VIE ET MILIEU - LIFE AND ENVIRONMENT, 2015, 65 (4): 187-200

GENOMICS OF AN ENVIRONMENTALLY-TRANSMITTED SYMBIOSIS: NEWLY SEQUENCED *VIBRIO FISCHERI* GENOMES FROM DIFFERENT HOST SQUIDS AND GEOGRAPHIC LOCATIONS

J. H. HOWARD[1], *B. L. WILLIAMS*[2], *M. BREEN*[3], *M. K. NISHIGUCHI* *[4]

[1] *Pennsylvania State University-Abington, Department of Biology, 1600 Woodland Road, Abington, Pennsylvania 19001-3990, USA*
[2] *Utah State University Uintah Basin, Department of Biology, 320 N 2000 W (Aggie Blvd), Vernal, Utah 84078-4228, USA*
[3] *Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, University Rd, Southampton SO17 1BJ, United Kingdom*
[4] *New Mexico State University, Department of Biology, BOX 30001, MSC 3AF, Las Cruces, New Mexico 88003-8001, USA*
** Corresponding author: nish@nmsu.edu*

VIBRIO FISCHERI
GENOME
SYMBIOSIS
BOBTAIL SQUID
EUPRYMNA
SEPIOLA

ABSTRACT. – Environmentally-transmitted symbiotic bacteria must balance selective pressures of host specificity and the abiotic environment. Here, we investigate genomes of three strains of *Vibrio fischeri* isolated from three different squid species living in various temperature and salinity environments. These sequences were compared to others from symbiotic and free-living vibrios for gene orthology, functional subsystem, selection, and phylogenetic analyses. A *V. cf. campbellii* bacterium isolated from seawater also was sequenced and provided an outgroup for the selection and phylogenetic analyses. This investigatory study provides the basis for further directed studies that may elucidate the genetic architecture underlying adaptation to both the free-living environment and host squids in *V. fischeri*. Symbiotic vibrios in this study had relatively conserved genomes with evidence of strong purifying selection and few duplication events. Horizontal gene transfer may provide a mechanism for the acquisition of host and environment-specific genes. Despite spending a majority of time in the free-living stage and a low probability of host-colonization for any one bacterium, symbiotic *V. fischeri* appear to have evolutionary histories molded by their squid hosts.

INTRODUCTION

Environmentally-transmitted symbiotic bacteria must be able to survive both within and external to their specific host. In opposition to vertically transmitted symbioses, hosts are axenic at birth and acquire symbionts from the environment. How symbionts actually balance these two environments has broad implications for pathogen transmission, coevolutionary theory, cooperation, and stability of symbiosis dynamics (Nishiguchi *et al*. 2008, Nyholm & Nishiguchi 2008, Sachs & Simms 2006, Sachs *et al*. 2011). Attempts to map the genetic basis of adaptation to either free-living or host environment have greatly increased with the ability to rapidly sequence whole genomes (Hudson 2008, Nadeau & Jiggins 2010, Stapley *et al*. 2010). For example, genetic signatures for pathogenicity in a *Streptococcus* strain have been identified (Suzuki *et al*. 2011), as have differences in metabolic capabilities in rumen symbionts *Prevotella* (Purushe *et al*. 2010), differences in cold tolerance in *Neurospora* (Ellison *et al*. 2011), and phosphorus acquisition in phosphorus-limited environments in *Pelagibacter* and *Prochlorococcus* (Coleman & Chisholm 2010). Not only can influential point mutations be identified, but whole genome patterns, such as duplication events, expansion or reduction of the core genome, composition of the accessory genome, and rearrangement of regulatory elements can be identified and investigated for their role in adaptation (Gilad *et al*. 2009, Levasseur & Pontarotti 2011, Medina & Sachs 2010).

We investigated an environmentally-transmitted symbiosis between bobtail squids and their *Vibrio* symbionts because of the breadth of host species and environmental conditions, as well as the wealth of prior knowledge available about this symbiotic complex. We examined three strains of *V. fischeri* isolated from different host species of squid from different geographic locations to catalog the genetic diversity of this complex system as a launching point for more in-depth comparative studies that might pinpoint the functional significance of genetic elements involved in adaptation to either host or free-living environment.

*The sepiolid squid-*Vibrio *model system*

Bobtail squids (Cephalopoda: Sepiolidae) harbor bioluminescent bacterial symbionts (γ-Proteobacteria: Vibrionaceae) in a specialized organ that directs light downwards to mask the squids' silhouette from benthic predators (Jones & Nishiguchi 2004). Axenic squid hatchlings are colonized through a "sieve" of host defenses that selects for only coevolved mutualism-competent strains (Nyholm & McFall-Ngai 2004). The squid and their colonists then enter a diel cycle (Boettcher *et al*. 1996), whereby the vibrios first grow and populate the light organ during the day while the squid is quiescent and hidden in the

sand. During the evening the squid emerges and bacterial bioluminescence, which is triggered by quorum sensing that occurs at high bacterial densities (Eberhard 1972, Fuqua *et al*. 1994, Nealson 1977), is directed ventrally by the squid to match downwelling moonlight (Jones & Nishiguchi 2004). Upon sunrise, the squid vents 90 % of the bacteria (Graf & Ruby 1998). The venting process seeds the surrounding seawater with a high number of potential vibrio symbionts capable of colonizing newly-hatched squids (Lee & Ruby 1994, Ruby & Lee 1998). Thus, symbiotic *Vibrio* must navigate both the host and free-living environments.

Coevolution of the symbionts and hosts has influenced colonization success. Some bacterial strains can efficiently colonize multiple squid species, as observed in the Mediterranean Sea where sympatric squid species live (Nishiguchi 2000, Nishiguchi *et al*. 2008, Nyholm & McFall-Ngai 2004, Nyholm & Nishiguchi 2008, Ruby 2008), while others are more host-specific, as demonstrated in the Indo-West Pacific where there is no geographical overlap among squid species (Nishiguchi 2000, Jones *et al*. 2006, Jones *et al*. 2007, Soto *et al*. 2009a, Zamborsky & Nishiguchi 2011). Fine sequence variation in genes required for the mutualism affect host specificity (Chavez-Dozal *et al*. 2014). In competition experiments using *Vibrio* isolates from different squid species, squids show preference for their native bacterial strains, and a hierarchy of competitive dominance exists (Chavez-Dozal *et al*. 2014, Jones *et al*. 2006, Nishiguchi *et al*. 1998, Nishiguchi 2002, Soto *et al*. 2014). Such dominance can be conferred by genes of major effect, which can result in rapid fixation or alter the evolutionary trajectory of a lineage (Nadeau & Jiggins 2010, Orr 2005). For example, introduction of a single regulatory gene (*rscS*) to a strain of *V. fischeri* symbiotic with a fish conferred the ability to colonize a novel host species (bobtail squid); this gene may have played a role in host-switching in this mutualism complex (Mandel *et al*. 2009). Experimentally evolving a strain native to *E. scolopes* in another squid species (*E. tasmanica*) resulted in an increase in colonization efficiency and competitive ability against native *E. tasmanica* strains in under 500 generations (Soto *et al*. 2012). Additionally, experimentally evolving both a free-living strain and a fish symbiotic strain in a squid species leads to a decrease in bioluminescense similar to what native strains exhibit (Schuster *et al*. 2010). In summary, strains appear to be uniquely adapted to their native squid hosts.

Mutualistic strains also illustrate adaptation to their free-living environment. Temperature and salinity tolerance varies among symbiotic strains of vibrios and matches native environmental conditions, which differ geographically (Soto *et al*. 2009a, 2009b). Adaptation to environmental variables can upset host-symbiont dynamics, e.g., native symbionts that are better adapted to colonize their host squid in laboratory settings may not be the most likely to colonize squids in nature, depending on the environmental conditions (Soto *et al*. 2012). In other words, strains that persist in high abundance because of adaptation to the free-living environment can colonize host squids even if other strains illustrated competitive dominance in a laboratory setting (Nishiguchi *et al*. 1998, Nishiguchi 2002, Jones *et al*. 2006, Nyholm & Nishiguchi 2008, Wollenberg & Ruby 2009, Soto *et al*. 2012). Similarly, abiotic factors and geography affect patterns of genetic diversity and colonization competency of *V. cholerae* in human hosts (Keymer *et al*. 2007). For symbionts, arriving at the host early can outweigh weaker colonization ability. In contrast, pleiotropic or epistatic responses in salinity tolerance occurred during experimental evolution in a novel host squid that may or may not be adaptive depending on the environment (Soto *et al*. 2012). Thus, the selective regime of free-living environment clearly produces adaptive responses in symbionts, but confounding variables may maintain the diversity of symbiotic strains that colonize any given host squid and thus also maintain diversity within the population of symbionts.

