


RESEARCH ARTICLE

Open Access



# Genomic and ecological attributes of marine bacteriophages encoding bacterial virulence genes

Cynthia B. Silveira<sup>1,2,3\*</sup> , Felipe H. Coutinho<sup>4</sup>, Giselle S. Cavalcanti<sup>1,2</sup>, Sean Benler<sup>1,2</sup>, Michael P. Doane<sup>1,2,5</sup>, Elizabeth A. Dinsdale<sup>1,2</sup>, Robert A. Edwards<sup>1,2</sup>, Ronaldo B. Francini-Filho<sup>6</sup>, Cristiane C. Thompson<sup>7</sup>, Antoni Luque<sup>2,8,9</sup>, Forest L. Rohwer<sup>1,2</sup> and Fabiano Thompson<sup>10</sup>

## Abstract

**Background:** Bacteriophages encode genes that modify bacterial functions during infection. The acquisition of phage-encoded virulence genes is a major mechanism for the rise of bacterial pathogens. In coral reefs, high bacterial density and lysogeny has been proposed to exacerbate reef decline through the transfer of phage-encoded virulence genes. However, the functions and distribution of these genes in phage virions on the reef remain unknown.

**Results:** Here, over 28,000 assembled viral genomes from the free viral community in Atlantic and Pacific Ocean coral reefs were queried against a curated database of virulence genes. The diversity of virulence genes encoded in the viral genomes was tested for relationships with host taxonomy and bacterial density in the environment. These analyses showed that bacterial density predicted the profile of virulence genes encoded by phages. The Shannon diversity of virulence-encoding phages was negatively related with bacterial density, leading to dominance of fewer genes at high bacterial abundances. A statistical learning analysis showed that reefs with high microbial density were enriched in viruses encoding genes enabling bacterial recognition and invasion of metazoan epithelium. Over 60% of phages could not have their hosts identified due to limitations of host prediction tools; for those which hosts were identified, host taxonomy was not an indicator of the presence of virulence genes.

**Conclusions:** This study described bacterial virulence factors encoded in the genomes of bacteriophages at the community level. The results showed that the increase in microbial densities that occurs during coral reef degradation is associated with a change in the genomic repertoire of bacteriophages, specifically in the diversity and distribution of bacterial virulence genes. This suggests that phages are implicated in the rise of pathogens in disturbed marine ecosystems.

**Keywords:** Marine phage, Virulence genes, Lysogeny, Virome, Bacterial pathogenicity

## Background

With a total estimated abundance of  $10^{31}$  particles, bacteriophages are the most abundant biological entities on Earth, and represent an untapped wealth of genetic information [1]. Bacteriophage genomes undergo frequent lateral gene transfers, and phage-encoded genes can be shared with microbial hosts and fixated under selective

pressure [2–4]. Viral genome size is constrained by the capsid volume and mutation rates, resulting in condensed genomes with frequent overlapping open reading frames [5–7]. Thus, the ubiquitous presence of genes encoding bacterial cellular functions in viral particles suggests that most of these genes bring adaptive advantage to the viruses [3, 4]. Yet, the environmental drivers of phage genomic composition just recently started to be described [3, 8, 9].

The expression of phage genes during infection confers new functions and modulates existing host functions [10–12]. Bacterial virulence genes are often carried by

\* Correspondence: [cynthiasilveira@gmail.com](mailto:cynthiasilveira@gmail.com)

<sup>1</sup>Department of Biology, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA

<sup>2</sup>Viral Information Institute, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA

Full list of author information is available at the end of the article



temperate phages, and lysogenic conversion (the change in bacterial phenotype as a result of phage integration) is a major mechanism for the emergence of pathogens [13]. The genus *Vibrio* includes several examples of virulence acquisition through phage integration, including the human pathogen *Vibrio cholerae* [14]. The CTX toxin in *V. cholerae* is a canonical example of phage-encoded pathogenicity through the direct acquisition of a toxicity function, but also through the regulation of the global bacterial transcriptome increasing the pathogen's fitness in the animal-associated environment [15]. Prophages inserted in the genome of the coral pathogen *Vibrio coralilyticus* show high nucleotide sequence identity and similar gene organization with virulence gene-encoding *V. cholerae* phages, suggesting that lysogenic conversion cause coral disease [16, 17]. Altogether, these studies suggest that phage-mediated bacterial virulence contribute to pathogenicity in many marine diseases. However, a community-level analysis of phage-encoded virulence genes in marine environments is still missing.

