt
5
6
7 Example output:
8
9 20150705_both_mutant_libraries_CONARTIST_MWUstat.2.consensus.txt.
   phage_removed

```

6. Insert the loci annotations prior to the MWUstats values.

(a) Open the script `insert_beginning_column.py`.

(b) Edit the user-defined portion.

(c) Run the script:

```

1 sys.argv[1] = file with column(s) to insert (e.g. locus IDs)
2 sys.argv[2] = file to insert column(s) into (e.g. _MWUstat.1.txt)

```

(d) Command:

```

1 python insert_beginning_column.py
2
3 allChrom_loci.txt.RAST.liberal.formatted
4
5 CONARTIST_mwustats_all_chrom_combined/20150519_CONARTIST_MWUstat.1.txt

```

(e) Output is the MWUstat files with annotation information in the first few columns:

```

1 20150519_CONARTIST_MWUstat.6.txt.with_locus.txt
2 20150519_CONARTIST_MWUstat.5.txt.with_locus.txt
3 20150519_CONARTIST_MWUstat.4.txt.with_locus.txt
4 20150519_CONARTIST_MWUstat.3.txt.with_locus.txt
5 20150519_CONARTIST_MWUstat.2.txt.with_locus.txt
6 20150519_CONARTIST_MWUstat.1.txt.with_locus.txt

```

### B.13.5 Pull out locus called significant in at least one experiment

Generates a table with just those loci that were called significant in at least one experiment.

1. Navigate to folder MWUstat.2 file to excerpt rows from.
2. Open script `MWUstat_to_excerpt_sig_in_at_least_one.py`.
3. Edit User-defined portion.



#### 4. Run script

```
1 sys.argv[1] = MWUstat.2 to excerpt from
2
3 $ python
4
5 MWUstat_to_excerpt_sig_in_at_least_one.py
6
7 20150705_both_mutant_libraries_CONARTIST_MWUstat.2.
8 -consensus.txt.phage_removed.with_locus.txt
```

5. Output is a file with excerpted rows, where each row has at least one '1' in it, indicating the significance of that locus in an experiment (file extension: `sig_in_any.txt`). For example,

```
1 file:
2 20150705_both_mutant_libraries_CONARTIST_MWUstat.2.consensus.txt.
3 phage_removed.with_locus.txt.sig_in_any.txt
4
5 head:
6 orig_order Locus RAST role RAST subsystem RAST subcategory RAST category
7 Alginate_a Alginate_b Alginate_c 2216_a 2216_b 2216_c filtered
8 Copepods_a filtered Copepods_b filtered Copepods_c Fucus_a Fucus_b Fucus_c
9 Glucose_a Glucose_b Glucose_c
10 5 IG_9CS106_Ch1_consensus_3 IG_9CS106_Ch1_consensus_3 IG_no_subsys
11 IG_no_subcat IG_no_cat 1.0 1.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 1.0 1.0 0.0
    1.0 1.0
```

6. Copy to an `.xlsx` file to manipulate, sort the table, add analysis of individual genes and pathways.

### B.13.6 Excerpt other MWUstats based on locus significance in any experiment

Generates a file with where the 1's in the `MWUstat.2...sig_in_any.txt` file are substituted with another MWUstat value of interest - like the average P-value (MWUstat.3) or the average read count ratio to indicate the locus' direction of enrichment (MWUstat.5).

1. Navigate to folder with the the excerpted MWUstat.2 file ( `.sig_in_any.txt`) and the other MWUstat file to get values from.
2. Open the script `sigvalue_to_MWUstat.py`.
3. Edit the User-defined portion.
4. Run the script.

```

1 sys.argv[1] = queries, the file with the table showing samples with
   significant
2 loci. e.g. 20150705_both_mutant_libraries_CONARTIST_MWUstat.2.consensus.txt.
3 with_locus.txt.sig_in_any_rep.txt
4
5 sys.argv[2] = database, the file with the MWUstat values to get. e.g.,
6 20150706_BMC_mutlib_CONARTIST_MWUstat.5.txt.with_locus.txt
7
8
9 For example,
10
11 python sigvalue_to_MWUstat.py
12
13 20150705_both_mutant_libraries_CONARTIST_MWUstat.2.consensus.txt.
14 with_locus.txt.sig_in_any_rep.txt
15
16 20150706_BMC_mutlib_CONARTIST_MWUstat.5.txt.with_locus.txt

```

5. Output is file with substituted MWUstat values. For example,

```

1
2 file: 20150705_both_mutant_libraries_CONARTIST_MWUstat.
3 2.consensus.sig_in_any_rep.BMC_mutlib.avg_read_count_ratio.txt
4
5 head:
6
7 orig_order Locus RAST role RAST subsystem RAST subcategory RAST category
8 Ph1_20140515 Ph6_20140515 Ph1_20140610 Ph6_20140610 Alginate_a
9 Alginate_b Alginate_c 2216_a 2216_b 2216_c filtered Copepods_a filtered
10 Copepods_b filtered Copepods_c Fucus_a Fucus_b Fucus_c Glucose_a Glucose_b
   Glucose_c
11 5 IG_9CS106_Chrl_consensus_3 IG_9CS106_Chrl_consensus_3 IG_no_subsys
12 IG_no_subcat IG_no_cat 0.0785 0.065 0.0506 0.0504 1.17 1.14 1.2 0 0 0
   0 0 0 0.947 1.01 0.953 0 1.22 1.23
13 6 9CS106_Chrl_consensus_3 Phosphoadenylyl-sulfate reductase [thioredoxin]
14 (EC 1.8.4.8) Cysteine Biosynthesis Lysine, threonine, methionine, and
   cysteine
15 Amino Acids and Derivatives 0.000131 0.000175 0.000104 0.000146 0.0784
   0.0487 0.0706 0 0 0 0 0 0 0 0 0 0.173 0.246 0.139
16 7 9CS106_Chrl_consensus_4 Sulfite reductase [NADPH] hemoprotein beta-
   component
17 (EC 1.8.1.2) Cysteine Biosynthesis Lysine, threonine, methionine, and
   cysteine
18 Amino Acids and Derivatives 0.000132 0.000175 0.000104 0.00278 0.036
   0.0237 0.0142 0 0 0 0 0 0 0 0.104 0 0.0109 0.0685 0.0613

```

### B.13.7 Pull out genes ‘core’ (consistent) or ‘flexible’ between replicates

Generates core and flexible calls with script `20150622find_core, flex.py` .

1. Ensure header names for the samples are in the format: `Treatment_replicate` (e.g. `Glucose_a`).
2. Open the script `find_core_flex.py` and edit the user-defined portion.
3. Run the script:

```
1 Input:
2 sys.argv[1] = file with MWUstat to sort by, e.g.
3 20150519_CONARTIST_MWUstat.2.txt.with_locus.txt
4
5 $ python find_core_flex.py 20150519_CONARTIST_MWUstat.2.txt.with_locus.txt
```

4. Output is `absent_`, `core_`, and `flexible_` counts (subsets of the MWUstat table) for each sample type.

```
1 absent_counts_filtered.Copepods.txt
2 core_counts_filtered.Copepods.txt
3 flexible_counts_filtered.Copepods.txt
4
5 absent_counts_Glucose.txt
6 core_counts_Glucose.txt
7 flexible_counts_Glucose.txt
```

5. If you are interested in looking at all the loci significant in at least one replicate, you can, after navigating into the folder with the core and flexible files, run the script `combine_core_and_flex_to_all.py`. The script will generate files `all_core_and_flex_counts_<MediaType>.txt`.

To run,

```
1 $ python combine_core_and_flex_to_all.py
```

### B.13.8 Look for significant loci shared and not shared between conditions

Generate lists of loci that are shared and non-shared between experimental conditions using the script `get_core_distinct.py`.

1. Open the script `get_core_distinct.py`.
2. Fill out the User-defined portion, which includes the input file path. Specify the `all_flex_and_core_counts_[Media type].txt` created above, and which type of sample you are comparing to.

3. Run script:

```
1 $ python get_core_distinct.py
```

4. Output is `.shared` (significant in both conditions) and `.notshared` (not significant in the second condition) subsets of the original input file, where the file name is `all_flex_and_core_counts_[Media type].txt.[Media compared to].[shared / notshared]`.

```
1 all_flex_and_core_counts_filtered.Copepods.txt.Glucose.notshared
2 all_flex_and_core_counts_filtered.Copepods.txt.Glucose.shared
3 all_flex_and_core_counts_Glucose.txt.2216.notshared
4 all_flex_and_core_counts_Glucose.txt.2216.shared
```

For example, `all_flex_and_core_counts_Glucose.txt.2216.notshared` has the loci that are significant in the glucose condition, but not in the 2216 condition.

### B.13.9 Pull out *P*-values and enrichment direction for significant loci

1. Navigate to directory with `'all_flex_and_core_counts*.txt'` files:

2. Run `sigloci_to_MWUstats.py`.

```
1 sys.argv[1] = directory with MWUstats, separated by the kind of stat,
2     with aggregated expts and locus IDs
3
4 For example:
5
6 $ python sigloci_to_MWUstats.py CONARTIST_mwustats_all_chrom_combined/
```

3. The output will be all MWUstat files. The ones of most interest will be the average *P*-value across the 100 simulations (`.average_pval_across_simulations.txt`) and the average read count ratio (`.average_read_ct_ratio.txt`), which indicates the direction of enrichment: ratios  $>1$  indicate enriched, and ratios  $<1$  are de-enriched loci.

```
1 all_flex_and_core_counts_filtered.Copepods.txt.
  average_pval_across_simulations.txt
2 all_flex_and_core_counts_filtered.Copepods.txt.average_read_ct_ratio.txt
3 all_flex_and_core_counts_filtered.Copepods.txt.proportion_called_sig.txt
4 all_flex_and_core_counts_filtered.Copepods.txt.proportion_met_user_cutoff.
  txt
5 all_flex_and_core_counts_filtered.Copepods.txt.std_dev_in_pval.txt
6 all_flex_and_core_counts_filtered.Copepods.txt.std_dev_read_ct.txt
```

### B.13.10 Link direction of enrichment: de-enriched or enriched

Make separate tables for MWU stats columns for each experiment. Ex.

```
1 20150519_G.a.trimmed.trimmed.sam.TAmap.  
2 9CS106_ALLChrom_consensus.totreads.vs20150519_mutlib.normalized_CONARTIST.txt
```

Run script

```
1 python mwustats_to_separate_files.py  
2  
3 output files  
4 20150519_CONARTIST_MWUstat.1.txt  
5 20150519_CONARTIST_MWUstat.2.txt  
6 20150519_CONARTIST_MWUstat.3.txt  
7 20150519_CONARTIST_MWUstat.4.txt  
8 20150519_CONARTIST_MWUstat.5.txt  
9 20150519_CONARTIST_MWUstat.6.txt
```

Copy in annotations for column 5 file by opening file, pasting in loci in order given in \_CONARTIST.txt in new excel file.

Grep the annotations want direction for.

Copy the direction of enrichment down in the summary table:

