Diversity and evolution of bacterial bioluminescence genes in the global ocean

Thomas Vannier San1,2,* Pascal Hingamp San1,2, Floriane Turrel1, Lisa Tanet1, Magali Lescot San1,2,* and Youri Timsit San1,2,*

ABSTRACT

Although bioluminescent bacteria are the most abundant and widely distributed of all light-emitting organisms, the biological role and evolutionary history of bacterial luminance are still shrouded in mystery. Bioluminescence has so far been observed in the genomes of three families of Gammaproteobacteria in the form of canonical lux operons that adopt the CDAB(F)E(G) gene order. LuxA and luxB encode the two subunits of bacterial luciferase responsible for light-emission. Our deep exploration of public marine environmental databases considerably expands this view by providing a catalog of new lux homolog sequences, including 401 previously unknown luciferase-related genes. It also reveals a broader diversity of the lux operon organization, which we observed in previously undescribed configurations such as CEDA, CAED and AxxCE. This expanded operon diversity provides clues for deciphering lux operon evolution and propagation within the bacterial domain. Leveraging quantitative tracking of marine bacterial genes afforded by planetary scale metagenomic sampling, our study also reveals that the novel lux genes and operons described herein are more abundant in the global ocean than the canonical CDAB(F)E(G) operon.

INTRODUCTION

Marine biodiversity and evolution are intimately related with biogeography and ecology (1–3). The Tara Oceans expedition recently provided a global picture of the complex interactions between marine micro-organisms and their environment (4–6). Bioluminescence, the chemical emission of visible light, is produced by a remarkable diversity of organisms and is particularly widespread in marine species (7–9). The luciferase enzymes that catalyze the emission of photons have evolved independently over 30 times, by convergence from non-luminescent enzymes (10,11). Although bioluminescent bacteria are the most abundant and widely distributed of all light-emitting organisms (7,12), certain functional and evolutionary aspects of bacterial luminescence still remain enigmatic, such as its biological role which remains a matter of debate (13). Early on bioluminescence was proposed to have evolved from ancient oxygen-detoxifying mechanisms (14–16). It has also been argued that stimulation of DNA repair through the activation of DNA photolyase may confer an advantage to luminous bacteria (17), although this hypothesis is still controversial (18). Yet another hypothesis is that bioluminescence is a visual attractant for zooplankton and fish that both provide ingested bacteria with growth medium and means for dispersal (19). Symbiosis with squid or fish is also an intriguing feature of specific bioluminescent bacteria (20,21).

To date, most of the few culturable light-emitting bacterial species that have been characterized fall within the Gammaproteobacteria class. These bacteria cluster phylogenetically in three families (Vibrionaceae, Shewanellaceae and Enterobacteriaceae) which all carry a highly conserved lux operon (12). Since its first identification 40 years ago, the canonical luxCDAB(F)E(G) organization has been systematically observed in all the bacterial bioluminescent genomes (22–24). The luxA and luxB genes encode the alpha and beta subunits of the luciferase heterodimer that emits light by the oxidation of FMNH2 and a long chain aldehyde substrate (24,27). LuxC, D and E together form a fatty acid reductase complex responsible for the synthesis of the long chain aldehyde substrate (24,27).

Despite a highly conserved core, some variations have been observed in the lux operon organization. Small differences in gene content, for instance the presence of an operon superstructure in gene content, for instance the presence of an operon superstructure in
tional riboflavin genes or luxF in *Photobacterium* species, have been observed (28,29). LuxG, which reduces FMN into FMNH2 in *Photobacterium* spp. whose operon also contains multiple insertions of ERIC sequences (30). Natural merodiploidy of the lux-rib operon has been also noticed in some strains in *Photobacterium leiognathi* (31). While the phylogeny of lux genes generally supports a vertical inheritance, multiple examples of instability of the lux locus and horizontal gene transfers (HGT) have been reported in different clades, at various taxonomic levels (32–34). Also, mutations or loss of the lux operon are frequently observed in non-luminous strains and appear to correlate with some environmental parameters (35–38).

Although bioluminescent bacteria are cosmopolite in the oceans and occupy a great diversity of ecological niches, including surface and deep waters, (39–43), many studies have revealed an intricate relationship between bacterial bioluminescent phenotype, lux operon diversity, environmental parameters and life style (29,44–50). Given the apparent ubiquity of bioluminescence in the ocean and the ease with which light emitting bacteria can be isolated from seawater, it has come as a surprise that bacterial bioluminescence has so far escaped detection by previous marine metagenomic studies (51). According to the pioneering authors, the unexpected absence of lux genes might have been explained by sampling protocols which filtered out size classes of potential interest, and to sequencing depth which might have been insufficient to catch bioluminescent bacteria if these were of low abundance (51).

In the present report, we surveyed the distribution of bacterial lux-related genes in a compilation of publicly available large-scale metagenomic environmental databases (*Tara Oceans* 2009–2013, Malaspina 2010, GOS and OSD2014) giving special care to screen the largest possible variety of organism size sampling fractions. Spanning a wide spectrum of marine bacteria diversity, including a majority of unculturables species, our study reveals new insights about distribution, diversity and evolution of marine lux-related genes and their operon organization at a planetary scale.

**MATERIALS AND METHODS**

**LuxA reference sequence dataset**

The coordinates of the *Vibrio harveyi* LuxA/B heterodimeric luciferase (pdbid: 3igc) were obtained from the Protein Data Bank (PDB) (52) (https://www.rcsb.org/, version 06/22/2019) and used as a reference luciferase structure. The corresponding *V. harveyi* LuxA protein sequence (UniProtKB—P07740) was used to query UniProtKB/Swiss-Prot (53,54) (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/swissprot.gz, version 06/14/2019) and Refseq (55) (ftp://ftp.ncbi.nih.gov/refseq/release-release-catalog/RefSeq-release93.catalog.gz) using blastp (56). Ten protein sequences with an E-value inferior to 1.0E-128 (threshold above which LuxB proteins are detected) were considered reliable LuxA homologs and formed the seed for the reference LuxA dataset. The ten seed reference LuxA sequences were then aligned with MAFFT with default parameters (57) (http://mafft.cbrc.jp/alignment/software/) and a hidden Markov model (HMM) profile was built using hmmbuild from HMMer v 3.0 with default parameters (58) (http://hmmer.org). The resulting LuxA HMM profile was used to search for additional luciferase homologs using hmmsearch in NR (59) (https://ftp.ncbi.nlm.nih.gov/blast/db/Fasta/nr.gz, version 06/15/2019) with an E-value threshold of 1.0E-186 set to avoid LuxB homologs. In addition, 55 draft whole genome shotgun (WSG) marine bacterial genomes (https://www.ncbi.nlm.nih.gov/genbank/wgs/, version 07/25/2018) and 334 marine bacterial complete genomes (https://www.ncbi.nlm.nih.gov/genbank/ genome/, version 07/25/2018) were screened using the LuxA HMM profile; after removing sequence redundancy, partial and synthetic sequences, we obtained a reference dataset of 129 LuxA protein sequences. This dataset was then used to compute a final LuxA HMM profile (Supplementary Table S1). A similar procedure was used for building the Lux B, C, D, E, G and F reference datasets.

