

Genomic signatures of obligate host dependence in the luminous bacterial symbiont of a vertebrate

Tory A. Hendry,^{1*†} Jeffrey R. de Wet² and Paul V. Dunlap¹

¹Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University Ave., Ann Arbor, MI 48109-1048, USA.

²Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, 100 Washtenaw Ave., Ann Arbor, MI 48109-2218, USA.

Summary

The majority of bacteria engaged in bioluminescent symbiosis are environmentally acquired and facultatively symbiotic. A few enigmatic bioluminescent symbionts have not been successfully cultured, which has led to speculation that they may be obligately dependent on their hosts. Here, we report the draft genome of the uncultured luminous symbiont of an anomalopid flashlight fish, '*Candidatus Photodesmus katoptron*'. The genome of the anomalopid symbiont is reduced by 80% compared with close relatives and lacks almost all genes necessary for amino acid synthesis and for metabolism of energy sources other than glucose, supporting obligate dependence on the host for growth. '*Candidatus Photodesmus katoptron*' is the first described obligate mutualistic symbiont of a vertebrate. Unlike most other obligate mutualists, the anomalopid symbiont genome has retained complete pathways for chemotaxis and motility as well as most genes involved in cell wall production, consistent with the hypothesis that these bacteria may be transmitted environmentally during an extra-host phase.

Introduction

Bioluminescent symbiosis is a mutualistic interaction in which hosts benefit from bacterial light production and provide nutrients to bacteria within specialized structures

called light organs. This type of association has evolved numerous times and is found in over 460 fish species from 21 families, but only a few known species of luminous bacteria take part in bioluminescent symbiosis (Herring and Morin, 1978; Dunlap *et al.*, 2007; Dunlap, 2009). All described luminous symbionts are members of the diverse *Gammaproteobacteria* family *Vibrionaceae* (Dunlap, 2009). Like other *Vibrionaceae*, luminous bacteria typically live in a variety of habitats and maintain large population sizes (Lee and Ruby, 1994; Thompson *et al.*, 2005; Reen *et al.*, 2006; Dunlap *et al.*, 2007; 2012; Preheim *et al.*, 2011). The described species known to engage in bioluminescent symbiosis are facultatively symbiotic and are environmentally acquired by hosts; each host generation acquires symbionts from free-living bacterial populations (Leis and Bullock, 1986; Lee and Ruby, 1994; Nyholm and McFall-Ngai, 2004; Dunlap *et al.*, 2007; 2008; 2012). Others exist, however, such as the bacteria symbiotic with anomalopid flashlight fish, for which little is known about their ecology or transmission because the bacteria cannot be cultured in laboratory media and because their hosts are relatively difficult to obtain (Morin *et al.*, 1975; Herring and Morin, 1978; Haygood and Distel, 1993). Flashlight fish are found in tropical locations around the world but are only common in a few locations. They are nocturnal and highly photophobic; during the night they swim in schools and use the bright light produced by bacteria in two under-eye light organs to hunt and avoid predators, and during the day they remain in groups in caves and crevices (Morin *et al.*, 1975; Herring and Morin, 1978). Although many attempts have been made to culture the luminous bacteria associated with these fish, such attempts have been unsuccessful, leading to the hypothesis that the bacteria may be obligately dependent on their hosts (Haygood, 1993).

Over time, obligate bacterial symbionts (here, the term symbiont is used exclusively in reference to mutualistic bacteria, not pathogens) tend to accumulate genomic changes compared with free-living relatives or facultative symbionts (Moran and Plague, 2004; McCutcheon and Moran, 2012). The presences of these genomic changes can be used to test for obligate dependence in symbionts with unknown ecologies, such as anomalopid symbionts. Because facultatively symbiotic luminous bacteria

Received 13 June, 2013; accepted 1 October, 2013. *For correspondence. E-mail thendry@umich.edu; Tel. (+1) 518 744 2890; Fax (+1) 734 763 0544. †Present address: The School of Plant Sciences, University of Arizona, 1140 E. South Campus Dr., Tucson, AZ 857210036, USA.

frequently live in a variety of habitats in addition to being symbiotic, their genomes are very similar in size and gene content to free-living, non-symbiotic relatives (Reen *et al.*, 2006). This is in contrast with the pattern commonly found across many independently evolved host-restricted bacterial lineages, such as some insect endosymbionts. These obligate, intracellular symbionts experience decreased selection on genes needed only outside the host (McCutcheon and Moran, 2012). Additionally, population bottlenecks during vertical transmission between host generations decrease the effective population size of symbiont lineages, leading to high levels of genetic drift and frequent nucleotide deletions and substitutions (Moran, 1996; Mira and Moran, 2002). As a result, the genomes of obligate symbionts frequently show a suite of evolutionary patterns including genome reduction, increased evolutionary rates, and high adenine-thymine (AT) substitution biases compared with relatives (Moran, 1996; Wernegreen and Moran, 1999; Woolfit and Bromham, 2003; Wernegreen and Funk, 2004).

Although anomalopid symbionts differ from other luminous symbionts in having not yet been successfully cultured, they share ecological similarities to other luminous symbionts. For instance, anomalopid symbionts are not intracellular or physically restricted to the host as obligate symbionts often are (Kessel, 1977; Haygood, 1993). Like other luminous symbionts, they are regularly released into seawater and remain intact and luminous for at least a period of hours (Haygood *et al.*, 1984). Additionally, multiple lines of evidence suggest that anomalopid symbionts are acquired from the environment, as occurs in other luminous symbionts, rather than transmitted vertically as would be predicted for obligate bacterial symbionts (Haygood, 1993). Here, we use the genome sequence of the bacterium '*Candidatus* Photodesmus katoptron,' the symbiont of the flashlight fish *Anomalops katoptron*, to test two long-standing hypotheses regarding flashlight fish symbiont evolutionary ecology – obligate host dependence and environmental transmission – by comparing it with the sequenced genomes of relatives (Hendry and Dunlap, 2011). If the anomalopid symbiont is obligately dependent on the host, then we expect to observe genome reduction through gene loss, particularly of genes used in nutrient synthesis because the host should supply nutrients to the symbiont. We also predict that the presence or absence of pathways within the genome will provide insight into which nutrients are provided to the bacteria by the host; pathways missing from the symbiont should need products supplemented by the host. Additionally, we use the genome sequence presented here to test the hypothesis that anomalopid symbionts are acquired from an environmental phase, as some genes, including those used in motility and cell wall

synthesis, are frequently lost in symbionts without an extra-host phase.

Results and discussion

Genome size

Consistent with the obligate host dependence hypothesis, the *de novo* assembled '*Ca. Photodesmus katoptron*' draft genome sequence is greatly reduced in size compared with relatives, although its gene content is similar to that of relatives. The 1 Mb '*Ca. Photodesmus katoptron*' genome is approximately one fifth the size of other *Vibrionaceae* genomes (mean: 4.8 Mb; range: 3.9–6.1 Mb). The assembly consists of a 14.5 kb complete circular predicted plasmid similar to relatives, an unfinished 0.56 Mb scaffold similar to *Vibrionaceae* chromosome I, an unfinished 0.43 Mb scaffold and six small contigs containing rRNA genes. The assembly of two large scaffolds, one of which has many genes with high similarity to those typically found on *Vibrionaceae* chromosome I, suggests that like other *Vibrionaceae* members, the anomalopid symbiont may have two chromosomes (Fig. 1) (Reen *et al.*, 2006). The '*Ca. Photodesmus katoptron*' genome is predicted to contain 903 genes as well as eight pseudogenes. Of the apparently functional genes 855 are protein coding genes, ten are rRNA genes (three 23S, four 16S and three 5S genes), and 38 are tRNA genes. Because the draft genome includes rRNA genes that are not fully assembled into scaffolds, this prediction may be an underestimate of the number of rRNA genes. In contrast, sequenced *Vibrionaceae* genomes range from 3839 to 6237 genes (Fig. 2). This assembly of the '*Ca. Photodesmus katoptron*' genome has a minimum read coverage of 70× and an average coverage of 165×, and contains predicted genes for all complete pathways considered necessary for life and used to test completeness of genome coverage (Table S2) (Raes *et al.*, 2007). Therefore, the low number of genes recovered is likely not due to insufficient sequencing. It is also unlikely that the '*Ca. Photodesmus katoptron*' genome represents the ancestral genome size of the family, as phylogenetic analyses do not resolve anomalopid symbionts as basal within the group (Fig. 3) (Hendry and Dunlap, 2011). The most parsimonious explanation for the small '*Ca. Photodesmus katoptron*' genome is that it has been reduced in size by gene loss. Such gene loss is commonly found in obligate symbionts and is consistent with obligate dependence of '*Ca. Photodesmus katoptron*' on the host. Additionally, the AT content of the '*Ca. Photodesmus katoptron*' genome (69%) is higher than the range typically found in *Vibrionaceae* (55–59%), another similarity to obligate endosymbionts (values taken from GenBank) (Benson *et al.*, 2013).