The rise of fleshy macroalgae (coral competitors) in degraded coral reefs fuels microbialization, the increase in bacterial biomass and energetic demands [18–21]. High bacterial densities are accompanied by increases in the abundance of temperate phages encoding bacterial virulence genes and the frequency of lysogenic infections, a dynamic named *Piggyback-the-Winner* (PtW) [20, 22–24]. During microbialization, the bacterial community also becomes dominated by super-heterotrophs, including Gammaproteobacteria and Bacteroidetes [13, 25–28]. If the phage-encoded virulence genes bring niche expansion and competitive advantage to the bacterial hosts during microbialization, the selection of these genes will lead to genomic adaptation observed as changes in the gene functions and relative abundances. These changes should be correlated with both bacterial densities and phage host taxonomy.

A meta-analysis of virome-assembled viral genomic sequences from coral reef boundary layers (water overlaying corals) in the Atlantic and Pacific was employed here to test these predictions. Phage-encoded virulence gene profiles were significantly predicted by microbial densities. However, there was only marginal evidence for a role of host taxonomy in virulence gene distribution. These findings indicate that phages represent a reservoir of bacterial virulence factors in marine environments that contributes to the rise of pathogens during microbialization.

## Results

### Viral community structure and diversity

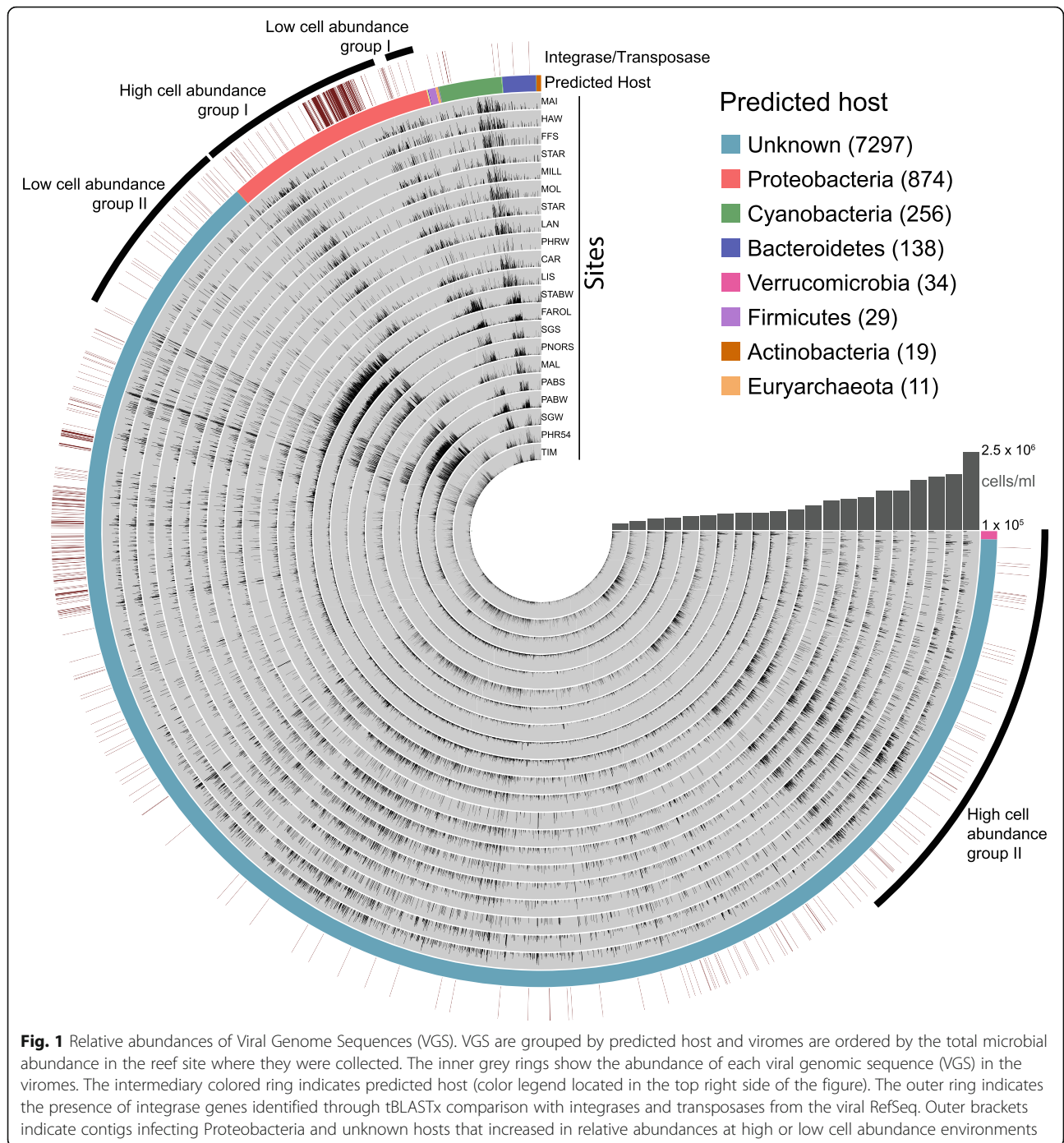
A total of 28,483 Viral Genomic Sequences (VGS) representing virome-assembled viral genomic sequences (herein referred to as viral genomes) composed the viral community in the coral reefs analyzed here, recruiting