**Diversity of lux operon**

Genbank bacterial genomes containing the reference and marine lux-like sequences are download from the NCBI web site (https://www.ncbi.nlm.nih.gov/genbank/). A syntenic graph representing the operon structural organization was done using Easyfig (60).

**OM-RGC Lux homologs search**

The *Tara Oceans* OM-RGC dataset (5,61) was screened with each of the Lux HMM profiles obtained above, using hmmsearch with an E-value threshold of 1.0E-10 (Supplementary Table S1). Further filtering based on alignment lengths eliminated incomplete Lux sequences. The length thresholds were set to 340, 300, 430, 275, 300 and 200 aa for the LuxA, B, C, D, E and F homologs, respectively.

**OM-RGC LuxA homolog structural filtering**

Bacterial luciferases and monooxygenases share a highly conserved TIM-barrel fold (25) that renders discrimination from each other difficult from primary sequence alignments alone. We therefore developed a specific procedure to help luciferase/monooxygenase discrimination based on three-dimensional (3D) structure modeling and comparison. The 3D coordinates of close structural homologs of bacterial luciferases were retrieved from PDB (52) and structurally superimposed with PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.) (Supplementary Table S2 and Figure S2a–i). This analysis selected a set of amino acids previously described to be specific of the luciferase active site (62). In a first filtration step, truncated sequences and sequences lacking the catalytic His44 were excluded from our selection. Moreover, a disordered loop from 262–290 (V. harveyi LuxA coordinates) that covers the active site is a specific signature of LuxA proteins (63). Thus, we used Disopred v3.1 (64) to systematically calculate the disorder content in all the aligned LuxA homologs. We then filtered out sequences displaying a disorder score below 0.2 in this region (defined by the multiple sequence alignment,
see below) from our final LuxA homolog dataset (Supplementary Table S3).

**Phylogenetic analyses of Lux-related sequences**

Multiple sequence alignments (MSA) for phylogenetic analyses were obtained with Clustal Omega (65) using default settings. In order to generate a phylogenetic tree that integrates LuxA-related, LuxB and selected monooxygenases (Supplementary Table S2), we first aligned each of these three sequence groups independently, and then processed each of the three MSAs with MaxAlign (66) to remove sequences with excess INDELS that reduce alignment area. TrimAl v1.2 with default setting (67) was then used for automated alignment trimming (removal of poorly aligned positions in the MSAs). The phylogenetic analysis was performed with RAxML (68) (parameters: -f a -# 10000 -m PROTGAMMAAUTO -p 12345 -x 12345). Trees were midpoint rooted and visualized with FigTree (https://github.com/rambaut/figtree/) and Evolvview v2 (69). Alignments were visualized and inspected with Jalview (70). A similar procedure was used for the phylogenetic analyses of the other Lux proteins.

**Abundance of lux genes in the global ocean**

We computed the abundance of the canonical and non-canonical lux operon genes (Supplementary Table S4 and File S2) by metagenomic read recruitment from large scale marine metagenomic data: *Tara Oceans* 2009–2013 (71–73) (PRJEB402, https://www.ncbi.nlm.nih.gov/; doi:10.1594/PANGAEA.859953, https://doi.pangaea.de/) Malaspina 2010 (74) (Gs0053074, https://img.jgi.doe.gov/), Malaspina, GOS and OSD 2014 (75,76) (PRJEB8968, https://www.ncbi.nlm.nih.gov/) Malaspina 2010 (74) (Gs0053074, https://img.jgi.doe.gov/), Malaspina, GOS and OSD 2014 (77) (PRJEB5129, https://www.ncbi.nlm.nih.gov/) (Supplementary Table S5). This procedure, which searches individual reads that align to a query sequence in each of the available raw unassembled metagenomes, is expected to be more sensitive than searching for homologs in the assembled metagenomes (e.g. OM-RGC). Indeed, the raw read mining procedure has the potential to detect even singleton homologous reads, whereas such singletons often don’t make it past the assembly stage of reduced datasets such as the OM-RGC. Metagenomic read counts per gene (RPKM_MG) correspond to the number of mapped reads per query gene divided by the total number of reads sequenced for each sample (in millions) divided by gene length (in kilobases). The reported relative abundance values correspond to: \[ \frac{\log(1 + \frac{\text{RPKM}_M \times 10^3)}{\log(3)}}{\text{RPKM}}. \] Metagenomic short reads were recruited with Bowtie2 2.1.0 aligner with default parameters (78). We discarded alignments below 80% identity and below 50% of query gene horizontal coverage.

**Geographic distribution of specific bacterial genomes**

We analyzed the global distribution of *Vibrio Campbellii* ATCC BAA-1116 (NCBI:txid338187) and *Alcanivorax bercatorium* GenoA1_TS13_700 (NCBI:txid2072747, Pacific ocean: South China 700m, 19.909 N, 115.243 E) genomes by metagenomic read recruitment from the same marine metagenomes as described above (*Tara Oceans, Malaspina, GOS and OSD*). For each genome, we computed relative genomic abundances as the number of reads mapped onto the genome normalized by the total number of reads sequenced for each sample (Supplementary Table S5) (79). We identified and removed from the read recruitment counts 267 and 528 outlier genes on *V. campbellii* and *Alcanivorax spp.* respectively, to reduce apparent biases due to atypical genomic regions with low sequence specificity. We generated the world maps with R version 3.2.3 using the packages maps_3.1.1, mapproj_1.2-4 and mapproj_1.5. The ‘free-living’ organismal size fractions were defined as the 0.22–1.6 μm and 0.22–3 μm filters of *Tara Oceans*, and the 0.2–0.8 μm filters of Malaspina. The ‘associated’ organismal size fraction was defined as the 0.8–2000 μm filters of *Tara Oceans*, and the 0.8–20 μm filters of Malaspina. Box plots and Mann–Whitney U-test were performed with R.