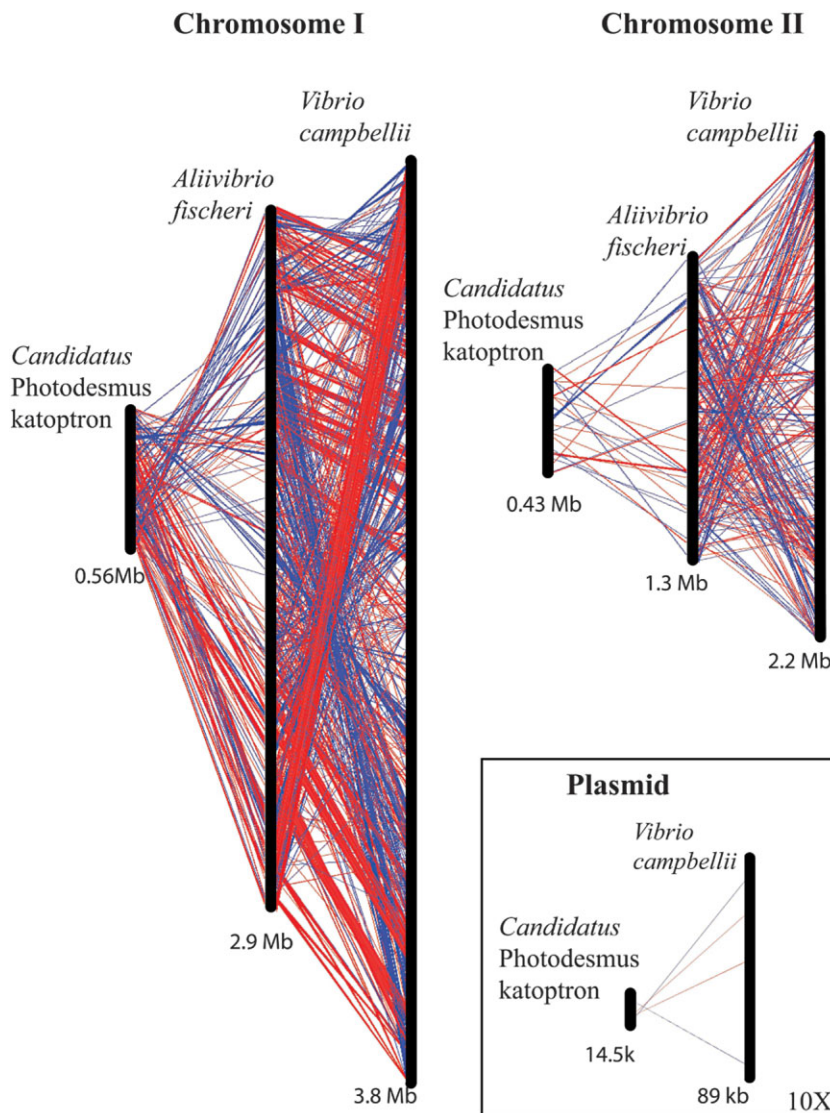


Fig. 1. Chromosome comparisons between 'Ca. Photodesmus katoptron' and *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116). Chromosomes are drawn to scale with alignments as determined by the Artemis comparison tool (Carver *et al.*, 2005). Each connecting line represents an area of high nucleotide sequence similarity, red lines indicate alignment of two positive strands, and blue lines represent similarity between positive and negative strands. Homology was determined with an *e* value cut-off of 1.0. Plasmid representations for 'Ca. Photodesmus katoptron' and *V. campbellii* (ATCC BAA-1116) are shown on a different scale.

Gene loss

In comparison with the close relatives *Aliivibrio fischeri*, a facultatively symbiotic luminous species, and the non-symbiotic luminous species *Vibrio campbellii* (previously classified as *V. harveyi*) (Lin *et al.*, 2010), 'Ca. Photodesmus katoptron' lacks genes from all functional categories analysed (Fig. 4). Categories with the smallest amount of loss were those likely to be necessary regardless of symbiotic state, such as genes involved in transcription, translation, and DNA synthesis (Fig. 4). As previous work has shown, genes needed for the symbiotic interaction, the luminescence genes, are also retained (Hendry and Dunlap, 2011). In contrast, metabolic categories of genes, including amino acid synthesis and energy metabolism genes, are some of the most reduced. Loss of metabolic genes is common in obligate symbionts, as

hosts provide nutrients to the bacteria, which allows the genes involved in the synthesis of these nutrients to be lost (McCutcheon and Moran, 2012), so this finding provides further support of obligate host dependence.

Genes required for amino acid synthesis show one of the most dramatic reductions compared with free-living relatives (Fig. 4, Table S3), indicating that 'Ca. Photodesmus katoptron' cannot synthesize a full complement of amino acids and must therefore acquire them from its host. Only three amino acid synthesis pathways appear to be complete in 'Ca. Photodesmus katoptron' (Fig. 5). Of the remaining amino acid synthesis genes, 13 of 18 are found in pathways that produce intermediate compounds used in other synthetic pathways, the synthesis of folate, ubiquinone and peptidoglycan, for which 'Ca. Photodesmus katoptron' has retained pathways (Fig. 5). This suggests that these 13 genes may have been

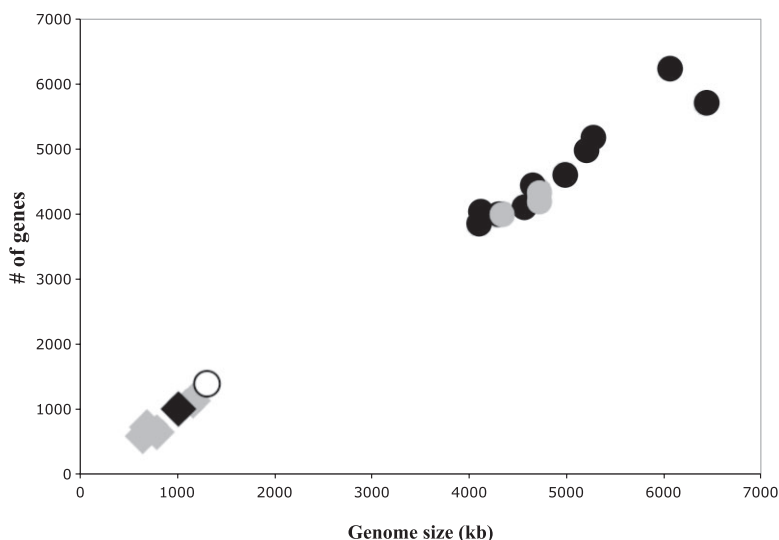


Fig. 2. Graph of genome size versus number of genes. '*Ca. Photodesmus katoptron*' (black diamond) has a genome similar in size to intracellular obligate symbionts ('*Ca. Baumannia cicadellicola*' Hc, '*Ca. Blochmannia pennsylvanicus*' BPEN, *Buchnera aphidicola* APS, '*Ca. Ruthia magnifica*' Cm, and *W. glossinida*), shown in grey diamonds. '*Candidatus Pelagibacter ubique*' SAR11 HTCC1062, often considered to have the smallest genome of a free-living bacterium, is shown in the white circle. Free-living representatives of *Vibrionaceae* [*A. fischeri* ES114, *A. salmonicida* LF11238, *P. profundum* SS9, *V. anguillarum* 775, *V. campbellii* ATCC BAA-1116, *V. cholera* O395, *V. parahaemolyticus* RIMD 2210633, *V. splendidus* LGP32, *V. vulnificus* YJ016 and the facultatively symbiotic fish symbiont *P. mandapamensis* svers.1.1 (Urbanczyk *et al.*, 2011)] are shown in black circles. Other free-living bacteria (*E. coli* K12 DH1, *Roseobacter denitrificans* OCh 114, and *Yersinia pestis* KIM 10) are shown in grey circles.