```
1 all_flex_and_core_counts_filtered.Copepods.txt.2216.shared.xlsx
```

THIS PAGE INTENTIONALLY LEFT BLANK

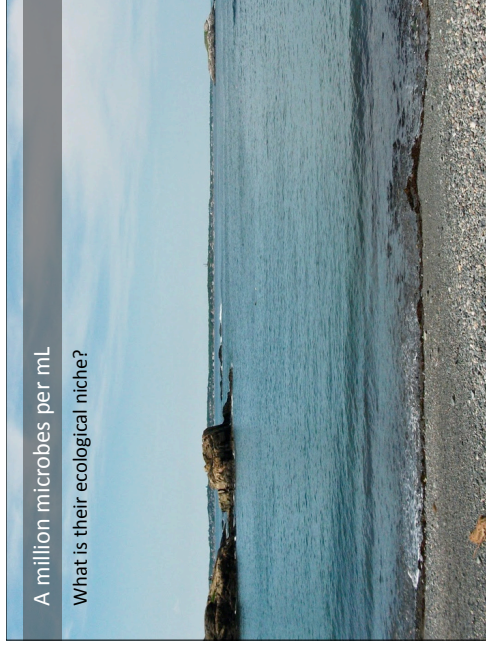
## Appendix C

# Thesis defense presentation

Because I loved sharing my science in person and in a graphical way, and because it was so helpful to have a copy of Jesse Shapiro's thesis defense slide deck while I was crafting my own, and if it can lessen the difficulty for future students, then I'm very glad to include my presentation here.

# Niche adaptations of the *Vibrionaceae*, from the coastal ocean to the laboratory

Alison Francesca Takemura  
Thesis defense  
Aug. 20, 2015



### Why care about microbes?

Biogeochemical cycling

Phytoplankton (Phototrophs)  
Plymouth Marine Laboratory  
Carbon fixation

Kathryn Kauffmann, Emma Nisum  
like *Vibrionaceae*  
1 µm

Particulate

Dissolved

Heterotrophs

CO<sub>2</sub>

Carbon remineralization

simplified from Azam, McIlvin, Nat. Rev. 2007

### Why care about microbes?

Disease

Cholera outbreak  
microscopy.org

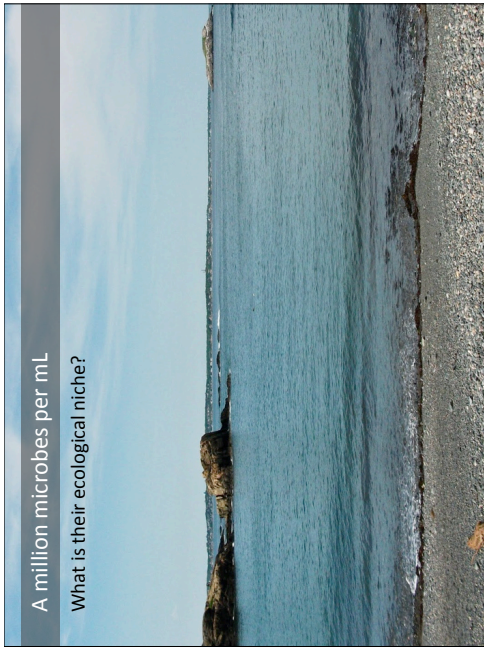
Some *Vibrio cholerae*

Coral disease

Consortia of at least four *Vibrio* species

Alma Cervino, Rice University





### How I approached

1. What can bulk environmental variables tell us about *Vibrio* niche? **meta-analysis**
2. How is a single *Vibrio* strain adapted to different habitats that are part of its niche? **mutant selection**

**Broad-scale**

Meta-analysis

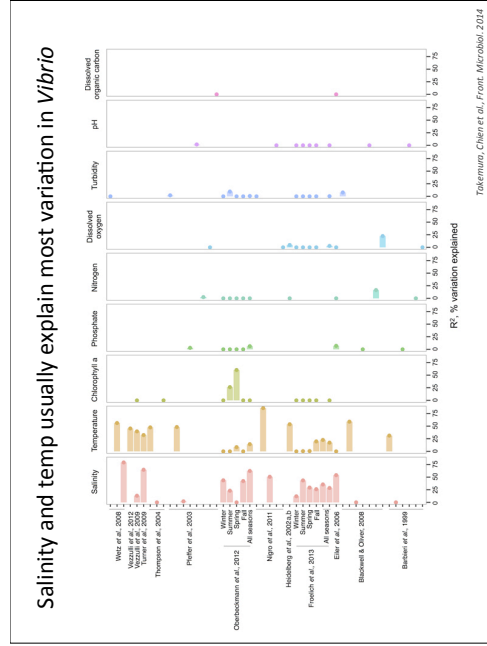
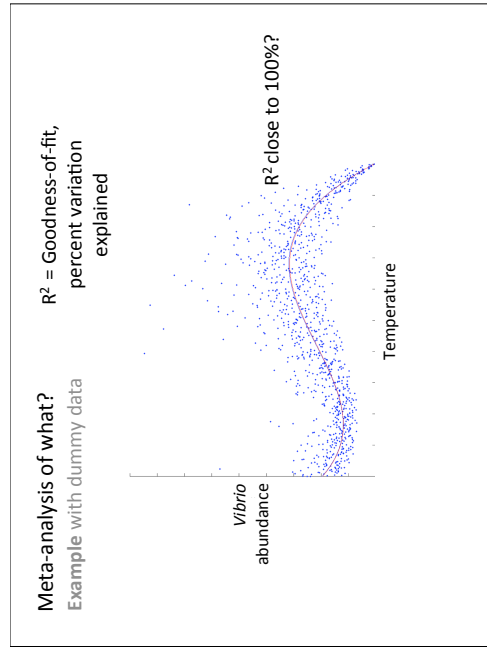
in environment in taxa in microorganisms

meter genus, species abundance

**Fine-scale**

Bulk of my work

$\mu$ m population, strain genome





**Heterotrophic microbes exploit both kinds of resources**  
 .. living attached and unattached

Nutrient gradients

Particles / Marine snow

Phytoplankton

Stocker, Seymour, Microb. Mol. Biol. Rev. 2012

**Vibrionaceae model system to study habitat adaptations**  
 Ecological populations with known habitat distributions

Sarah Preheim

Observed planktonic growth!

Brown

Zooplankton: Copepod *Apocyclops oyi*

Preheim Thesis, 2010

What genes make F13 fit on different resources?

**Habitat compositions: protein and saccharides**  
 (Dependent on age, season)

	<i>Fucus</i>	<i>Apocyclops</i>
Protein	1.4%	57%
Saccharide	65%	24%
Alginate	10-40%	
Laminarin	2-34%	
Fucoidan	5-20%	
Mannitol	5-30%	

**Vibrio sp. F13 can grow on alginate**  
 A third resource

Jan-Hendrik Heilmann

OD<sub>600</sub> (~Cell density)

Time (hours)

OD<sub>600</sub> (~Cell density)

Time (hours)

How *Vibrio* sp. F13 adapted to three resources?  
 Describe quantitatively?

### To identify adaptations, select collection (library) of mutants

Mutant frequencies indicate fitness (Tn-seq)

Determine mutant abundance by sequencing

Ex: normal **boom**ial gene = normal gene

Lost gene needed for growth

### Prepare selection conditions: dissolved habitat resources

Habitat and reference media

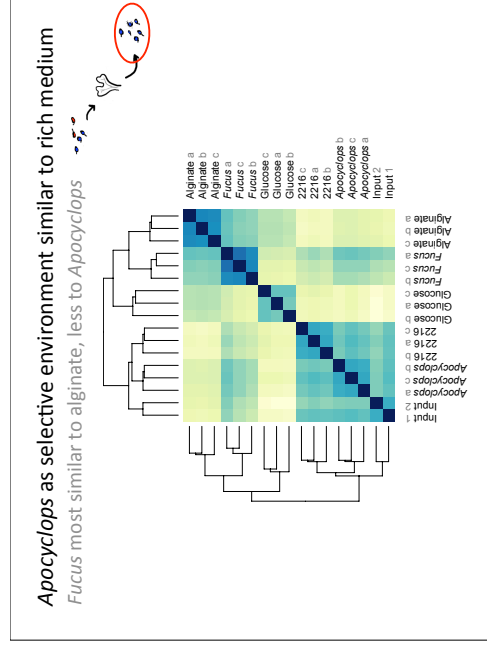
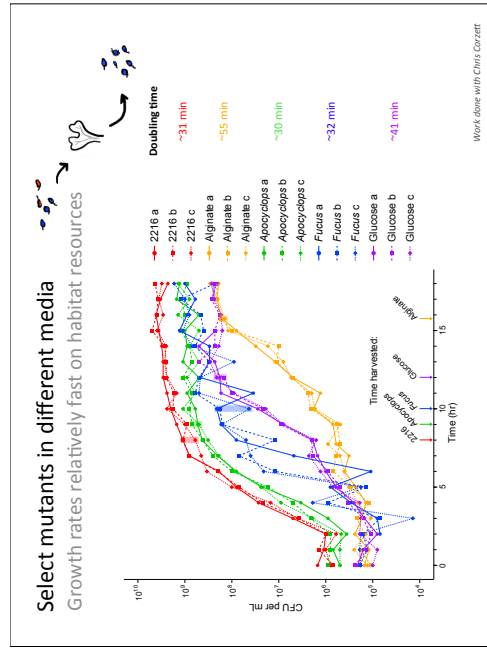
Apocyclops ophi

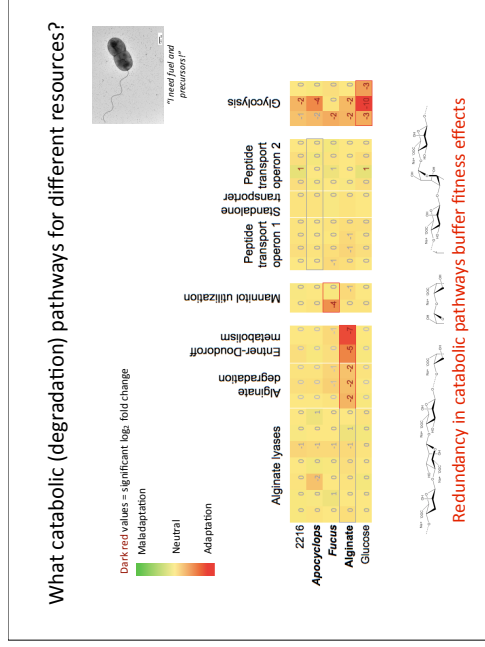
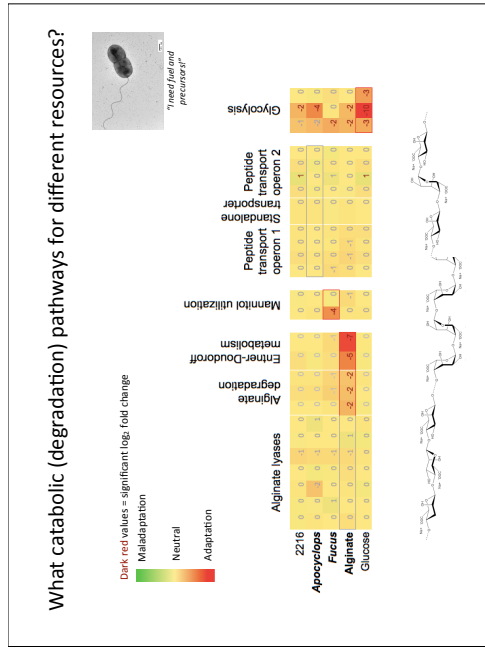
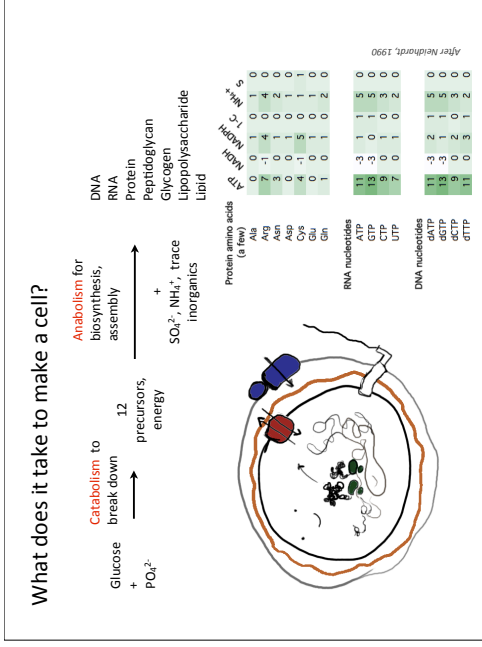
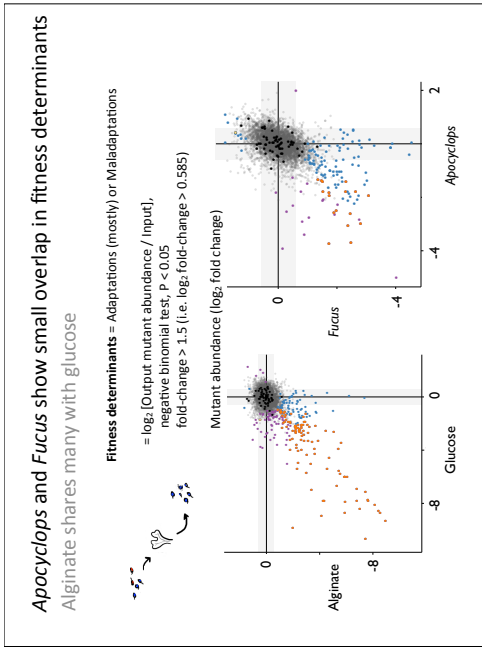
Fucus vesiculosus

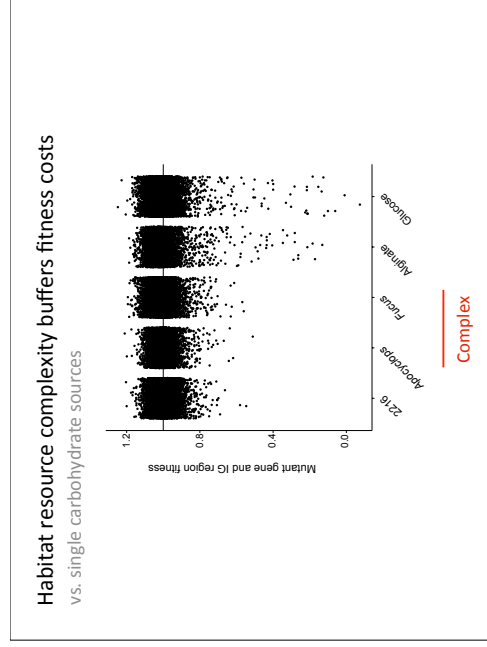
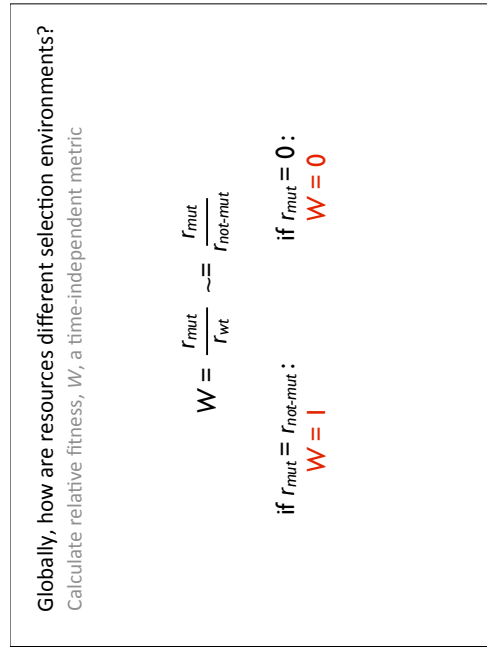
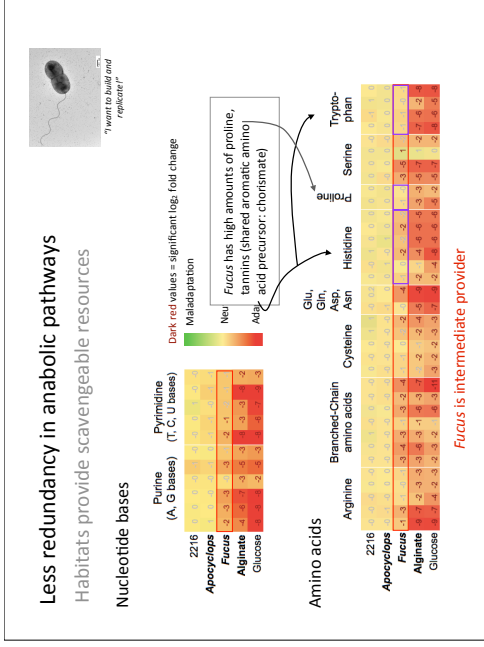
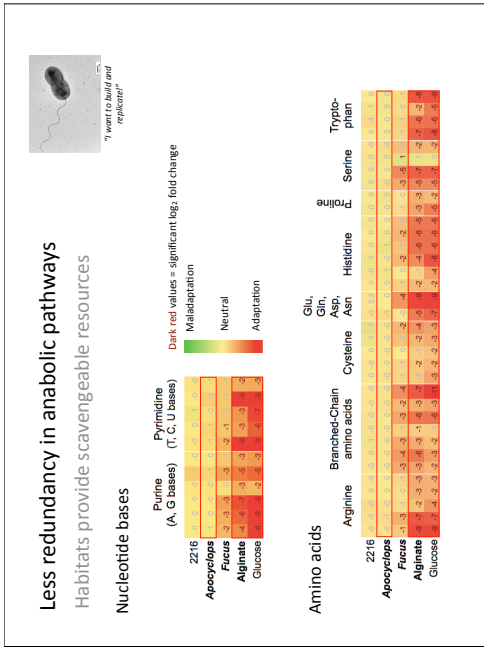
Both ground up

Reference media

Apocyclops resources







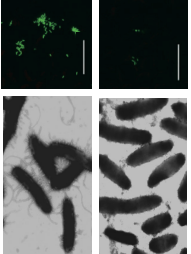
### Evidence of a fitness tradeoff?

Mannose-sensitive hemagglutinin (MSHA) mutants increase slightly in *Fucus* medium

**What is MSHA?**

Used in attachment:

- to cellulose (*V. cholerae*)
- to green alga (*Pseudodictyonema tunicatol*);




**Type IV pilus MSHA synthesis**


Z216	Apocyclops	1.3801	2
	<i>Fucus</i>	1.3801	2
	Alginate	1.3801	2
	Glucose	1.3801	2

W/ adaptation: Green  
Neutral: Yellow  
Adaptation: Red

**Normally?:**



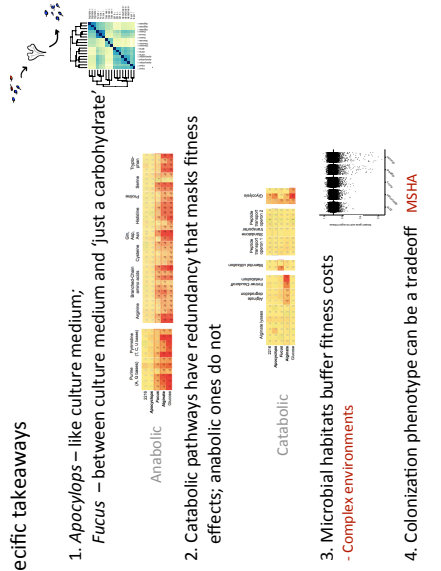
**Occurring in expt?:**



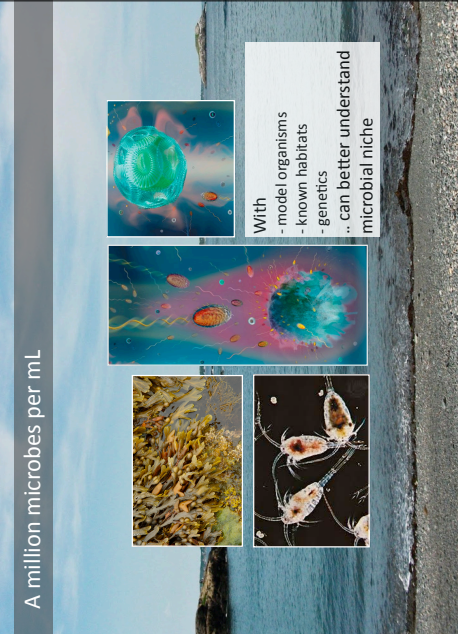
Dattay et al., Microb., 2006

### Specific takeaways

1. Apocyclops – like culture medium; *Fucus* – between culture medium and ‘just a carbohydrate’
2. Catabolic pathways have redundancy that masks fitness effects; anabolic ones do not
3. Microbial habitats buffer fitness costs  
- Complex environments
4. Colonization phenotype can be a tradeoff **MSHA**



### A million microbes per mL



**With**

- model organisms
- known habitats
- genetics

.. can better understand microbial niche

### Acknowledgements

<b>Committee members</b>	Jan-Hendrik Hehemann	Phil Gschwend
Martin Polz	Hong Xue	John MacFarlane
Eric Alm	<b>Phil Azevalo</b>	<b>Microbiology</b>
Jeff Gore	Dave Vaninsberghe	Bonnielee Whang
Matt Johnson	Joseph Elshirbini	Kris Prather
Ed DeLong	Joy Yang	Mike Laub
Polz Lab	Sana AlAttas	
<b>Sarah Preheim</b>	Matt Skalak	
Michael Cutler	Yixin Li	
Antonio Martin-Platero	Miriam Yoo	
Otto Cordero	<b>Roscoff Institute</b>	
<b>Chris Corzett</b>	<b>Frederique LeRoux</b>	
Andreas Henschel		
Kathryn Kauffman	<b>Palsson Lab</b>	
Fatma Husain	esp. Chisholm, Alm,	
Diana Chien	Thompson, DeLong labs	
Manshi Datta	Shiela Frankel	
Nate Cermak	Jim Long	





33



# Bibliography

- Abd, H., Saeed, A., Weintraub, A., Nair, G. B., and Sandström, G. (2007). *Vibrio cholerae* O1 strains are facultative intracellular bacteria, able to survive and multiply symbiotically inside the aquatic free-living amoeba *Acanthamoeba castellanii*. *FEMS Microbiology Ecology*, 60(1):33–39.
- Abd, H., Valeru, S. P., Sami, S. M., Saeed, A., Raychaudhuri, S., and Sandström, G. (2010). Interaction between *Vibrio mimicus* and *Acanthamoeba castellanii*. *Environmental microbiology reports*, 2(1):166–171.
- Abd, H., Weintraub, A., and Sandström, G. (2005). Intracellular survival and replication of *Vibrio cholerae* O139 in aquatic free-living amoebae. *Environmental Microbiology*, 7(7):1003–1008.
- Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L., and Polz, M. F. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature*, 430(6999):551–554.
- Aiso, K., Simidu, U., and Hasuo, K. (1968). Microflora in the digestive tract of inshore fish in Japan. *Journal of General Microbiology*, 52(3):361–364.
- Almagro-Moreno, S., Kim, T. K., Skorupski, K., and Taylor, R. K. (2015). Proteolysis of Virulence Regulator ToxR Is Associated with Entry of *Vibrio cholerae* into a Dormant State. *PLoS Genet*, 11(4):e1005145.
- Amaro, C., Biosca, E. G., Fouz, B., Alcaide, E., and Esteve, C. (1995). Evidence that water transmits *Vibrio vulnificus* biotype 2 infections to eels. *Applied and Environmental Microbiology*, 61(3):1133–1137.
- Andrews, J. H. and Harris, R. F. (2000). The Ecology and Biogeography of Microorganisms on Plant Surfaces. *Annual Review of Phytopathology*, 38(1):145–180.
- Ankrah, N. Y. D., May, A. L., Middleton, J. L., Jones, D. R., Hadden, M. K., Gooding, J. R., LeClerc, G. R., Wilhelm, S. W., Campagna, S. R., and Buchan, A. (2014). Phage

- infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *The ISME Journal*, 8(5):1089–1100.
- Armada, S. P., Farto, R., Pérez, M. J., and Nieto, T. P. (2003). Effect of temperature, salinity and nutrient content on the survival responses of *Vibrio splendidus* biotype I. *Microbiology*, 149(2):369–375.
- Arnold, T. M. and Targett, N. M. (2002). Marine Tannins: The Importance of a Mechanistic Framework for Predicting Ecological Roles. *Journal of Chemical Ecology*, 28(10):1919–1934.
- Arnosti, C. (2014). Patterns of Microbially Driven Carbon Cycling in the Ocean: Links between Extracellular Enzymes and Microbial Communities. *Advances in Oceanography*, 2014:e706082.
- Asplund, M., Rehnstam-Holm, A.-S., Atnur, V., Raghunath, P., Saravanan, V., Härnström, K., Collin, B., Karunasagar, I., and Godhe, A. (2011). Water column dynamics of *Vibrio* in relation to phytoplankton community composition and environmental conditions in a tropical coastal area. *Environmental Microbiology*, 13(10):2738–2751.
- Azam, F. (1998). Microbial Control of Oceanic Carbon Flux: The Plot Thickens. *Science*, 280(5364):694–696.
- Azam, F. and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, 5(10):782–791.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1):75.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2:2006.0008.
- Bagwell, C. E., Piceno, Y. M., Ashburne-Lucas, A., and Lovell, C. R. (1998). Physiological Diversity of the Rhizosphere Diazotroph Assemblages of Selected Salt Marsh Grasses. *Applied and Environmental Microbiology*, 64(11):4276–4282.
- Baran, R., Bowen, B. P., and Northen, T. R. (2011). Untargeted metabolic footprinting reveals a surprising breadth of metabolite uptake and release by *Synechococcus* sp. PCC 7002. *Molecular BioSystems*, 7(12):3200–3206.

- Barquist, L., Boinett, C. J., and Cain, A. K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biology*, 10(7):1161–1169.
- Barrett, R. D. H., MacLean, R. C., and Bell, G. (2005). Experimental evolution of *Pseudomonas fluorescens* in simple and complex environments. *The American Naturalist*, 166(4):470–480.
- Bassler, B. L., Yu, C., Lee, Y. C., and Roseman, S. (1991). Chitin utilization by marine bacteria. Degradation and catabolism of chitin oligosaccharides by *Vibrio furnissii*. *Journal of Biological Chemistry*, 266(36):24276–24286.
- Beardsley, C., Pernthaler, J., Wosniok, W., and Amann, R. (2003). Are readily culturable bacteria in coastal North Sea waters suppressed by selective grazing mortality? *Applied and environmental microbiology*, 69(5):2624–2630.
- Belasco, J. G. and Brawerman, G. (2012). *Control of Messenger RNA Stability*. Elsevier.
- Bell, W. and Mitchell, R. (1972). Chemotactic and growth responses of marine bacteria to algal extracellular products. *Biological Bulletin*, pages 265–277.