**Biogeography and taxonomic distribution of the OM-RGC luxA homologs**

We used the Ocean Gene Atlas web service (80) (http://tara-oceans.mio.osupytheas.fr/) to describe the geographic and taxonomic distributions of the 3 groups of OM-RGC LuxA homologs (OLAHs). Abundance and location were visualized using the *Tara Oceans* metagenomic samples corresponding to size fractions 0.22–1.6 and 0.22–3 μm sampled either at surface or mesopelagic depths.

**RESULTS**

**Diversity of lux gene homologs**

In order to survey *luxA, B, C, D* and *E* homologs in the global ocean, we first inventoried all Lux sequences deposited in public reference databases (NR, RefSeq, UniProtKB/Swiss-Prot) to produce a set of reference lux genes (Supplementary Table S1). As expected for genes forming an operon, a similar number of occurrences of each *lux* gene was found in public reference databases. Surprisingly, this search also revealed a set of *lux* homologs in the reference genomes of six bacteria that had not previously been described as bioluminescent: *Alcanivorax spp.*, *Rhizobacter spp.* (Gammaproteobacteria); *Enhygromyxa salina*, *Plesiocystis pacifica* (Deltaproteobacteria); *Leptospira santuarosi* (Spirochaetes); *Actinomyces spp.* (Actinobacteria) (Supplementary Table S1). These six bacterial reference genomes possess *luxA, C, D* and *E* of the *lux* operon, with the notable exception of *luxB*. This thus revealed the presence of *lux*-like operons in three bacterial classes never described as capable of producing light: Deltaproteobacteria, Spirochaetes and Actinobacteria. Genes in these *luxB*-less operons will henceforth be referred to as *lux*-like genes.

We then used this set of reference sequences to search for *lux* homologs in the most recent microbial gene catalog of *Tara Oceans* (OM-RGC), which represents assembled genetic material recovered from the free-living mostly bacterial size fractions (0.2–3 μm) collected in the global ocean (Supplementary File S1). Consistent with the current knowledge of bacterial bioluminescence distribution, we detected the presence in the OM-RGC of *lux* genes from
several well-known bioluminescent bacteria such as *Vibrio campbelli*, *Vibrio cholerae*, *Aliivibrio fischeri*, *Shewanella woodyi* and *P. leiognathi* (Figure 1 and Supplementary Table S1).

An initial search for LuxA proteins provided a set of 800 marine candidates having both high similarity score and size close to our reference LuxA sequences (Supplementary Figure S1). Further filtering according to the presence of a correctly positioned structurally disordered region as well as the presence of specific active site residues generated a set of 401 OM-RGC LuxA homologs (OLAHs) (Supplementary Table S3).

By combining LuxA, LuxA-like and LuxB reference sequences together with the 401 marine OLAHs and some selected monoxygenases of known structure, we inferred a phylogenetic tree of LuxA homologs (Figure 1). Phylogeny and disorder content of LuxA sequences are highly congruent (Supplementary Figure S3) as indicated by the clustering of *bona fide* reference LuxA sequences that display a similar disordered loop. The LuxA reference sequences are grouped together in a monophyletic branch well separated from the the LuxB monophyletic branch, both of which are then connected to a group of homologs we annotated as LuxA-like sequences. The bulk of the marine homologs then cluster into three branches, forming OLAH groups 1–3 (Figure 1). However, OLAH groups 1 and 2, as well as LuxA-like sequences all share a similar consensus in the catalytic site with LuxA reference sequences (Supplementary Figure S4). The monoxygenases and OLAH group 3 appear to form a separate branch, coherent with a distinct active site logo signature (Supplementary Figure S4).

Interestingly, the few LuxA-like sequences belonging to bacteria that have not been previously described as luminescent organisms form a separate but close branch to the LuxA reference group (bootstrap value of 100). We should note that the LuxA sequences of *Alcanivorax* spp. and *P. pacifica* found in the OM-RGC dataset are grouped with the LuxA-like branch (Figure 1). Two monoxygenase sequences (pdbid: 3i7g and 4uwm) are found within OLAH groups 2 and 3, respectively (Figure 1 and Supplementary Table S2). Phylogenetic analysis of marine LuxB, C, D, E homologs similarly show distinct groups consisting of Lux references, Lux-like and OM-RGC distant relatives (Supplementary Figure S5).

**Diversity of lux operons**

Reference and marine lux-like sequences differed not only in their phylogeny, but also in their operon organization. Indeed, we observed operon structural organizations very different from the canonical CDAB(F)E(G) order (Figure 2 and Supplementary Figure S6). In particular, our study revealed a novel CEDA order in which luxB is missing, in both Gammaproteobacteria and Deltaproteobacteria. Moreover, the *Leptospira* spp. that belongs to the more distant Spirochaete class encodes a CAED operon, while the *Actinomyces* AxxCE lux-like operon structure is even more atypical. Importantly, we observed a congruence between the operon organization and the phylogenetic distance in the LuxA sequence tree (Figure 2). Thus, our findings extend the known diversity of the *lux* operon organizations and reveal the existence of non-canonical *lux*-like operons in new taxons.

The similarities between homologous genes in the different operons have been also analyzed (Figure 2 and Supplementary Figure S7). As expected, each *lux* gene is highly similar to its relative in close taxons. Interestingly, the genomes of the two distinct myxobacteria genera *Enhydrogymnus* and *Pleisiocistis* (81,82), exhibit full *lux* operons more than 90% identical. High similarities are also found between the *Alcanivorax* and *Rhizobacter* *lux*-like operons (61.1%). These findings suggest a lateral transfer of the CEDA *lux*-like operon between these species.

**Abundance and distribution of lux-related genes in the global ocean**

In order to investigate the distribution of the *lux*-related genes (in canonical or non-canonical operons), all reference and OLAH sequences were used as queries to recruit raw metagenomic reads from public large scale marine metagenomes (*Tara* Oceans, Malaspina, GOS and OSD) (Supplementary Table S5; see materials and methods). In addition, to differentiate *lux* genes likely harboured in free living bacteria from those likely associated with particles or larger organisms (e.g. through symbiosis), we analyzed their relative abundances in samples obtained by filtering through filters of pore sizes of either 0.22–3 μm (free-living) or 0.8–2000 μm (associated). The abundance of *lux* genes are summarized in Figure 3 and detailed in Supplementary Table S7. We have also estimated the abundance of the canonical *lux* operons of all the well-known bioluminescent bacteria (Figure 2). Rather surprisingly, the *lux*-operon of *V. campbelli* was the only detectable known bioluminescent bacterium, which furthermore was detected in *Tara* Oceans samples only, thus suggesting its higher abundance relative to other bioluminescent bacteria. Reference *lux* genes were not observed in either free-living or associated size fractions collected during the Malaspina expedition (which systematically sampled bathypelagic layers), nor the GOS expedition, nor on OSD2014 solstice day (mostly coastal waters). An important finding was that the reference *V. campbelli* *lux*A gene was preferentially detected in associated size fractions (i.e. non free-living) collected in just five *Tara* Oceans coastal and surface samples (TARA005, TARA007, TARA008 in the Mediterranean sea, and TARA123, TARA125 close to the Marquise islands) (Supplementary Table S7). This shows that the *V. campbelli* canonical *lux* operons are mainly associated with particles or larger planktonic species. In contrast, Figure 3 shows that the genes of the *Alcanivorax* non-canonical CEDA operon are abundantly distributed in both *Tara* Oceans and Malaspina samples, in all size fractions (from 0.2 to 2000 μm), both in surface and deep ocean depths.

