Ongoing Transposon-Mediated Genome Reduction in the Luminous Bacterial Symbionts of Deep-Sea Ceratioid Anglerfishes

Tory A. Hendry,[a] Lindsay L. Freed,[b] Dana Fader,[a] Danté Fenolio,[c] Tracey T. Sutton,[b] Jose V. Lopez[b]

[a]Department of Microbiology, Cornell University, Ithaca, New York, USA
[b]Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, Dania Beach, Florida, USA
[c]Department of Conservation and Research, San Antonio Zoo, San Antonio, Texas, USA

ABSTRACT   Diverse marine fish and squid form symbiotic associations with extracellular bioluminescent bacteria. These symbionts are typically free-living bacteria with large genomes, but one known lineage of symbionts has undergone genomic reduction and evolution of host dependence. It is not known why distinct evolutionary trajectories have occurred among different luminous symbionts, and not all known lineages previously had genome sequences available. In order to better understand patterns of evolution across diverse bioluminescent symbionts, we *de novo* sequenced the genomes of bacteria from a poorly studied interaction, the extracellular symbionts from the "lures" of deep-sea ceratioid anglerfishes. Deep-sea anglerfish symbiont genomes are reduced in size by about 50% compared to free-living relatives. They show a striking convergence of genome reduction and loss of metabolic capabilities with a distinct lineage of obligately host-dependent luminous symbionts. These losses include reductions in amino acid synthesis pathways and abilities to utilize diverse sugars. However, the symbiont genomes have retained a number of categories of genes predicted to be useful only outside the host, such as those involved in chemotaxis and motility, suggesting that they may persist in the environment. These genomes contain very high numbers of pseudogenes and show massive expansions of transposable elements, with transposases accounting for 28 and 31% of coding sequences in the symbiont genomes. Transposon expansions appear to have occurred at different times in each symbiont lineage, indicating either independent evolutions of reduction or symbiont replacement. These results suggest ongoing genomic reduction in extracellular luminous symbionts that is facilitated by transposon proliferations.

IMPORTANCE   Many female deep-sea anglerfishes possess a "lure" containing luminous bacterial symbionts. Here we show that unlike most luminous symbionts, these bacteria are undergoing an evolutionary transition toward small genomes with limited metabolic capabilities. Comparative analyses of the symbiont genomes indicate that this transition is ongoing and facilitated by transposon expansions. This transition may have occurred independently in different symbiont lineages, although it is unclear why. Genomic reduction is common in bacteria that only live within host cells but less common in bacteria that, like anglerfish symbionts, live outside host cells. Since multiple evolutions of genomic reduction have occurred convergently in luminous bacteria, they make a useful system with which to understand patterns of genome evolution in extracellular symbionts. This work demonstrates that ecological factors other than an intracellular lifestyle can lead to dramatic gene loss and evolutionary changes and that transposon expansions may play important roles in this process.

Bioluminescent symbiosis between bacteria and marine fishes and squid has evolved independently many times, with 2 origins in squid (1) and at least 17 origins across fishes, for a total of over 460 bacterially luminous host species (2, 3). Repeated evolutions of bioluminescent symbiosis are likely due to the benefits derived by hosts, which use bacterially produced light to avoid predators and find prey and for intraspecific signaling (4). Hosts maintain luminous bacteria in specialized structures called light organs and provide them with nutrients (5–7). Diverse host species include coral-reef-dwelling, coastal, and pelagic fishes, as well as poorly studied species from the deep sea, such as deep-sea anglerfishes of the suborder Ceratioidei (Teleostei: Lophiiformes). On a global scale, Ceratioidei is notable in that it is the most speciose fish taxon in the bathypelagic zone (oceanic waters deeper than 1,000 m), which is by far Earth's largest ecosystem (8, 9). A common feature of the taxon is that females of the majority of species have light organs in the form of lures, or escae, extending from the illicium, a modified dorsal fin ray on their heads. These light organs contain dense, extracellular populations of bacteria and emit light through transparent tissue (Fig. 1) (8, 10). However, ceratioids remain the least-characterized group of bacterially luminous fishes, and little is known about their symbionts.

The majority of bacterially luminous fish species engage in symbiosis with just six species of bacteria in the genera *Aliivibrio* and *Photobacterium* from the *Gammaproteobacteria* family *Vibrionaceae* (2, 11). In contrast, ceratioids associate with luminous bacteria that are related to other *Vibrionaceae* species, but may be distinct lineages (12). *Aliivibrio* and *Photobacterium* symbiont species are all facultatively symbiotic and form free-living populations in multiple habitats in addition to engaging in symbiosis with hosts (2). In keeping with this habitat diversity, members of *Vibrionaceae* have relatively large (~5 Mb) and diverse genomes (13). Although bioluminescent symbiosis appears to have evolved several times in different *Vibrionaceae* lineages (11), most luminous

FIG 1 Female deep-sea anglerfish. Shown are adult female specimens of *M. johnsonii* (top) and *C. couesii* collected on DEEPEND Consortium cruises in similar locations to those of our samples. (Photo credit: Danté Fenolio, San Antonio Zoo.)

symbionts show only small-scale genomic changes in response to host interactions (14–16).

One exception to this pattern are the luminous symbionts in the genus "*Candidatus* Photodesmus," which are obligately dependent on their hosts, flashlight fishes (Anomalopidae), for growth (7, 17). These bacteria are extracellular (18, 19) and appear to colonize new hosts from the environment rather than directly from host-associated populations (5, 17). However, their genomes are reduced by 80% compared to those of free-living relatives, and their loss of metabolic genes indicates that they are unable to establish free-living populations in most environments (7, 17). "*Ca.* Photodesmus" genomes also show very similar molecular patterns to intracellular, vertically transmitted symbionts, including high substitution rates and relaxed purifying selection (20, 21).

Intracellular bacterial symbionts, including the mutualists of many insects, show a striking pattern of convergently derived genome reduction (22–25). Physical restriction to host cells, sharing metabolic products, and vertical transmission between host generations are thought to lead to extreme gene loss in multiple ways. First, restriction to a host cell relaxes selection on functions only needed outside the host, allowing for the loss of genes underlying those functions. Second, host-restricted bacteria experience reductions in effective population sizes (N_e) due to a lack of opportunities for recombination and horizontal gene transfer, as well as population bottlenecks during transmission between host generations. This decrease in N_e increases the effect of genetic drift and the rate of nucleotide substitutions and gene loss (26–28). Together these processes are thought to lead to relatively fast genomic degeneration immediately following host restriction and continued reduction to minimal genomes (24, 29).

Recently derived intracellular symbionts sometimes show remnants of this ongoing genomic degeneration, such as high numbers of pseudogenes or transposable elements (TEs) (29–33). This proliferation of TEs is thought to be caused by relaxed selection on TE regulation, as well as relaxed purifying selection on genes throughout the genome that serve as potential insertion sites. TEs may facilitate the process of genomic reduction by inserting within and disrupting genes (30, 31). These genomic elements are eventually lost from genomes by deletion, as they are not typically seen in relatively long-term intracellular symbionts (24, 34). They also do not tend to accumulate in free-living bacteria, presumably because of selection against insertion into required genes (35).

Extracellular symbionts, particularly those that are horizontally transmitted or acquired from the environment, are not physically restricted to hosts and therefore do not typically undergo gene loss (25, 36, 37), with a few known exceptions (7, 38, 39). Like all known luminous bacteria, ceratioid symbionts are extracellular, and pores present on the surface of light organs may allow movement of the bacteria between light organs and the environment (19). However, previous work has stated that anglerfish symbionts are unculturable, suggesting the potential for obligate dependence on hosts and possibly genomic patterns of evolution similar to those of flashlight fish symbionts (12). To resolve the evolutionary histories of deep-sea anglerfish symbionts, we generated and analyzed *de novo* genome sequences for the light organ symbionts from two distantly related ceratioid host species, two individuals of *Cryptopsaras couesii* (Ceratiidae) and one individual *Melanocetus johnsonii* (Melanocetidae).