Study aims

The sepiolid squid-*Vibrio* symbiosis is well suited to explore the genetic scaffolding that enables adaptation to either the host or free-living environment. This well-studied model system has already revealed genetic mechanisms underlying adaptive phenotypes, such as several genes and regulatory elements involved in biofilm formation, which is necessary for host colonization (Yildiz & Visick 2009, Morris & Visick 2010, Chavez-Dozal & Nishiguchi 2011, Chavez-Dozal *et al*. 2012, Chavez-Dozal *et al*. 2014). Some genes and promoters involved in response to nitric oxide (NO) have been found using this model system. NO is a host-derived signal that may function as a population control strategy against the bacterial symbionts (Davidson *et al*. 2004, Wang *et al*. 2010). Other genes involved in circumventing host defenses or facilitating persistence in the free-living environment have been revealed by several transcriptional studies (Chun *et al*. 2008, Guerrero-Ferreira & Nishiguchi 2010, Jones & Nishiguchi 2006, Wier *et al*. 2010). These genet-

Table I. – Strain designations and ecological data for *Vibrio* species whose genomes were sequenced in this study.

Strain	Host	Locality	Temp	Temp °C	Salinity	Salinity ppt
CB37	None	SE Australia	Med	12–25	Med	20.0-35.5
ETJB5C	*Euprymna tasmanica*	SE Australia	Med	12–25	Med	20.0-35.5
EM17	*Euprymna morsei*	Japan	Low	2–17	Low	32.2-34.0
SA1G	*Sepiola affinis*	Mediterranean	Med	12–24	High	37.0-38.0

ic mechanisms appear to play parallel roles in other symbioses as well. Therefore, we used the natural diversity in squid hosts and environments to compare three mutualistic strains from three host squid species, from two temperature ranges, and three salinity levels (Table I). This allowed us to identify unique genes from each genome, differences in functional subsystems, genes currently under selection, point mutations in homologs, and investigate phylogenetic relationships among symbionts. Our findings will enable further studies targeting these points of interest with increased sample sizes and comparative methods focused on one variable (e.g., host, temperature, or salinity).

MATERIALS AND METHODS

Four bacterial genomes were sequenced via 454-pyrosequencing (Supplementary Table I). We chose this platform because longer reads and coverage depth are advantageous for assembly of small genomes like those of bacteria (Hudson 2008, Mardis 2008, Metzker 2010). Three strains of *V. fischeri* were isolated from different geographically located squid hosts (EM17 from *Euprymna morsei* in Tokyo Bay, Japan; ETJB5C from *E. tasmanica* in Jervis Bay, Australia; and SA1G from *Sepiola affinis* in Banyuls-sur-Mer, France; Table I). For comparison, we also sequenced a free-living, non-symbiotically competent *Vibrio* strain, *V. cf. campbellii* (CB37 isolated from Coogee Bay, Australia; Table I) and used it as our outgroup for phylogenetic and selection analyses. High-quality reads were obtained; sequence reads smaller than 40 nucleotides were discarded from analyses (Supplementary Tables I, II).

Pairwise and site-wise comparisons of homologs were completed and core and accessory genomes were identified (orthology analysis; Supplementary Table III). Analysis of these data enabled us to generate a list of genes potentially involved in adaptation to either the symbiotic or free-living state. Comparisons were made with other published genomes for orthology (symbiotic *V. fischeri* strains only) and selection analyses: ES114, a symbiont of *E. scolopes* (Ruby *et al.* 2005); MJ11, a symbiont of the fish *Monocentris japonica* (Mandel *et al.* 2009); SR5, a symbiont of a Mediterranean squid *Sepiola robusta* (Gyllborg *et al.* 2012); *V. campbellii* ATCC BAA-1116, a *Vibrio* strain isolated from a green barrel tunicate (Bassler *et al.* 2007, Lin *et al.* 2010); and two other vibrios, *Photobacterium profundum* 3TCK (Bartlett *et al.* 2006) and *Vibrio* sp. EJY3 (Roh *et al.* 2012). For phylogenetic analyses, we included our four sequenced genomes, and *V. fischeri* SR5 and MJ11.

Orthology analysis of the amino acid sequences from each *Vibrio* strain was performed using OrthoMCL database (Chen *et al.* 2006). Putative proteins encoded by the sequences were categorized as orthologs, paralogs, or orphans. Proteins failing to fall into clusters were analyzed with less stringent parameters (e-value of $1e^{-5}$ and minimum coverage of 50 %) using NCBI's BlastClust package and applying a reciprocal pBlast approach. Clusters were aligned with MUSCLE (Edgar 2004) using default parameters. We identified core clusters, defined as one orthologous sequence per strain, to calculate nucleotide diversity (π) and construct a phylogeny. The parameter Π, measured as the number of nucleotide differences between strains divided by the total length of a nucleotide sequence, was calculated for each set of orthologs in pairwise comparisons. We then averaged the values for all sets of pairwise comparisons to calculate the degree of polymorphism between strains. To visualize results identifying core and accessory genomes, we implemented PERL scripts within the publicly available Venn Diagram tool from VIB / UGent (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Annotated genetic elements were assigned to functional groups such as "Virulence, Disease, and Defense," "Cell Wall and Capsule," and "Stress Response" using Rapid Annotation Using Subsystem Technology (RAST) (Overbeek *et al.* 2014). The number of genes in each subsystem were summarized to highlight differences in allocation to subsystems among our symbionts, a model organism (ES114; (Ruby *et al.* 2005)), and the free-living *Vibrio* strain we sequenced. Annotated genes are provided in Supplementary Table IV.

The proportions of GC content in each open reading frame (ORF) were calculated, including the minimum, maximum, and SD of GC content per genome. Any ORF with GC proportions outside of ± 2 SDs of the mean were noted (Supplementary Table V). Normality of the distribution of GC proportions in the ORFs per genome was examined with Shapiro-Wilk tests.

The Codeml program was implemented from the PAML 4 package (Yang 2007) in a batch mode process on all clusters containing two or more strains to calculate selection ratios, both pairwise and sitewise, on all orthologs. Proteins and sites with Ka/Ks ratios exceeding 1.0 were classified as being under positive selection.

Maximum likelihood phylogenetic analysis was performed on multiple sequence alignments constructed at the protein level then concatenated to form orthologous alignments for the maximum likelihood program PhyML 3.0 (Guindon *et al.* 2010). We adopted the LG amino acid replacement matrix model (Le & Gascuel 2008) and utilized the SPR option, which provides the slowest but most accurate tree-topology search. A bootstrap analysis using 100 replicates also was performed.

RESULTS

Quality of sequence reads

We performed Roche 454 sequencing on the genomes of three *Vibrio fischeri* squid light organ isolates and on a free-living *Vibrio cf. campbellii* strain to increase our understanding of genetic differences that underlie symbionts of different squid species and different marine environments. Genomic DNA was sequenced in NMSU's sequencing facility as described in the Methods. Sequence read lengths of less than 40 nt were discarded. After standard filtering programs were implemented, we obtained good quality reads ranging from 47.9 % (EM17)

Table II. – Strain designations and information for *Vibrio* spp. and a *Photobacterium* sp. whose genomes were obtained from GenBank and used in orthology, positive selection, and sequence comparison analyses.

Species	Strain	Host	Geography	Accession No(s).	Reference
V. fischeri	ES114	*Euprymna scolopes*	Hawaii	NC_006840.2 NC_006841.2 NC_006842.1	(Ruby *et al.* 2005)
V. fischeri	MJ11	*Monocentris japonica*	Sea of Japan	NC_011184.1 NC_011185.1 NC_011186.1	(Mandel *et al.* 2009)
V. fischeri	SR5	*Sepiola robusta*	Mediterranean	CM001400.1 CM001401.1	(Gyllborg *et al.* 2012)
V. harveyi	ATCC BAA-1116	Free-living	Unknown	NC_009783.1 NC_009784.1	(Bassler *et al.* 2007)
Vibrio sp.	EJY3	*Grapsidae* (crab)	South Korea	NC_016613.1 NC_016614.1	(Roh *et al.* 2012)
P. profundum	3TCK	Free-living	San Diego Bay	NZ_AAPH00000000.1	(Bartlett *et al.* 2006)

to 78.4 % (ETJB5C), and averaging 59.0 %, of raw data, with an average read length of 480 nt (Supplementary Table II).

Comparison of Roche 454 sequencing reads to reference genomes

Reference genomes are provided in Table II. The reference genomes used to map sequences from the three *Vibrio fischeri* strains included in this genome-sequencing project were those of *V. fischeri* ES114 (Ruby *et al.* 2005) and *V. fischeri* MJ11 (Mandel *et al.* 2009). Genomic data from *V. campbellii* ATCC BAA-1116 (Lin *et al.* 2010), which originally was identified in GenBank data as a *V. harveyi,* was used as a reference for genome sequencing of the out-group used in analyses, *V. cf. cambellii* CB37. Our three symbiotic strains of *V. fischeri* have an average of 3898 genes, demonstrating slightly reduced genomes compared to an average genome size of 5173 genes found in the other non-*V. fischeri* vibrios used in our analysis (Table II). Overall, genes located on Chromosome I of the symbiotic strains were more conserved than genes located on Chromosome II. On average, the symbiotic strains shared 94.55 % sequence identity with *V. fischeri*'s Chromosome I, while Chromosome II averaged 89.89 %.