retained for reasons other than amino acid synthesis. This gene loss contrasts with other *Vibrionaceae*, which can synthesize a full complement of amino acids, and it is striking compared with obligate and intracellular bacteria, which can typically synthesize three or more amino acids, even if they do not provide amino acids to the host (Yu *et al.*, 2009). The inability to synthesize amino acids implies that the fish host must be providing these nutrients to its symbiont in the light organ environment. The fact that '*Ca. Photodesmus katoptron*' has retained some predicted amino acid transport proteins is consistent with this hypothesis (Table S3).

Energy metabolism genes are reduced in '*Ca. Photodesmus katoptron*' compared with relatives (Fig. 4). '*Candidatus Photodesmus katoptron*' has retained complete pathways for glycolysis and the citric acid cycle, the genes for all necessary components of ATP synthase, and numerous genes involved in electron transport (Table S2, Table S3). However, compared with *A. fischeri* and *V. campbellii*, '*Ca. Photodesmus katoptron*' has lost many genes necessary to catabolize amino acids and carbohydrates other than glucose (Table S3). Additionally, the '*Ca. Photodesmus katoptron*' genome contains only one set of phosphotransferase system (PTS) genes, which are used to transport sugar sources (Table S3). Based on similarity, these PTS genes are specific to glucose. Typically, members of *Vibrionaceae* are metabolically diverse and can utilize many carbon/energy sources (Reen *et al.*, 2006). The loss of energy metabolism genes and lack of other sugar transport genes suggests that '*Ca. Photodesmus katoptron*' can only utilize glucose as a carbon/energy source, indicating that the host is providing glucose to the symbiont. Glucose has been previously suggested, although not demonstrated, as a likely carbohydrate for fish to provide luminous symbionts (Nealson,

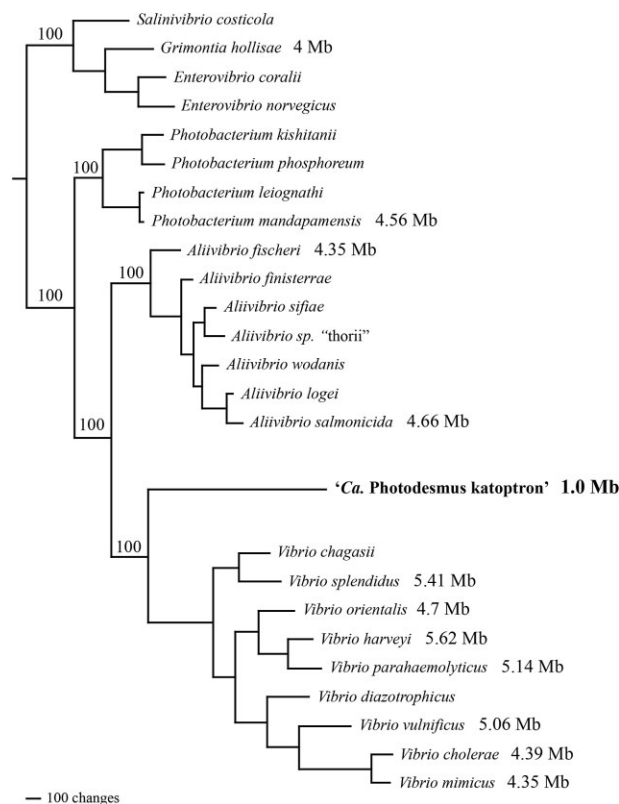


Fig. 3. Phylogenetic tree with genome sizes of '*Ca. Photodesmus katoptron*' and relatives. The tree is modified from Hendry and Dunlap (2011). Briefly, the tree was obtained from maximum likelihood analysis using eight housekeeping genes (16S, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA* and *topA*) and six luminescence genes (*luxCDABEG*). Bootstrap values >70% are shown above nodes. Genome sizes are shown to the right of species for which the data are available. Genome sizes were taken from GenBank (Benson *et al.*, 2013) and are averages when a sequenced genome was available for more than one strain per species.

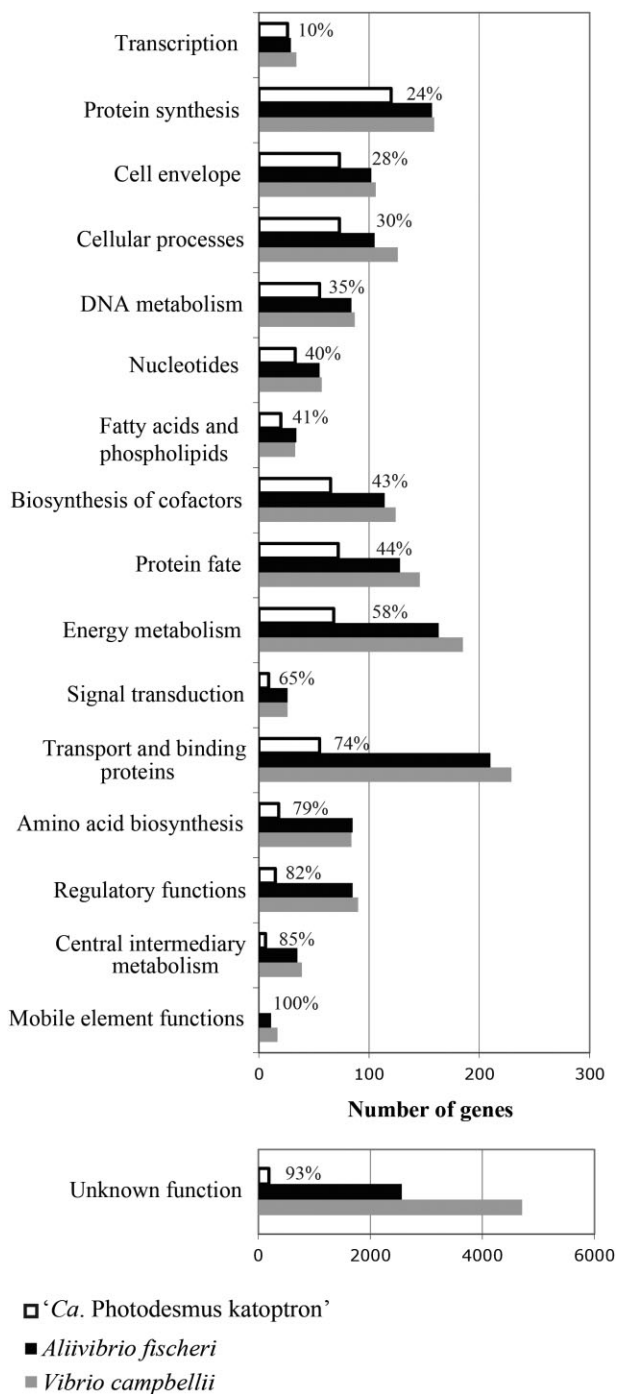


Fig. 4. Genes in functional categories for 'Ca. Photodesmus katoptron,' *A. fischeri* and *V. campbellii*. Genes in TIGRFams for *A. fischeri* (ES114) and *V. campbellii*. (ATCC BAA-1116) were obtained from IMG database (Markowitz *et al.*, 2012). Percentages show the amount of reduction in the 'Ca. Photodesmus katoptron' genome compared with *A. fischeri*. Genes of unknown function are shown on a separate scale.