RESULTS AND DISCUSSION

Genetic diversity, evolutionary relationships, and rates. Symbiont genome sequences from the two ceratioid fish species show a high degree of host specificity and little genetic diversity. The symbiont genomes from separate *C. couseii* specimens (CC26 and CC32) were extremely similar, sharing 99.9% nucleotide sequence identity across entire genome alignments. Furthermore, previously reported 16S rRNA gene sequences from *C. couseii* and *M. johnsonii* specimens caught in 1990 near the Canary Islands (12, 40) were each 100% identical across 1,416 bp to those sequences from the *C. couseii* and *M. johnsonii* (MJ02) symbionts sequenced here. We also observed very little intralight organ diversity within each sample. The majority (>90%) of possible alternate

bases found in the reads from the CC32 and MJ02 libraries were present in less than 0.05% of the read depth at each site, suggesting that they are sequencing errors (see Fig. S1 in the supplemental material). Additionally, alternate bases present in greater than 1% of the read depth, which might represent actual intralight organ variation, were found at a rate of only 1.6 potential polymorphisms per kilobase. Although these samples sizes are low and this pattern could change with further sampling, the low level of genetic diversity found within each anglerfish symbiont species and the apparent symbiont specificity to a host species are unusual compared to those of free-living luminous symbionts of fishes (11, 21) and very similar to those of previously reported patterns in flashlight fish symbionts (21) and obligate symbionts (41, 42). This pattern could result from small, possibly monoclonal founding populations of bacterial cells within the esca and/or low genetic diversity across symbiont populations.

Phylogenetic analyses recovered anglerfish symbiont genotypes from *C. couesii* and *M. johnsonii* as sisters to each other with high support (Fig. 2; see Fig. S2 in the supplemental material). The symbionts were placed within a clade containing members of the genus *Enterovibrio*, closely related to *Enterovibrio calviensis* and *Enterovibrio norvegicus*. The relatively short branch separating the anglerfish symbiont clade from *Enterovibrio* strains, as well as their position nested within the genus, supports the placement of the symbionts within the genus *Enterovibrio*. However, the long branches separating the symbionts of different fish species, as well as their low average nucleotide identity (ANI), which was only 72.4% among shared coding loci, supports the separation of the symbionts from each fish species into separate bacterial species. We propose the names "*Candidatus* Enterovibrio luxaltus" ("deep light") and "*Candidatus* Enterovibrio escacola" ("esca [bait] dwelling") for the CC26/CC32 and MJ02 symbionts from *C. couesii* and *M. johnsonii*, respectively.

The "*Ca.* Enterovibrio escacola" and "*Ca.* Enterovibrio luxaltus" lineages are each separated from other taxa by long branches, suggesting that they may be evolving at an elevated rate compared to relatives. To test this, we compared the likelihoods of molecular clock models that assumed distinct relative substitution rates for anglerfish symbionts and relatives. Allowing anglerfish symbionts to evolve at a distinct rate led to a significant increase in likelihood over the null hypothesis of a global clock (Table 1). The same result was found in this analysis for flashlight fish symbionts, which have previously been shown to be evolving more quickly than free-living relatives (20), and was true for both nucleotide and amino acid sequence data. This was not the case for clades containing free-living relatives, either facultative symbionts or nonsymbiotic species, which were generally evolving more slowly than the rest of the tree (Table 1). The anglerfish symbiont clade was estimated to be evolving at 3.5 times the rate of relatives at the nucleotide level and 4.8 times the rate of relatives in protein sequence, higher rates than the estimates for flashlight fish symbiont evolution compared to relatives. These results confirm that the anglerfish symbionts are evolving at an increased rate compared to free-living relatives and have accumulated a higher relative number of nonsynonymous substitutions.

Genome assemblies. The genome from the *C. couesii* CC26 sample assembled into two plasmids recovered as circular (CC26 P1 and P2) and two large contigs possibly matching to chromosomes I and II of *Vibrionaceae* taxa (Table 2). The CC32 assembly was highly similar (discussed above) but contained more contigs, so we will present data for just the better-assembled CC26 genome hereafter. The *M. johnsonii* MJ02 symbiont assembly contained 39 contigs, including four circular plasmids. Based on assembly coverage depth, all plasmids had a similarly low copy number (1 to 3 copies per genome). All three symbiont assemblies were found to contain conserved protein coding genes typically used to assess genome completeness (43, 44), indicating that they are nearly fully complete in sequence. The "*Ca.* Enterovibrio luxaltus" genome assembled at 2.1 Mb, and the "*Ca.* Enterovibrio escacola" genome totaled 2.6 Mb. Both genomes had a high number of predicted pseudogenes (785 and 974, respectively), a low number of rRNA genes (1 operon plus an additional 5S rRNA gene in "*Ca.*
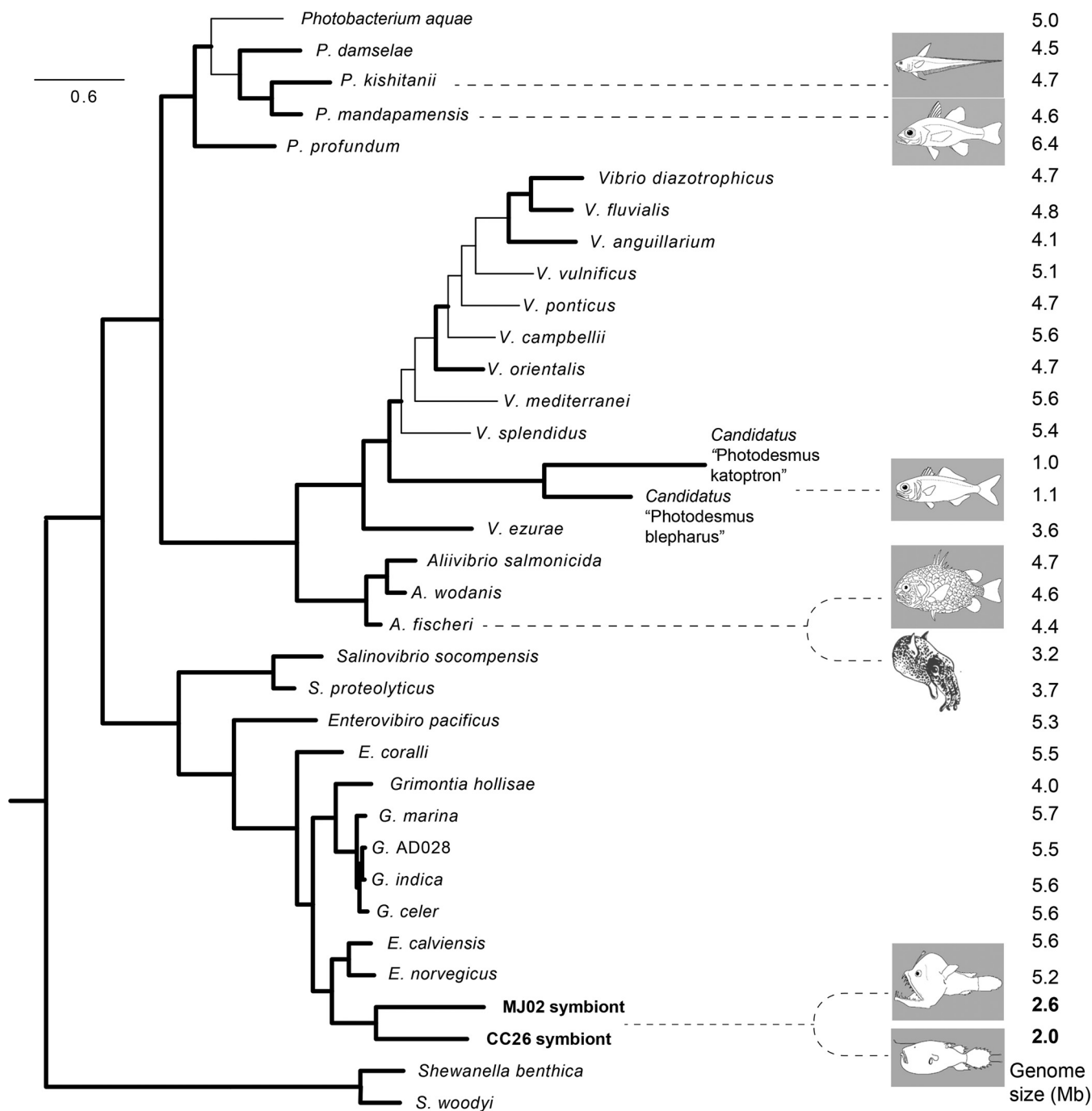
FIG 2 Relationship of anglerfish symbionts to major relative clades. Shown is a maximum likelihood phylogenomic tree based on 253 conserved protein sequences obtained from PhyloPhlAn. Nodes with bootstrap support values over 95% are shown with bold branches. Representative fish and squid host pictures are shown to indicate luminous symbionts. Genome size values are indicated in the far right column and were taken from GenBank. The anglerfish symbiont sequences from CC26 and MJ02 are shown in bold.