$49.8 \pm 2.2\%$  (mean  $\pm$  SD) of virome reads per site (Fig. 1). The host of most of these viruses could not be predicted (24,297 genomes recruiting 64.5% of all hits, on average across all samples), followed by viruses predicted to infect Proteobacteria (2281 genomes with 21.8% of hits), Cyanobacteria (1084 genomes with 11.5% of hits), and others (821 genomes with 1.98% of hits). The phage community structure, defined by the relative abundances of phage genomes, was significantly predicted by microbial densities at the reef site (high and low cell abundance groups in Fig. 1 and non-Metric Multidimensional Scaling analysis in Additional file 1: Figure S1, permutational linear model  $p = 0.001$ , pseudo- $F_{1,19} = 5.42$  using the relative abundances of genomes in each virome as response and  $\text{Log}_{10}$  of cell abundance as predictor variable).

The rank-abundance curve built with mean relative abundances of viral genomes across all 21 viromes indicated that the community was highly diverse (Fig. 2 and Additional file 1: Figure S1). Only two members displayed abundances above 1%. Site-specific diversity was  $7.47 \pm 0.19$  for Shannon index (mean  $\pm$  SE),  $14,589 \pm 1481$  for species abundance, and  $0.79 \pm 0.01$  for evenness (Additional file 1: Table S1 shows diversity indexes for each site). The Shannon diversity had a negative relationship with microbial density in each site (linear regression  $p = 0.04$ ,  $R^2 = 0.18$ , Additional file 1: Figure S3A). Species abundance estimates were also negatively related with microbial abundances, having a steeper and significant negative slope (linear regression  $p = 4.53e-05$ ,  $R^2 = 0.59$ , Additional file 1: Figure S3B). The steep decrease in viral species abundance with increasing microbial abundance led to no change in community evenness despite the decrease in Shannon diversity (linear regression between evenness and microbial abundance  $p = 0.63$ ).