1979). The dependence of 'Ca. Photodesmus katoptron' on glucose and amino acids is evidence for an obligate relationship with the host; marine environments are typically nutrient poor, and habitats other than the host light organ are unlikely to provide a consistent supply of these nutrients. This evidence of nutrient dependence in 'Ca. Photodesmus katoptron' provides the first indication of which nutrients symbiotically luminous fish provide to their symbiotic bacteria.

Genomic signatures of adaptation to a restricted environment

Hierarchical clustering analyses based on the gene content, in contrast with phylogenetic analysis, cluster the 'Ca. Photodesmus katoptron' genome more closely with oligotrophic (adapted to stable, low-nutrient conditions) or intracellular bacteria than with copiotrophic (adapted to variable, high-nutrient conditions) relatives (Fig. 6). In gene content comparisons of 'Ca. Photodesmus katoptron' with the genomes of various species representing different ecological lifestyles (obligate endosymbionts, facultative copiotrophic symbionts, non-symbiotic copiotrophs and non-symbiotic oligotrophs) taxa grouped together based on their characterization as obligate symbionts, oligotrophs or copiotrophs rather than based on phylogenetic relationships. This is consistent with known genomic patterns within these groups (Lauro *et al.*, 2009). 'Candidatus Photodesmus katoptron' is more similar in gene content to unrelated obligate symbionts or free-living oligotrophs than to close relatives, both symbiotic and not. Compared with close relatives, the 'Ca. Photodesmus katoptron' genome displays characteristics linked to slow growth in a stable environment. These traits include a low number of rRNA operons, a small number of transport proteins, and low numbers of regulatory and signal transduction genes (Fig. 7) (Lauro *et al.*, 2009). These signatures of adaptation to a stable environment are predicted to occur if 'Ca. Photodesmus katoptron' exists predominantly in the host light organ rather than in multiple habitats, such as in ocean water or associated with other hosts.

Additional support for restriction to a stable environment comes from the types of cell surface proteins found in 'Ca. Photodesmus katoptron,' which also has implications for the way the symbiont interacts with the fish host. Both facultative and obligate symbionts, including *A. fischeri*, use surface-binding pilin proteins and secretion proteins for transfer to new hosts (Dale *et al.*, 2001; Nyholm and McFall-Ngai, 2004; Ruby *et al.*, 2005). 'Candidatus Photodesmus katoptron', however, has lost all pilin genes found in relatives as well as many secretion systems (Table S3). The lack of pilin genes implies that not only is 'Ca. Photodesmus katoptron' interacting with a

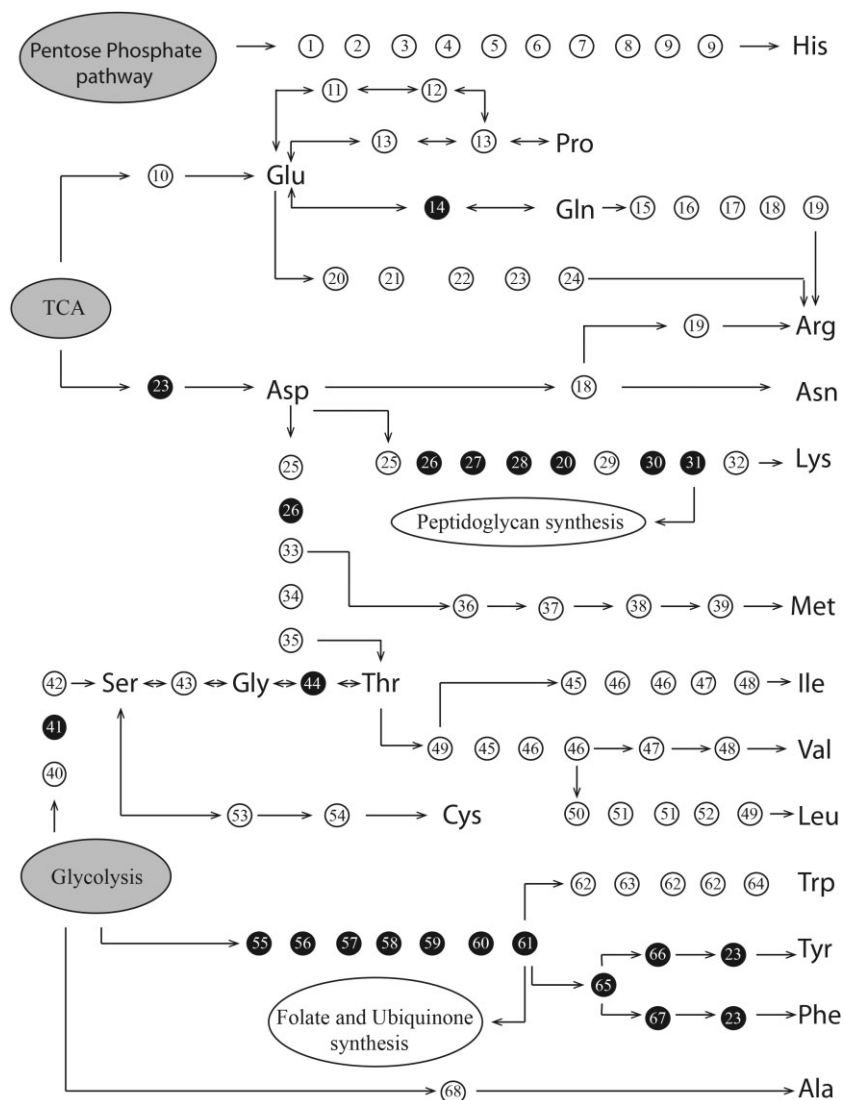


Fig. 5. Amino acid synthesis pathways showing gene loss in 'Ca. Photodesmus katoptron'. Pathways for the synthesis of the 20 standard amino acids encoded by the universal genetic code were taken from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database for taxa *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116) (Kanehisa *et al.*, 2012). Each circle represents an enzyme. Numbers within circles correspond to KEGG pathway database Enzyme Commission (EC) numbers and can be found matched to gene names in the Table S3. Darkened circles represent enzymatic transformations that 'Ca. Photodesmus katoptron' is theoretically capable of carrying out based on gene presence, whereas white circles indicate missing enzymatic steps. Grey ovals show pathways that contribute starting products for amino acid synthesis. White ovals show other pathways that utilize amino acid synthesis genes. Pathways for which 'Ca. Photodesmus katoptron' has retained a majority of genes (synthesis of lysine, tyrosine and phenylalanine) are also used in the synthesis of peptidoglycan, folate and ubiquinone.

limited number of surface types and organisms but also that it might use novel means of colonizing new host generations.

Retention of genes possibly needed outside the host

Certain groups of genes frequently lost in obligate symbionts and possibly linked to persistence outside of the host remain mostly intact in 'Ca. Photodesmus katoptron'. These include genes needed for chemotaxis, motility and production of the cell envelope. Of the functional categories of genes most reduced in obligate symbionts, 'Ca. Photodesmus katoptron' has retained a surprising number of genes involved in cellular processes and synthesis of the cell envelope and related proteins (Fig. 7A). A more fine-scale analysis reveals that the majority of cellular process genes retained in the anomalopid symbiont are those needed for chemotaxis

and motility. Specifically, 'Ca. Photodesmus katoptron' has retained the exact gene complements used for flagellar synthesis and chemotaxis in relatives *V. campbellii* and *A. fischeri* (Fig. 7B, Table S3). The retention of these genes suggests that 'Ca. Photodesmus katoptron' is capable of both producing a functional flagellum and navigating based on chemical attractants and/or repellants, and is consistent with both earlier reports of anomalopid symbiont motility in seawater and electron micrographs of the bacteria showing polar flagella (Kessel, 1977; Haygood *et al.*, 1984). In contrast, motility genes are frequently lost in obligate symbionts (Fig. 7B), including the extracellular stinkbug symbiont *Ishikawaella capsulata*, which is not active outside of the host (Kikuchi *et al.*, 2009). However, notable exceptions exist in cases where chemotaxis and motility genes have been retained for transmission or host interaction. For example, *Buchnera aphidicola* has retained flagellar genes that are apparently