- Bengtsson-Palme, J., Rosenblad, M. A., Molin, M., and Blomberg, A. (2014). Metagenomics reveals that detoxification systems are underrepresented in marine bacterial communities. *Bmc Genomics*, 15:749. WOS:000341533400001.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.
- Bergholz, P. W., Bagwell, C. E., and Lovell, C. R. (2001). Physiological Diversity of Rhizoplane Diazotrophs of the Saltmeadow Cordgrass, *Spartina patens*: Implications for Host Specific Ecotypes. *Microbial Ecology*, 42(3):466–473.
- Bertuzzo, E., Mari, L., Righetto, L., Gatto, M., Casagrandi, R., Rodriguez-Iturbe, I., and Rinaldo, A. (2012). Hydroclimatology of dual-peak annual cholera incidence: Insights from a spatially explicit model. *Geophysical Research Letters*, 39(5):L05403.
- Berube, P. M., Biller, S. J., Kent, A. G., Berta-Thompson, J. W., Roggensack, S. E., Roache-Johnson, K. H., Ackerman, M., Moore, L. R., Meisel, J. D., Sher, D., Thompson, L. R., Campbell, L., Martiny, A. C., and Chisholm, S. W. (2015). Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *The ISME Journal*, 9(5):1195–1207.
- Biosca, E. G., Llorens, H., Garay, E., and Amaro, C. (1993). Presence of a capsule in *Vibrio vulnificus* biotype 2 and its relationship to virulence for eels. *Infection and Immunity*, 61(5):1611–1618.

- Black, W. (1953). Constituents of the marine algae. *Chem. Soc. London*, 50(Ann. Rept. on Progr. Chem.):322–35.
- Blackwell, K. D. and Oliver, J. D. (2008). The ecology of *Vibrio vulnificus*, *Vibrio cholerae*, and *Vibrio parahaemolyticus* in North Carolina Estuaries. *The Journal of Microbiology*, 46(2):146–153.
- Blanch, A., Hispano, C., Bultó, P., Ballesté, E., González-López, J., and Vilanova, X. (2009). Comparison of *Vibrio* spp. populations found in seawater, in exhibition aquaria, in fish intestine and in fish feed. *Journal of Applied Microbiology*, 106(1):57–65.
- Bold, H. and Wynne, M. (1985). *Introduction to the algae*. Prentice-Hall, Englewood Cliffs, NJ.
- Boucher, Y., Cordero, O. X., Takemura, A., Hunt, D. E., Schliep, K., Baptiste, E., Lopez, P., Tarr, C. L., and Polz, M. F. (2011). Local Mobile Gene Pools Rapidly Cross Species Boundaries To Create Endemicity within Global *Vibrio cholerae* Populations. *mBio*, 2(2).
- Broza, M. and Halpern, M. (2001). Pathogen reservoirs: Chironomid egg masses and *Vibrio cholerae*. *Nature*, 412(6842):40–40.
- Burge, C. A., Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Petes, L. E., Prager, K. C., Weil, E., Willis, B. L., Ford, S. E., and Harvell, C. D. (2014). Climate Change Influences on Marine Infectious Diseases: Implications for Management and Society. *Annual Review of Marine Science*, 6(1):249–277.
- Böer, S. I., Heinemeyer, E.-A., Luden, K., Erler, R., Gerdts, G., Janssen, F., and Brennholt, N. (2013). Temporal and Spatial Distribution Patterns of Potentially Pathogenic *Vibrio* spp. at Recreational Beaches of the German North Sea. *Microbial Ecology*, 65(4):1052–1067.
- Caburlotto, G., Haley, B., Lleò, M., Huq, A., and Colwell, R. (2010). Serodiversity and ecological distribution of *Vibrio parahaemolyticus* in the Venetian Lagoon, Northeast Italy. *Environmental Microbiology Reports*, 2(1):151–157.
- Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., Krause, D. J., and Whitaker, R. J. (2012). Patterns of Gene Flow Define Species of Thermophilic Archaea. *PLoS Biol*, 10(2):e1001265.
- Chan, E. C. S. and McManus, E. A. (1969). Distribution, characterization, and nutrition of marine microorganisms from the algae *Polysiphonia lanosa* and *Ascophyllum nodosum*. *Canadian Journal of Microbiology*, 15(5):409–420.
- Chiang, S. L. and Rubin, E. J. (2002). Construction of mariner-based transposon for epitope-tagging and genomic targeting. *Gene*, 296(1):179–185.

- Chiavelli, D. A., Marsh, J. W., and Taylor, R. K. (2001). The Mannose-Sensitive Hemagglutinin of *Vibrio cholerae* Promotes Adherence to Zooplankton. *Applied and Environmental Microbiology*, 67(7):3220–3225.
- Chimetto, L. A., Brocchi, M., Thompson, C. C., Martins, R. C., Ramos, H. R., and Thompson, F. L. (2008). Vibrios dominate as culturable nitrogen-fixing bacteria of the Brazilian coral *Mussismilia hispida*. *Systematic and Applied Microbiology*, 31(4):312–319.
- Colwell, R. R. (1996). Global Climate and Infectious Disease: The Cholera Paradigm\*. *Science*, 274(5295):2025–2031.
- Cook, D. W., Bowers, J. C., and DePaola, A. (2002). Density of total and pathogenic (tdh+) *Vibrio parahaemolyticus* in Atlantic and Gulf coast molluscan shellfish at harvest. *Journal of Food Protection*, 65(12):1873–1880. WOS:000179754600006.
- Cordero, O. X., Ventouras, L.-A., DeLong, E. F., and Polz, M. F. (2012a). Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 109(49):20059–20064.
- Cordero, O. X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., Le Roux, F., Mincer, T., and Polz, M. F. (2012b). Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science*, 337(6099):1228–1231.
- Craig, S. A., Carpenter, C. D., Mey, A. R., Wyckoff, E. E., and Payne, S. M. (2011). Positive Regulation of the *Vibrio cholerae* Porin OmpT by Iron and Fur. *Journal of Bacteriology*, 193(23):6505–6511.
- Criminger, J. D., Hazen, T. H., Sobczyk, P. A., and Lovell, C. R. (2007). Nitrogen fixation by *Vibrio parahaemolyticus* and its implications for a new ecological niche. *Applied and environmental microbiology*, 73(18):5959–5961.
- Cunneen, M. M., Liu, B., Wang, L., and Reeves, P. R. (2013). Biosynthesis of UDP-GlcNAc, UndPP-GlcNAc and UDP-GlcNAcA Involves Three Easily Distinguished 4-Epimerase Enzymes, Gne, Gnu and GnaB. *PLoS ONE*, 8(6):e67646.
- Dalisay, D. S. (2006). A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. *Microbiology*, 152(10):2875–2883.
- Davis, C. L. and Robb, F. T. (1985). Maintenance of Different Mannitol Uptake Systems during Starvation in Oxidative and Fermentative Marine Bacteria. *Applied and Environmental Microbiology*, 50:743–748.
- Davis, J. W. and Sizemore, R. K. (1982). Incidence of *Vibrio* species associated with blue crabs (*Callinectes sapidus*) collected from Galveston Bay, Texas. *Applied and environmental microbiology*, 43(5):1092–1097.

- Davis, T. A., Volesky, B., and Mucci, A. (2003). A review of the biochemistry of heavy metal biosorption by brown algae. *Water Research*, 37(18):4311–4330.
- De León, L. F., Podos, J., Gardezi, T., Herrel, A., and Hendry, A. P. (2014). Darwin’s finches and their diet niches: the sympatric coexistence of imperfect generalists. *Journal of Evolutionary Biology*, 27(6):1093–1104.
- Deepanjali, A., Kumar, H. S., Karunasagar, I., and Karunasagar, I. (2005). Seasonal Variation in Abundance of Total and Pathogenic *Vibrio parahaemolyticus* Bacteria in Oysters along the Southwest Coast of India. *Applied and Environmental Microbiology*, 71(7):3575–3580.
- DeLoney-Marino, C. R., Wolfe, A. J., and Visick, K. L. (2003). Chemoattraction of *Vibrio fischeri* to Serine, Nucleosides, and N-Acetylneuraminic Acid, a Component of Squid Light-Organ Mucus. *Applied and Environmental Microbiology*, 69(12):7527–7530.
- DeLong, E. F. (2009). The microbial ocean from genomes to biomes. *Nature*, 459(7244):200–206.
- Demarre, G., Guérout, A.-M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P., and Mazel, D. (2005). A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncP) conjugative machineries and their cognate *Escherichia coli* host strains. *Research in Microbiology*, 156(2):245–255.
- DePaola, A., Hopkins, L. H., Peeler, J. T., Wentz, B., and McPhearson, R. M. (1990). Incidence of *Vibrio parahaemolyticus* in US coastal waters and oysters. *Applied and Environmental Microbiology*, 56(8):2299–2302.
- DePaola, A., Nordstrom, J. L., Bowers, J. C., Wells, J. G., and Cook, D. W. (2003). Seasonal Abundance of Total and Pathogenic *Vibrio parahaemolyticus* in Alabama Oysters. *Applied and Environmental Microbiology*, 69(3):1521–1526.
- Deutschbauer, A., Price, M. N., Wetmore, K. M., Shao, W., Baumohl, J. K., Xu, Z., Nguyen, M., Tamse, R., Davis, R. W., and Arkin, A. P. (2011). Evidence-Based Annotation of Gene Function in *Shewanella oneidensis* MR-1 Using Genome-Wide Fitness Profiling across 121 Conditions. *PLoS Genet*, 7(11):e1002385.
- Dobretsov, S. V. and Qian, P.-Y. (2002). Effect of Bacteria Associated with the Green Alga *Ulva reticulata* on Marine Micro- and Macrofouling. *Biofouling*, 18(3):217–228.
- Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology*, 2(5):414–424.

- Doolittle, W. F. and Papke, R. T. (2006). Genomics and the bacterial species problem. *Genome Biology*, 7(9):116.
- Duan, D., Xu, L., Fei, X., and Xu, H. (1995). Marine organisms attached to seaweed surfaces in Jiaozhou Bay, China. *World Journal of Microbiology and Biotechnology*, 11(3):351–352.
- Duan, J. and Su, Y.-C. (2005). Occurrence of *Vibrio parahaemolyticus* in Two Oregon Oyster-growing Bays. *Journal of Food Science*, 70(1):M58–M63.
- Edwards, D. M., Reed, R. H., Chudek, J. A., Foster, R., and Stewart, W. D. P. (1987). Organic solute accumulation in osmotically-stressed *Enteromorpha intestinalis*. *Marine Biology*, 95(4):583–592.
- Eiler, A., Gonzalez-Rey, C., Allen, S., and Bertilsson, S. (2007). Growth response of *Vibrio cholerae* and other *Vibrio* spp. to cyanobacterial dissolved organic matter and temperature in brackish water. *FEMS Microbiology Ecology*, 60(3):411–418.
- Eiler, A., Johansson, M., and Bertilsson, S. (2006). Environmental Influences on *Vibrio* Populations in Northern Temperate and Boreal Coastal Waters (Baltic and Skagerrak Seas). *Applied and Environmental Microbiology*, 72(9):6004–6011.
- Eilers, H., Pernthaler, J., and Amann, R. (2000). Succession of pelagic marine bacteria during enrichment: a close look at cultivation-induced shifts. *Applied and environmental microbiology*, 66(11):4634–4640.