The plasmid content carried in the symbiotic genomes is extremely variable. Only 20.06 % of the genes in ES114's plasmid are found in the plasmid of the Australian isolate, ETJB5C; roughly half as many are found in the Mediterranean isolate SA16, and none are present in the plasmid of the Japanese Sea isolate, EM17. Instead, the latter two strains share approximately 90 % of their plasmid content with the fish symbiont (MJ11) plasmid, while a very small percentage (less than 2 %) of ETJB5C's plasmid is the same as in MJ11 (Supplementary Table I).

Orthology analysis

We identified the genes shared among all of the six genomes, thereby identifying the core *V. fischeri* genome. Including our three newly sequenced genomes, as well as the three reference genomes (ES114 (Ruby *et al.* 2005), SR5 (Gyllborg *et al.* 2012), and MJ11 (Mandel *et al.* 2009)), there are 3018 genes common, listed in Supplemental Table III. This represents 78.4 % of the average genome size. Eliminating the MJ11 genome from analysis increases the core genome for squid symbionts to 3091, whereas eliminating SR5 but leaving MJ11 in increases the core genome size to 3138. Our three newly sequenced genomes alone share 3186 orthologous genes. The overlap in genomic content among different groups of *V. fischeri* strains, along with *V. cf. campbellii* CB37, is depicted in Venn diagrams (Fig. 1). Accessory genomes of all six symbiotic strains, which include strain-specific genes as well as orthologs shared by some but not all of the symbiotic strains, contain on average 833 genes, with each strain possessing an average of 226 unique genes. Orthologs found in more than two but less than six genomes also include varying numbers of paralogous genes, ranging from 20 genes that have undergone at least one duplication

Table III. – Nucleotide diversity (π) of our three symbiotic *V. fischeri* strains and free-living strain CB37, including comparisons to previously-sequenced *V. fischeri* strains ES114, MJ11, and SR5.

	SR5	ETJB5C	MJ11	SA1G	EM17	ES114	CB37
SR5	–	0.0296	0.0286	0.0291	0.0277	0.0301	0.2882
ETJB5C		–	0.0344	0.0364	0.0191	0.0185	0.2883
MJ11			–	0.0176	0.0332	0.0354	0.2883
SA1G				–	0.0359	0.0382	0.2883
EM17					–	0.0195	0.2884
ES114						–	0.2885
CB37							–

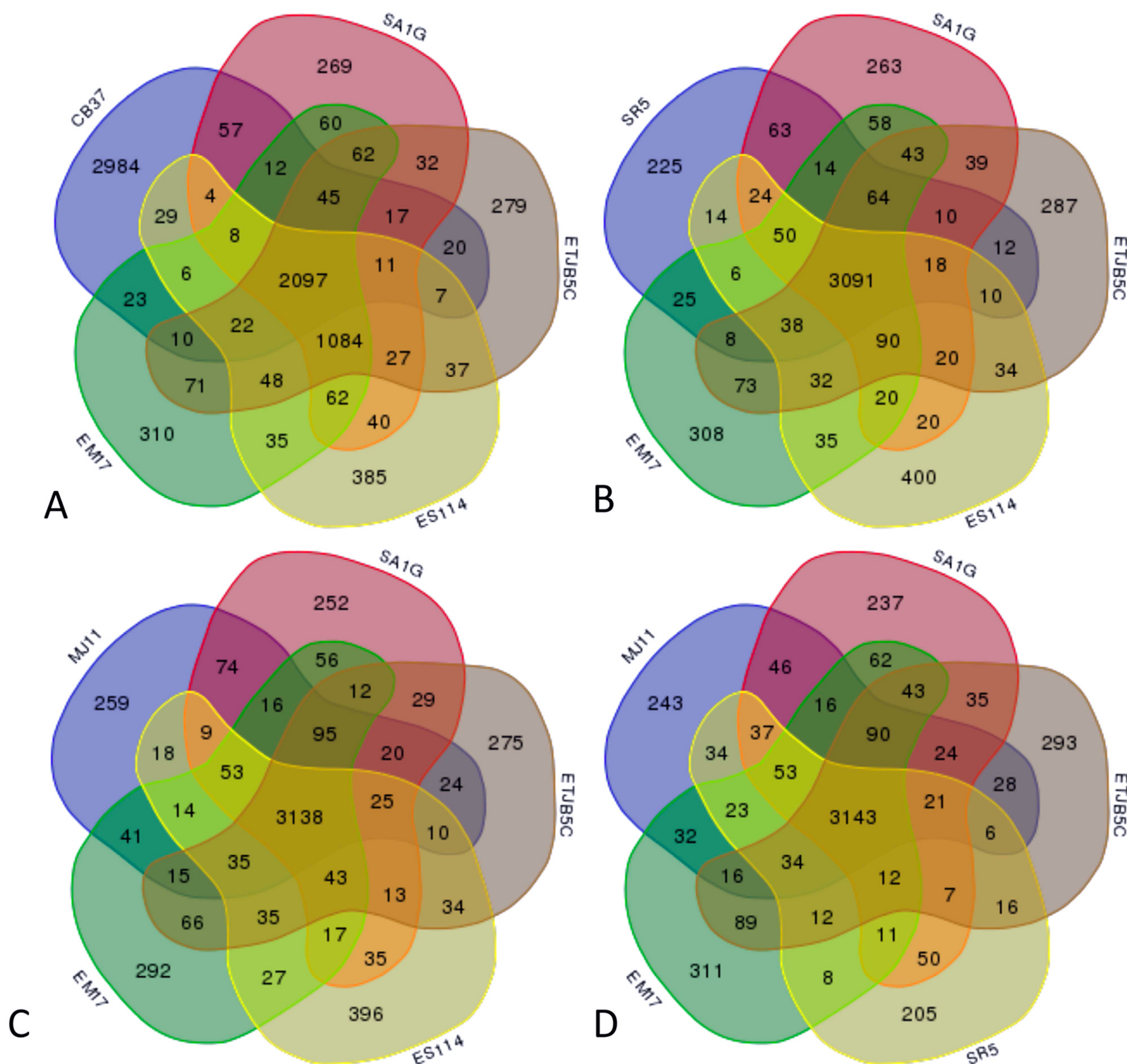


Fig. 1. – Venn diagrams from the Bioinformatics and Evolutionary Genomics Tool depict shared gene content graphically for up to 5 genomes. A. Venn diagram depicting our three sequenced strains of symbiotically competent *V. fischeri* (SA1G, EM17, ETJB5C), a model organism (ES114), and the free-living *V. cf. campbellii* CB37. Here, we compare our free-living strain, CB37, to our symbiotic strains. B. Venn diagram depicting symbiotically competent *V. fischeri* that colonize squid hosts, including our three strains (SA1G, EM17, ETJB5C), a model organism (ES114), and another previously published squid symbiont (SR5). Here, we compare our symbiotic strains to other squid symbiotic strains. C. Venn diagram depicting four strains of symbiotically competent *V. fischeri* with squid hosts (SA1G, EM17, ETJB5C, ES114) and a fish host (MJ11). D. Venn diagram depicting four strains of symbiotically competent *V. fischeri* with squid hosts (SA1G, EM17, ETJB5C, SR5) and a fish host (MJ11). Note that the core genomes when including the fish symbionts (panels C, D) are actually larger (3138 vs. 3143) than the core genome for just the squid symbionts (panel B; 3091), indicating that the fish symbiont is not the most divergent of the symbiotic strains.

event in SR5 to 51 found in EM17 (Supplemental Table VI).

In pairwise comparisons of genetic diversity among the symbiotic *V. fischeri* strains, the greatest difference was seen between SA1G and ES114, with their orthologous genes differing at approximately 3.8 % of nucleotide sites (Table III). The most similar pairing was between SA1G and MJ11, where approximately 1.8 % of their orthologs differed in nucleotide sequence (Table III).

Functional groups analysis

The number of genes assigned to different functional groups varies between symbiotic genomes and the free-living CB37 (RAST analysis; Table IV). Annotations of all specific genes recovered from this RAST analysis are given in Supplementary Table IV.

GC analysis

Distributions of GC proportions by gene for all four genomes sequenced in this study were not normal but were skewed towards ORFs with lower GC content than the genome mean (ETJB5C, D = 0.065, $p < 0.01$; SA1G, D = 0.062, $p < 0.01$; EM17, D = 0.060, $p < 0.01$; CB37, D = 0.088, $p < 0.01$; Table V). A high frequency of genes at the edges of the distribution of GC proportions (more than 2 SDs from the mean) for each genome were spatially clustered; that is, they were proximate in the genome (Table V, Supplementary Table V).