In order to further investigate these intriguing findings, we extended this *lux* only metagenomic read recruitment analysis to the distribution of *Vibrio* spp. and *Alcanivorax* spp. full length genomes in the global ocean. Figure 4 summarizes their geographic, water column and free-living/associated distributions (quantitative data are reported in Supplementary Table S6). *Vibrio* genomes are widely distributed in temperate regions, in surface, meso-
Figure 1. Phylogenetic tree of LuxA and related sequences found in the global ocean. Sequences are annotated as follows: LuxA in brown; LuxB in yellow; LuxA-like in red; OM-RGC LuxA Homologs (OLAH) group 1 (128 sequences), 2 (209 sequences) and 3 (63 sequences) from light to dark blue. Reference monooxygenases clustered together in a separate branch are colored purple; the two purple circles correspond to PDB monooxygenase sequences, 4uwm (13) and 2r7g (14) that branch in OLAH groups 3 and 2, respectively. The pink circles correspond to the OM-RGC assigned as LuxA reference sequences of *Vibrio campbellii* (1), *Vibrio cholerae* (2), *Aliivibrio fischeri* (3), *Shewanella woodyi* (4) and *Photobacterium leiognathi* (5). The red circles correspond to the sequences we annotate as LuxA-like of *Plesiocystis pacifica* (6) and *Alcanivoracaceae* bacterium (7). The yellow circles correspond to the OM-RGC assigned as LuxB reference sequences of *S. woodyi* (8), *A. fischeri* (9), *P. leiognathi* (10), *V. campbellii* (11), *V. cholerae* (12). The origin of the sequences used for tree inference are presented in Supplementary Table S3, and OLAH amino acid sequences are provided in fasta format (Supplementary File S1).
and bathypelagic depths and are not observed in polar-
regions (Figure 4A). In contrast to the preference of indi-
vidual lux genes for associated fractions (Figure 3), Vib-
rio full genomes were detected in both free-living and as-
associated fractions, albeit significantly more abundant in
the free-living bacterial fractions than in the associated
fractions (Mann–Whitney U-test $P < 0.0002$) (Figure 4B).
Consequently, our study suggests that free-living Vibrios
tend to be devoid of the lux operon.

Alcanivorax genomes showed a very different distribu-
tion (Figure 4C): they were highly abundant (more than 10
times the abundance of vibrio genomes), and except for one
station (West of the North Atlantic gyre), they were exclu-
sively localized in tropical and sub-tropical zones of the Pa-
cific Ocean. They were present in a wide range of depths
from surface to bathypelagic layers. Similarly to Vibrio,
Alcanivorax genomes were present in both the free-living and
the associated-size factions (Figure 4D), which, contrary
to observations with Vibrios, was consistent with the distribu-
tion of the Alcanivorax lux-like operon (Figure 3).

To gain further insight in the distribution of sequences
corresponding to OLAH group 1 (Figures 5A and B) and
OLAH groups 2–3 (Supplementary Figures S6 and 7) of the
LuxA phylogenetic tree (Figure 1 and Supplementary Fig-
ure S3), we also analyzed their ocean distributions in the
Tara Oceans samples. Group 1 sequences showed a particu-
larly interesting distribution along the water column. Al-
though they were observed in all oceans and latitudes, they
were systematically much more abundant in mesopelagic
(at 200–1000 m) than in surface depths (Figure 5C).
These sequences clustered into three known groups of bac-
teria: Entotheonella spp., Actinobacteria and Proteobacteria
(Figure 5D). Interestingly, the differences in abundances be-
tween surface and mesopelagic depths gradually decreased
from group 1 to group 3 (Figure 5C; Supplementary Figures
S8c and 9c) and a more diversified taxonomic distribution
was also observed in group 3 (Figure 5D; Supplementary
Figures S7d and 8d).

DISCUSSION

Many studies have shown that bacterial luminescence is re-
lated to environmental parameters and/or life-style, such as
symbiotic associations (35–38). Bacterial bioluminescence
has so far been observed as a canonical lux operon that fol-
ows the CDAB(F)E(G) gene order, in genomes of gram-
negative bacteria that group phylogenetically in three fami-
lies of Gammaproteobacteria: the Vibrionaceae, Enterobac-
teriaceae and Shewanellaceae (12). However, minor varia-
tions around this common theme have been observed in vari-
ous species (12,28–31). Our deep exploration of public ge-
nomic and metagenomic marine databases expands the lux
operon catalog (Figure 2). It also reveals a larger distribu-
tion of lux operons among bacterial taxa than previously
Figure 3. Abundance of *lux* genes in the global ocean. RPKM values of canonical and non-canonical lux genes are listed when at least one gene of the operon has significant matches with reads of the sample (Supplementary Table S7). (A) RPKM values of the *Vibrio campbellii* canonical lux operon. (B) RPKM values of the *Alcanivoracaceae spp.* non-canonical lux-like operon. RPKM values are color-coded with increasing abundance (white, yellow, green). Size fractions are indicated with blue squares for free-living fractions and red squares for particle-attached and symbiotic fractions. Samples are organized geographically: South Atlantic Ocean (SAO), South Pacific Ocean (SPO), Mediterranean Sea (MS), Indian Ocean (IO), North Atlantic Ocean (NAO) and Southern Ocean (SO). Nucleotide sequences of *lux* operon genes are provided in fasta format (Supplementary File S2). Samples used for the metagenomic read recruitment are listed in Supplementary Table S5.
known, as well as a highly variable lux operon organization repertoire (Figure 2). We describe the existence of a new lux-like CEDA operon, lacking the luxB gene, in both Gammaproteobacteria and Deltaproteobacteria (83).