Enterovibrio luxaltus" and 1 operon in "*Ca.* Enterovibrio escacola"), and a low number of tRNA genes (Table 2).

Consistent with their role as luminous symbionts, both genomes contain luminescence genes *luxCDABEG* in the same operon structure seen in other luminous bacteria (2). We were unable to determine if these genes may be regulated by quorum sensing, as is true in some other luminous species such as *Aliivibrio fischeri* and *Vibrio harveyi*, because we found no orthologs of known luminescence regulatory genes in the

TABLE 1 Comparison of molecular clock models and relative substitution rates using either 7 housekeeping gene sequences or 253 conserved protein sequences[a]

Tree	df	Clade(s) with distinct rate(s)	LR	Null (−lnL0)	Alternative (−lnLA)	P[b]	Relative substitution rate
Housekeeping	1	Anglerfish symbionts	269.87	33,576.04	33,441.11	0.0000*	3.5
	1	Flashlight fish symbionts	264.18	33,576.04	33,443.95	0.0000*	2.5
Protein	1	Anglerfish symbionts	488.81	57,332.26	57,087.86	0.0000*	4.8
	1	Flashlight fish symbionts	340.63	57,332.26	57,161.95	0.0000*	2.4
Housekeeping	1	*E. calviensis* and *E. norvegicus*	12.26	33,576.04	−33,569.91	0.0005*	0.7
Protein	1	*E. calviensis* and *E. norvegicus*	55.59	57,332.26	−57,304.47	0.0000*	0.6
Housekeeping	1	*Grimontia*	0	33,576.04	−33,576.04	1.0000	0.6
Protein	1	*Grimontia*	8.54	57,332.26	−57,327.99	0.0040*	0.7
Housekeeping	1	*Aliivibrio*	0	33,576.04	−33,576.04	1.0000	0.6
Protein	1	*Aliivibrio*	80.44	57,332.26	−57,292.05	0.0000*	0.6
Housekeeping	1	*Photobacterium*	126.52	33,576.04	−33,513.78	0.0000*	0.5
Protein	1	*Photobacterium*	96.04	57,332.26	−57,284.25	0.0000*	0.6

[a]Abbreviations: df, degrees of freedom; LR, likelihood ratio; −lnL0, log likelihood null model; −lnLA, log likelihood alternative model.
[b]*, significant likelihood ratio test result.

symbiont genomes (45, 46). No luminous strains have previously been reported in the early branching *Vibrionaceae* genera *Enterovibrio*, *Grimontia*, or *Salinivibrio*. Phylogenetic analysis of the anglerfish symbiont *lux* genes does not show high relatedness with *lux* genes from other taxa, which could suggest horizontal gene transfer (see Fig. S3 in the supplemental material), although we cannot exclude this possibility. The inclusion of luminous anglerfish symbionts within the genus *Enterovibrio* suggests a possible earlier evolution of luminescence than previously thought and supports the hypothesis that luminescence arose ancestrally in *Vibrionaceae* and has been lost in many lineages (47).

Genome reduction. The total genome size for both symbiont species (2 and 2.6 Mb) is reduced by about 50% compared to the genomes of free-living relatives (Fig. 2). The closest relatives to the symbionts, free-living *Enterovibrio* strains, have 5- to 5.5-Mb genomes, and the average across free-living *Vibrionaceae* species is about 5 Mb (values from GenBank). This reduction is even more extreme when considering just functional protein coding genes (non-pseudogenes). The three most closely related *Enterovibrio* strains with sequenced genomes average about 5,200 predicted functional protein coding genes, whereas the anglerfish symbionts have 1,662 (CC26) and 2,316 (MJ02), 68% and 55% reductions in predicted functional coding sequence number. The phylogenetic reconstruction (Fig. 2) shows that the genomic reductions in anglerfish symbionts and flashlight fish symbionts represent two convergent evolutions among luminous symbionts. All other known groups of luminous symbionts are free-living and have larger (~4.5 Mb) genomes. Genome reduction can be caused by relaxed purifying selection at a genome-wide scale, as is found in bacteria that have become obligately associated with hosts (24, 48). Similar to obligate symbionts, the anglerfish symbiont genomes are reduced, evolving at an elevated evolutionary rate compared to relatives, and contain a large number of pseudogenes. These changes are consistent with genome-wide relaxed selection and high genetic drift in the anglerfish symbionts and suggest that they may be undergoing genome reduction due to the host association, rather than due to genome streamlining, as can be found in some marine bacteria (49, 50).

Gene content and inferred ecology. The overall gene content of anglerfish symbiont genomes (Fig. 3) is dramatically different from most free-living *Vibrionaceae* species, including closely related *Enterovibrio* species. One possible exception to this pattern is the genomes of *Salinivibrio* species (Fig. 3). This genus is typically isolated from hypersaline environments such as salt lakes or salted meats (51). Because these strains are the free-living *Vibrionaceae* taxa with the smallest known genomes, and because they share some potential similarities with anglerfish symbionts, such as gene loss and increases in transposable element numbers (discussed below), we performed some focused comparisons between *Salinivibrio costicola* subsp. *costicola* and angler-

TABLE 2 Overview of genomic features of the symbionts of three anglerfish specimens (two *C. couesii* and one *M. johnsonii*) compared to the genomes of symbionts from two flashlight fish (*Anomalops katoptron* and *Photoblepharon palpebratus*), two free-living, nonsymbiotic relatives, and a facultative symbiont of *Euprymna scolopes*[a]

Host	Bacterium	Specimen or strain	Genome size (bp)	Contigs/plasmids	Ecology	Coverage	N$_{50}$ (reference)	G+C %	Total CDS	Complete CDS	Pseudogenes	rRNA/tRNA
C. couesii	E. luxaltus	CC26	2,143,356	4/2	Obligate symbiont	14x	1,604,691 (this study)	37.7	2,447	1,662	785	4/36
C. couesii	E. luxaltus	CC32	2,013,547	11/1	Obligate symbiont	81x	249,450 (this study)	37.6	NA	NA	NA	NA
M. johnsonii	E. escacola	MJ02	2,645,619	39/4	Obligate symbiont	58x	113,883 (this study)	39.8	3,290	2,316	974	3/39
A. katoptron	P. katoptron	Akat8	1,015,921	NA/1	Obligate symbiont	NA	NA (17)	30.8	933	873	13	15/32
P. palpebratus	P. blepharus	Ppalp1	1,112,309	NA/2	Obligate symbiont	NA	NA (17)	35.6	1,003	932	23	15/33
NA	V. campbellii	ATCC BAA-1116	6,058,377	NA/1	Free-living	NA	NA (75)	45.4	5,798	5,210	435	32/121
NA	E. norvegicus	FF-33	5,160,129	NA/unknown	Free-living	NA	NA (76)	47.4	4,664	4,929	188	8/69
E. scolopes	A. fischeri	ES114	4,272,718	NA/1	Facultative symbiont	NA	NA (52)	38.3	3,814	3,654	5	37/118