Positive selection analysis

A total of 38 genes show evidence in at least one pair-wise-comparison of positive selection with ratios greater than 1.0: twenty two in EM17, ten in ES114, four in ETJB5C, and one each in MJ11 and SA1G (Table VI).

Phylogenetic analysis

The maximum likelihood approach of PhyML was used to create a phylogeny of our *V. fischeri* strains. We collected 1,870 orthologs from our three symbiotically competent *V. fischeri* strains (EM17, ETJB5C, SA1G), one free-living strain used as an outgroup (CB37), and three other symbiotic *V. fischeri* strains (ES114 (Ruby *et al*. 2005), MJ11 (Mandel *et al*. 2009), SR5 (Gyllborg *et al*. 2012). These orthologs were aligned and concatenated at the protein level to generate multiple sequence alignments. The best-supported tree identified two clades, one displaying EM17 as sister to ETJB5C and ES114, and the other with SR5 as sister to SA1G and MJ11 (Fig. 2), with *V. cf. campbellii* CB37 as the outgroup. High bootstrap values were obtained for all tree nodes.

Table IV. – Variance in genes comprising functional groups of newly sequenced genomes of *Vibrio* species and one model organism *V. fischeri* ES114.

Subsystem feature counts	CB37	ETJB5C	EM17	SA1G	ES114
Amino Acids and Derivatives	451	324	328	329	329
Carbohydrates	576	375	371	414	414
Cell Division and Cell Cycle	33	31	35	33	33
Cell Wall and Capsule	183	205	176	191	191
Cofactors, vitamins, prosthetic groups, pigments	254	202	202	205	205
DNA metabolism	149	134	151	132	132
Dormancy and Sporulation	7	7	4	4	4
Fatty Acids, Lipids, and Isoprenoids	131	118	107	124	124
Iron Acquisition and Metabolism	83	50	55	72	72
Membrane transport	227	194	200	199	199
Metabolism of aromatic compounds	14	4	4	4	4
Miscellaneous	193	175	170	184	184
Motility and chemotaxis	164	107	108	101	101
Nitrogen metabolism	50	38	38	39	39
Nucleosides and nucleotides	110	95	99	94	94
Phages, prophages, transposable elements, plasmids	9	7	19	9	9
Phosphorus metabolism	58	37	38	36	36
Potassium metabolism	58	53	51	52	52
Protein metabolism	270	232	238	241	241
Regulation and cell signaling	119	99	92	93	93
Respiration	185	131	129	130	130
RNA metabolism	220	141	142	141	141
Secondary metabolism	0	0	0	0	0
Stress response	228	185	184	188	188
Sulfur metabolism	30	32	28	29	29
Virulence, disease, and defense	106	76	75	73	73
Total open reading frames	3908	3052	3044	3117	3117

Table V. – Mean GC content within genomes of four *Vibrio* strains sequenced in this study. Distributions were skewed toward more open reading frames (ORFs) with low GC content (those with GC proportions farther than 2 standard deviations (SDs) from the mean). Clusters, defined as more than 3 ORFs in proximity, with high or low GC content are provided in Supplementary Table V.

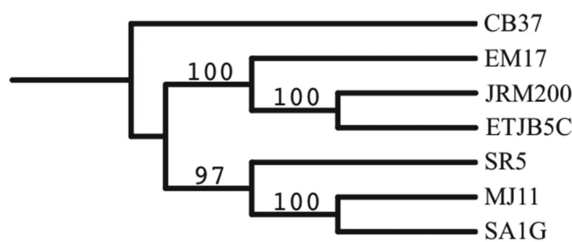
Strain	Mean ORF length (bp)	Mean GC proportion	Total ORFs	MAX GC proportion	MIN GC proportion	SD GC proportion	ORFs GC content above mean	ORFs GC content below mean	Number of ORFs in clusters	Proportion of ORFs in clusters
ETJB5C	952	0.3872	3864	0.5014	0.1795	0.0349	40	156	77	0.3929
SA1G	968	0.3873	3885	0.4987	0.2027	0.0352	42	140	31	0.1703
EM17	953	0.3883	3946	0.5014	0.2068	0.0353	48	146	32	0.1649
CB37	934	0.4535	5340	0.5464	0.2650	0.0318	17	250	78	0.2921

Table VI. – Positively-selected genes identified through pairwise comparative analyses of symbiotic strains.

Strain ID	Compared to:	Gene Annotation	Gene ID	Known or putative function	Reference
ES114	ETJB5C	Hypothetical protein	VF_A0138	Unknown	–
	MJ11	FhuC ferric hydroxamate uptake	VF_A0158	Iron transport	–
	ETJB5C	Hypothetical protein	VF_A0755	Unknown	–
	SA1G	RplP 50S ribosomal subunit protein L16	VF_0243	Translation	–
	SA1G	Crp/Fnr family transcriptional regulator	VF_0318	Environmental stress response	(Soto & Nishiguchi 2014)
	EM17	ProW glycine betaine/proline ABC transporter permease	VF_0786	Osmoprotection	(Chavez-Dozal *et al.* 2014)
	SA1G	Hypothetical protein	VF_1745	Unknown	–
	ETJB5C	Hypothetical protein	VF_1916	Unknown	–
	ETJB5C	AtpI ATP synthase subunit I	VF_2571	Respiration	–
	SA1G	PanB pantothenate hydroxymethyltransferase	VF_2169	Pantothenate biosynthesis (FA, TCA metabolism)	(Jones *et al.* 1993)
ETJB5C	MJ11	ProtC transcriptional regulatory protein	277	Possible antibiotic resistance	(Wietzorrek & Bibb 1997)
	EM17	PriA primosome assembly protein	665	Transcription	–
	ES114	Hypothetical protein	2268	Unknown	–
	EM17; ES114	Phage shock protein	3285	Possible extracytoplasmic stress response	(Maxson & Darwin 2004)
EM17	MJ11	Glutathione S-transferase	190	Detoxification	(Vuilleumier 1997)
	ES114	RpoZ DNA-directed RNA polymerase subunit omega	691	Transcription	–
	ES114	ProQ solute/DNA competence effector	933	Osmoprotectant regulation	(Browne-Silva & Nishiguchi 2008)
	ES114	Hypothetical protein	1289	Unknown	–
	SA1G	Hypothetical protein	1433	Unknown	–
	ETJB5C	Hypothetical protein	1936	Unknown	–
	ETJB5C	Surface protein	1992	Unknown	–
	ES114	Hypothetical protein/possible peptidase	2476	Unknown	–
	ES114; ETJB5C	RpsU 30S ribosomal protein S21	2576	Translation	–
	ETJB5C	PilP pili assembly protein	2620	Conjugation/twitching motility	–
	ES114; ETJB5C	tRNA dihydrouridine synthase B	2709	Translation	–
	ETJB5C	RplU 50S ribosomal protein L21	2794	Translation	–
	ES114	RaiA stationary phase translation inhibitor and ribosome stability factor	3022	Translation	–
	ETJB5C	Type VI secretion system lysozyme-related protein	3088	Possible competition role	(Soto *et al.* 2014)
	ETJB5C	Putative penicillin-binding protein 1C	3124	Antibiotic resistance	–
	ES114	RpsF 30S ribosomal protein S6	3179	Translation	–
	ETJB5C	Hypothetical protein	3250	Unknown	–
	ETJB5C	Hypothetical protein	3804	Unknown	–
	SA1G	RhlE RNA helicase	1433	Transcription	–

Table VI. – Continued.

Strain ID	Compared to:	Gene Annotation	Gene ID	Known or putative function	Reference
	MJ11	Putative lipoprotein	1644	Membrane function/ environmental sensing	–
	ETJB5C	DksA RNA polymerase binding protein	2388	Transcription	–
	ETJB5C	MshO mannose-sensitive hemagglutinin type IV pilus assembly	2722	Host colonization	–
SA1G	MJ11	PilZ Type IV pilus assembly protein	3870	Motility/biofilm formation	(Mattick 2002)
MJ11	ETJB5C	Beta-lactamase	4686	Antibiotic resistance	(Ambler 1980)



Key: CB37 is a free-living *Vibrio harveyi* collected in Australia; EM17 is a *Vibrio fischeri* isolated from *Euprymna morsei* from Japan; JRM200 is a *V. fischeri* isolated from *E. scolopes*, Hawaii; ETJB5C is a *V. fischeri* isolated from *E. tasmanica*, Australia; SR5 is a *V. fischeri* isolated from *Sepiola robusta* from the Mediterranean Sea; MJ11 is a *V. fisheri* fish symbiont isolated from *Monocentris japonica*, Japan; and, SA1G is a *V. fischeri* isolated from *S. affinis* from the Mediterranean Sea.