In Gammaproteobacteria, the CEDA lux-like operon configuration is observed in Pseudomonales (Rhizobacter) and Oceanospirales (Alcanivorax) (84). Since the canonical CDAB(F)E(G) form of the lux operon is highly conserved in three Gammaproteobacteria families (7,12), it is somewhat surprising to find a distinct lux gene order among the Gammaproteobacteria. Remarkably, the finding of two different lux operon organizations within the Gammaproteobacteria appears to be congruent with the phylogeny of this clade. Indeed, according to Gammaproteobacteria species phylogeny (83), the CDAB(F)E(G) and CEDA lux-like operon configuration correspond to the divergence of two distinct groups, the PO (Pseudomonadales and Oceanospirillales) and VAAP/Entero (Vibrionales, Alteromonadales, Aeromonadales and Pasteurellales/Enterobacterials), respectively (Figure 2). Despite an absence of homology around the operon in the genomes of the two Gammaproteobacteria families harbouring the CEDA lux-like operon, the lux-like genes display a high similarity, suggesting a lateral transfer of the lux operon between the two remote Rhizobacter and Alcanivorax genera (Supplementary Figure S7). Such
lateral transfer might have occurred in rhizosphere micro-
biomes shared by these two species (85). In addition to mar-
ine environments (84) _Acanthopaxillus_ spp. has been indeed ob-
served in coastal areas where it thrives in oil-polluted soils
and rhizospheres (85).

In Deltaproteobacteria, we have also found the non-
canonical CEDA _lux_ operon form in two _myxococcale_ ma-
rine bacteria (86), _Enhygromyxa salina_ (81) and _P. pacifica_ (82). The close homology between the CEDA operon
of these two distant species that share the same eco-
logical niche strongly supports a lateral transfer origin (81,82)
(Supplementary Figure S7). Remarkably, two distant bacte-
rial classes share the CEDA configuration: the Gammapro-
teobacteria and the Deltaproteobacteria. Our findings
cannot yet distinguish between an HGT or/and a vertical in-
heritance of this operon configuration from a common an-
cestor of these two classes. In the latter case, the CEDA
configuration would correspond to an ancestral _lux_ operon
that predated the _luxA/luxB_ duplication, congruent with
the LuxA inferred phylgetic tree. On the another hand,
an HGT scenario is not uncommon between distant taxons
when two species share a common niche (87).

We also show that _Leptospira_ spp. belonging to a bac-
terial taxon phylogenetically distant from proteobacteria,
the _Spirochaetes_, also carries an even more distinctive _lux_
operon configuration: the CAED form. _Leptospira_ genera
occupy a great variety of habitats (88) and display highly
plastic genomes consistent with frequent HGT events (89).
A yet more distant form AxxCE is observed in
_Actinomyces_, a gram-positive _Actinobacteria_ (90). Interestingly, it
has been reported that some _Mycobacteria_ spp. contain a
locus, _mel2_, very similar to the _lux_ operon (91–93). Strik-
ingly, the _mel2_ locus organization follows a similar AxxCE
order as found here in _Actinomyces_. Consistent with the ab-
fluence of the _luxB_ gene from the _Actinobacter lux_ operon,
the analysis of the relatedness of _MelF_ to _luxA_ and _luxB_ placed
_MelF_ on an independent branch equally related to both of
them (91). Our results suggest that the non-canonical _lux-

Figure 5. The OM-RGC LuxA homolog (OLAH) group 1 sequences are observed in all oceans and are more abundant in the mesopelagic layer. Geographic
distribution and abundance of OLAH group 1 sequences in the _Tara_ Oceans samples for the 0.22–1.6 µm and 0.22–3 µm size fractions at (A) surface and (B) mesopelagic depths. Gray crosses indicates _Tara_ Oceans sampling stations. No mesopelagic sampling was done in the Mediterranean sea and in the Indian
Ocean. (C) Abundance according to sampling depths. The size of the circles indicate the relative genomic abundances of the OLAH group 1 sequences.
(D) Pie chart depicting percent abundance of OLAH group 1 sequences according to taxonomic classification. The list of OLAH group 1 identifiers used
to retrieve biogeography from the Ocean Gene Atlas web service (http://tara-oceans.mio.osupytheas.fr/) are provided in Supplementary Table S3.
like operon of gram-positive bacteria may correspond to an ancestral form and support the hypothesis of Kasai et al. suggesting that the lux operon originated in gram-positive bacteria (32).

In order to better understand the link between the genetic diversity, evolution and the ecology of the lux operon, we have searched independently each lux genes in the global ocean and collected information about their geographic distribution, their localization in the water column and their presence in either free-living or associated size fractions. Two distinct scanning procedures of the metagenomic data provided complementary results.

First, we investigated the presence of lux genes in the assembled metagenomic data (OM-RGC) by protein homology search (Figure 1), thus providing a qualitative view of the presence of lux genes in the global ocean. Previous analyses did not recover lux sequences from assembled metagenomic DNA of planktonic marine microorganisms (51), thus suggesting that luminous Photobacterium, Alitivibrio and Vibrio were present at abundances so low they went undetected with afforded sequencing efforts (29). In the present study however, the higher sequencing deepness detected several well-known bioluminescent bacteria such as V. campbellii, V. cholera, A. fischeri, S. woodyi and P. leiognathi. We infer that the other well-known marine bioluminescent bacteria that are not detected here are either not represented in the Tara Oceans sampling locations, or in abundances too low to allow their assembly. Interestingly, the OM-RGC assembly does contain lux-like genes belonging to the non-canonical CEDA lux-like operon of Alcanivorax spp. (84) and P. pacifica (82), two species that have been recently characterized as marine bacteria.

Second, we quantitatively explored the biogeographic abundances of known reference lux operons by metagenomic read recruitment, a method that provides a measure of geographic abundance of lux operons in the global ocean. Only V. campbellii and Alcanivorax spp. lux genes were detected, indicating either that they have a higher abundance than other known bacteria containing lux or the lux-like genes, or that the niche habitats of other bioluminescent bacteria were not sampled by Tara Oceans, Malaspina, GOS and OSD. The comparison of their full genome distributions in the global ocean shows contrasting stability of the lux operon in V. campbellii compared to Alcanivorax spp. The Vibrios appear to contain an accessory lux operon preferentially found in symbiotic or particle associated strains localized in coastal environments (Figures 3 and 4). This is consistent with previous studies showing that non-luminescent Vibrios lacking the lux operon are free-living (35,36,38). In contrast, most of the detected Alcanivorax strains stably possess the CEDA-lux operon in their genomes and display a homogenous distribution restricted to the Pacific Ocean. Alcanivorax spp. are ~10× time more abundant than V. campbellii, and are slightly more frequently observed in free living fractions (Figures 3 and 4). The two distinct patterns of bacterial/lux operon relationships suggest that the different lux operon configuration may have different functions and genomic stability. Although several Alcanivorax spp. have been characterized as hydrocarbon degrading bacteria, to our knowledge, their bioluminescent phenotypes are not documented.