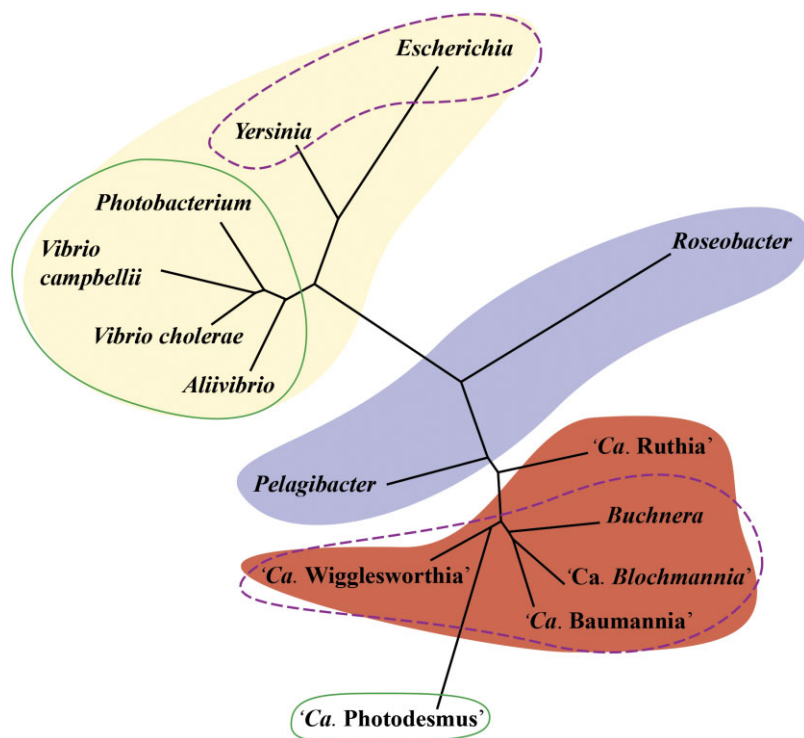


Fig. 6. Genomic clustering of '*Ca. Photodesmus katoptron*' and representatives of various ecological lifestyles. This hierarchical clustering dendrogram is based on gene presence/absence, not phylogenetic relationships between taxa. Shaded areas show bacteria with similar lifestyles (red: obligate intracellular symbionts; blue: free-living oligotrophs; and yellow: free-living copiotrophs, symbiotic and not). Outlines show phylogenetic relationships of taxonomically classified *Gammaproteobacteria* (purple dashed: family *Enterobacteriaceae*; green solid: family *Vibrionaceae*). The symbiont *Candidatus Ruthia magnifica* is an unclassified member of *Gammaproteobacteria*. Oligotrophic taxa, *Roseobacter* and *Pelagibacter*, are *Alphaproteobacteria* classified in separate families and distantly related to other taxa analysed.

used for protein secretion (Maezawa *et al.*, 2006; Toft and Fares, 2008). In another example, the tsetse fly symbiont *Wigglesworthia glossinidia* has retained functional flagella that are necessary for transmission to new host generations (Rio *et al.*, 2012). Chemotaxis genes are almost universally lost in obligate symbionts; to our knowledge, they have only been found in the secondary symbiont of the tsetse fly, *Sodalis glossinidius*, where they are thought to be necessary for a portion of the symbiont's within host life cycle (Toh *et al.*, 2012).

In the category of genes involved in cell envelope synthesis, '*Ca. Photodesmus katoptron*' has retained 86% of genes used in relatives for peptidoglycan synthesis and 90% of genes for synthesis of lipopolysaccharide, and appears to have complete pathways to synthesize these two major components of the cell wall (Table S3). The retention of cell wall synthesis genes is consistent with previous electron micrographs detailing cell wall components (Kessel, 1977). These categories are both reduced in other obligate symbionts, even the extracellular *I. capsulata*, and their retention suggests that '*Ca. Photodesmus katoptron*' may be able to synthesize a robust cell wall, a feature shown to be lacking in many obligate symbionts (McCutcheon and Moran, 2012). The loss of cell wall synthesis genes in *I. capsulata* suggests that it is not the extracellular lifestyle of the anomalopid symbiont alone that may have selected for retention of cell wall synthesis genes but perhaps something specific to the '*Ca. Photodesmus katoptron*' habitat or lifestyle.

Given the amount of gene loss in the anomalopid symbiont, it is unlikely that complete gene pathways would have been retained if they were not required by the bacterium. Chemotaxis and flagellar genes are unlikely to be used for motility within the light organ because of the high density of bacterial cells (Kessel, 1977). One possibility is that '*Ca. Photodesmus katoptron*' uses these genes for a portion of its lifecycle within the host or for host interaction, as is seen in the *Wigglesworthia* example mentioned earlier. Alternatively, it is possible that chemotaxis and motility genes are used outside of the host. Haygood and colleagues (1984) previously showed that symbionts are regularly released from pores on anomalopid light organ surfaces and persist in seawater for at least several hours. The retention of chemotaxis, motility and cell wall genes may be related to the symbiont's ability to persist and navigate outside of the host. Given the genome reduction demonstrated here, however, '*Ca. Photodesmus katoptron*' seems unlikely to establish significant populations outside of the host. Therefore, if chemotaxis and motility genes are necessary outside of the host, these traits may be used to search for new hosts rather than to establish free-living populations.

Conclusions and evolutionary speculations

'*Candidatus Photodesmus katoptron*' represents the first mutualistic bacterium symbiotic with a vertebrate and the first luminous symbiont to show a strong genomic pattern

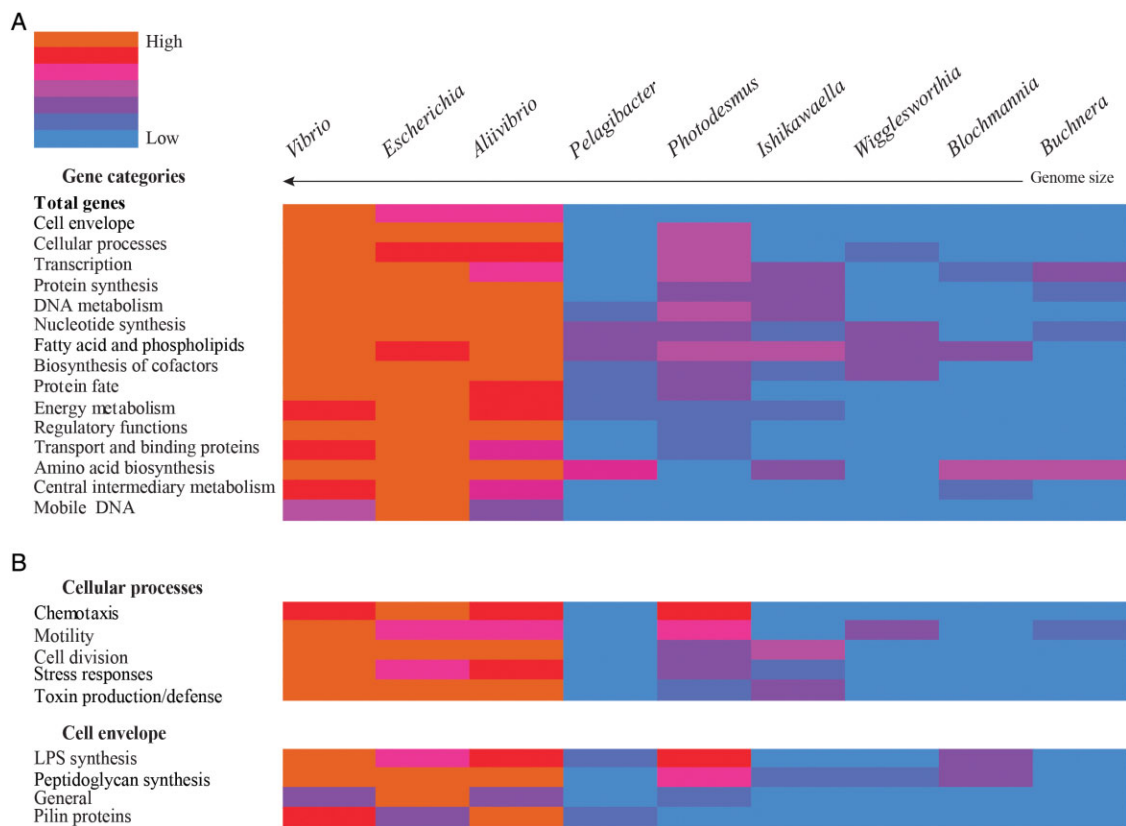


Fig. 7. Gene content of ‘*Ca. Photodesmus katoptron*’ compared with representatives with different ecologies. Numbers of genes in each category are represented by colours along a red-to-blue spectrum. Copiotrophic species are represented by *V. campbellii*, *E. coli* and *A. fischeri*. *Pelagibacter ubique* is oligotrophic. Obligate symbionts included here are *I. capsulata*, *W. glossinidia*, *Blochmannia pennsylvanicus* and *Buchnera aphidicola*. *Ishikawaella capsulata* is extracellular, whereas other obligate species are intracellular. A. Total gene number and broad functional categories. B. Subcategories within cellular processes and cell envelope synthesis.