- Eisenhut, M., Huege, J., Schwarz, D., Bauwe, H., Kopka, J., and Hagemann, M. (2008). Metabolome Phenotyping of Inorganic Carbon Limitation in Cells of the Wild Type and Photorespiratory Mutants of the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. *Plant Physiology*, 148(4):2109–2120.
- Epstein, P. R. (1993). Algal blooms in the spread and persistence of cholera. *Biosystems*, 31(2–3):209–221.
- Ferdous, Z. (2009). Survival of *Vibrio cholerae* O139 in Association with *Anabaena variabilis* in Four Different Microcosms. *American Journal of Applied Sciences*, 6(3):439–444.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, 42(D1):D222–D230.
- Flårdh, K., Cohen, P. S., and Kjelleberg, S. (1992). Ribosomes exist in large excess over the apparent demand for protein synthesis during carbon starvation in marine *Vibrio* sp. strain CCUG 15956. *Journal of bacteriology*, 174(21):6780–6788.

- Fontanez, K. M., Eppley, J. M., Samo, T. J., Karl, D. M., and DeLong, E. F. (2015). Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. *Frontiers in Microbiology*, 6:469. WOS:000356254300001.
- Fraley, C. and Raftery, A. E. (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical report, Department of Statistics University of Washington.
- Franco, S., Swenson, G., and Long, R. (2012). Year round patchiness of *Vibrio vulnificus* within a temperate Texas bay. *Journal of Applied Microbiology*, 112(3):593–604.
- Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., and Hanage, W. P. (2009). The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity. *Science*, 323(5915):741–746.
- Frischkorn, K., Stojanovski, A., and Paranjpye, R. (2013). *Vibrio parahaemolyticus* type IV pili mediate interactions with diatom-derived chitin and point to an unexplored mechanism of environmental persistence. *Environmental Microbiology*, 15(5):1416–1427.
- Froelich, B., Ayrapetyan, M., and Oliver, J. D. (2012). Integration of *Vibrio vulnificus* into Marine Aggregates and Its Subsequent Uptake by *Crassostrea virginica* Oysters. *Applied and environmental microbiology*, 79(5):1454–1458.
- Froelich, B., Bowen, J., Gonzalez, R., Snedeker, A., and Noble, R. (2013). Mechanistic and Statistical Models of Total *Vibrio* Abundance in the Neuse River Estuary. *Water Research*.
- Froelich, B. and Oliver, J. D. (2013). The Interactions of *Vibrio vulnificus* and the Oyster *Crassostrea virginica*. *Microbial Ecology*, 65(4):1–10.
- Froelich, B., Ringwood, A., Sokolova, I., and Oliver, J. (2010). Uptake and depuration of the C- and E-genotypes of *Vibrio vulnificus* by the Eastern Oyster (*Crassostrea virginica*). *Environmental Microbiology Reports*, 2(1):112–115.
- Fu, Y., Waldor, M. K., and Mekalanos, J. J. (2013). Tn-Seq Analysis of *Vibrio cholerae* Intestinal Colonization Reveals a Role for T6ss-Mediated Antibacterial Activity in the Host. *Cell Host & Microbe*, 14(6):652–663. WOS:000330853500007.
- Gamble, M. D., Bagwell, C. E., LaRocque, J., Bergholz, P. W., and Lovell, C. R. (2010). Seasonal Variability of Diazotroph Assemblages Associated with the Rhizosphere of the Salt Marsh Cordgrass, *Spartina alterniflora*. *Microbial Ecology*, 59(2):253–265.
- Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V., and Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proceedings of the National Academy of Sciences*, 106(38):16422–16427.



- Giannoukos, G., Ciulla, D. M., Huang, K., Haas, B. J., Izard, J., Levin, J. Z., Livny, J., Earl, A. M., Gevers, D., Ward, D. V., Nusbaum, C., Birren, B. W., and Gnirke, A. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology*, 13(3):r23.
- Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A. C., Knight, R., Joint, I., Somerfield, P., Fuhrman, J. A., and Field, D. (2012). Defining seasonal marine microbial community dynamics. *The ISME Journal*, 6(2):298–308.
- Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappé, M. S., Short, J. M., Carrington, J. C., and Mathur, E. J. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science*, 309(5738):1242–1245.
- Goecke, F., Labes, A., Wiese, J., and Imhoff, J. F. (2010). Chemical interactions between marine macroalgae and bacteria. *Marine Ecology Progress Series*, 409:267–299.
- González-Ballester, D., Casero, D., Cokus, S., Pellegrini, M., Merchant, S. S., and Grossman, A. R. (2010). RNA-Seq Analysis of Sulfur-Deprived *Chlamydomonas* Cells Reveals Aspects of Acclimation Critical for Cell Survival. *The Plant Cell*, 22(6):2058–2084.
- Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., and Gordon, J. I. (2009). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, 6(3):279–289.
- Gosink, K. K., Kobayashi, R., Kawagishi, I., and Häse, C. C. (2002). Analyses of the roles of the three cheA homologs in chemotaxis of *Vibrio cholerae*. *Journal of bacteriology*, 184(6):1767–1771.
- Grant, P. R. and Grant, B. R. (2006). Evolution of Character Displacement in Darwin’s Finches. *Science*, 313(5784):224–226.
- Grimes, D. J., Johnson, C. N., Dillon, K. S., Flowers, A. R., Noriega, N. F., and Berutti, T. (2009). What Genomic Sequence Information Has Revealed About *Vibrio* Ecology in the Ocean—A Review. *Microbial Ecology*, 58(3):447–460.
- Grisez, L., Reyniers, J., Verdonck, L., Swings, J., and Ollevier, F. (1997). Dominant intestinal microflora of sea bream and sea bass larvae, from two hatcheries, during larval development. *Aquaculture*, 155(1):387–399.
- Grote, J., Thrash, J. C., Huggett, M. J., Landry, Z. C., Carini, P., Giovannoni, S. J., and Rappé, M. S. (2012). Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade. *mBio*, 3(5):e00252–12.

- Halpern, M., Gancz, H., Broza, M., and Kashi, Y. (2003). *Vibrio cholerae* Hemagglutinin/Protease Degrades Chironomid Egg Masses. *Applied and Environmental Microbiology*, 69(7):4200–4204.
- Halpern, M., Raats, D., Lavion, R., and Mittler, S. (2006). Dependent population dynamics between chironomids (nonbiting midges) and *Vibrio cholerae*. *FEMS Microbiology Ecology*, 55(1):98–104.
- Halpern, M., Senderovich, Y., and Izhaki, I. (2008). Waterfowl—The Missing Link in Epidemic and Pandemic Cholera Dissemination? *PLoS Pathog*, 4(10):e1000173.
- Hamdan, L. J. and Fulmer, P. A. (2011). Effects of COREXIT® EC9500a on bacteria from a beach oiled by the Deepwater Horizon spill. *Aquatic microbial ecology*, 63(2):101.
- Haygood, M. G. and Distel, D. L. (1993). Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. *Nature*, 363(6425):154–156.
- Heidelberg, J. F., Heidelberg, K. B., and Colwell, R. R. (2002a). Bacteria of the  $\gamma$ -Subclass Proteobacteria Associated with Zooplankton in Chesapeake Bay. *Applied and Environmental Microbiology*, 68(11):5498–5507.
- Heidelberg, J. F., Heidelberg, K. B., and Colwell, R. R. (2002b). Seasonality of Chesapeake Bay Bacterioplankton Species. *Applied and Environmental Microbiology*, 68(11):5488–5497.
- Hellweger, F. L., van Sebille, E., and Fredrick, N. D. (2014). Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science*, 345(6202):1346–1349. WOS:000341483800065.
- Hollants, J., Leliaert, F., Clerck, O., and Willems, A. (2013). What we can learn from sushi: a review on seaweed–bacterial associations. *FEMS microbiology ecology*, 83(1):1–16.
- Hood, M. and Winter, P. (1997). Attachment of *Vibrio cholerae* under various environmental conditions and to selected substrates. *FEMS Microbiology Ecology*, 22(3):215–223.
- Hood, M. A., Guckert, J. B., White, D. C., and Deck, F. (1986). Effect of nutrient deprivation on lipid, carbohydrate, DNA, RNA, and protein levels in *Vibrio cholerae*. *Applied and Environmental Microbiology*, 52(4):788–793.
- Hsieh, J. L., Fries, J. S., and Noble, R. T. (2007). *Vibrio* and phytoplankton dynamics during the summer of 2004 in a eutrophying estuary. *Ecological Applications*, 17(sp5):S102–S109.
- Hsieh, J. L., Fries, J. S., and Noble, R. T. (2008). Dynamics and predictive modelling of *Vibrio* spp. in the Neuse River Estuary, North Carolina, USA. *Environmental Microbiology*, 10(1):57–64.

- Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., and Polz, M. F. (2008a). Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science*, 320(5879):1081–1085.
- Hunt, D. E., Gevers, D., Vahora, N. M., and Polz, M. F. (2008b). Conservation of the chitin utilization pathway in the Vibrionaceae. *Applied and environmental microbiology*, 74(1):44–51.
- Huq, A., Colwell, R. R., Rahman, R., Ali, A., Chowdhury, M. A., Parveen, S., Sack, D. A., and Russek-Cohen, E. (1990). Detection of *Vibrio cholerae* O1 in the aquatic environment by fluorescent-monoclonal antibody and culture methods. *Applied and environmental microbiology*, 56(8):2370–2373.
- Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., Morris, J. G., Khan, M. N. H., Siddique, A. K., Yunus, M., Albert, M. J., Sack, D. A., and Colwell, R. R. (2005). Critical Factors Influencing the Occurrence of *Vibrio cholerae* in the Environment of Bangladesh. *Applied and Environmental Microbiology*, 71(8):4645–4654.
- Huq, A., Small, E. B., West, P. A., Huq, M. I., Rahman, R., and Colwell, R. R. (1983). Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Applied and Environmental Microbiology*, 45(1):275–283.
- Huq, A., West, P. A., Small, E. B., Huq, M. I., and Colwell, R. R. (1984). Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Applied and Environmental Microbiology*, 48(2):420–424.
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22:415–427.
- Høi, L., Larsen, J. L., Dalsgaard, I., and Dalsgaard, A. (1998). Occurrence of *Vibrio vulnificus* Biotypes in Danish Marine Environments. *Applied and Environmental Microbiology*, 64(1):7–13.