Fig. 2. – Maximum-likelihood bootstrap consensus tree constructed using PhyML 3.0 and showing relationships among *V. fischeri* strains sequenced in this study. The tree was generated using concatenated amino acid sequences from 1870 orthologous genes shared by all strains. We adopted the LG amino acid replacement matrix model (Le & Gascuel 2008) and utilized the SPR option, which provides the slowest but most accurate tree-topology search. Bootstrap values for nodes were generated using 100 replications.

DISCUSSION

Overview

The plethora of genes divergent among our three symbiotic strains potentially reflect different environmentally-selective regimes and different hosts as well as several divergent between all symbiotic and obligately free-living vibrios (Table III). Rapid Annotation using Subsystem Technology (RAST) uses its own algorithms to discover open reading frames and bin them into functional groups/ subsystems (Overbeek *et al*. 2014); different strains have different numbers of genes in each subsystem, which may be indicative of varying environments. Genes under positive selection were also detected and may encompass both host- and free-living effects (Table VI). Within the three symbiotic strains examined, genomes are relatively conserved with evidence of strong purifying selection and few duplication events. Interestingly, a reduced genome size, compared to that of other vibrios we used as reference (mean 3898 vs. 5173 genes), may be indicative of tradeoffs to allow rapid reproduction within host light organs, despite the fact that *V. fischeri* must cycle through a free-living stage. This is common in vertically-transmitted, obligate mutualisms, where bacterial symbionts are streamlined to the specific capabilities that are encompassed during symbiosis (Moran & Mira 2001, Moran 2002, Toh *et al*. 2006, McCutcheon & Moran 2012).

Symbiotic strains demonstrate substantial variation in their accessory genomes presumably due to the necessity of adapting to both the specific host species and unique abiotic environment faced by each *V. fischeri* strain (Fig. 1). Some of these may be acquired via horizontal gene transfer (HGT) as suggested by GC signatures that deviate significantly from mean genome proportions (Table V). The proximity or clustering of ORFs with similar GC frequencies that deviate > 2 SDs from the mean of the genome illustrates that genes with lower than average GC content are not randomly distributed throughout the genome. These blocks of ORFs may be of interest to assess HGT, which may be indicated by either abnormally high or low GC content (Garcia-Vallvé *et al*. 2000, Marcus *et al*. 2000), and which often occurs in the form of pathogenicity islands where several ORFs of an operon are transferred in one event. However, other factors may influence GC content (Hildebrand *et al*. 2010, Hayek 2013), and a further analysis of these blocks is warranted and should be addressed in a separate, more in-depth analysis. Other forces beyond vertical inheritance are also indicated by the phylogenetic analysis, which lacks a biogeographic signature (Fig. 2). Each symbiont's genome contains unique elements that may be shaped by selection from both host and environment; details are summarized in the sections below.

Genome level summary

Our three symbiotic strains of *V. fischeri* exhibited 3,186 orthologous genes (approx. 81.7 %), which comprise the core genome. Our strains combined with two other published squid symbionts had a core genome of 3,091 orthologs; when fish symbiont MJ11 is added to the analysis, there are 3018 (approx. 78 %) shared orthologs (Fig. 1). In comparison, four strains of *V. vulnificus* shared 3,459 genes (Gulig *et al*. 2010), possibly suggesting that pathogenicity in vibrio strains requires a slightly larger suite of conserved genes than does mutualism or

that adaptation to different host squids caused more divergence in our mutualistic genomes.

The accessory genomes uncovered in this study, which are by definition everything exclusive of the core genome, contain an average 833 genes, with each strain possessing an average of 226 unique genes; the remainder were paralogs. The strain-specific genes have no close homologs (> 60 % similarity over 50 % of length) in any of the other five *V. fischeri* strains we examined. Accessory genomes have been hypothesized to be reflective of environmental and host heterogeneity and to equip each strain with the suite of genes most adaptive to its unique ecological niche (Read & Ussery 2006, Mira *et al*. 2010). Strain-specific genes are postulated to arise either from gene duplication, leading to divergence and evolution of new or altered functions, or from lateral gene transfer, in which novel genes are appropriated from other organisms (Zhang 2003, Gevers *et al*. 2004, Treangen *et al*. 2009). However, a more recent analysis found that gene duplication may have a greater effect on gene dosage and less effect on gene neofunctionalization for shaping bacterial genomes (Treangen & Rocha 2011). The strain-specific genes identified through this study are keys to deciphering the relative significance of environmental and host-driven parameters in future studies of *V. fischeri* evolution.

In contrast to findings of significant gene duplication and lineage-specific expansion of protein families within Vibrionaceae (Gu *et al*. 2009), we saw little evidence of widespread gene duplication within symbiotic *V. fischeri* strains. *Vibrio fischeri* EM17 had the most paralogs contained in both its core and accessory genomes (260), while *V. fischeri* SR5 had the fewest (171). The average number of paralogs was 215, representing 5.6 % of the average genome size. Some duplicated genes within each symbiotic genome may have adaptive significance distributed between host and environmental effects, e.g., quadruplets of anaerobic glycerol subunits in SA1G shown to be upregulated during symbiosis (Wier *et al*. 2010) and duplicates of cold shock proteins in MJ11 possibly retained because of environmental effects (Supplementary Table VI).

The proportion of open reading frames associated with different subsystems as assigned during RAST analysis further demonstrates the evolutionary divergence between free-living and symbiotic strains. For example, the free-living *V. cf. campbellii* CB37 has a much larger genome, including many more coding regions associated with nutrient and energy acquisition (Table IV). The symbiotic genomes, in contrast, are streamlined and fairly conserved among each other. In addition, there are fewer elements associated with virulence, disease, and defense in the symbiotic genomes. This analysis showed conservation among the symbiotic genomes in broad functional categories, though some differences are apparent. "Iron acquisition and metabolism" may be limiting in SA1G and ES114 in comparison to ETJB5C and EM17 (Table IV). For example, the former two strains have genes for paraquat-inducible protein A (heme and hemin uptake), while the latter strains lack this gene (Supplementary Table IV). Both SA1G and ES114 have 17 and 22 more genes for "iron acquisition and metabolism" respectively, than EM17, and 22 and 27 more than ETJB5C (Supplementary Table IV). In contrast, the free-living *Vibrio* CB 37 had a total of 83 genes in that subcategory, suggesting that strain experiences the highest selection pressure to sequester a rare resource.

To highlight current levels of natural selection operating on symbiotic *V. fischeri* strains, we calculated ratios of non-synonymous to synonymous (K_a/K_s) substitutions among all possible pairwise comparisons of orthologous proteins. Positive selection was indicated for twenty-two genes in EM17, ten in ES114, four in ETJB5C, and one each in MJ11 and SA1G (Table VI). However, out of a total 19,762 pairwise comparisons, the vast majority of these exhibit very low K_a/K_s ratios, suggesting strong purifying selection occurring on most of the orthologous proteins. Many of the genes exhibiting evidence of positive selection are hypothetical proteins, with no known function. However, the positively-selected genes for which annotations are available include some that are likely influential in adaptation to both host and environment, including mannose-sensitive hemagglutinin *mshO* in *V. fischeri* EM17, expressed during host infection and colonization, and, in *V. fischeri* SA1G, type IV pilus assembly gene *pilZ*, important for motility and biofilm formation (Mattick 2002); see Table VI).

Comparisons with other *Vibrio* species

By comparing our sequence database with genomic sequences available on GenBank of *Vibrio fischeri* and six other vibrios (Table II), we were able to identify genes that are strongly conserved in the symbiotic strains but highly divergent or missing in free-living or pathogenic strains. These genes potentially represent core "symbiotic" genes that are indispensable for host colonization and are inviting candidates for mutational analyses to decipher specific functions that enable *V. fischeri*'s beneficial associations with eukaryotic hosts. Some of these conserved genes include sequences from MJ11, the fish symbiont, while others only include the squid symbionts and may be specific to this host type.

Among the set of conserved orthologs that include MJ11 sequences is *asc1B*, which encodes arylsulfatase regulator and has been shown to be important in other bacterial colonization events (Cheng *et al*. 1992, Morgan *et al*. 2004). Genes in which the MJ11 sequence varies by just one codon include *ntrC*, a gene encoding nitrogen regulatory response regulator/sigma 54 interaction protein that in *V. vulnificus* is involved in membrane saccharide synthesis, biofilm formation, and possibly carbohydrate metabolism (Kim *et al*. 2007) and in *V. fischeri*

is somehow involved in colonization (Hussa *et al.* 2007); and *mshJ*, part of the mannose-sensitive hemagglutinen type IV pilus operon. These genes have functions in the colonization and eventual formation of biofilm, which is crucial for the vibrio community that is established inside the squid light organ (Yip *et al.* 2005, Visick *et al.* 2007, Browne-Silva & Nishiguchi 2008, Geszvain & Visick 2008, Ariyakumar & Nishiguchi 2009, Yildiz & Visick 2009, Chavez-Dozal *et al.* 2012).