of obligate host dependence. This genomic pattern is consistent with previous suggestions of obligate dependence such as the increased evolutionary rate found in anomalopid symbionts (Hendry and Dunlap, 2011). This work also highlights how comparative genomics can be used to infer the ecology of uncultured bacteria. The loss of pathways from the ‘*Ca. Photodesmus katoptron*’ genome suggests which nutrients are likely provided to the symbiont by the host; this is the first observation of this sort in symbiotically luminous fish. Some of these nutrients are different and more energy-rich than those that are provided to *A. fischeri* by the squid *Euprymna scolopes*, the only other bioluminescent symbiosis for which data are available. Although *E. scolopes* also seems to supply symbiotic bacteria with at least some amino acids (Graf and Ruby, 1998), the main carbon/energy sources provided by the squid appear to be glycerol and chitin (Wier *et al.*, 2010), carbohydrates that are more difficult to breakdown and may provide less net energy than glucose. The extent of nutrient provisioning by the host has implications for the evolution of depend-

ence, and we speculate that high-quality nutrient provisioning of ‘*Ca. Photodesmus katoptron*’ by the host may have allowed for initial gene loss and initiation of obligate dependence.

Another insight from these genomic comparisons is that the anomalopid symbionts have retained some pathways commonly lost in obligate symbionts undergoing genome reduction, such as chemotaxis and motility genes. The amount of genome reduction observed in ‘*Ca. Photodesmus katoptron*’ as well as the specific loss of amino acid synthesis and energy metabolism capabilities is consistent with obligate dependence on the host for growth. However, the retention of genes commonly lost in obligate symbionts, chemotaxis, motility and cell wall genes, suggests a specific ecological use for these genes in the anomalopid symbiont, possibly an important extra-host phase of their life cycle. Previous work, including failed attempts to discover vertical transmission of symbionts and studies of larval development, lead to the hypothesis that like other luminous symbionts, ‘*Ca. Photodesmus katoptron*’ is acquired by larval fish from

populations released from the light organs of adults (Haygood, 1993; Haygood and Distel, 1993). Polymerase chain reaction (PCR) assays, which are able to detect symbionts from the immediate surroundings of adult fish, have been negative for symbionts in both the gonads and ripe gametes of adults, contradicting the possibility of direct vertical inheritance (Haygood, 1993). Furthermore, studies of larvae demonstrate that like other symbiotically luminous fish, larval anomalopids hatch without light organs and that the structures form later in development (Leis and Bullock, 1986; Colin, 1989; Baldwin and Johnson, 1995; Konishi and Okiyama, 1997; Dunlap *et al.*, 2008; 2009; 2012). The timing of light organ development indicates that hosts likely cannot be colonized until at least a week after hatching. The genomic findings of this study are supportive of the hypothesis that the symbionts are transmitted during an extra-host phase to larvae post-hatching. However, it is also possible that the symbionts are vertically transmitted through a currently unknown mechanism.

If the symbionts are transmitted outside of the host, it raises the question of how obligate host dependence and genome reduction evolved in this system without an obvious form of vertical transmission, which is common to all other known obligate symbionts (Bright and Bulgheresi, 2010; Sachs *et al.*, 2011). Vertical transmission may take place directly via gametes or indirectly through contact with or provisioning by adults (Bright and Bulgheresi, 2010; Sachs *et al.*, 2011). In the latter case, pseudovertical transmission, obligate symbionts may be inactive in the environment before host colonization, such as with the bacterial capsules provided with the eggs of some stinkbug species (Kikuchi *et al.*, 2009; Hosokawa *et al.*, 2010). Alternatively, close proximity of adults and offspring may allow for transmission of active bacteria without dispersal, as with entomopathogenic nematodes and symbiotic *Xenorhabdus nematophila*, which is acquired inside insect hosts (Bright and Bulgheresi, 2010). It should be noted, however, that pseudovertical transmission does not necessarily lead to genomic reduction in a symbiont. For instance, *X. nematophila* does not have a reduced genome.

An intriguing possibility suggested by Haygood is that the anomalopid behaviour of forming aggregations in caves during the day may allow their symbionts to build up higher density in the water surrounding fish populations and colonize larval fish without having to disperse or persist for extended periods in open water (Morin *et al.*, 1975; Meyer-Rochow, 1976; Haygood, 1993; Baldwin and Johnson, 1995). This hypothesis has been strengthened by observations indicating that anomalopid eggs are not pelagic but likely remain near adults and that larvae are rarely found away from adults (Meyer-Rochow, 1976; Baldwin and Johnson, 1995). The development of larvae around a high density of adults would ensure that

symbionts regularly encounter new hosts and it could lessen the selection on symbiont genes needed in non-host environments. This hypothesized transmission mode would mimic vertical transmission but would lead to symbiont-host codivergence at the level of the population rather than the individual and is testable using population genetic approaches. This mode of transmission could still lead to small symbiotic populations and population bottlenecks between host generations, factors that have been found to increase gene loss because of genetic drift in intracellular symbionts (Moran, 1996; Mira and Moran, 2002). The importance of population bottlenecks on the evolution of the bacteria could also be tested using population genetic approaches.

To our knowledge, other possible instances of genome reduction because of local enrichment of symbionts have not been reported, although other symbionts are also likely to build up density in the environment surrounding adults. For example, an increase in local symbiont density has been reported for the luminous symbiont *A. fischeri* in the habitat of adult and hatchling squid (Lee and Ruby, 1994). A key difference between these two systems may be that the anomalopid host provides higher quality carbohydrates to symbionts than the squid host, making the light organ environment better relative to other possible environments and allowing for the loss of genes not needed in the light organ, such as genes needed to metabolize carbohydrates other than glucose. Findings from the '*Ca. Photodesmus katoptron*' genome suggest that other factors besides transmission mode, such as high-nutrient provisioning by the host, may play a role in causing genome reduction in cases without a fairly direct form of vertical transmission.

Experimental procedures

Sequencing and assembly

DNA for genome sequencing came from five light organs of four fish specimens of *Anomalops katoptron* caught near the island of Cebu in the Philippines. DNA was extracted as in Hendry and Dunlap (2011), and combined for sequencing. Sequencing was performed by the DNA Sequencing Core at the University of Michigan on half of a plate of a Roche 454FLX Genome Sequencer. Reads were assembled in MIRA3 (Chevreux *et al.*, 1999) into 23 contigs with at least 70× coverage and an average coverage of 165×. At this coverage cut-off, all contigs represented '*Ca. Photodesmus katoptron*' sequences rather than host or contaminant sequences (based on BLAST searches and nucleotide signatures). Contigs were further assembled by PCR and Sanger sequencing into one small (plasmid) and two large scaffolds (Table S1). Because of the repetitive nature of rRNA operons, six contigs containing rRNA genes could not be assigned to scaffolds. The high level of sequencing coverage as well as the number of intact conserved pathways and genes (Table S2) (Raes *et al.*, 2007) strongly suggests

that the entire genome of 'Ca. Photodesmus katoptron' is represented by these sequences. To ensure that no low coverage contigs were missed, contigs of greater than 2500 bases in length but less than 70× coverage were checked by BLAST; they were found to be contaminant sequences of host DNA. To confirm that the bacterial sequences were not contaminated, all contigs (both greater than and less than 70× coverage) were checked for similarity of genomic nucleotide signatures to each other versus relatives (*Escherichia coli* K12, *V. vulnificus* CMCP6, *Photobacterium profundum* SS9, *V. campbellii* ATCC BAA-1116, *A. fischeri* MJ11, *V. splendidus* LGP32, and *V. cholerae* MJ 1236). All greater than 70× contigs mapped to the 'Ca. Photodesmus katoptron' genome except rRNA and plasmid sequences, which typically have different nucleotide signatures than other areas of a genome (Dick *et al.*, 2009). The 'Ca. Photodesmus katoptron' genome has been deposited in GenBank under the accession AMSD00000000.