- Islam, M., Rahim, Z., Alam, M., Begum, S., Moniruzzaman, S., Umeda, A., Amako, K., Albert, M., Sack, R., Huq, A., and Colwell, R. (1999). Association of *Vibrio cholerae* O1 with the cyanobacterium, *Anabaena* sp., elucidated by polymerase chain reaction and transmission electron microscopy. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(1):36–40.
- Islam, M. S., Drasar, B., and Bradley, D. (1989). Attachment of toxigenic *Vibrio cholerae* O1 to various freshwater plants and survival with a filamentous green alga, *Rhizoclonium fontanum*. *Journal of Tropical Medicine and Hygiene*, 92(6):396–401.

- Islam, M. S., Drasar, B. S., and Bradley, D. J. (1988). Survival and attachment of toxigenic *Vibrio cholerae* O1 in association with four marine algae. *Bangladesh J Microbiol*, 5:41Á.
- Islam, M. S., Drasar, B. S., and Bradley, D. J. (1990a). Long-term persistence of toxigenic *Vibrio cholerae* O1 in the mucilaginous sheath of a blue-green alga, *Anabaena variabilis*. *Journal of Tropical Medicine and Hygiene*, 93(2):133–139.
- Islam, M. S., Drasar, B. S., and Bradley, D. J. (1990b). Survival of toxigenic *Vibrio cholerae* O1 with a common duckweed, *Lemna minor*, in artificial aquatic ecosystems. *Transactions of the Royal Society of tropical medicine and hygiene*, 84(3):422–424.
- Islam, M. S., Goldar, M. M., Morshed, M. G., Bakht, H. B. M., and Sack, D. A. (2006). Chemotaxis between *Vibrio cholerae* O1 and a blue-green alga, *Anabaena* sp. *Epidemiology and Infection*, 134(3):645.
- Islam, M. S., Goldar, M. M., Morshed, M. G., Khan, M. N., Islam, M. R., and Sack, R. B. (2002). Involvement of the hap gene (mucinase) in the survival of *Vibrio cholerae* O1 in association with the blue-green alga, *Anabaena* sp. *Canadian journal of microbiology*, 48(9):793–800.
- Jiang, S. C. and Fu, W. (2001). Seasonal Abundance and Distribution of *Vibrio cholerae* in Coastal Waters Quantified by a 16s-23s Intergenic Spacer Probe. *Microbial Ecology*, 42(4):540–548. ArticleType: research-article / Full publication date: Nov., 2001 / Copyright © 2001 Springer.
- Johnson, C. N. (2013). Fitness Factors in Vibrios: a Mini-review. *Microbial Ecology*, 65(4):826–851.
- Johnson, C. N., Bowers, J. C., Griffitt, K. J., Molina, V., Clostio, R. W., Pei, S., Laws, E., Paranjpye, R. N., Strom, M. S., Chen, A., Hasan, N. A., Huq, A., Noriea, N. F., Grimes, D. J., and Colwell, R. R. (2012). Ecology of *Vibrio parahaemolyticus* and *Vibrio vulnificus* in the Coastal and Estuarine Waters of Louisiana, Maryland, Mississippi, and Washington (United States). *Applied and Environmental Microbiology*, 78(20):7249–7257.
- Johnson, C. N., Flowers, A. R., Noriea, N. F., Zimmerman, A. M., Bowers, J. C., DePaola, A., and Grimes, D. J. (2010). Relationships between Environmental Factors and Pathogenic Vibrios in the Northern Gulf of Mexico. *Applied and Environmental Microbiology*, 76(21):7076–7084.
- Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W. (2006). Niche Partitioning Among *Prochlorococcus* Ecotypes Along Ocean-Scale Environmental Gradients. *Science*, 311(5768):1737–1740.
- Jones, S. and Summer-Brason, B. (1998). Incidence and detection of pathogenic *Vibrio* sp. in a northern new England Estuary, USA. *Journal of Shellfish Research*, 17(5):1665–1669.

- Juhas, M., Eberl, L., and Glass, J. I. (2011). Essence of life: essential genes of minimal genomes. *Trends in Cell Biology*, 21(10):562–568.
- Julie, D., Solen, L., Antoine, V., Jaufrey, C., Annick, D., and Dominique, H.-H. (2010). Ecology of pathogenic and non-pathogenic *Vibrio parahaemolyticus* on the French Atlantic coast. Effects of temperature, salinity, turbidity and chlorophyll a. *Environmental Microbiology*, 12(4):929–937.
- Jutla, A., Akanda, A. S., Huq, A., Faruque, A. S. G., Colwell, R., and Islam, S. (2013). A water marker monitored by satellites to predict seasonal endemic cholera. *Remote Sensing Letters*, 4(8):822–831.
- Kanagasabhapathy, M., Sasaki, H., and Nagata, S. (2008). Phylogenetic identification of epibiotic bacteria possessing antimicrobial activities isolated from red algal species of Japan. *World Journal of Microbiology and Biotechnology*, 24(10):2315–2321.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42(D1):D199–D205.
- Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R. R., Stocker, R., Follows, M. J., Stepanauskas, R., and Chisholm, S. W. (2014). Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild *Prochlorococcus*. *Science*, 344(6182):416–420. WOS:000334867800042.
- Kassen, R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. *Journal of Evolutionary Biology*, 15(2):173–190.
- Kauffman, A. K. M. (2014). *Demographics of lytic viral infection of coastal ocean vibrio*. Thesis, Massachusetts Institute of Technology. Thesis: Ph. D., Joint Program in Biological Oceanography (Massachusetts Institute of Technology, Department of Civil and Environmental Engineering; and the Woods Hole Oceanographic Institution), 2014.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6):845–858.
- Kirchman, D. L. (2012). *Processes in Microbial Ecology*. OUP Oxford.
- Kirschner, A. K., Schlesinger, J., Farnleitner, A. H., Hornek, R., Süß, B., Golda, B., Herzig, A., and Reitner, B. (2008). Rapid growth of planktonic *Vibrio cholerae* non-O1/non-O139 strains in a large alkaline lake in Austria: dependence on temperature and dissolved organic carbon quality. *Applied and environmental microbiology*, 74(7):2004–2015.

- Kirschner, A. K. T., Schauer, S., Steinberger, B., Wilhartitz, I., Grim, C. J., Huq, A., Colwell, R. R., Herzig, A., and Sommer, R. (2011). Interaction of *Vibrio cholerae* non-O1/non-O139 with Copepods, Cladocerans and Competing Bacteria in the Large Alkaline Lake Neusiedler See, Austria. *Microbial Ecology*, 61(3):496–506.
- Knirel, Y. A., Paredes, L., Jansson, P.-E., Weintraub, A., Widmalm, G., and Albert, M. J. (1995). Structure of the Capsular Polysaccharide of *Vibrio cholerae* O139 Synonym Bengal Containing D-galactose 4,6-Cyclophosphate. *European Journal of Biochemistry*, 232(2):391–396.
- Koenig, J. E., Bourne, D. G., Curtis, B., Dlutek, M., Stokes, H. W., Doolittle, W. F., and Boucher, Y. (2011). Coral-mucus-associated *Vibrio* integrons in the Great Barrier Reef: genomic hotspots for environmental adaptation. *The ISME Journal*, 5(6):962–972.
- Koren, O. and Rosenberg, E. (2006). Bacteria Associated with Mucus and Tissues of the Coral *Oculina patagonica* in Summer and Winter. *Applied and Environmental Microbiology*, 72(8):5254–5259.
- Kramer, J. G. and Singleton, F. L. (1992). Variations in rRNA content of marine *Vibrio* spp. during starvation-survival and recovery. *Applied and environmental microbiology*, 58(1):201–207.
- Krediet, C. J., Ritchie, K. B., Alagely, A., and Teplitski, M. (2013). Members of native coral microbiota inhibit glycosidases and thwart colonization of coral mucus by an opportunistic pathogen. *The ISME Journal*, 7(5):980–990.
- Kumazawa, N. H., Fukuma, N., and Komoda, Y. (1991). Attachment of *Vibrio parahaemolyticus* strains to estuarine algae. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science*, 53(2):201–205.
- Kvennefors, E. C. E., Sampayo, E., Ridgway, T., Barnes, A. C., and Hoegh-Guldberg, O. (2010). Bacterial Communities of Two Ubiquitous Great Barrier Reef Corals Reveals Both Site- and Species-Specificity of Common Bacterial Associates. *PLoS ONE*, 5(4):e10401.
- Lack, D. (1947). *Darwin's Finches*. Cambridge University Press, Cambridge, UK.
- Lakshmanaperumalsamy, P. and Purushothaman, A. (1982). Heterotrophic bacteria associated with seaweed. *Proceedings: Plant Sciences*, 91(6):487–493.
- Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J., and Robertson, H. M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. *Proceedings of the National Academy of Sciences*, 96(20):11428–11433.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3):R25.

- Langridge, G. C., Phan, M.-D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J., and Turner, A. K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, 19(12):2308–2316.
- Lapota, D., Galt, C., Losee, J. R., Huddell, H. D., Orzech, J. K., and Neelson, K. H. (1988). Observations and measurements of planktonic bioluminescence in and around a milky sea. *Journal of Experimental Marine Biology and Ecology*, 119(1):55–81.
- Larocque, J. R., Bergholz, P. W., Bagwell, C. E., and Lovell, C. R. (2004). Influence of host plant-derived and abiotic environmental parameters on the composition of the diazotroph assemblage associated with roots of *Juncus roemerianus*. *Antonie van Leeuwenhoek*, 86(3):249–261.
- Larsen, M. H., Blackburn, N., Larsen, J. L., and Olsen, J. E. (2004). Influences of temperature, salinity and starvation on the motility and chemotactic response of *Vibrio anguillarum*. *Microbiology*, 150(5):1283–1290.
- Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., DeMaere, M. Z., Ting, L., Ertan, H., Johnson, J., Ferriera, S., Lapidus, A., Anderson, I., Kyrpides, N., Munk, A. C., Detter, C., Han, C. S., Brown, M. V., Robb, F. T., Kjelleberg, S., and Cavicchioli, R. (2009). The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences*, 106(37):15527–15533.
- Laycock, R. A. (1974). Bacteria associated with the surface of *Laminaria* fronds. *Marine Biology*, 25(3):223–231.
- Le Roux, F., Binesse, J., Saulnier, D., and Mazel, D. (2007). Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene *vsm* by use of a novel counterselectable suicide vector. *Applied and Environmental Microbiology*, 73(3):777–784.
- Leplae, R., Hebrant, A., Wodak, S. J., and Toussaint, A. (2004). ACLAME: A CLAssification of Mobile genetic Elements. *Nucleic Acids Research*, 32(suppl 1):D45–D49.
- Lewis, D. H. and Smith, D. C. (1967). Sugar Alcohols (polyols) in Fungi and Green Plants. *New Phytologist*, 66(2):143–184.
- Lipp, E. K., Rodriguez-palacios, C., and Rose, J. B. (2001). Occurrence and distribution of the human pathogen *Vibrio vulnificus* in a subtropical Gulf of Mexico estuary. *Hydrobiologia*, 460(1-3):165–173.