Finally, genes conserved within squid symbionts but more divergent in MJ11 include *aepA* encoding an exoenzyme regulatory protein with a putative hydrolase function (Murata *et al.* 1994) and a NADH-dependent flavin oxireductase gene involved in producing the substrate for luciferase (Duane & Hastings 1975). In the phytopathogen *Erwinia carotovora*, *aepA* is necessary for host infection and is up-regulated in response to quorum-sensing signals. Flavin oxireductase, which is instrumental for bioluminescence, has significant sequence divergence from the other *Vibrio* symbionts and is particularly divergent in SA1G, with otherwise close sequence similarity to MJ11. Presumably there are functional attributes of bioluminescence shared between SA1G and MJ11 but distinct from the other symbiotic strains. Whether these attributes relate to an ecological component common between SA1G and MJ11 or is phylogenetic convergence is not presently known.

Phylogenetic analysis

Phylogenetic analysis of the six symbiotic *V. fischeri* strains, using CB37 as the outgroup, revealed a pattern inconsistent with geography, host, or environmental parameters alone and suggests a complex evolutionary history within *V. fischeri* (Fig. 2). One clade consisted of EM17 as sister to ETJB5C and ES114, while a second clade contained SR5 as sister to SA1G and MJ11. These relationships do not have strictly biogeographical signatures; for example, both EM17 and MJ11 are Sea of Japan isolates, yet fall into separate clades. Similarly, the Mediterranean squid symbiotic strain SA1G is more closely related to MJ11 than it is to the other Mediterranean strain SR5. Presumably, different ecological constraints have imposed isolating barriers between *V. fischeri* populations geographically close to one another, despite their hosts sharing multiple species of *Vibrio* bacteria (Fidopiastis *et al.* 1998, Nishiguchi 2000, Zamborsky & Nishiguchi 2011).

Certainly, more extensive taxon sampling might help elucidate evolutionary relationships among symbiotic *V. fischeri* strains, but what our analysis suggests is that divergence of orthologous proteins follows neither clear biogeographical routes nor obvious environmental factors. Instead, evolution of this group of bacteria may operate in a fragmented manner, similar to what has been found in other groups of closely related bacteria, where evolutionary independence of orthologous genes contributes to patchy retention of genes acquired through homologous recombination with proximal populations of *V. fischeri* (Retchless & Lawrence 2010).

CONCLUSION

A selective balance presumably exists between environmental and biogeographical factors that shape the evolution of *V. fischeri* symbiotic strains, likely creating a selection mosaic dependent upon specific interactions between the bacteria, their hosts, and heterogenous environments (Fierer & Jackson 2006, Horner-Devine *et al.* 2004, Yannarell & Triplett 2005). Other researchers have demonstrated similar findings when examining microbial distributions and evolutionary relationships. For example, bacterial communities are driven primarily by environmental heterogeneity rather than geographic distance (Horner-Devine *et al.* 2004); diversity in soil bacterial communities is primarily controlled by edaphic factors (Fierer & Jackson 2006); and, bacterial community composition in shallow lake systems is strongly influenced by local environmental factors (Yannarell & Triplett 2005). The identification of genes that are unique to each strain used in our study, as well as those that exhibit positive selection signatures, is the first step toward disentangling the respective roles that the divergent environments within and outside the host have had on *V. fischeri* evolution.

However, the process of teasing apart environmental and host effects is complicated by the fact that many genes have multiple functions or are upregulated under numerous scenarios. For example, EM17 3088, a Type VI secretion system lysozyme-related protein, could potentially be used against competitors during host colonization (Brooks *et al.* 2013) or against predators in the water column (Pernthaler 2005). In the legume-rhizobia symbiosis, fully one-third of known symbiosis genes are involved in multiple pathways connected to cell metabolism, transcription, signal transduction, and protein modification and regulation (Tian *et al.* 2012). A similar cross-functionalization of symbiotic genes is likely to be valid in the squid-*Vibrio* mutualism.

In summary, *Vibrio fischeri* evolution is likely a patchwork of host- and abiotic environment-driven adaptation. Future work needs to employ comparative methods sampling more strains that differ in only one variable (host, temperature, salinity) but targeting one or a few of the genes/genetic elements we have identified here. While genomic and transcriptional studies have been extremely informative for this symbiosis complex thus far, the proposed strategy will maximize sequencing efficiency and provide more comparative power than previously available.

ACKNOWLEDGEMENTS. – Data are archived in GenBank as BioProject IDs PRJNA212805 (CB37); PRJNA212806 (EM17);

PRJNA212810 (SA1G); and PRJNA212811 (ETJB5C)). We would like to thank NMSU's 454 sequencing staff for help with this project, in particular Dr A Tchourbanov for his help with genome assemblies and Dr A Chavez-Dozal for the DNA preparations. We thank Dr D Houstan for comments on an earlier draft. This research was funded by grants to MKN from the National Institutes of Health NIAID 1SC1AI081659-01 and NIAID ARRA 3SC1AI081659-02S1, and National Science Foundation IOS-074449.

REFERENCES

Ambler RP 1980. The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci* 289(1036): 321-331.

Ariyakumar DS, Nishiguchi MK 2009. Characterization of two host-specific genes, mannose-sensitive hemagglutinin (*mshA*) and uridyl phosphate dehydrogenase (UDPDH) that are involved in the *Vibrio fischeri-Euprymna tasmanica* mutualism. *Fems Microbiol Lett* 299(1): 65-73.

Bartlett DH, Valle G, Lauro FM, Vezzi A, Simonato F, Eloe E, Vitulo N, Stratton TK, D'angelo M, Ferriera S, Johnson J, Kravitz S, Beeson K, Sutton G, Rogers Y, Friedman R, Frazier M, Venter JC 2006. *Photobacterium profundum* 3TCK, whole genome shotgun sequencing project. J Craig Venter Institute, Rockville, MD 20850, USA.

Bassler B, Clifton SW, Fulton L, Delehaunty K, Fronick C, Harrison M, Markivic C, Fulton R, Tin-Wollam AM, Shah N, Pepin K, Nash W, Thiruvilangam P, Bhonagiri V, Waters C, Tu KC Irgon J, Wilson RK 2007. The *Vibrio harveyi* Genome Sequencing Project. Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA.

Boettcher KJ, Ruby EG, McFall-Ngai MJ 1996. Bioluminescence in the symbiotic squid *Euprymna scolopes* is controlled by a daily biological rhythm. *J Comp Physiol A Sens Neural Behav Physiol* 179(1): 65-73.

Brooks TM, Unterweger D, Bachmann V, Kostiuk B, Pukatzki S 2013. Lytic activity of the *Vibrio cholerae* type VI secretion toxin VgrG-3 is inhibited by the antitoxin TsaB. *J Biol Chem* 288(11): 7618-7625.

Browne-Silva J, Nishiguchi M 2008. Gene sequences of the pil operon reveal relationships between symbiotic strains of *Vibrio fischeri*. *Int J Syst Evol Microbiol* 58(6): 1292-1299.

Chavez-Dozal A, Hogan D, Gorman C, Quintanal-Villalonga A, Nishiguchi MK 2012. Multiple *Vibrio fischeri* genes are involved in biofilm formation and host colonization. *FEMS Microbiol Ecol* 81(3): 562-573.

Chavez-Dozal AA, Gorman C, Lostroh CP, Nishiguchi MK 2014. Gene-swapping mediates host specificity among symbiotic bacteria in a beneficial symbiosis. *PLoS ONE* 9(7): e101691.

Chavez-Dozal A, Nishiguchi MK 2011. Variation in biofilm formation among symbiotic and free-living strains of *Vibrio fischeri*. *J Basic Microbiol* 51(5): 452-458.

Chen F, Mackey AJ, Stoeckert CJ, Jr Roos DS 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34(Database issue): D363-368.

Cheng Q, Hwa V, Salyers AA 1992. A locus that contributes to colonization of the intestinal tract by *Bacteroides thetaiotaomicron* contains a single regulatory gene (*chuR*) that links 2 polysaccharide utilization pathways. *J Bacteriol* 174(22): 7185-7193.

Chun CK, Troll JV, Koroleva I, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE, Bonaldo MD, Casavant TL, Soares MB, Ruby EG, McFall-Ngai MJ 2008. Effects of colonization, luminescence, and autoinducer on host transcription during development of the squid-*Vibrio* association. *Proc Nat Acad Sci USA* 105(32): 11323-11328.

Coleman ML, Chisholm SW 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Nat Acad Sci USA* 107(43): 18634-18639.

Davidson SK, Koropatnick TA, Kossmehl R, Sycuro L, McFall-Ngai MJ 2004. NO means 'yes' in the squid-*Vibrio* symbiosis: nitric oxide (NO) during the initial stages of a beneficial association. *Cell Microbiol* 6(12): 1139-1151.