Annotation and comparative analyses

The 'Ca. Photodesmus katoptron' genome was annotated by the Institute for Genome Sciences Annotation Engine at the University of Maryland. The annotation was manually curated to confirm start and stop sites by alignment to similar sequences and to identify pseudogenes. If an open reading frame (ORF) represented less than 60% of the length of orthologous sequences, it was classified as a pseudogene. Because rRNA gene sequences are frequently identical between operons, the number of contigs containing rRNA genes does not necessarily equate to the number of rRNA operons. Therefore, the number of rRNA genes was estimated by counting only genes with unique flanking regions. Genes were assigned into TIGRfams categories of genes by the IGS Annotation Engine, and these assignments were also checked manually. Whenever possible genes in the categories of unclassified, unknown function, and hypothetical proteins were placed in functional categories based on the TIGRfam database (Haft *et al.*, 2013). Additionally, all genes already placed in functional categories were checked for accuracy of placement against the TIGRfam database (Haft *et al.*, 2013). Genes of unknown function account for only 21% of the 'Ca. Photodesmus katoptron' genome.

For comparisons of gene content, genes classified into TIGRfams were obtained from genomes in the Integrated Microbial Genomes (IMG) database of the Department of Energy Joint Genome Institute (Markowitz *et al.*, 2012). For *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116), each gene in every category was checked for similarity to genes in 'Ca. Photodesmus katoptron' categories. When a gene was found to be present in the 'Ca. Photodesmus katoptron' genome and not in the TIGRfam genes of *A. fischeri* or *V. campbellii*, the genomes of these species were checked for the gene by BLAST. It is therefore possible that the *A. fischeri* or *V. campbellii* genomes actually contain more possible TIGRfam genes than shown here, but it is not possible a gene is present in 'Ca. Photodesmus katoptron' but missed in the relatives analysed.

For hierarchical clustering analysis, genes in TIGRfams were obtained from IMG for the following species: *A. fischeri* ES114, 'Ca. Baumannia cicadellicola' Hc, 'Ca. Blochmannia

pennsylvanicus' BPEN, *Buchnera aphidicola* APS, *E. coli* K12 DH1, 'Ca. Pelagibacter ubique' SAR11 HTCC1062, *P. profundum* SS9, *Roseobacter denitrificans* OCh 114, 'Ca. Ruthia magnifica' Cm, *V. cholerae* O395, *V. campbellii* ATCC BAA-1116, *W. glossinidia* and *Yersinia pestis* KIM 10 (Markowitz *et al.*, 2012). Genes present in the 'Ca. Photodesmus katoptron' genome were then added to this matrix. Genes present in only one taxon were eliminated. Genes were coded by copy number as absent (0), present with 1–5 copies (1), present with 6–11 copies (2) or present with >11 copies (3). These characters were considered ordered. An unrooted dendrogram was constructed using the neighbour-joining method in PAUP (Swofford, 1998). Other hierarchical clustering methods were also tried, including k-means and average linkage clustering and topology shown in (Fig. 6) was robust.

For comparison of gene content by category between 'Ca. Photodesmus katoptron' and copiotrophic, oligotrophic and obligate bacteria (Fig. 7), genes in TIGRfams were taken from IMG (Markowitz *et al.*, 2012) except in the case of *I. capsulata*, for which numbers were obtained from Nikoh and colleagues (2011). The following species were used: *A. fischeri* ES114, 'Ca. Blochmannia pennsylvanicus' BPEN, *Buchnera aphidicola* APS, *E. coli* K12 DH1, 'Ca. Pelagibacter ubique' SAR11 HTCC1062, *V. campbellii* ATCC BAA-1116, *W. glossinidia*, and *I. capsulata*. All categories were manually checked for missing genes or incorrect category assignments. The heatmap used in Fig. 7 was constructed by dividing up the range of values observed within each category and assigning each division a colour along a red-to-blue colour scale.

Acknowledgements

Specimens were obtained with assistance from LiveAquaria .com, Quality Marine Wholesale Suppliers, and Fish Doctors Aquarium of Ypsilanti, MI, USA. Genome annotation was performed by the Annotation Engine at the Institute for Genome Sciences, University of Maryland School of Medicine. We thank Katherine E. Dougan for assistance manually curating gene annotations. Analysing sequence reads for genomic nucleotide signatures was done by Sunit Jain, and we thank him and Greg J. Dick for assistance. Computational assistance for hierarchical clustering analysis was provided by Susanna M. Messinger. We also thank Greg J. Dick for discussions about the paper. This project was funded by Department of Ecology and Evolutionary Biology Block Grants from the University of Michigan and by Rackham Graduate Student Research Grants from the University of Michigan Rackham Graduate School.

References

- Baldwin, C.C., and Johnson, G.D. (1995) A larva of the Atlantic flashlight fish, *Kryptoptahanaron alfredi* (Beryciformes: Anomalopidae), with a comparison of beryciform and stephanoberyciform larvae. *Bull Mar Sci* **56**: 1–24.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsh-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013) GenBank. *Nucleic Acids Res* **41**: D36–D42.
- Bright, M., and Bulgheresi, S. (2010) A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol* **8**: 218–230.

- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. (2005) ACT: the Artemis comparison tool. *Bioinformatics* **21**: 3422–3423.
- Chevreaux, B., Wetter, T., and Suhai, S. (1999) Genome sequence assembly using trace signals and additional sequence information. *Comput Sci Biol Proc Ger Conf Bioinformatics (GCB)* **99**: pp. 45–56.
- Colin, P.L. (1989) Anomalopidae – lanterneyes, flashlightfishes. In *The Larvae of Indo-Pacific Shoreline Fishes*. Leis, J.M., and Trnksi, T. (eds). Honolulu, HI, USA: University of Hawaii Press, pp. 83–87.
- Dale, C., Young, S.A., Haydon, D.T., and Welburn, S.C. (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *Proc Natl Acad Sci USA* **98**: 1883–1888.
- Dick, G.J., Andersson, A.F., Baker, B.J., Sheri, L., Yelton, A.P., and Banfield, J.F. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.
- Dunlap, P.V. (2009) Microbial bioluminescence. In *Encyclopedia of Microbiology*. Schaechter, M. (ed.). Oxford, UK: Elsevier, pp. 45–61.
- Dunlap, P.V., Ast, J.C., Kimura, S., Fukui, A., Yoshino, T., and Endo, H. (2007) Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbiosis. *Cladistics* **23**: 507–532.
- Dunlap, P.V., Davis, K.M., Tomiyama, T., Fujino, M., and Fukui, A. (2008) Developmental and microbiological analysis of the inception of bioluminescent symbiosis in the marine fish *Nuchequula nuchalis* (Perciformes: Leiognathidae). *Appl Environ Microbiol* **74**: 7471–7481.
- Dunlap, P.V., Kojima, Y., Nakamura, S., and Nakamura, M. (2009) Inception of formation and early morphogenesis of the bacterial light organ of the sea urchin cardinalfish, *Siphamia versicolor* (Perciformes: Apogonidae). *Mar Biol* **156**: 2011–2020.
- Dunlap, P.V., Gould, A.L., Wittenrich, M.L., and Nakamura, M. (2012) Initiation of bioluminescent symbiosis in the bacterially luminous coral reef fish *Siphamia versicolor* (Perciformes: Apogonidae). *J Fish Biol* **81**: 1340–1356.
- Graf, J., and Ruby, E.G. (1998) Host-derived amino acids support the proliferation of symbiotic bacteria. *Proc Natl Acad Sci USA* **95**: 1818–1822.
- Haft, D.H., Selengut, J.D., Richter, R.A., Harkins, D., Basu, M.K., and Beck, E. (2013) TIGRFAMs and genome properties in 2013. *Nucleic Acids Res* **41**: D387–D395.
- Haygood, M.G. (1993) Light organ symbioses in fishes. *Crit Rev Microbiol* **19**: 191–216.
- Haygood, M.G., and Distel, D.L. (1993) Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. *Nature* **363**: 154–156.
- Haygood, M.G., Tebo, B.M., and Nealson, K.H. (1984) Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and 2 anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*) – populations sizes and growth within the light organs, and rates of release into the seawater. *Mar Biol* **78**: 249–254.
- Hendry, T.A., and Dunlap, P.V. (2011) The uncultured luminous symbiont of *Anomalops katoptron* (Beryciformes: Anomalopidae) represents a new bacterial genus. *Mol Phylogenet Evol* **61**: 834–843.
- Herring, P.J., and Morin, J.G. (1978) Bioluminescence in fishes. In *Bioluminescence in action*. Herring, P.J. (ed.). London, UK: Academic Press, pp. 273–329.
- Hosokawa, T., Kikuchi, Y., Nikoh, N., Meng, X.-Y., Hironaka, M., and Fukatsu, T. (2010) Phylogenetic position and peculiar genetic traits of a midgut bacterial symbiont of the stinkbug *Parastrachia japonensis*. *Appl Environ Microbiol* **76**: 4130–4135.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109–D114.
- 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.
- Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X.-Y., Kamagata, Y., and Fukatsu, T. (2009) Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol* **7**: 2.
- Konishi, Y., and Okiyama, M. (1997) Morphological development of four trachichthyoid larvae (Pisces: Beryciformes), with comments on trachichthyoid relationships. *Bull Mar Sci* **60**: 66–88.
- Lauro, F.M., McDougald, D., Thomas, T., Timothy, J., Egan, S., Rice, S., et al. (2009) The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106**: 15527–15533.
- Lee, K., and Ruby, E.G. (1994) Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. *Appl Environ Microbiol* **60**: 1565–1571.
- Leis, J.M., and Bullock, S. (1986) The luminous cardinalfish *Siphamia* (Pisces, Apogonidae): development of larvae and the luminous organ. In *Indo-Pacific Fish Biology: Proceedings of the Second International Conference on Indo-Pacific Fish*. Uyeno, T., Arai, R., Taniuchi, T., and Matsuura, K. (eds). Tokyo, Japan: Ichthyological Society of Japan, pp. 703–714.
- Lin, B., Wang, Z., Malanoski, A.P., O'Grady, E.A., Wimpee, C.F., Uddhakul, V., et al. (2010) Comparative genomic analyses identify the *Vibrio harveyi* genome sequenced strains BAA-1116 and HY01 as *Vibrio campbellii*. *Environ Microbiol Rep* **2**: 81–89.
- McCutcheon, J.P., and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13–26.
- Maezawa, K., Shigenobu, S., Taniguchi, H., Kubo, T., Aizawa, S., and 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.
- Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* **40**: D115–D122.
- Meyer-Rochow, V.B. (1976) Some observations on spawning and fecundity in luminescence fish *Photoblepharon palpebratus*. *Mar Biol* **37**: 325–328.

- Mira, A., and Moran, N.A. (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol* **44**: 137–143.
- Moran, N.A. (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* **93**: 2873–2878.
- Moran, N.A., and Plague, G.R. (2004) Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev* **14**: 627–633.
- Morin, J.G., Harrington, A., Neelson, K., Krieger, N., Baldwin, T.O., and Hastings, J.W. (1975) Light for all reasons – versatility in behavioral repertoire of flashlight fish. *Science* **190**: 74–76.
- Neelson, K.N. (1979) Alternative strategies of symbiosis of marine luminous fishes harboring light-emitting bacteria. *Trends Biochem Sci* **4**: 105–110.
- Nikoh, N., Hosokawa, T., Oshima, K., Hattori, M., and Fukatsu, T. (2011) Reductive evolution of bacterial genome in insect gut environment. *Genome Biol Evol* **3**: 702–714.
- Nyholm, S.V., and McFall-Ngai, M.J. (2004) The winnowing: establishing the squid-*Vibrio* symbiosis. *Nat Rev Microbiol* **2**: 632–642.
- Preheim, S.P., Boucher, Y., Wildschutte, H., David, L.A., Veneziano, D., Alm, E.J., and Polz, M.F. (2011) Metapopulation structure of *Vibrionaceae* among coastal marine invertebrates. *Environ Microbiol* **13**: 265–275.
- Raes, J., Korb, J.O., Lercher, M.L., von Mering, C., and Bork, P. (2007) Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**: R10.
- Reen, F.J., Almagro-Moreno, S., Ussery, D., and Boyd, E.F. (2006) The genomic code: inferring *Vibrionaceae* niche specialization. *Nat Rev Microbiol* **4**: 697–704.
- Rio, R.V.M., Symula, R.E., Wang, J., Lohs, C., Wu, Y., Snyder, A.K., *et al.* (2012) Insight into the transmission biology of species-specific functional capabilities of Tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *MBio* **3**: e00240–11.
- Ruby, E.G., Urbanowski, M., Campbell, J., Dunn, A., Faini, M., Gunsalus, R., *et al.* (2005) Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. *Proc Natl Acad Sci USA* **102**: 3004–3009.
- Sachs, J.L., Essenberg, C.J., and Turcotte, M.M. (2011) New paradigms for the evolution of beneficial infections. *Trends Ecol Evol* **26**: 202–209.
- Swofford, D.L. (1998) *PAUP* 4.0b10: Phylogenetic analysis using parsimony*. Sunderland, MA, USA: Sinauer.
- Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., *et al.* (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313.
- Toft, C., and Fares, M.A. (2008) The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol Biol Evol* **25**: 2069–2076.
- Toh, H., Weiss, B.L., Perkin, S.A.H., Yamashita, A., Oshima, K., Hattori, M., and Aksoy, S. (2012) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* **16**: 149–156.
- Urbanczyk, H., Ogura, Y., Hendry, T.A., Gould, A.L., Kiwaki, N., Atkinson, J.T., *et al.* (2011) Genome sequence of *Photobacterium mandapamensis* svers.1.1, the bioluminescent symbiont of the cardinalfish *Siphamia versicolor*. *J Bacteriol* **193**: 3144–3145.
- Wernegreen, J.J., and Funk, D.J. (2004) Mutation exposed: a neutral explanation for extreme base composition of an endosymbiont genome. *J Mol Evol* **59**: 849–858.
- Wernegreen, J.J., and Moran, N.A. (1999) Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. *Mol Biol Evol* **16**: 83–97.
- Wier, A.M., Nyholm, S.V., Mandel, M.J., Massengo-Tiasse, R.P., Schaefer, A.L., Koroleva, I., *et al.* (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 USA* **107**: 2259–2264.
- Woolfit, M., and Bromham, L. (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol Biol Evol* **20**: 1545–1555.
- Yu, X.J., Walker, D.H., Liu, Y., and Zhang, L.H. (2009) Amino acid biosynthesis deficiency in bacteria associated with human and animal hosts. *Infect Genet Evol* **9**: 514–517.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Contig and scaffold assembly information including contig sizes (bp) and read coverage depth.

Table S2. Functionally conserved genes typically found in most bacterial genomes and also found in the 'Ca. Photodesmus katoptron' genome.

Table S3. Genes by functional categories that are typically found in free-living *Vibrionaceae* and reduced in number in 'Ca. Photodesmus katoptron'.