- Liston, J. (1954). A group of luminous and nonluminous bacteria from the intestine of flatfish. *Journal of General Microbiology*, 12(1):i.

- Liston, J. (1957). The occurrence and distribution of bacterial types on flatfish. *Journal of General Microbiology*, 16(1):205–216.
- Liu, Z. and Liu, J. (2013). Evaluating bacterial community structures in oil collected from the sea surface and sediment in the northern Gulf of Mexico after the Deepwater Horizon oil spill. *MicrobiologyOpen*, 2(3):492–504.
- Lizárraga-Partida, M. L., Mendez-Gómez, E., Rivas-Montaño, A. M., Vargas-Hernández, E., Portillo-López, A., González-Ramírez, A. R., Huq, A., and Colwell, R. R. (2009). Association of *Vibrio cholerae* with plankton in coastal areas of Mexico. *Environmental Microbiology*, 11(1):201–208.
- Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A. S. G., and Colwell, R. (2000). Climate and Infectious Disease: Use of Remote Sensing for Detection of *Vibrio Cholerae* by Indirect Measurement. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4):1438–1443. ArticleType: research-article / Full publication date: Feb. 15, 2000 / Copyright © 2000 National Academy of Sciences.
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.
- Lovell, C. R., Decker, P. V., Bagwell, C. E., Thompson, S., and Matsui, G. Y. (2008). Analysis of a diverse assemblage of diazotrophic bacteria from *Spartina alterniflora* using DGGE and clone library screening. *Journal of Microbiological Methods*, 73(2):160–171.
- Lyons, M. M., Lau, Y.-T., Carden, W. E., Ward, J. E., Roberts, S. B., Smolowitz, R., Vallino, J., and Allam, B. (2007). Characteristics of Marine Aggregates in Shallow-water Ecosystems: Implications for Disease Ecology. *EcoHealth*, 4(4):406–420.
- Magny, G. C. d., Mozumder, P. K., Grim, C. J., Hasan, N. A., Naser, M. N., Alam, M., Sack, R. B., Huq, A., and Colwell, R. R. (2011). Role of Zooplankton Diversity in *Vibrio cholerae* Population Dynamics and in the Incidence of Cholera in the Bangladesh Sundarbans. *Applied and Environmental Microbiology*, 77(17):6125–6132.
- Magny, G. C. d., Murtugudde, R., Sapiano, M. R. P., Nizam, A., Brown, C. W., Busalacchi, A. J., Yunus, M., Nair, G. B., Gil, A. I., Lanata, C. F., Calkins, J., Manna, B., Rajendran, K., Bhattacharya, M. K., Huq, A., Sack, R. B., and Colwell, R. R. (2008). Environmental signatures associated with cholera epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46):17676–17681.
- Mahowald, M. A., Rey, F. E., Sedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henrissat, B., Crock, L. W., Russell, A., Verberkmoes, N. C., Hettich, R. L., and Gordon, J. I. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proceedings of the National Academy of Sciences*, 106(14):5859–5864.



- Martens, E. C., Kelly, A. G., Tauzin, A. S., and Brumer, H. (2014). The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *Journal of Molecular Biology*, 426(23):3851–3865. WOS:000345880200004.
- Martinez-Urtaza, J., Blanco-Abad, V., Rodriguez-Castro, A., Ansedo-Bermejo, J., Miranda, A., and Rodriguez-Alvarez, M. X. (2012). Ecological determinants of the occurrence and dynamics of *Vibrio parahaemolyticus* in offshore areas. *The ISME Journal*, 6(5):994–1006.
- Materna, A. C., Friedman, J., Bauer, C., David, C., Chen, S., Huang, I. B., Gillens, A., Clarke, S. A., Polz, M. F., and Alm, E. J. (2012). Shape and evolution of the fundamental niche in marine *Vibrio*. *The ISME Journal*, 6(12):2168–2177.
- Matz, C., McDougald, D., Moreno, A. M., Yung, P. Y., Yildiz, F. H., and Kjelleberg, S. (2005). Biofilm formation and phenotypic variation enhance predation-driven persistence of *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16819–16824.
- Mayr, E. (1982). *The growth of biological thought: diversity, evolution and inheritance*. Harvard.
- Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1):31–46.
- Miller, V. L. and Mekalanos, J. J. (1984). Synthesis of cholera toxin is positively regulated at the transcriptional level by *toxR*. *Proceedings of the National Academy of Sciences*, 81(11):3471–3475.
- Mishra, A., Taneja, N., and Sharma, M. (2012). Environmental and epidemiological surveillance of *Vibrio cholerae* in a cholera-endemic region in India with freshwater environs. *Journal of Applied Microbiology*, 112(1):225–237.
- Mizanur, R. M., Islam, M. S., Khan, S. I., and Rahim, Z. (2002). The Chemotactic Response of *Vibrio cholerae* 0139 to The Mucilaginous Sheath of Blue-Green Algae Is Mediated by a Combination of Sheath Components. *Microbes and Environments*, 17(1):18–25.
- Mizanur, R. M., Khan, S. I., Islam, M. S., Sultana, I., Rahim, Z., and Albert, M. J. (2001). Role of Chemotaxis in the attachment of *Vibrio cholerae* 0139 with different aquatic flora and fauna. *Pak. J. Biol. Sci.*, 4:1395–1399.
- Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., and Rocap, G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *The ISME Journal*, 4(5):673–685.

- Motes, M. L., DePaola, A., Cook, D. W., Veazey, J. E., Hunsucker, J. C., Garthright, W. E., Blodgett, R. J., and Chirtel, S. J. (1998). Influence of Water Temperature and Salinity on *Vibrio vulnificus* in Northern Gulf and Atlantic Coast Oysters (*Crassostrea virginica*). *Applied and Environmental Microbiology*, 64(4):1459–1465.
- Mouriño-Pérez, R. R., Worden, A. Z., and Azam, F. (2003). Growth of *Vibrio cholerae* O1 in Red Tide Waters off California. *Applied and Environmental Microbiology*, 69(11):6923–6931.
- Moxley, K. and Schmidt, S. (2010). Preliminary characterization of an estuarine, benzoate-utilizing *Vibrio* sp. isolated from Durban Harbour, South Africa. *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*, pages 1249–1254.
- Mueller, R. S., McDougald, D., Cusumano, D., Sodhi, N., Kjelleberg, S., Azam, F., and Bartlett, D. H. (2007). *Vibrio cholerae* Strains Possess Multiple Strategies for Abiotic and Biotic Surface Colonization. *Journal of Bacteriology*, 189(14):5348–5360.
- Muroga, K., Higashi, M., and Keitoku, H. (1987). The isolation of intestinal microflora of farmed red seabream (*Pagrus major*) and black seabream (*Acanthopagrus schlegeli*) at larval and juvenile stages. *Aquaculture*, 65(1):79–88.
- Murphree, R. L. and Tamplin, M. L. (1995). Uptake and retention of *Vibrio cholerae* O1 in the Eastern oyster, *Crassostrea virginica*. *Applied and environmental microbiology*, 61(10):3656–3660.
- Nakanishi, K., Nishijima, M., Nishimura, M., Kuwano, K., and Saga, N. (1996). Bacteria that Induce Morphogenesis in *Ulva Pertusa* (Chlorophyta) Grown Under Axenic Conditions. *Journal of phycology*, 32(3):479–482.
- Nealson, K. H. and Hastings, J. W. (2006). Quorum Sensing on a Global Scale: Massive Numbers of Bioluminescent Bacteria Make Milky Seas. *Applied and Environmental Microbiology*, 72(4):2295–2297.
- Neidhardt, F. C. and Umberger, H. E. (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, volume 1. ASM Press, Washington, D.C., 2nd edition.
- Neogi, S. B., Islam, M. S., Nair, G. B., Yamasaki, S., and Lara, R. J. (2012). Occurrence and distribution of plankton-associated and free-living toxigenic *Vibrio cholerae* in a tropical estuary of a cholera endemic zone. *Wetlands Ecology and Management*, 20(3):271–285.
- Nesvizhskii, A. I. (2010). A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics*, 73(11):2092–2123.

- Newman Jr, J. T., Cosenza, B. J., and Buck, J. D. (1972). Aerobic microflora of the bluefish (*Pomatomus saltatrix*) intestine. *Journal of the Fisheries Board of Canada*, 29(3):333–336.
- Nigro, O., Hou, A., Vithanage, G., Fujioka, R., and Steward, G. (2011). Temporal and spatial variability in culturable pathogenic vibrio spp. in Lake Pontchartrain, Louisiana, following hurricanes katrina and rita. *Applied and Environmental Microbiology*, 77(15):5384–5393.
- Nishina, T., Wada, M., Ozawa, H., Hara-Kudo, Y., Konuma, H., Hasegawa, J., and Kumagai, S. (2004). Growth kinetics of *Vibrio parahaemolyticus* O3: K6 under varying conditions of pH, NaCl concentration and temperature. *Shokuhin eiseigaku zasshi. Journal of the Food Hygienic Society of Japan*, 45(1):35.
- Oberbeckmann, S., Fuchs, B. M., Meiners, M., Wichels, A., Wiltshire, K. H., and Gerdts, G. (2012). Seasonal Dynamics and Modeling of a *Vibrio* Community in Coastal Waters of the North Sea. *Microbial Ecology*, 63(3):543–551.
- Paerl, H. W. and Gallucci, K. K. (1985). Role of Chemotaxis in Establishing a Specific Nitrogen-Fixing Cyanobacterial-Bacterial Association. *Science*, 227(4687):647–649.
- Paerl, H. W. and Pinckney, J. L. (1996). A mini-review of microbial consortia: Their roles in aquatic production and biogeochemical cycling. *Microbial Ecology*, 31(3):225–247.
- Passalacqua, K. D., Varadarajan, A., Ondov, B. D., Okou, D. T., Zwick, M. E., and Bergman, N. H. (2009). Structure and complexity of a bacterial transcriptome. *Journal of Bacteriology*, 191(10):3203–3211.
- Patel, P., Callow, M. E., Joint, I., and Callow, J. A. (2003). Specificity in the settlement–modifying response of bacterial biofilms towards zoospores of the marine alga *Enteromorpha*. *Environmental Microbiology*, 5(5):338–349.
- Patra, T., Koley, H., Ramamurthy, T., Ghose, A. C., and Nandy, R. K. (2012). The Entner-Doudoroff Pathway Is Obligatory for Gluconate Utilization and Contributes to the Pathogenicity of *Vibrio cholerae*. *Journal of Bacteriology*, 194(13):3377–3385.
- Percival, E. and McDowell, R. (1967). *Chemistry and Enzymology of Marine Algal Polysaccharides*. Academic Press, London, UK.
- Pernthaler, A., Pernthaler, J., Eilers, H., and Amann, R. (2001). Growth patterns of two marine isolates: adaptations to substrate patchiness? *Applied and environmental microbiology*, 67(9):4077–4083.