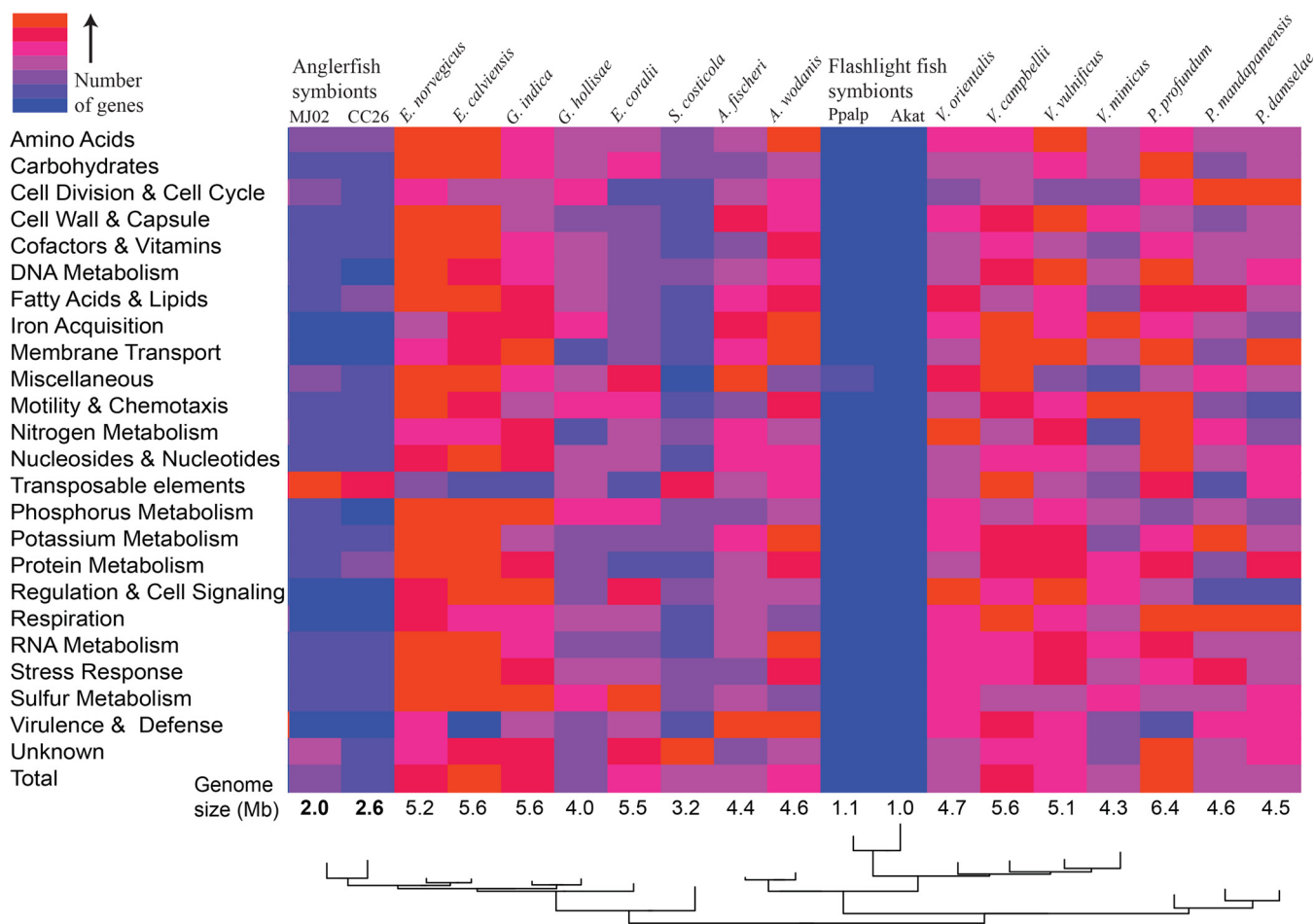[a]CDSs, coding DNA sequences; NA, not applicable.

FIG 3 Numbers of genes by functional category. Shown is a heat map of gene content in the genomes of anglerfish symbionts, flashlight fish symbionts, and free-living *Vibrionaceae* members (all other columns). Columns are ordered by relatedness based on a phylogeny (adapted from Fig. 2). Colors were assigned based on the distribution of genes within each functional category. Genomes were taken from GenBank, and categories are based on RAST classifications. For specific numbers of genes, see the table version of this figure (Fig. S5).

fish symbionts. Although the anglerfish symbiont genomes look somewhat similar to *S. costicola* subsp. *costicola* at a broad scale, there are many notable differences at a finer scale (Table 3). *Salinivibrio* species have similar numbers of metabolic genes, including phosphotransferase system (PTS) genes for sugar uptake, and genes involved in amino acid synthesis and energy metabolism, compared to other free-living relatives. Anglerfish symbiont genomes have only one PTS, which is specific to glucose, and have a marked reduction in complete amino acid synthesis pathways (see Fig. S4 in the supplemental material) and total numbers of energy metabolism genes. The anglerfish symbiont genomes have also lost most of the DNA repair and recombination genes typically found in relatives, a pattern that is widely shared among obligate and intracellular symbiotic bacteria (52).

Due to the small genome size of anglerfish luminous symbionts compared to close relatives, we assume that genes inferred to be functional within the genomes have been retained because they are necessary for growth or are ecologically beneficial, whereas genes common in relatives but lacking in the symbiont genomes have been lost due to decreased selection. This pattern of gene retention allows for some inference about the likely ecological lifestyle of symbionts, although we cannot rule out the possibility that any individual gene may have been retained by chance. For instance, the limited metabolic capabilities of the anglerfish symbionts suggest that they must acquire glucose and many amino acids from the environment and that anglerfishes must be supplying these nutrients to their symbionts (7). These metabolic

TABLE 3 Numbers of genes in highly reduced or retained categories in the anglerfish symbionts (CC26 and MJ02 samples), flashlight fish symbiont (*Anomalops katoptron* symbiont Akat8), and free-living relatives (*E. calviensis*, *S. costicola* subsp. *costicola*, and *A. fischeri*)

	No. of genes in category for:					
	Free-living relative			Symbiont		
				Flashlight		
Category	*Enterovibrio*	*Salinivibrio*	*Aliivibrio*	fish	CC26	MJ02
Reduced categories						
PTS (type)	12 (varied)	11 (varied)	13 (varied)	1 (glucose)	1 (glucose)	1 (glucose)
Amino acid synthesis (complete pathways)	91 (20)	74 (19)	85 (20)	18 (4)	26 (7)	34 (9)
Energy metabolism	661	202	312	49	83	84
DNA repair and recombination	12	11	12	12	2	3
Methyl-accepting chemotaxis proteins	36	20	43	2	1	3
Retained categories						
Cell wall synthesis	253	120	232	76	103	116
Motility and chemotaxis	150	102	131	57	89	89

limitations further suggest some degree of host dependence or restriction to marine environments where glucose and amino acids would be regularly available, which may be scarce in the deep sea. Patterns in other functional categories that are typically underrepresented in bacteria adapted to stable environmental conditions also support this hypothesis (53). Genes in the categories of membrane transport, regulation and cell signaling, and stress responses are all highly reduced in anglerfish symbionts compared to free-living relatives (see Fig. S5 in the supplemental material). Members of the *Vibrionaceae* are typically metabolically diverse; they are found in many marine habitats and associate with many types of hosts (13, 54). The evolutionary switch to a more limited environmental range and possible dependence on hosts for growth appears to be rare among host-associated or bioluminescently symbiotic *Vibrionaceae* species, having only been previously described in the luminous symbionts of flashlight fish, which also rely on hosts for glucose and amino acids (7).

The genomic evidence suggests that despite some host dependence, like all other known luminous symbionts, these bacteria may leave the light organ and persist outside the host. Although all broad functional categories show some reduction in gene number, some pathways that are typically lost in obligate or intracellular symbionts are retained in the anglerfish symbionts. Notably, a relatively large number of genes involved in cell wall synthesis are found in the anglerfish symbiont genomes, suggesting that they can synthesize a robust cell wall. These are typically lost in host-dependent symbionts, even those that are extracellular (38). Furthermore, the bacteria have retained genes suggesting that they are chemotactic and motile. This includes all nearly 60 genes necessary for production of flagella and transmission of chemotaxis signals. In contrast, many accessory genes, which are not required for these functions, have been lost (Table 3). Although components of these pathways are sometimes retained, full pathways for chemotaxis and motility are universally lost in known obligate symbionts (17, 55–58). Electron micrographs from inside anglerfish light organs show densely packed bacterial populations where motility and chemotaxis are unlikely to be useful (10). Therefore, these functions may be used primarily outside the host. In light of the large-scale loss of genetic pathways across the symbiont genomes, the apparent selection to retain pathways useful mainly outside the host suggests that an extrahost phase may be an important part of the symbiont's lifestyle. This is also seen in flashlight fish symbionts, which have retained genes in the categories of cell wall synthesis and motility and are known to persist and be motile in seawater outside the host (17).