To obtain finer grain resolution of bacterial bioluminescence evolution and biogeography, we carried out a comprehensible analysis of LuxA homologs, carefully selected on the basis of both sequence homology, structural and disorder properties. Interestingly, disorder content is congruent with phylogeny and shows distinct patterns in the three well separated OLAH groups (Figure 1 and Supplementary Figure S3). We believe that this set of sequences may provide new insights to decipher luciferase evolution pathways. LuxA belongs to the group C family of two component monooxygenases, which adopt a very similar TIM-barrel fold and probably share a common ancestor (94–96). However, a comparison of LuxA and the closest non-luminous monooxygenase structures (pdk_id: 4us5 and 4uw) (97,98) revealed that their active sites have both different sequences and structures (Supplementary Table S2 and Figure S2a–i). A first interesting result is the discovery of two OLAH groups (1 and 2) that share many common features with reference LuxA that seem to distinguish them from monooxygenases (Supplementary Figures S3 and 4). Remarkably, the phylogeny and the ocean distribution of OLAH sequences also show a congruent pattern. Indeed, sequences of the cosmopolite OLAH group 1 are systematically much more abundant in mesopelagic layers (Figure 5A–C). In contrast, the two OLAH groups 2 and 3 that include known reference monooxygenases are uniformly distributed along the water column (Supplementary Figures S8a–d and S9a–d). The fact that the group 1 is preferentially found in dark environments is coherent with the hypothesis that they could represent a new class of luciferases. An alternative hypothesis explaining higher abundance deep in the ocean could be a specific adaptation to high-pressure of luminescent or non-luminescent proteins. Interestingly, more than one third of OLAH group 1 sequences are assigned to Entotheonella. These bacteria have been recently described to form symbiotic associations with sponges and are characterized by an over-representation of FMN-monooxygenase genes (99), that are closely related to luciferases. Such reservoirs of FMN-monooxygenase genes may have contributed to the evolution of luciferases.

How and why bacterial luciferases have evolved to catalyze light-emission is still an open question. To date, the diversity of bioluminescent bacteria is predominantly based on the study of culturable Gammaproteobacteria harbouring a highly conserved lux operon organization. Our metagenomic study provides new insights about lux operon diversity and distribution in the global ocean, as well as clues for future exploration on luciferase evolution and functions in marine bacteria including unknown and uncultivable species.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NARGAB Online.

**ACKNOWLEDGEMENTS**

The authors would like to thank Gérald Grégori for sharing his enthusiasm about bioluminescence. We are also grateful to Laurie Casalot, Gwenola Simon and Christian Tamburini for useful discussions. Thanks to Simon Roux for sharing his Easyfig script.
The bulk of genomic computations were performed using the OSU Pythées cluster with the help of the SIP members and the Airain HPC machine provided through GENC1- [TGCC/CINES/IDRIS] (grants t2011076389, t2012076389, t2013036389, t2014036389, t2015036389 and t2016036389).

Tara Oceans (which includes both the Tara Oceans and Tara Oceans Polar Circle expeditions) would not exist without the support of the Tara Expeditions Foundation and the continuous support of 23 institutes (http://oceans.taraexpeditions.org). We further thank the commitment of the following sponsors: CNRS (in particular Groupement de Recherche GDR3280 and the Research Portion of Global Ocean Systems Ecology, FR2022), Federation for the study of Global Ocean Systems Ecology (GDR), and European Molecular Biology Laboratory (EMBL), Genoscope/CEA, The French Ministry of Research and the French Government ‘Investissements d’Avenir’ programmes, FRANCE GENOMIQUE, MEMO LIFE and PSL* Research University. We also thank the support and commitment of agnès b. and Etienne Bourgois, the Prince Albert II de Monaco Foundation, the Veolia Foundation, Region Bretagne, Lorient Agglomeration, Serge Ferrari, Worldcourier and KAUST. The global sampling effort was enabled by countless scientists and crew who sampled aboard the Tara from 2009–2013, and we thank MERCATOR-CORIOLIS and ACRI-ST for providing daily satellite data during the expeditions. We are also grateful to the countries who graciously granted sampling permissions. The authors declare that all data reported herein are freely available from the date of publication, with no restrictions, and that all of the analyses, publications and ownership of data are free from legal entanglement or restriction by the various nations whose waters the Tara Oceans expeditions sampled in. This article is contribution number 101 of Tara Oceans.

FUNDING

French Government ‘Investissements d’Avenir’ programmes OCEANOMICS [ANR-11-BTBR-0008]; FRANCE GENOMIQUE [ANR-10-INSB-09-08]; MEMO LIFE [ANR-10-LABX-54]; PSL* Research University [ANR-11-IDEX-0001-02]; GENC1 [TGCC/CINES/IDRIS, t2011076389, t2012076389, t2013036389, t2014036389, t2015036389, t2016036389].

Conflict of interest statement. We declare no conflict of interest and confirm the material is original research and has not been previously published or submitted for publication elsewhere.