- Pfeffer, C. S., Hite, M. F., and Oliver, J. D. (2003). Ecology of *Vibrio vulnificus* in Estuarine Waters of Eastern North Carolina. *Applied and Environmental Microbiology*, 69(6):3526–3531.

- Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics*, 29(3):170–175.
- Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Research*, 21(4):599–609.
- Poretzky, R. S., Bano, N., Buchan, A., LeClerc, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A., and Hollibaugh, J. T. (2005). Analysis of Microbial Gene Transcripts in Environmental Samples. *Applied and Environmental Microbiology*, 71(7):4121–4126.
- Prasanthan, V., Udayakumar, P., Sarathkumar, and Ouseph, P. (2011). Influence of abiotic environmental factors on the abundance and distribution of *Vibrio* species in coastal waters of Kerala, India. *Indian Journal of Marine Sciences*, 40(4):587–592.
- Preheim, S. P. (2010). *Ecology and population structure of vibronaceae in the coastal ocean*. Thesis, Massachusetts Institute of Technology.
- Preheim, S. P., Boucher, Y., Wildschutte, H., David, L. A., Veneziano, D., Alm, E. J., and Polz, M. F. (2011a). Metapopulation structure of Vibrionaceae among coastal marine invertebrates. *Environmental Microbiology*, 13(1):265–275.
- Preheim, S. P., Timberlake, S., and Polz, M. F. (2011b). Merging Taxonomy with Ecological Population Prediction in a Case Study of Vibrionaceae. *Applied and Environmental Microbiology*, 77(20):7195–7206.
- Pritchard, J. R., Chao, M. C., Abel, S., Davis, B. M., Baranowski, C., Zhang, Y. J., Rubin, E. J., and Waldor, M. K. (2014). ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet*, 10(11):e1004782.
- Rakhuba, D. V., Kolomiets, E. I., Dey, E. S., and Novik, G. I. (2010). Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish Journal of Microbiology / Polskie Towarzystwo Mikrobiologów = The Polish Society of Microbiologists*, 59(3):145–155.
- Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., Shah, M., Hettich, R. L., and Banfield, J. F. (2005). Community Proteomics of a Natural Microbial Biofilm. *Science*, 308(5730):1915–1920.
- Ramaiah, N., Ravel, J., Straube, W. L., Hill, R. T., and Colwell, R. R. (2002). Entry of *Vibrio harveyi* and *Vibrio fischeri* into the viable but nonculturable state. *Journal of applied microbiology*, 93(1):108–116.
- Ramirez, G., Buck, G., Smith, A., Gordon, K., and Mott, J. (2009). Incidence of *Vibrio vulnificus* in estuarine waters of the south Texas Coastal Bend region. *Journal of Applied Microbiology*, 107(6):2047–2053.

- Randa, M. A., Polz, M. F., and Lim, E. (2004). Effects of Temperature and Salinity on *Vibrio vulnificus* Population Dynamics as Assessed by Quantitative PCR. *Applied and Environmental Microbiology*, 70(9):5469–5476.
- Rao, B. M., III, F. L. I., and Holmes, L. D. (2013). Bioluminescence and Chitinase Production during Chitin Fermentation by *Vibrio harveyi*. *Journal of Life Sciences*, 7(5):491–494.
- Rawlings, T. K., Ruiz, G. M., and Colwell, R. R. (2007). Association of *Vibrio cholerae* O1 El Tor and O139 Bengal with the Copepods *Acartia tonsa* and *Eurytemora affinis*. *Applied and environmental microbiology*, 73(24):7926–7933.
- Reed, R., Davison, I., Chudek, J., and Foster, R. (1985). The Osmotic Role of Mannitol in the Phaeophyta - an Appraisal. *Phycologia*, 24(1):35–47. WOS:A1985ADN1800002.
- Rioux, L. E., Turgeon, S. L., and Beaulieu, M. (2007). Characterization of polysaccharides extracted from brown seaweeds. *Carbohydrate Polymers*, 69(3):530–537.
- Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pašić, L., Thingstad, T. F., Rohwer, F., and Mira, A. (2009). Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology*, 7(11):828–836.
- Rowbotham, T. J. (1980). Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *Journal of clinical pathology*, 33(12):1179–1183.
- Ruby, E. G. and Lee, K.-H. (1998). The *Vibrio fischeri*-*Euprymna scolopes* light organ association: current ecological paradigms. *Applied and environmental microbiology*, 64(3):805–812.
- Ruby, E. G. and Nealson, K. H. (1976). Symbiotic association of *Photobacterium fischeri* with the marine luminous fish *Monocentris japonica*: a model of symbiosis based on bacterial studies. *The Biological Bulletin*, 151(3):574–586.
- Saito, K. and Matsuda, F. (2010). Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. *Annual Review of Plant Biology*, 61(1):463–489.
- Saliba, A.-E., Westermann, A. J., Gorski, S. A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research*, page gku555.
- Sandström, G., Saeed, A., and Abd, H. (2010). *Acanthamoeba polyphaga* is a possible host for *Vibrio cholerae* in aquatic environments. *Experimental parasitology*, 126(1):65–68.
- Sawabe, T. (2006). The Mutual Partnership between *Vibrio haliotocoli* and Abalones. In Thompson, F. L., Austin, B., and Swings, J., editors, *Biology of Vibrios*, pages 219–230. Amer Soc Microbiology, Washington. WOS:000278637800016.

- Schneider, T. and Riedel, K. (2010). Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics*, 10(4):785–798.
- Scholl, D., Adhya, S., and Merrill, C. (2005). Escherichia coli K1's Capsule Is a Barrier to Bacteriophage T7. *Applied and Environmental Microbiology*, 71(8):4872–4874.
- Scholl, D., Rogers, S., Adhya, S., and Merrill, C. R. (2001). Bacteriophage K1-5 Encodes Two Different Tail Fiber Proteins, Allowing It To Infect and Replicate on both K1 and K5 Strains of Escherichia coli. *Journal of Virology*, 75(6):2509–2515.
- Schwahnhäuser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342.
- Schwarz, D., Orf, I., Kopka, J., and Hagemann, M. (2013). Recent Applications of Metabolomics Toward Cyanobacteria. *Metabolites*, 3(1):72–100.
- Sera, H., Ishida, Y., and Kadota, H. (1972). Bacterial flora in the digestive tracts of marine fish. IV. Effect of H<sup>+</sup> concentration and gastric juices on the indigenous bacteria. *Bull. Jap. Soc. Sci. Fish*, 38:859–863.
- Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., Polz, M. F., and Alm, E. J. (2012). Population Genomics of Early Events in the Ecological Differentiation of Bacteria. *Science*, 336(6077):48–51.
- Sharma, A. K., Becker, J. W., Ottesen, E. A., Bryant, J. A., Duhamel, S., Karl, D. M., Cordero, O. X., Repeta, D. J., and DeLong, E. F. (2014). Distinct dissolved organic matter sources induce rapid transcriptional responses in coexisting populations of Prochlorococcus, Pelagibacter and the OM60 clade. *Environmental Microbiology*, 16(9):2815–2830.
- Sharon, G. and Rosenberg, E. (2008). Bacterial growth on coral mucus. *Current microbiology*, 56(5):481–488.
- Shen, X., Cai, Y., Liu, C., Liu, W., Hui, Y., and Su, Y.-C. (2009). Effect of temperature on uptake and survival of *Vibrio parahaemolyticus* in oysters (*Crassostrea plicatula*). *International journal of food microbiology*, 136(1):129–132.
- Shi, Y., Tyson, G. W., and DeLong, E. F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature*, 459(7244):266–269.
- Shu, H.-Y., Fung, C.-P., Liu, Y.-M., Wu, K.-M., Chen, Y.-T., Li, L.-H., Liu, T.-T., Kirby, R., and Tsai, S.-F. (2009). Genetic diversity of capsular polysaccharide biosynthesis in *Klebsiella pneumoniae* clinical isolates. *Microbiology*, 155(12):4170–4183.
- Simidu, U., Ashino, K., and Kaneko, E. (1971). Bacterial flora of phyto-and zoo-plankton in the inshore water of Japan. *Canadian journal of microbiology*, 17(9):1157–1160.

- Sjoblad, R. D. and Mitchell, R. (1979). Chemotactic responses of *Vibrio alginolyticus* to algal extracellular products. *Canadian Journal of Microbiology*, 25(9):964–967.
- Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, advance online publication.
- Smriga, S., Sandin, S. A., and Azam, F. (2010). Abundance, diversity, and activity of microbial assemblages associated with coral reef fish guts and feces. *FEMS Microbiology Ecology*, 73(1):31–42.
- Sobrinho, P. d. S. C., Destro, M. T., Franco, B. D. G. M., and Landgraf, M. (2010). Correlation between Environmental Factors and Prevalence of *Vibrio parahaemolyticus* in Oysters Harvested in the Southern Coastal Area of Sao Paulo State, Brazil. *Applied and Environmental Microbiology*, 76(4):1290–1293.
- Sochard, M. R., Wilson, D. F., Austin, B., and Colwell, R. R. (1979). Bacteria associated with the surface and gut of marine copepods. *Applied and Environmental Microbiology*, 37(4):750–759.
- Spira, W. M., Huq, A., Ahmed, Q. S., and Saeed, Y. A. (1981). Uptake of *Vibrio cholerae* Biotype eltor from Contaminated Water by Water Hyacinth (*Eichornia crassipes*). *Applied and Environmental Microbiology*, 42(3):550–553.
- Stabb, E. V. (2006). The *Vibrio fischeri*-*Euprymna scolopes* light organ symbiosis. *The biology of vibrios*. ASM Press, Washington, DC, pages 204–218.
- Stephens, E. L., Molina, V., Cole, K. M., Laws, E., and Johnson, C. N. (2013). In situ and in vitro impacts of the Deepwater Horizon oil spill on *Vibrio parahaemolyticus*. *Marine Pollution Bulletin*, 75(1–2):90–97.
- Stewart, J. R., Gast, R. J., Fujioka, R. S., Solo-Gabriele, H. M., Meschke, J. S., Amaral-Zettler, L. A., del Castillo, E., Polz, M. F., Collier, T. K., and Strom, M. S. (2008). The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. *Environ Health*, 7(Suppl 2):S3.
- Stocker, R. (2012). Marine Microbes See a Sea of Gradients. *Science*, 338(6107):628–633.
- Sugita, H., Ushioka, S., Kihara, D., and Deguchi, Y. (1985). Changes in the bacterial composition of water in a carp rearing tank. *Aquaculture*, 44(3):243–247.
- Sussman, A. J. and Gilvarg, C. (1971). Peptide Transport and Metabolism in Bacteria. *Annual Review of Biochemistry*, 40(1):397–408.
- Sussman, M. (1997). *Escherichia Coli: Mechanisms of Virulence*. Cambridge University Press.

- Suttle, C. A. (2007). Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10):801–812.