Although the anglerfish symbionts have intact chemotaxis and motility pathways, they show a large reduction in genes coding for methyl-accepting chemotaxis proteins

(MCPs), cell surface proteins that detect chemical signals in the environment and elicit a motility response (59, 60). Whereas most *Vibrionaceae* species, including *Salinivibrio* and *Enterovibrio* species, have 20 to 50 MCP genes specific to varied ligands, the "*Ca.* Enterovibrio luxaltus" and "*Ca.* Enterovibrio escacola" genomes contain only one or three functional MCP genes, respectively (Table 3). The predicted function of the anglerfish symbiont MCP genes is not apparent by comparison to genes of known function. Two of the MCP genes found in the MJ02 genome, as well as the one CC26 MCP gene, cluster very closely in phylogenetic analysis and are related to an MCP gene with unknown function retained in a flashlight fish symbiont genome (see Fig. S6 in the supplemental material). The additional MJ02 MCP gene is more distantly related. The anglerfish symbionts appear to have relatively restricted chemotaxis abilities, presumably detecting only a small number of attractants or repellants in the environment. This may indicate that the bacteria are not actively searching for nutrient-rich habitats but may focus on chemicals associated with their specific habitat, such as chemical cues from hosts.

The anglerfish symbionts have also retained genes for carbon storage and utilization in the form of polyhydroxybutyrate (PHB). Host-associated bacteria typically lose pathways involved in carbon storage (61), but flashlight fish symbionts have retained multiple types of carbon storage pathways (*phbCAB* in both species, glycogen storage in one), which have been hypothesized to be useful for their known persistence outside the host (17). Both flashlight fish symbionts and anglerfish symbionts from an *M. johnsonii* specimen show occlusions in electron micrographs of host light organs that appear to be PHB granules, supporting the expression of the genes for carbon storage within the host environment (10, 18). Under this model, the fish host supplies an excess of carbon (in the form of glucose), as well as other nutrients to the bacteria in the light organ. The symbionts store excess carbon as PHB granules, which can later be used as a carbon and energy source outside the host. However, the metabolic limitations of the anglerfish symbionts mean that after release from the light organ, finding another habitat that can supply glucose and amino acids, such as another fish host, is imperative. This model then also explains the retention of pathways for chemotaxis and motility, which may be useful in finding new hosts or habitats.

Transposable element expansion. Both the "*Ca.* Enterovibrio luxaltus" and "*Ca.* Enterovibrio escacola" genomes contain extremely high numbers of transposable elements (TEs), specifically transposons (Table 4). The "*Ca.* Enterovibrio luxaltus" genome contains 691 TE genes, all but two of which are transposases, and the "*Ca.* Enterovibrio escacola" genome contains 921 TE genes, most of which are transposases. These genes account for 28% and 31%, respectively, of the total coding sequences in the genomes, the highest percentage per bacterial genome that we found previously reported (29–33, 35, 62). In comparison, free-living *Vibrionaceae* species, like other free-living bacteria, have relatively few TE genes. Here we focused on *Salinivibrio costicola* subsp. *costicola*, which is closely related to the genus *Enterovibrio* and has the largest number of TE genes of the genomes analyzed here (excluding anglerfish symbionts), yet those genes account for only 2% of the coding sequence in the genome (Table 4).

These results suggest an expansion of TEs in the anglerfish symbiont genomes. In order to better understand this pattern we categorized the identified transposase genes by insertion sequence (IS) family. In both symbiont genomes, transposase genes fell into a relatively small number of families (Table 4). Within the CC26 genome assembly, almost all transposases were members of IS family IS*5*, while in the MJ02 symbiont genome transposases were predominately from families IS*5*, IS*982*, and others. The transposase genes were relatively evenly spread across contigs within the well-assembled CC26 genome (see Fig. S7 in the supplemental material), with the exception of a few regions (>20 kb) with no insertions. These regions without transposons contained gene clusters or operons presumed to be necessary to the bacteria, including genes involved in cell division, cell wall synthesis, chemotaxis and motility,

TABLE 4 Transposable element features of the anglerfish symbionts (CC26 and MJ02 samples), flashlight fish symbiont (*A. katoptron* symbiont Akat8), and *S. costicola* subsp. *costicola* genomes

	Result for:			
	Symbiont			
Parameter	CC26	MJ02	Flashlight fish	*Salinivibrio* genome
Protein coding genes, no.	2,447	3,290	873	4,442
Total TEs, no.	691	921	0	87
Transposase genes, no.	689	888	NA[a]	80
Other TEs (phage genes), no.	2	33	NA	7
Complete transposase genes, no.	0	0	NA	49
TE % of CDS	28	31	NA	2
Transposases by IS family, no.				
IS5	672	198		5
IS982		333		
IS256		75		
ISL3		57		
Tn3		25		
IS200/605		60		
ISAs1				12
IS66		29		9
IS6	8			
Other and unknown	9	111		54

[a]NA, not applicable.

the tricarboxylic acid (TCA) cycle, and the luminescence (*lux*) gene operon. This pattern is consistent with large-scale expansions of a few transposons around the genomes, with some selection against insertion in necessary genes. The location of these transposons also suggests that their expansion may have facilitated gene loss in the genomes. All 55 pseudogenized functional genes (non-tranposase genes) within the MJ02 symbiont genome are adjacent to or interrupted by transposase fragments.