REFERENCES

1. DeLong EF. (2009) The microbial ocean from genomes to biomes. Nature, 459: 200–206.
2. Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Hersh, G.J. (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proc. Natl. Acad. Sci. U.S.A., 103, 12115–12120.
3. Orcutt, B.N., Sylvan, J.B., Knab, N.J. and Edwards, K.J. (2011) Microbial ecology of the Dark Ocean above, at, and below the Seafloor. Microbiol. Mol. Biol. Rev., 75, 361–422.
4. de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I. et al. (2015) Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science, 348, 1261605.
5. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. et al. (2015) Ocean plankton. Structure and function of the global ocean microbiome. Science, 348, 1261359.
6. Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F. et al. (2015) Determinants of community structure in the global plankton interactome. Science, 348, 1262073.
7. Meighen, E.A. (1994) Genetics of bacterial bioluminescence. Annu. Rev. Genet., 28, 117–159.
8. Haddock, S.H.D., Mahé, F., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F. et al. (2015) Determinants of community structure in the global plankton interactome. Science, 348, 1262073.
9. Meighen, E.A. (1994) Genetics of bacterial bioluminescence. Annu. Rev. Genet., 28, 117–159.
10. Waldenmaier, H.E., Oliveira, A.G. and Stevani, C.V. (2012) Thoughts on the diversity of convergent evolution of bioluminescence on earth. Int. J. Astrobiol., 11, 335–343.
11. Delrotte, J., Ullrich-Lüter, E., Blaue, S., Ortega-Martínez, O., Eeckhaut, I., Flamman, P. and Mallefet, J. (2017) A puzzling homology: a brittle star using a putative cnidarian-type luciferase for bioluminescence. Open Biol., 7, 163003.
12. Dunlap, P. (2014) Biochemistry and Genetics of Bacterial Bioluminescence. In: Thouand, G and Marks, R (eds). Bioluminescence: Fundamentals and Applications in Biotechnology, Vol. 1, Springer, Berlin, Heidelberg. pp. 37–64.
13. Peat, S.M. and Adams, B.J. (2008) Natural selection on the luxA gene of bioluminescent bacteria. Symbiosis, 46, 101–108.
14. McElroy, W.D. and Seliger, H.H. (1962) Origin and evolution of bioluminescence. In: Kash, M and Pullman, B (eds). In Horizons in Biochemistry. Academic Press, NY. pp. 91–101.
15. Rees, J.F., de Wergifosse, B., Noiset, O., Dubuisson, M., Janssens, B. and Thompson, E.M. (1998) The origins of marine bioluminescence: turning oxygen defence mechanisms into deep-sea communication tools. J. Exp. Biol., 201, 1211–1221.
16. Timmins, G.S., Jackson, S.K. and Swartz, H.M. (2001) The evolution of bioluminescent oxygen consumption as an ancient oxygen detoxification mechanism. J. Mol. Evol., 52, 321–332.
17. Czyz, A., Wrobel, B. and Wehrzy, G. (2000) Vibrio harveyi bioluminescence plays a role in stimulation of DNA repair. Microbiology (Reading, Engl.), 146, 283–288.
18. Walker, E.L., Bose, J.L. and Stabb, E.V. (2006) Photolyase confers resistance to UV light but does not contribute to the symbiotic benefit of bioluminescence in Vibrio fisheki ES114. Appl. Environ. Microbiol., 72, 6600–6606.
19. Zarubin, M., Belkin, S., Ionescu, M. and Genin, A. (2012) Bacterial bioluminescence as a lure for marine zooplankton and fish. Proc. Natl. Acad. Sci. U.S.A., 109, 853–857.
20. Schwartzman, J.A. and Ruby, E.G. (2016) A conserved chemical dialog of mutualism: lessons from squid and vibrio. Microb. Biogeosci., 18, 1–10.
21. Miyashiro, T. and Ruby, E.G. (2012) Shedding light on bioluminescence regulation in Vibrio fischeri. Cell, 151, 1–10.
22. Engebretson, J., Nealon, K. and Silverman, M. (1983) Bacterial bioluminescence: Isolation and genetic analysis of functions from Vibrio fischeri. Cell, 22, 773–781.
23. Engebretson, J. and Silverman, M. (1984) Identification of genes and gene products necessary for bacterial bioluminescence. Proc. Natl. Acad. Sci. U.S.A., 81, 4154–4158.
24. Meighen, E.A. (1991) Molecular biology of bacterial bioluminescence. Microbiol. Mol. Biol. Rev., 55, 123–142.
25. Baldwin, T.O., Christopher, J.A., Rauschel, F.M., Sinclair, J.F., Ziegler, M.M., Fisher, A.J. and Rayment, I. (1995) Structure of bacterial luciferase. Curr. Opin. Struct. Biol., 5, 798–809.
26. Sparks, J.M. and Baldwin, T.O. (2001) Functional implications of the unstructured loop in the βαβ8 barrel structure of the bacterial luciferase a subunit. Biochemistry, 40, 15436–15443.
27. Brodin, E., Winkler, A. and Machereux, P. (2018) Molecular mechanisms of bacterial bioluminescence. Comput. Struct. Biotechnol. J., 16, 551–564.
28. Lin, J.-W., Chao, Y.-F. and Weng, S.-F. (2001) Riboflavin synthesis genes ribE, ribB, ribH, ribA reside in the lux operon of
Photobacterium leiognathi. *Biochem. Biophys. Res. Commun.*, 284, 587–595.

29. Urbanczyk,H., Ast,J.C. and Dunlap,P.V. (2011) Phylogeny, genomics, and symbiosis of Photobacterium. *FEMS Microbiol. Rev.*, 35, 324–342.

30. Forst,S. and Neilson,K. (1996) Molecular biology of the symbiotic-pathogenic bacteria Xenorhabdus spp. and Photorhabdus spp. *Microbiol. Rev.*, 60, 21–43.

31. Ast,J.C., Urbanczyk,H. and Dunlap,P.V. (2007) Natural Merodiploidy of the lux-rib operon of photobacterium leiognathi from coastal waters of honshu, Japan. *J. Bacteriol.*, 189, 6148–6158.

32. Kasai,S., Okada,K., Hoshino,A., Iida,T. and Honda,T. (2007) Natural horizontal transfer in vibriaceae. *J. Bacteriol.*, 190, 3494–3504.

33. Urbanczyk,H., Furukawa,T., Yamamoto,Y. and Dunlap,P.V. (2012) Natural replacement of vertically inherited lux-rib genes of Photobacterium aquimaris by horizontally acquired homologues. *Environ. Microbiol. Rep.*, 4, 412–416.

34. O’Grady,E.A. and Wimpee,CF. (2008) Mutations in the lux operon of Vibrio cholerae. *Appl. Environ. Microbiol.*, 74, 61–66.

35. Zo,Y.-G., Chokesajjwatee,N., Grim,C., Taviani,E., Alam,M., Huq,A., Sack,R.B. and Arakawa,E., Watanabe,H. (2000) Temperature affects species distribution in marine fishes: a study of their distribution, densities, and dispersion. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 14181–14184.

36. Miller,S.D., Haddock,S.H.D., Elvidge,C.D. and Lee,T.F. (2005) Detection of a bioluminescent milky sea from space. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 14181–14184.

37. Grimes,D.J., Ford,T.E., Colwell,R.R., Baker-Austin,C., Lee,K.H. and Ruby,EG. (1994) Effect of the squid host on the drivers from orbit. *Microb. Ecol.*, 67, 489–500.

38. Capella-Gutiérrez,S., Silla-Martínez,J.M. and Gabaldón,T. (2009) SWISS-PROT: protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, 28, 45–48.

39. O’Leary,N.A., Wright,M.W., Brister,J.R.,Ciufo,S., Haddad,D., McVeigh,R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjie,D., et al. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 44, D733–D745.

40. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410.

41. Sullivan,M.J., Petty,N.K. and Beatson,S.A. (2011) Easyfig: a genome visualization tool. *Bioinformatics*, 27, 1009–1010.

42. Jones,B.J., Thompson,T.B., Shogen,J.B., Baldwin,T.O. and Raymont,I. (1996) The 1.5-A resolution crystal structure of bacterial luciferase alpha subunit. *Biochemistry*, 35, 11255–11262.

43. Karsenti,E., Acinas,S.G., Bork,P., Bowler,C., De Vargas,C., Raes,J., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2013) GenBank. *Nucleic Acids Res.*, 41, D36–D42.

44. Jones,D.T. and Cozzetto,D. (2015) DISOPRED3: precise disorder predictions with annotated protein-binding activity. *Bioinformatics*, 31, 857–863.

45. Tamburini,C., Canals,M., de,M.X.D., Houpert,L., Lefèvre,D., O’Grady,E.A. and Wimpee,CF. (2008) Phylogenetic analysis of the incidence of lux gene horizontal transfer in vibriaceae. *Appl. Environ. Microbiol.*, 74, 607–611.

46. Jones,B.W., Maruyama,A., Ouverney,C.C. and Nishiguchi,M.K. (2000) Phylogenetic analysis of the incidence of lux gene horizontal transfer in vibrionaceae. *Microbiol. Rev.*, 64, 38–59.
metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci. Data*, 4, 170093.

74. Pernice,M.C., Giner,C.R., Logares,R., Perera-Bel,J., Acinas,S.G., Duarte,C.M., Gasol,J.M. and Massana,R. (2016) Large variability of bathypelagic microbial eukaryotic communities across the world’s oceans. *ISME J.*, 10, 945–958.

75. Rusch,D.B., Halpern,A.L., Sutton,G., Heidelberg,K.B., Williamson,S., Yooseph,S., Wu,D., Eisen,J.A., Hoffman,J.M., Remington,K. et al. (2007) The Sorcerer II Global Ocean sampling Expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.*, 5, e77.

76. Williamson,S.J., Allen,L.Z., Lorenzi,H.A., Fadrosh,D.W., Brami,D., Thiagarajan,M., McClure,J.P., Towchigrechko,A., Yooseph,S. and Venter,J.C. (2012) Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One*, 7, e42047.

77. Kopf,A., Bicak,M., Kottmann,R., Schnetzer,J., Kostadinov,I., Lehmann,K., Fernandez-Guerra,A., Jeantongh,C., Rahav,E., Ullrich,M. et al. (2015) The ocean sampling day consortium. *Gigascience*, 4, 27.

78. Langmead,B. and Salzberg,SL. (2012) Fast gapped-read alignment of short DNA sequences. *Gigascience*, 14, 10–11.

79. Venter,J.C. (2012) Metagenomic exploration of viruses throughout the oceans. *Science*, 334, 44–47.

80. Caro-Quintero,A. and Konstantinidis,K.T. (2015) Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. *ISME J.*, 9, 958–967.

81. Vincent,A.T., Schiettekatte,O., Goarant,C., Neela,V.K., Bernet,E., Thibeaux,R., Ismail,N., Mohd Khalid,M.K.N., Amran,F., Masuzawa,T. et al. (2019) Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. *PLoS Negl. Trop. Dis.*, 13, e0007270.

82. Xu,Y., Zhu,Y., Wang,Y., Chang,Y.-F., Zhang,Y., Jiang,X., Zhaah,X., Zhu,Y., Zhang,J., Zeng,L. et al. (2016) Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic Leptospira. *Sci. Rep.*, 6, 20020.

83. Perna,M., Chiu,china,M., Waite,D.W., Hinke,C., Skarshewski,A., Chaumeil,P.-A. and Hugenholtz,P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.*, 36, 996–1004.

84. Subbian,S., Mehta,P.K., Cirillo,S.L. and Cirillo,J.D. (2007) The Mycobacterium marinum me2 locus displays similarity to bacterial bioluminescence systems and plays a role in defense against reactive oxygen and nitrogen species. *BMC Microbiol.*, 7, 4.

85. Janagama,H.K., Tounkang,S., Cirillo,S.L.G., Zinniel,D.K., Barletta,R.G. and Cirillo,J.D. (2013) Molecular analysis of the Mycobacterium tuberculosis lux-lace me2 operon. *Tuberculosis*, 93, 583–587.

86. El-Etr,S.H., Subbian,S., Cirillo,S.L.G. and Cirillo,J.D. (2004) Identification of two Mycobacterium marinum loci that affect interactions with macrophages. *Infect. Immun.*, 72, 6902–6913.

87. Ellis,H.R. (2010) The F4MN-dependent two-component monoxygenase systems. *Arch. Biochem. Biophys.*, 497, 1–12.

88. Chaumeil,P.-A. and Hugenholtz,P. (2018) A standardized structural and functional marker gene-based phylogeny of the genus Leptospira through the prism of genomics. *PLoS Negl. Trop. Dis.*, 12, e0007270.

89. Lackner,G., Peters,E.E., Helfrich,E.N. and Piel,J. (2017) Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc. Natl. Acad. Sci. U.S.A.*, 114, E347–E356.

90. Parks,D.H., Chuvochina,M., Waite,D.W., Rinke,C., Skarshewski,A., Barletta,R.G. and Cirillo,J.D. (2013) Molecular analysis of the Mycobacterium marinum mel2 locus displays similarity to bacterial bioluminescence systems and plays a role in defense against reactive oxygen and nitrogen species. *BMC Microbiol.*, 7, 4.

91. Hugenholtz,P. (2012) Metagenomic exploration of viruses throughout the oceans. *Science*, 334, 44–47.

92. Schneiker,S., Martins dos Santos,V.A.P., Bartels,D., Bekel,T., Brecht,M., Buhrmester J., Chernikova,T.N., Denaro,R., Ferrer,M., Gertler,C. et al. (2006) Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium Alcanivorax borkumensis. *Nat. Biotechnol.*, 24, 997–1004.

93. Shibata,A.K. and Robert,F.M. (2009) Shifts in alkane-degrading bacteria genotypes during bioremediation of a vegetated coastal soil. *World J. Microbiol. Biotechnol.*, 25, 1657–1657.

94. Dávila-Céspedes,A., Hufendiek,P., Crusell,M., Schäferle,T.F. and König,G.M. (2016) Marine-derived myxobacteria of the suborder Nannocystineae: an underexplored source of structurally intriguing and biologically active metabolites. *CellMot.J. Org. Chem.*, 12, 969–984.