Dong TG, Ho BT, Yoder-Himes DR, Mekalanos JJ 2013. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *PNAS* 110(7): 2623-2628.

Duane W, Hastings JW 1975. Flavin mononucleotide reductase of luminous bacteria. *Mol Cell Biochem* 6(1): 53-64.

Eberhard A 1972. Inhibition and activation of bacterial luciferase synthesis. *J Bacteriol* 109(3): 1101-1105.

Edgar RC 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res* 32(5): 1792-1797.

Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *PNAS USA* 108(7): 2831-2836.

Fidopiastis PM, von Boletzky S, Ruby EG 1998. A new niche for *Vibrio logei*, the predominant light organ symbiont of squids in the genus *Sepiola*. *J Bacteriol* 180(1): 59-64.

Fierer N, Jackson RB 2006. The diversity and biogeography of soil bacterial communities. *PNAS USA* 103(3): 626-631.

Fuqua WC, Winans SC, Greenberg EP 1994. Quorum sensing in bacteria - the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* 176(2): 269-275.

Garcia-Vallvé S, Romeu A, Palau J 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res* 10(11): 1719-1725.

Geszvain K, Visick KL 2008. The hybrid sensor kinase RscS integrates positive and negative signals to modulate biofilm formation in *Vibrio fischeri*. *J Bacteriol* 190(13): 4437-4446.

Gevers D Vandepoele K Simillion C Van de Peer Y 2004. Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol* 12(4): 148-154.

Gilad Y, Pritchard JK, Thornton K 2009. Characterizing natural variation using next-generation sequencing technologies. *Trends Genet* 25(10): 463-471.

Graf J, Ruby EG 1998. Host-derived amino acids support the proliferation of symbiotic bacteria. *PNAS USA* 95(4): 1818-1822.

Gu JY, Neary J, Cai H, Moshfeghian A, Rodriguez SA, Lilburn TG, Wang YF 2009. Genomic and systems evolution in Vibrionaceae species. *BMC Genomics* 10(Suppl 1): S11-24.

Guerrero-Ferreira RC, Nishiguchi MK 2010. Differential gene expression in bacterial symbionts from loliginid squids demonstrates variation between mutualistic and environmental niches. *Environ Microbiol Rep* 2(4): 514-523.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3): 307-321.

Gulig PA, de Crecy-Lagard V, Wright AC, Walts B, Telonis-Scott M, McIntyre LM 2010. SOLiD sequencing of four *Vibrio vulnificus* genomes enables comparative genomic analysis and identification of candidate clade-specific virulence genes. *BMC Genomics* 11: 512.

Gyllborg MC, Sahl JW, Cronin DC 3rd, Rasko, DA Mandel MJ 2012. Draft genome sequence of *Vibrio fischeri* SR5, a strain isolated from the light organ of the Mediterranean squid *Sepiola robusta*. *J Bacteriol* 194(6): 1639.

Haardt M, Bremer E 1996. Use of phoA and lacZ fusions to study the membrane topology of ProW, a component of the osmoregulated ProU transport system of *Escherichia coli*. *J Bacteriol* 178(18): 5370-5381.

Hayek N 2013. Lateral transfer and GC content of bacterial resistance genes. *Front Microbiol* 4: 41.

Hildebrand F, Meyer A, Eyre-Walker A 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6(9): e1001107.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJ 2004. A taxa-area relationship for bacteria. *Nature* 432(7018): 750-753.

Hudson ME 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Resour* 8(1): 3-17.

Hussa EA, O'Shea, TM Darnell, CL Ruby, EG Visick KL 2007. Two-component response regulators of *Vibrio fischeri*: identification, mutagenesis, and characterization. *J Bacteriol* 189(16): 5825-5838.

Jones BW, Nishiguchi MK 2004. Counterillumination in the hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca: Cephalopoda). *Mar Biol* 144(6): 1151-1155.

Jones BW, Lopez JE, Huttenburg J, Nishiguchi MK 2006. Population structure between environmentally transmitted vibrios and bobtail squids using nested clade analysis. *Mol Ecol* 15(14): 4317-4329.

Jones BW, Nishiguchi MK 2006. Differentially expressed genes reveal adaptations between free-living and symbiotic niches of *Vibrio fischeri* in a fully established mutualism. *Can J Microbiol* 52(12): 1218-1227.

Jones BW, Maruyama A, Ouverney CC, Nishiguchi MK 2007. Spatial and temporal distribution of the Vibrionaceae in coastal waters of Hawaii, Australia, and France. *Microb Ecol* 54(2): 314-323.

Jones CE, Brook JM, Buck D, Abell C, Smith AG 1993. Cloning and sequencing of the *Escherichia coli* panB gene, which encodes ketopantoate hydroxymethyltransferase, and overexpression of the enzyme. *J Bacteriol* 175(7): 2125-2130.

Keymer DP, Miller MC, Schoolnik GK, Boehm AB 2007. Genomic and phenotypic diversity of coastal *Vibrio cholerae* strains is linked to environmental factors. *Appl Environ Microbiol* 73(11): 3705-3714.

Kim HS, Lee MA, Chun SJ, Park SJ, Lee KH 2007. Role of *ntrC* in biofilm formation via controlling expression of the gene encoding an ADP-glycero-manno-heptose-6-epimerase in the pathogenic bacterium, *Vibrio vulnificus*. *Mol Microbiol* 63(2): 559-574.

Le SQ Gascuel O 2008. An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7): 1307-1320.

Lee KH, Ruby EG 1994. Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. *Appl Environ Microbiol* 60(5): 1565-1571.

Levasseur A, Pontarotti P 2011. The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics. *Biol Direct* 6: 11.

Lin B, Wang Z, Malanoski AP, O'Grady EA, Wimpee CF, Vuddhakul V, Alves N Jr, Thompson FL, Gomez-Gil B Vora GJ 2010. Comparative genomic analyses identify the *Vibrio harveyi* genome sequenced strains BAA-1116 and HY01 as *Vibrio campbellii*. *Environ Microbiol Rep* 2(1): 81-89.

Mandel MJ, Wollenberg MS, Stabb EV, Visick KL, Ruby EG 2009. A single regulatory gene is sufficient to alter bacterial host range. *Nature* 458(7235): 215-218.

Marcus SL, Brumell JH, Pfeifer CG, Finlay BB 2000. *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes Infect* 2(2): 145-156.

Mardis ER 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet* 24(3): 133-141.

Mattick JS 2002. Type IV pili and twitching motility. *Annu Rev Microbiol* 56: 289-314.

Maxson ME, Darwin AJ 2004. Identification of inducers of the *Yersinia enterocolitica* phage shock protein system and comparison to the regulation of the RpoE and Cpx extracytoplasmic stress responses. *J Bacteriol* 186(13): 4199-4208.

McCutcheon JP, Moran NA 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10(1): 13-26.

Medina M, Sachs JL 2010. Symbiont genomics, our new tangled bank. *Genomics* 95(3): 129-137.

Metzker ML 2010. Sequencing technologies - the next generation. *Nat Rev Genet* 11(1): 31-46.

Mira A, Martin-Cuadrado AB, D'Auria G, Rodriguez-Valera F 2010. The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol* 13(2): 45-57.

Moran NA, Mira A 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* 2(12): RESEARCH0054.

Moran NA 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108(5): 583-586.

Morgan E, Campbell JD, Rowe SC, Bispham J, Stevens MP, Bowen AJ, Barrow PA, Maskell DJ, Wallis TS 2004. Identification of host-specific colonization factors of *Salmonella enterica* serovar *Typhimurium*. *Mol Microbiol* 54(4): 994-1010.

Morris AR, Visick KL 2010. Control of biofilm formation and colonization in *Vibrio fischeri*: a role for partner switching? *Environ Microbiol* 12(8): 2051-2059.

Murata H, Chatterjee A, Liu Y, Chatterjee AK 1994. Regulation of the production of extracellular pectinase, cellulase, and protease in the soft rot bacterium *Erwinia carotovora* subsp. *carotovora*: evidence that *aepH* of *E. carotovora* subsp. *carotovora* 71 activates gene expression in *E. carotovora* subsp. *carotovora*, *E. carotovora* subsp. *atroseptica*, and *Escherichia coli*. *Appl Environ Microbiol* 60(9): 3150-3159.

Nadeau NJ, Jiggins CD 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet* 26(11): 484-492.

Nealson KH 1977. Autoinduction of bacterial luciferase. Occurrence, mechanism and significance. *Arch Microbiol* 112(1): 73-79.

Nishiguchi MK, Ruby EG, McFall-Ngai MJ 1998. Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in sepiolid squid-vibrio symbioses. *Appl Environ Microbiol* 64(9): 3209-3213.

Nishiguchi MK 2000. Temperature affects species distribution in symbiotic populations of Vibrio spp. *Appl Environ Microbiol* 66(8): 3550-3555.

Nishiguchi MK 2002. Host recognition is responsible for symbiont composition in environmentally transmitted symbiosis. *Microb Ecol* 44(1): 10-18.