All identified transposase genes in both anglerfish symbiont genomes were classified as pseudogenes, as they were all either missing the inverted-repeat regions typically found in the specific types of transposases or were truncated in aligned length compared to orthologs from close relatives. The very high number of transposase genes coupled with the fact that they are no longer functional suggests that each symbiont genome displays remnants of previous large-scale transposon proliferations. In order to investigate the relative timing of these events, we aligned transposases from each genome by family and performed phylogenetic analyses. IS5 family transposases were expanded in both CC26 and MJ02 symbiont genomes, and IS982 and IS256 were expanded in just the MJ02 genome. IS5 transposases from both genomes (Fig. 4) and IS256 in the MJ02 genome (not shown) show a phylogenetic pattern consistent with a single expansion within each genome, with rapid diversification taking place over a short time scale, compared to typical rates of divergence of the same transposase groups in close relatives, and subsequent degeneration. IS982 in the MJ02 genome shows a very similar pattern, although with possibly two expansions from distinct ancestral transposases within the genome (see Fig. S8 in the supplemental material). Such expansions could have occurred when a change in selective pressure on the bacterial genomes, such as the initiation of the symbiotic interaction between the bacteria and the host, allowed for a loss of regulation on transposable elements and therefore proliferation that was unchecked by purifying selection (24). This proliferation must have eventually been selected against after a large expansion, since all remaining transposases in the genome are nonfunctional. Both symbiont genomes appear to be at the tail end of a recent TE expansion, where constraint on remaining genes has halted continued transposase function, but the remnants of these genes remain in the bacterial genome.
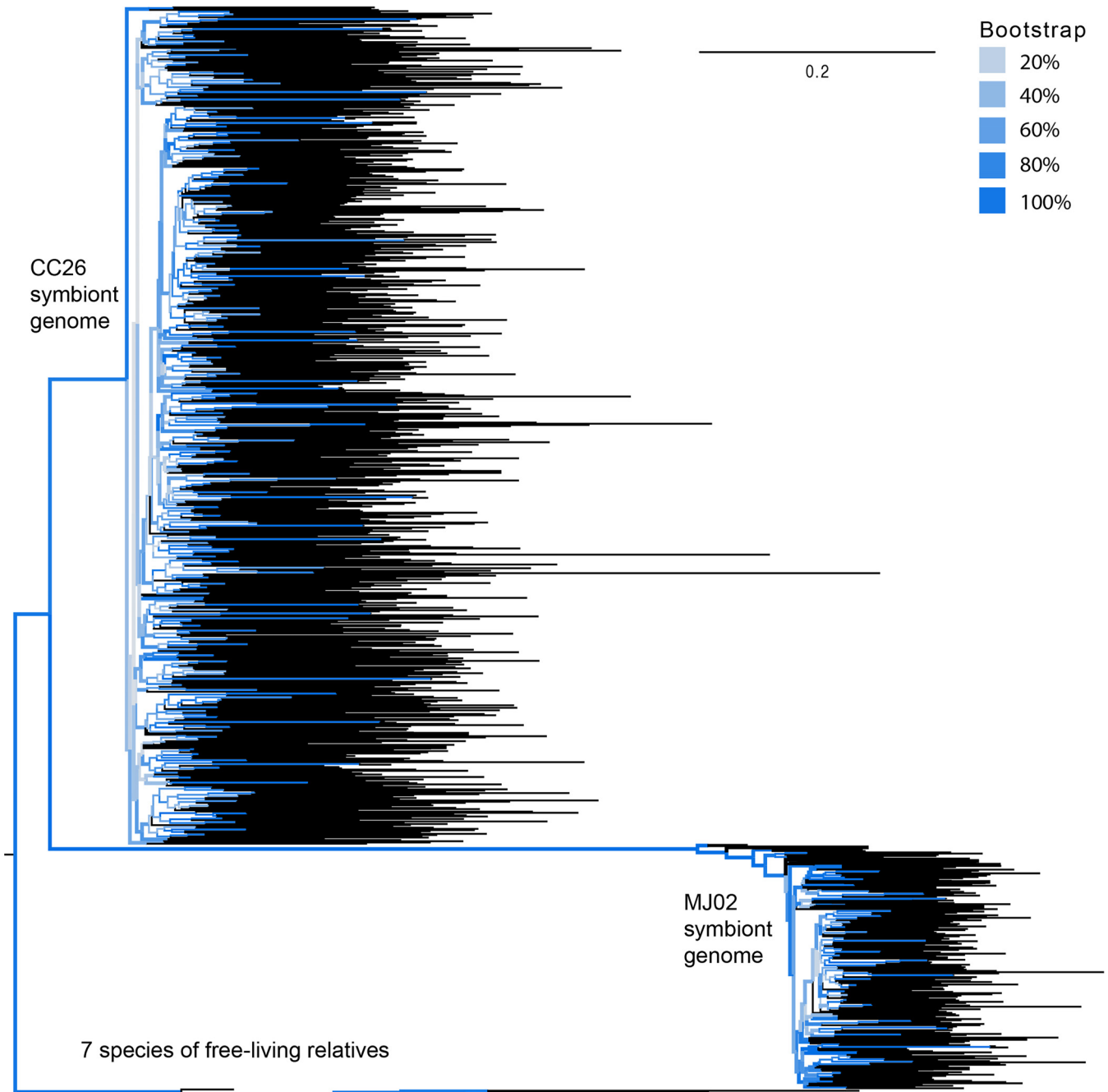
FIG 4 Phylogeny of IS5 family transposases. Relationships among IS5 family transposase genes and pseudogenes from the CC26 and MJ02 symbiont genomes, as well as 7 species of free-living relatives (*Vibrio alginolyticus* XSHD3, *Vibrio anguillarum* 775, *Vibrio mimicus* 6, *Vibrio vulnificus* YJ016, *Vibrio* sp. strain RC341, and *Shewanella oneidensis* MR-1) were reconstructed by maximum likelihood analysis. Bootstrap values are color coded, showing that the backbone of the tree has high bootstrap support. Tips are colored black.

Phylogenetic analysis of IS5 family transposases found in both bacterial genomes suggests that the two anglerfish symbiont lineages may have independently undergone TE expansions. The IS5 transposon expansion appears to have occurred at different time points within the "*Ca.* Enterovibrio luxaltus" and "*Ca.* Enterovibrio escacola" genomes rather than having occurred in the common ancestor of these symbionts. Compared to divergence between orthologs from different species of related bacteria, very long branches separate the IS5 transposases from each symbiont (Fig. 4), with the expansion of IS5 within "*Ca.* Enterovibrio escacola" possibly occurring

more recently. Since anglerfish symbionts are not physically restricted inside host cells, genomic degeneration and TE expansions in the bacteria could have occurred separately from the initial establishment of the symbiosis, possibly at different times in each lineage. Alternatively, one or both of these symbiont lineages could be the result of a more recent symbiont replacement and subsequent genomic reduction.

Conclusions. A striking feature of the anglerfish symbiont genomes is the convergence in gene content and genomic patterns with flashlight fish symbionts. Like other luminous symbionts, these two groups of bacteria are maintained extracellularly yet have independently undergone a process of large-scale gene loss leading to likely host dependence. This raises the question of why other luminous symbionts have not evolved similarly and what features may be shared between these fish groups that could influence the evolution of their symbionts. One possibility is that opportunities for vertical transmission of symbionts could allow for genome reduction to begin. The known schooling behavior of flashlight fishes presents a possible mechanism for pseudovertical transmission of their symbionts through the environment to larvae developing near adults (5, 7, 17, 63). Ceratioid fish are nonschooling and lay buoyant eggs that hatch near the ocean surface (8). Ceratioid larvae do not develop near adults but have an ontogenic vertical migration in which the light organ gradually develops (8, 10). In this system, vertically transmitted bacteria would need to be transferred to eggs by adults and persist until the larval light organ is developed enough to be colonized or juveniles would need to encounter a colonized adult in the deep sea. Alternatively, ceratioids may acquire their symbionts from environmental populations, as do other deep-sea fish species (11, 64–66). It is not clear from the genome sequences of the symbionts which of these scenarios is more likely, but bacterial operational taxonomic units (OTUs [i.e., OTUs based on the V4 variable region of the 16S rRNA]) matching both "*Ca.* Enterovibrio luxaltus" and "*Ca.* Enterovibrio escacola" have been found in water samples taken from the same locations sampled here (77), consistent with environmental persistence and possible environmental acquisition of symbionts.

The extremely high number of transposase remnants in both anglerfish symbiont genomes demonstrates that the process of genome reduction is still ongoing in these lineages. Furthermore, the pattern of transposon expansions in each genome suggests these proliferations occurred at different times in each lineage, either because genomic reduction and TE expansions occurred in anglerfish symbiont lineages at different times rather than occurring coincidentally with bacteria becoming host associated, as is seen in intracellular symbionts, or because symbiont replacements have occurred in one or both lineages. These results highlight the importance of studying diverse symbiotic systems in order to better understand which patterns may be shared across systems and when symbioses may break with expectations.

MATERIALS AND METHODS

Anglerfish specimens were collected during DEEPEND cruises in the Gulf of Mexico. Specimens were identified after collection by Tracey Sutton, and lures were immediately removed with a sterile scalpel and placed in ethanol. Specimens were stored at −80°C until processed by the Microbiology and Genetics Laboratory at Nova Southeastern University's Halmos College of Natural Sciences and Oceanography. All microbial DNA isolations were conducted following the Earth Microbiome Project protocol with the Mo Bio PowerLyzer PowerSoil kit. Illumina sequencing libraries were made from samples CC32 and MJ02 using a NexteraXT kit, and a library for CC26 was constructed using a Swift Biosciences PCR-free kit. Paired-end libraries were sequenced with a 250-bp read length on an Illumina HiSeq2500 instrument at the Cornell University Institute of Biotechnology Biotechnology Resource Center Genomics Facility. Genome assembly was done using the Discovar *de novo* assembler. Contigs were then binned by tetranucleotide frequency and coverage depth in MetaBAT (67). The binned genomic contigs for CC26 and MJ02 were then annotated in RAST (68). All coding sequences predicted by RAST were then compared to the most recent UniRef90 database release (March 2017) (69). Coding sequences for which the RAST annotation differed from the UniRef best hit were manually checked. Loci were considered possible pseudogenes if they were <60% of the length of the best UniRef hit or showed <30% amino acid similarity. All possible pseudogenes were checked manually by BLAST in UniRef90. Phylogenetic trees were constructed using Bayesian or maximum likelihood methods with conserved housekeeping genes (20, 70, 71) or genome-wide conserved protein sequences (72). Tests of evolutionary rate were performed in PAML (73). Transposase sequences were identified using the ISfinder database (74).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .01033-18.