- Sutton, D. C. and Clements, K. D. (1988). Aerobic, heterotrophic gastrointestinal microflora of tropical marine fishes. In *Proceedings of the 6th International Coral Reef Symposium*, volume 3, pages 185–190, Townsville, Australia.
- Szabo, G., Preheim, S. P., Kauffman, K. M., David, L. A., Shapiro, J., Alm, E. J., and Polz, M. F. (2013). Reproducibility of Vibrionaceae population structure in coastal bacterioplankton. *The ISME Journal*, 7(3):509–519.
- Tait, K., Joint, I., Daykin, M., Milton, D. L., Williams, P., and Camara, M. (2005). Disruption of quorum sensing in seawater abolishes attraction of zoospores of the green alga *Ulva* to bacterial biofilms. *Environmental Microbiology*, 7(2):229–240.
- Takemura, A. F., Chien, D. M., and Polz, M. F. (2014). Associations and dynamics of Vibrionaceae in the environment, from the genus to the population level. *Frontiers in Microbiology*, 5.
- Tamplin, M. L., Gauzens, A. L., Huq, A., Sack, D. A., and Colwell, R. R. (1990). Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. *Applied and Environmental Microbiology*, 56(6):1977–1980.
- Tarsi, R. and Pruzzo, C. (1999). Role of Surface Proteins in *Vibrio cholerae* Attachment to Chitin. *Applied and Environmental Microbiology*, 65(3):1348–1351.
- Teschler, J. K., Zamorano-Sánchez, D., Utada, A. S., Warner, C. J. A., Wong, G. C. L., Lington, R. G., and Yildiz, F. H. (2015). Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *Nature Reviews Microbiology*, 13(5):255–268.
- Thekin, K. H. and Taylor, R. K. (1996). Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. *Infection and Immunity*, 64(7):2853–2856.
- Thompson, J. R., Marcelino, L. A., and Polz, M. F. (2005). Diversity, Sources, and Detection of Human Bacterial Pathogens in the Marine Environment. In Belkin, S. and Colwell, R. R., editors, *Oceans and Health: Pathogens in the Marine Environment*, pages 29–68. Springer US.
- Thompson, J. R. and Polz, M. F. (2006). Dynamics of *Vibrio* populations and their role in environmental nutrient cycling. *The biology of vibrios*. ASM Press, Washington, DC, pages 190–203.
- Thompson, J. R., Randa, M. A., Marcelino, L. A., Tomita-Mitchell, A., Lim, E., and Polz, M. F. (2004). Diversity and Dynamics of a North Atlantic Coastal *Vibrio* Community. *Applied and Environmental Microbiology*, 70(7):4103–4110.



- Tibbles, B. J. and Rawlings, D. E. (1994). Characterization of nitrogen-fixing bacteria from a temperate saltmarsh lagoon, including isolates that produce ethane from acetylene. *Microbial Ecology*, 27(1):65–80.
- Tomlinson, S. and Taylor, P. W. (1985). Neuraminidase associated with coliphage E that specifically depolymerizes the Escherichia coli K1 capsular polysaccharide. *Journal of Virology*, 55(2):374–378.
- Turner, J. W., Good, B., Cole, D., and Lipp, E. K. (2009). Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME Journal*, 3(9):1082–1092. WOS:000269618300008.
- Turner, K. H., Everett, J., Trivedi, U., Rumbaugh, K. P., and Whiteley, M. (2014). Requirements for Pseudomonas aeruginosa Acute Burn and Chronic Surgical Wound Infection. *PLoS Genet*, 10(7):e1004518.
- Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L., and Whiteley, M. (2015). Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. *Proceedings of the National Academy of Sciences*, page 201419677.
- Tyson, G. W. and Banfield, J. F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*, 10(1):200–207.
- Urdaci, M. C., Stal, L. J., and Marchand, M. (1988). Occurrence of nitrogen fixation among Vibrio spp. *Archives of Microbiology*, 150(3):224–229.
- Vahl, O. (1972). Efficiency of particle retention in Mytilus edulis L. *Ophelia*, 10(1):17–25.
- van Opijnen, T., Bodi, K. L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, 6(10):767–772.
- van Opijnen, T. and Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, 11(7):435–442.
- van Opijnen, T., Lazinski, D. W., and Camilli, A. (2014). Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. *Current Protocols in Molecular Biology*, pages 7–16.
- Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L., and Jansson, J. K. (2009). Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal*, 3(2):179–189.
- Vezzulli, L., Pezzati, E., Moreno, M., Fabiano, M., Pane, L., and Pruzzo, C. (2009). Benthic ecology of Vibrio spp. and pathogenic Vibrio species in a coastal Mediterranean environment (La Spezia Gulf, Italy). *Microbial Ecology*, 58(4):808–818.

- Vinga, I., São-José, C., Tavares, P., and Santos, M. (2006). Bacteriophage entry in the host cell. In *Modern Bacteriophage Biology and Biotechnology*, pages 163–203. Kerala, INC : Research Signpost.
- Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012). Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle. *PLoS ONE*, 7(8):e43432.
- Waldor, M. K., Colwell, R., and Mekalanos, J. J. (1994). The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24):11388–11392.
- Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N. S., Kim, P. B., Cooper, S. R., Raisner, R. M., Herman, A., Sivitz, A. B., Lakshmanaswamy, A., Kashiya, Y., Baker, D., and Yoshikuni, Y. (2012). An Engineered Microbial Platform for Direct Biofuel Production from Brown Macroalgae. *Science*, 335(6066):308–313.
- Warner, E. and Oliver, J. D. (2008). Population Structures of Two Genotypes of *Vibrio vulnificus* in Oysters (*Crassostrea virginica*) and Seawater. *Applied and Environmental Microbiology*, 74(1):80–85.
- Watnick, P. I., Fullner, K. J., and Kolter, R. (1999). A role for the mannose-sensitive hemagglutinin in biofilm formation by *Vibrio cholerae* El Tor. *Journal of Bacteriology*, 181(11):3606–3609.
- West, P. A., Okpokwasili, G. C., Brayton, P. R., Grimes, D. J., and Colwell, R. R. (1984). Numerical taxonomy of phenanthrene-degrading bacteria isolated from the Chesapeake Bay. *Applied and environmental microbiology*, 48(5):988–993.
- Wetz, J. J., Blackwood, A. D., Fries, J. S., Williams, Z. F., and Noble, R. T. (2008). Trends in total *Vibrio* spp. and *Vibrio vulnificus* concentrations in the eutrophic Neuse River Estuary, North Carolina, during storm events. *Aquatic Microbial Ecology*, 53(1):141.
- Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.*, 75:39–68.
- Williams, H. T. (2013). Phage-induced diversification improves host evolvability. *BMC Evolutionary Biology*, 13(1):17.
- Worden, A. Z., Seidel, M., Smriga, S., Wick, A., Malfatti, F., Bartlett, D., and Azam, F. (2006). Trophic regulation of *Vibrio cholerae* in coastal marine waters. *Environmental Microbiology*, 8(1):21–29.

- Wright, A. C., Simpson, L. M., Oliver, J. D., and Morris, J. G. (1990). Phenotypic evaluation of acapsular transposon mutants of *Vibrio vulnificus*. *Infection and Immunity*, 58(6):1769–1773.
- Wyres, K. L., Gorrie, C., Edwards, D. J., Wertheim, H. F. L., Hsu, L. Y., Van Kinh, N., Zadoks, R., Baker, S., and Holt, K. E. (2015). Extensive Capsule Locus Variation and Large-Scale Genomic Recombination within the *Klebsiella pneumoniae* Clonal Group 258. *Genome Biology and Evolution*, 7(5):1267–1279.
- Xing, M., Hou, Z., Yuan, J., Liu, Y., Qu, Y., and Liu, B. (2013). Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (*Scophthalmus maximus*). *FEMS Microbiology Ecology*, 86(3):432–443.
- Xing, P., Hahnke, R. L., Unfried, F., Markert, S., Huang, S., Barbeyron, T., Harder, J., Becher, D., Schweder, T., Gloeckner, F. O., Amann, R. I., and Teeling, H. (2015). Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME Journal*, 9(6):1410–1422. WOS:000354786700012.
- Yang, Y., Hu, M., Yu, K., Zeng, X., and Liu, X. (2015). Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein & Cell*, 6(4):265–274.
- Yawata, Y., Cordero, O. X., Menolascina, F., Hehemann, J.-H., Polz, M. F., and Stocker, R. (2014). Competition–dispersal tradeoff ecologically differentiates recently speciated marine bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 111(15):5622–5627.
- Ymele-Leki, P., Houot, L., and Watnick, P. I. (2013). Mannitol and the Mannitol-Specific Enzyme IIB Subunit Activate *Vibrio cholerae* Biofilm Formation. *Applied and Environmental Microbiology*, 79(15):4675–4683.
- Yooseph, S., Nealson, K. H., Rusch, D. B., McCrow, J. P., Dupont, C. L., Kim, M., Johnson, J., Montgomery, R., Ferriera, S., and Beeson, K. (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature*, 468(7320):60–66.
- Yoshihara, Y., Kiyosue, T., Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1997). Regulation of Levels of Proline as an Osmolyte in Plants under Water Stress. *Plant and Cell Physiology*, 38(10):1095–1102.
- Yoshimizu, M. and Kimura, T. (1976). Study on the intestinal microflora of salmonids. *Fish pathology*, 10(2):243–259.
- Yu, C., Bassler, B. L., and Roseman, S. (1993). Chemotaxis of the marine bacterium *Vibrio furnissii* to sugars. A potential mechanism for initiating the chitin catabolic cascade. *Journal of Biological Chemistry*, 268(13):9405–9409.

- Zarubin, M., Belkin, S., Ionescu, M., and Genin, A. (2012). Bacterial bioluminescence as a lure for marine zooplankton and fish. *Proceedings of the National Academy of Sciences*, 109(3):853–857.
- Zettler, E. R., Mincer, T. J., and Amaral-Zettler, L. A. (2013). Life in the "plastisphere": microbial communities on plastic marine debris. *Environmental Science & Technology*, 47(13):7137–7146.
- Zhang, Y. J., Ioerger, T. R., Huttenhower, C., Long, J. E., Sasseti, C. M., Sacchettini, J. C., and Rubin, E. J. (2012). Global Assessment of Genomic Regions Required for Growth in *Mycobacterium tuberculosis*. *PLoS Pathog*, 8(9):e1002946.
- Zimmerman, A., DePaola, A., Bowers, J., Krantz, J., Nordstrom, J., Johnson, C., and Grimes, D. (2007). Variability of total and pathogenic vibrio parahaemolyticus densities in Northern Gulf of Mexico water and oysters. *Applied and Environmental Microbiology*, 73(23):7589–7596.