Nishiguchi MK, DeVinney R, Hirsch AM, Riley M, Mansky L, Vendatum G 2008. Perspective: Evolution of Virulence: Deciphering the mechanisms between pathogenic and benign symbioses. *Vie Milieu* 58(2): 87-106.

Nyholm SV, McFall-Ngai MJ 2004. The winnowing: Establishing the squid-*Vibrio* symbiosis. *Nat Rev Microbiol* 2(8): 632-642.

Nyholm SV, Nishiguchi MK 2008. The evolutionary ecology of a sepiolid squid-*Vibrio* association: From cell to environment. *Vie Milieu* 58(2): 175-184.

Orr HA 2005. The genetic theory of adaptation: A brief history. *Nat Rev Genet* 6(2): 119-127.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S Parrello B, Shukla M 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42(D1): D206-D214.

Pernthaler J 2005. Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* 3(7): 537-546.

Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM, Henrissat B, Nelson KE 2010. Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. *Microb Ecol* 60(4): 721-729.

Read TD, Ussery DW 2006. Opening the pan-genomics box. *Curr Opin Microbiol* 9(5): 496-498.

Retchless AC, Lawrence JG 2010. Phylogenetic incongruence arising from fragmented speciation in enteric bacteria. *PNAS USA* 107(25): 11453-11458.

Roh H, Yun EJ, Lee S, Ko HJ, Kim S, Kim BY, Song H, Lim KI, Kim KH, Choi IG 2012. Genome sequence of *Vibrio* sp. strain EJY3, an agarolytic marine bacterium metabolizing 3,6-anhydro-L-galactose as a sole carbon source. *J Bacteriol* 194(10): 2773-2774.

Ruby EG, Lee KH 1998. The *Vibrio fischeri-Euprymna scolopes* light organ association: Current ecological paradigms. *Appl Environ Microbiol* 64(3): 805-812.

Ruby EG, Urbanowski M, Campbell J, Dunn A, Faini M, Gunsalus R, Lostroh P, Lupp C McCann J, Millikan D, Schaefer A, Stabb E, Stevens A, Visick K, Whistler C, Greenberg EP 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. *PNAS USA* 102(8): 3004-3009.

Ruby EG 2008. Symbiotic conversations are revealed under genetic interrogation. *Nat Rev Microbiol* 6(10): 752-762.

Sachs JL, Simms EL 2006. Pathways to mutualism breakdown. *Trends Ecol Evol* 21(10): 585-592.

Sachs JL, Russell JE, Hollowell AC 2011. Evolutionary instability of symbiotic function in *Bradyrhizobium japonicum*. *PLoS ONE* 6(11): e26370.

Schuster BM, Perry LA, Cooper VS, Whistler CA 2010. Breaking the language barrier: experimental evolution of non-native *Vibrio fischeri* in squid tailors luminescence to the host. *Symbiosis* 51(1): 85-96.

Smith MN, Crane RA, Keates RA Wood JM 2004. Overexpression, purification, and characterization of ProQ, a posttranslational regulator for osmoregulatory transporter ProP of *Escherichia coli*. *Biochemistry* 43(41): 12979-12989.

Soto W, Gutierrez J, Remmenga MR, Nishiguchi MK 2009a. Salinity and temperature effects on physiological responses of *Vibrio fischeri* from diverse ecological niches. *Microb Ecol* 57(1): 140-150.

Soto W, Lostroh CP, Nishiguchi MK 2009b. Physiological responses to stress in the Vibrionaceae: aquatic microorganisms frequently affiliated with hosts. Springer, New York.

Soto W, Punke EB, Nishiguchi MK 2012. Evolutionary perspectives in a mutualism of sepiolid squid and bioluminescent bacteria: combined usage of microbial experimental evolution and temporal population genetics. *Evolution* 66(5): 1308-1321.

Soto W, Rivera FM, Nishiguchi MK 2014. Ecological diversification of *Vibrio fischeri* serially passaged for 500 generations in novel squid host *Euprymna tasmanica*. *Microb Ecol* 67(3): 700-721.

Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J 2010. Adaptation genomics: the next generation. *Trends Ecol Evol* 25(12): 705-712.

Suzuki H, Lefebure T, Hubisz MJ, Bitar PP, Lang P, Siepel A, Stanhope MJ 2011. Comparative genomic analysis of the *Streptococcus dysgalactiae* species group: gene content, molecular adaptation, and promoter evolution. *Genome Biol Evol* 3(168-185.

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR 2012. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *PNAS* 109(22): 8629-8634.

Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* 16(2): 149-156.

Treangen TJ, Abraham AL, Touchon M, Rocha EPC 2009. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol Rev* 33(3): 539-571.

Treangen TJ, Rocha EP 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* 7(1): e1001284.

Visick KL, O'Shea TM, Klein AH, Geszvain K, Wolfe AJ 2007. The sugar phosphotransferase system of *Vibrio fischeri* inhibits both motility and bioluminescence. *J Bacteriol* 189(6): 2571-2574.

Vuilleumier S 1997. Bacterial glutathione S-transferases: What are they good for? *J Bacteriol* 179(5): 1431-1441.

Wang YL, Dunn AK, Wilneff J, McFall-Ngai MJ, Spiro S Ruby EG 2010. *Vibrio fischeri* flavohaemoglobin protects against nitric oxide during initiation of the squid-*Vibrio* symbiosis. *Mol Microbiol* 78(4): 903-915.

Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiasse RP, Schaefer AL, Koroleva I, Splinter-BonDurant S, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE, Bonaldo MD, Casavant TL, Soares MB, Cronan JE, Reed JL, Ruby EG, McFall-Ngai MJ 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. *PNAS USA* 107(5): 2259-2264.

Wietzorrek A, Bibb M 1997. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an *ompR*-like DNA-binding fold. *Mol Microbiol* 25(6): 1181-1184.

Wollenberg MS, Ruby EG 2009. Population structure of *Vibrio fischeri* within the light organs of *Euprymna scolopes* squid from two Oahu (Hawaii) populations. *Appl Environ Microbiol* 75(1): 193-202.

Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8): 1586-1591.

Yannarell AC, Triplett EW 2005. Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microbiol* 71(227-239.

Yildiz FH, Visick KL 2009. *Vibrio* biofilms: so much the same yet so different. *Trends Microbiol* 17(3): 109-118.

Yip ES, Grublesky BT, Hussa EA, Visick KL 2005. A novel, conserved cluster of genes promotes symbiotic colonization and sigma-dependent biofilm formation by *Vibrio fischeri*. *Mol Microbiol* 57(5): 1485-1498.

Zamborsky DJ, Nishiguchi MK 2011. Phylogeographical patterns among Mediterranean Sepiolid squids and their *Vibrio* symbionts: environment drives specificity among sympatric species. *Appl Environ Microbiol* 77(2): 642-649.

Zhang JZ 2003. Evolution by gene duplication: an update. *Trends Ecol Evol* 18(6): 292-298.

Zhou A, Chen YI, Zane GM, He Z, Hemme CL, Joachimiak MP, Baumohl JK, He Q, Fields MW, Arkin AP 2012. Functional characterization of Crp/Fnr-type global transcriptional regulators in *Desulfovibrio vulgaris* Hildenborough. *Appl Environ Microbiol* 78(4): 1168-1177.

Received on June 9, 2015
Accepted on August 24, 2015
Associate editor: A Chenuil

These supplementary tables are available on our website:

Supplementary Table I. – Genomic data summary and percent coverage of reference strains available on GenBank corresponding with newly sequenced genomes. NE = not examined.

Supplementary Table II. – Data concerning Roche 454 sequencing read quality.

Supplemental Table III. – List of core genes shared by symbiotic strains of *Vibrio fischeri*. The gene numbers are based on those for the V. fischeri MJ11 reference genome.

Supplementary Table IV. – Genetic elements identified in three symbiotic strains of *Vibrio fischeri* and one free-living *Vibrio* sp. sequenced in this study, in addition to one model organism, *V. fischeri* ES114 (Ruby *et al*. 2005). Annotations provided by RAST (Overbeek *et al*. 2014).

Supplementary Table V. – Patterns of GC content in Vibrio genomes sequenced in this study. Clusters of open reading frames (ORFs) that deviate from mean GC content are provided in blocks separated by spaces. Clusters are defined as 3 or more ORFs in proximity. High or low GC content is defined as GC proportions that fall more the 2 standard deviations from the mean for that genome.

Supplementary Table VI. – Gene duplication within symbiotic *Vibrio fischeri* and the non-symbiotic *V. cf. campbellii* strains. From the orthology analysis, amino acid sequences demonstrating closest-match sequence similarity to other genes within the sequence dataset are listed below. Lengths of the sequences are denoted by nucleotide number. Similarities to most closely-matched reference genomes are shown, as well as putative gene functions for these duplicated gene sequences.