TEXT S1, DOCX file, 0.2 MB.

FIG S1, DOCX file, 1 MB.

FIG S2, DOCX file, 0.4 MB.

FIG S3, DOCX file, 0.4 MB.

FIG S4, DOCX file, 0.3 MB.

FIG S5, DOCX file, 1.4 MB.

FIG S6, DOCX file, 0.7 MB.

FIG S7, DOCX file, 0.4 MB.

FIG S8, DOCX file, 0.5 MB.

TABLE S1, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. 2014. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. Proc Natl Acad Sci U S A 111:E4736–E4742. https://doi.org/10.1073/pnas.1416574111.

2. Dunlap PPV, Urbanczyk H. 2013. Luminous bacteria, p 495–528. *In* Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), The prokaryotes. Springer, Berlin, Germany.

3. Davis MP, Sparks JS, Smith WL. 2016. Repeated and widespread evolution of bioluminescence in marine fishes. PLoS One 11:e0155154. https://doi.org/10.1371/journal.pone.0155154.

4. Herring PJ, Morin JG. 1978. Bioluminescence in fishes, p 273–329. *In* Herring PJ (ed), Bioluminescence in action. Academic Press, London, United Kingdom.

5. Haygood MG. 1993. Light organ symbioses in fishes. Crit Rev Microbiol 19:191–216. https://doi.org/10.3109/10408419309113529.

6. Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiassé RP, Schaefer AL, Koroleva I, Splinter-Bondurant S, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE, Bonaldo MdF, 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. Proc Natl Acad Sci U S A 107:2259–2264. https://doi.org/10.1073/pnas.0909712107.

7. Hendry TA, de Wet JR, Dunlap PV. 2014. Genomic signatures of obligate host dependence in the luminous bacterial symbiont of a vertebrate. Environ Microbiol 16:2611–2622. https://doi.org/10.1111/1462-2920.12302.

8. Pietsch TW. 2009. Oceanic anglerfishes: extraordinary diversity in the deep sea. University of California Press, Oakland, CA.

9. Miya M, Pietsch TW, Orr JW, Arnold RJ, Satoh TP, Shedlock AM, Ho HC, Shimazaki M, Yabe M, Nishida M. 2010. Evolutionary history of anglerfishes (Teleostei: Lophiiformes): a mitogenomic perspective. BMC Evol Biol 10:58. https://doi.org/10.1186/1471-2148-10-58.

10. Munk O. 1999. The escal photophore of ceratioids (Pisces; Ceratioidei)—a review of structure and function. Acta Zool 80:265–284. https://doi.org/10.1046/j.1463-6395.1999.00023.x.

11. Dunlap PV, Ast JC, Kimura S, Fukui A, Yoshino T, Endo H. 2007. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. Cladistics 23:507–532. https://doi.org/10.1111/j.1096-0031.2007.00157.x.

12. Haygood MG, Distel DL. 1993. Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. Nature 363:154–156. https://doi.org/10.1038/363154a0.

13. Gu J, Neary J, Cai H, Moshfeghian A, Rodriguez SA, Lilburn TG, Wang Y. 2009. Genomic and systems evolution in Vibrionaceae species. BMC Genomics 10:S11. https://doi.org/10.1186/1471-2164-10-S1-S11.

14. Mandel MJ, Wollenberg MS, Stabb EV, Visick KL, Ruby EG. 2009. A single regulatory gene is sufficient to alter bacterial host range. Nature 458:215–218. https://doi.org/10.1038/nature07660.

15. Brooks JF, Gyllborg MC, Cronin DC, Quillin SJ, Mallama CA, Foxall R, Whistler C, Goodman AL, Mandel MJ. 2014. Global discovery of colonization determinants in the squid symbiont *Vibrio fischeri*. Proc Natl Acad Sci U S A 111:17284–17289. https://doi.org/10.1073/pnas.1415957111.

16. Bongrand C, Koch EJ, Moriano-Gutierrez S, Cordero OX, McFall-Ngai M, Polz MF, Ruby EG. 2016. A genomic comparison of 13 symbiotic *Vibrio fischeri* isolates from the perspective of their host source and colonization behavior. ISME J 10:2907–2917. https://doi.org/10.1038/ismej.2016.69.

17. Hendry TA, de Wet JR, Dougan KE, Dunlap PV. 2016. Genome evolution in the obligate but environmentally active luminous symbionts of flashlight fish. Genome Biol Evol 8:2203–2213. https://doi.org/10.1093/gbe/evw161.

18. Kessel M. 1977. The ultrastructure of the relationship between the luminous organ of the teleost fish *Photoblepharon palpebratus* and its symbiotic bacteria. Cytobiologie 15:145–158.

19. Haygood MG, Tebo BM, Nealson KH. 1984. Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and two anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*): population sizes and growth within the light organs, and rates of release into the seawater. Mar Biol 78:249–254. https://doi.org/10.1007/BF00393010.

20. Hendry TA, Dunlap PV. 2011. The uncultured luminous symbiont of *Anomalops katoptron* (Beryciformes: Anomalopidae) represents a new

bacterial genus. Mol Phylogenet Evol 61:834–843. https://doi.org/10.1016/j.ympev.2011.08.006.

21. Hendry TA, Dunlap PV. 2014. Phylogenetic divergence between the obligate luminous symbionts of flashlight fishes demonstrates specificity of bacteria to host genera. Environ Microbiol Rep 6:331–338. https://doi.org/10.1111/1758-2229.12135.

22. Moran NA. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci U S A 104:8627–8633. https://doi.org/10.1073/pnas.0611659104.

23. Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190. https://doi.org/10.1146/annurev.genet.41.110306.130119.

24. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10:13–26. https://doi.org/10.1038/nrmicro2670.

25. Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. Nat Rev Microbiol 8:218–230. https://doi.org/10.1038/nrmicro2262.

26. Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A 93:2873–2878. https://doi.org/10.1073/pnas.93.7.2873.

27. Wernegreen JJ, Moran NA. 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. Mol Biol Evol 16:83–97. https://doi.org/10.1093/oxfordjournals.molbev.a026040.

28. Woolfit M, Bromham L. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol 20:1545–1555. https://doi.org/10.1093/molbev/msg167.

29. Manzano-Marín A, Latorre A. 2016. Snapshots of a shrinking partner: genome reduction in Serratia symbiotica. Sci Rep 6:32590. https://doi.org/10.1038/srep32590.

30. Plague GR, Dunbar HE, Tran PL, Moran NA. 2008. Extensive proliferation of transposable elements in heritable bacterial symbionts. J Bacteriol 190:777–779. https://doi.org/10.1128/JB.01082-07.

31. Burke GR, Moran NA. 2011. Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. Genome Biol Evol 3:195–208. https://doi.org/10.1093/gbe/evr002.

32. Schmitz-Esser S, Penz T, Spang A, Horn M. 2011. A bacterial genome in transition—an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont Amoebophilus asiaticus. BMC Evol Biol 11:270. https://doi.org/10.1186/1471-2148-11-270.

33. Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, Vallier A, Jackson DG, Latorre A, Weiss RB, Heddi A, Moya A, Dale C. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. Genome Biol Evol 6:76–93. https://doi.org/10.1093/gbe/evt210.

34. Tamas I, Klasson L, Canbäck B, Näslund AK, Eriksson AS, Wernegreen JJ, Sandström JP, Moran NA, Andersson SGE. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376–2379. https://doi.org/10.1126/science.1071278.

35. Newton ILG, Bordenstein SR. 2011. Correlations between bacterial ecology and mobile DNA. Curr Microbiol 62:198–208. https://doi.org/10.1007/s00284-010-9693-3.

36. Sachs JL, Essenberg CJ, Turcotte MM. 2011. New paradigms for the evolution of beneficial infections. Trends Ecol Evol 26:202–209. https://doi.org/10.1016/j.tree.2011.01.010.

37. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017. The evolution of host-symbiont dependence. Nat Commun 8:15973. https://doi.org/10.1038/ncomms15973.

38. Nikoh N, Hosokawa T, Oshima K, Hattori M, Fukatsu T. 2011. Reductive evolution of bacterial genome in insect gut environment. Genome Biol Evol 3:702–714. https://doi.org/10.1093/gbe/evr064.

39. Kenyon LJ, Meulia T, Sabree ZL. 2015. Habitat visualization and genomic analysis of "Candidatus Pantoea carbekii," the primary symbiont of the brown marmorated stink bug. Genome Biol Evol 7:620–635. https://doi.org/10.1093/gbe/evv006.

40. Haygood MG, Distel DL, Herring PJ. 1992. Polymerase chain reaction and 16S rRNA gene sequences from the luminous bacterial symbionts of two deep-sea anglerfishes. J Mar Biol Assoc 72:149–159. https://doi.org/10.1017/S0025315400048852.

41. Funk DJ, Wernegreen JJ, Moran NA. 2001. Intraspecific variation in symbiont genomes: bottlenecks and the aphid-Buchnera association. Genetics 157:477–489.

42. Abbot P, Moran NA. 2002. Extremely low levels of genetic polymorphism in endosymbionts (Buchnera) of aphids (Pemphigus). Mol Ecol 11:2649–2660. https://doi.org/10.1046/j.1365-294X.2002.01646.x.

43. Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. 2007. Prediction of effective genome size in metagenomic samples. Genome Biol 8:R10. https://doi.org/10.1186/gb-2007-8-1-r10.

44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

45. Tu KC, Waters CM, Svenningsen SL, Bassler BL. 2008. A small-RNA-mediated negative feedback loop controls quorum-sensing dynamics in Vibrio harveyi. Mol Microbiol 70:896–907. https://doi.org/10.1111/j.1365-2958.2008.06452.x.

46. Miyashiro T, Ruby EG. 2012. Shedding light on bioluminescence regulation in Vibrio fischeri. Mol Microbiol 84:795–806. https://doi.org/10.1111/j.1365-2958.2012.08065.x.

47. Urbanczyk H, Ast JC, Kaeding AJ, Oliver JD, Dunlap PV. 2008. Phylogenetic analysis of the incidence of lux gene horizontal transfer in Vibrionaceae. J Bacteriol 190:3494–3504. https://doi.org/10.1128/JB.00101-08.

48. Sabater-Muñoz B, Toft C, Alvarez-Ponce D, Fares MA. 2017. Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. ISME J 11:1291–1304. https://doi.org/10.1038/ismej.2017.18.

49. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. ISME J 8:1553–1565. https://doi.org/10.1038/ismej.2014.60.

50. Sun Z, Blanchard JL. 2014. Strong genome-wide selection early in the evolution of Prochlorococcus resulted in a reduced genome through the loss of a large number of small effect genes. PLoS One 9:e88837. https://doi.org/10.1371/journal.pone.0088837.

51. Gorriti MF, Dias GM, Chimetto LA, Trindade-Silva AE, Silva BS, Mesquita MM, Gregoracci GB, Farias ME, Thompson CC, Thompson FL. 2014. Genomic and phenotypic attributes of novel salinivibrios from stromatolites, sediment and water from a high altitude lake. BMC Genomics 15:473. https://doi.org/10.1186/1471-2164-15-473.

52. Sloan DB, Moran NA. 2013. The evolution of genomic instability in the obligate endosymbionts of whiteflies. Genome Biol Evol 5:783–793. https://doi.org/10.1093/gbe/evt044.

53. Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J, Ferriera S, Lapidus A, Anderson I, Kyrpides N, Munk AC, Detter C, Han CS, Brown MV, Robb FT, Kjelleberg S, Cavicchioli R. 2009. The genomic basis of trophic strategy in marine bacteria. Proc Natl Acad Sci U S A 106:15527–15533. https://doi.org/10.1073/pnas.0903507106.

54. 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. Proc Natl Acad Sci U S A 102:3004–3009. https://doi.org/10.1073/pnas.0409900102.

55. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. Nat Genet 32:402–407. https://doi.org/10.1038/ng986.

56. Maezawa K, Shigenobu S, Taniguchi H, Kubo T, Aizawa S, Morioka M. 2006. Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium Buchnera aphidicola sp. strain APS. J Bacteriol 188:6539–6543. https://doi.org/10.1128/JB.00561-06.

57. Toh H, Weiss BL, Perkin SAH, 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 16:149–156. https://doi.org/10.1101/gr.4106106.

58. Rio RVM, Symula RE, Wang J, Lohs C, Wu YN, Snyder AK, Bjornson RD, Oshima K, Biehl BS, Perna NT, Hattori M, Aksoy S. 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont Wigglesworthia. mBio 3:e00240-11. https://doi.org/10.1128/mBio.00240-11.

59. Brennan CA, DeLoney-Marino CR, Mandel MJ. 2013. Chemoreceptor VfcA mediates amino acid chemotaxis in Vibrio fischeri. Appl Environ Microbiol 79:1889–1896. https://doi.org/10.1128/AEM.03794-12.

60. Nishiyama S, Suzuki D, Itoh Y, Suzuki K, Tajima H, Hyakutake A, Homma M, Butler-Wu SM, Camilli A, Kawagishi I. 2012. Mlp24 (McpX) of Vibrio cholerae implicated in pathogenicity functions as a chemoreceptor for multiple amino acids. Infect Immun 80:3170–3178. https://doi.org/10.1128/IAI.00039-12.

61. Henrissat B, Deleury E, Coutinho PM. 2002. Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? Trends Genet 18: 437–440. https://doi.org/10.1016/S0168-9525(02)02734-8.

62. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2:E69. https://doi.org/10.1371/journal.pbio.0020069.

63. Morin JG, Harrington A, Nealson K, Krieger N, Baldwin TO, Hastings JW. 1975. Light for all reasons: versatility in the behavioral repertoire of the flashlight fish. Science 190:74–76. https://doi.org/10.1126/science.190.4209.74.

64. Ast JC, Dunlap PV. 2005. Phylogenetic resolution and habitat specificity of members of the *Photobacterium phosphoreum* species group. Environ Microbiol 7:1641–1654. https://doi.org/10.1111/j.1462-2920.2005.00859.x.

65. Fukui A, Takami M, Tsuchiya T, Sezaki K, Igarashi Y, Kinoshita S, Watabe S. 2010. Pelagic eggs and larvae of *Coelorinchus kishinouyei* (Gadiformes: Macrouridae) collected from Suruga Bay, Japan. Ichthyol Res 57: 169–179. https://doi.org/10.1007/s10228-009-0144-8.

66. Dunlap PV, Takami M, Wakatsuki S, Hendry TA, Sezaki K, Fukui A. 2014. Inception of bioluminescent symbiosis in early developmental stages of the deep-sea fish, *Coelorinchus kishinouyei* (Gadiformes: Macrouridae). Ichthyol Res 61:59–67. https://doi.org/10.1007/s10228-013-0374-7.

67. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.

68. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

69. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932. https://doi.org/10.1093/bioinformatics/btu739.

70. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029.

71. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.

72. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304. https://doi.org/10.1038/ncomms3304.

73. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088.

74. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.

75. Lin B, Wang Z, Malanoski AP, O'Grady EA, Wimpee CF, Vuddhakul V, Jr, Alves N, 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 2:81–89. https://doi.org/10.1111/j.1758-2229.2009.00100.x.

76. Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, Le Roux F, Mincer T, Polz MF. 2012. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. Science 337:1228–1231. https://doi.org/10.1126/science.1219385.

77. Freed LL. 2018. Characterization of the bioluminescent symbionts from ceratioids collected in the Gulf of Mexico. Masters thesis. Halmos College of Natural Sciences and Oceanography, Nova Southeastern University.