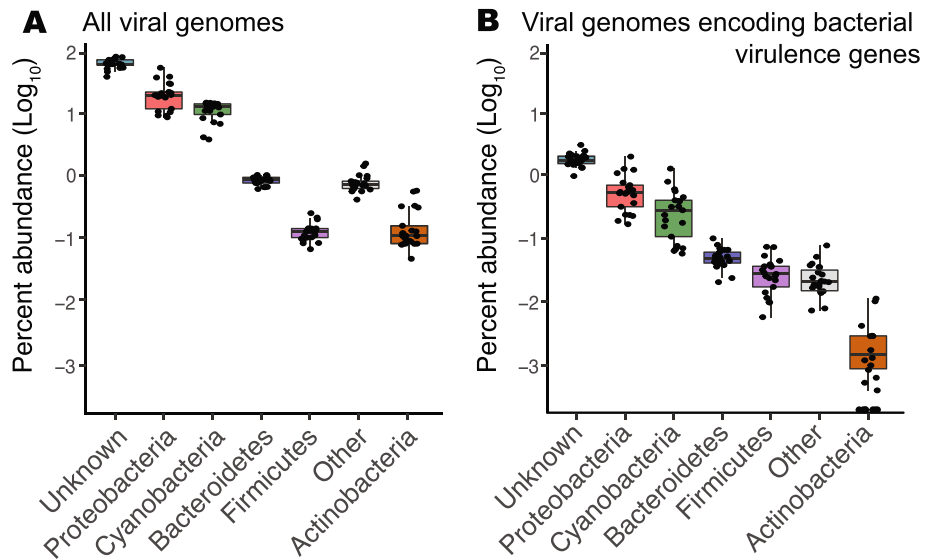
### Virulence gene profile

A total of 1149 viral genomes accounting for 2 to 4% of the viral community encoded at least one bacterial virulence gene (Fig. 1 and Additional file 1: Figure S1). There was a trend for higher frequency and number of copies of virulence genes in low abundance viruses, although the relationship was not significant (Additional file 1: Figure S1, inlet; linear regression  $p = 0.08$ ,  $a = 0.14$ ). Most of the virulence-encoding viral genomes infected unknown hosts (63%), followed by those predicted to infect Proteobacteria (21%), Cyanobacteria (11%), and Bacteroidetes (2%) (Fig. 2b). This profile is similar to the host prediction of the whole viral community, with the exception of viruses infecting Firmicutes, which were over-represented in the community encoding virulence genes relative to the whole community, and those infecting Actinobacteria, which displayed the opposite pattern (Fig. 2a).

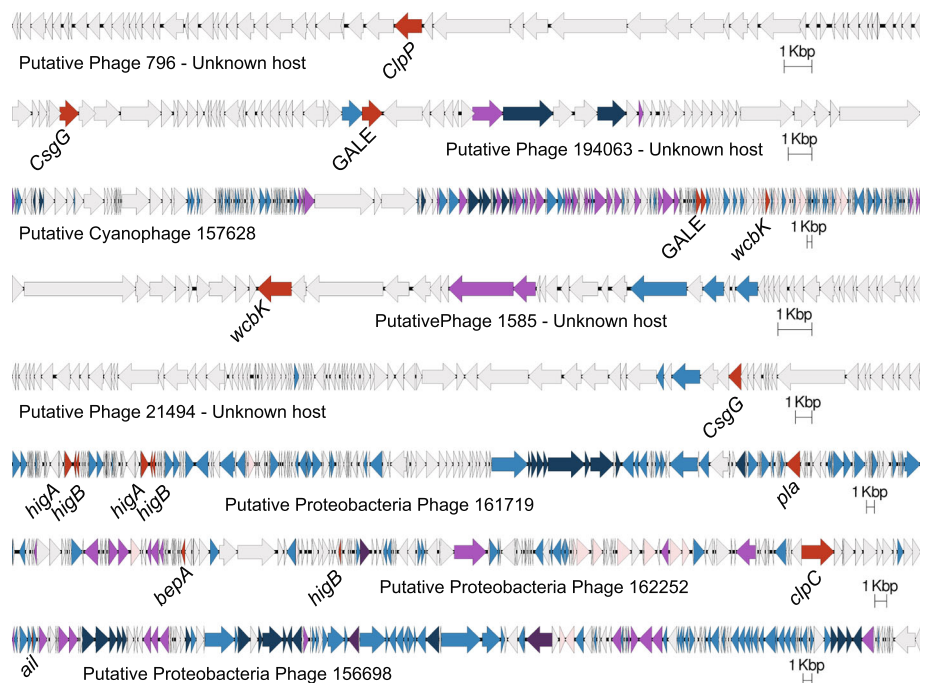


The protein annotations and genome composition of the 30 most abundant viral genomes encoding bacterial virulence genes showed that these genomic sequences varied from 5.4 to 190 Kbp in length and were predicted to infect unknown hosts (13), Proteobacteria (11) and Cyanobacteria (6). Their relative abundances and annotations are provided in Additional file 1: Table S2. About 70% of the open reading frames (ORFs) in these genomes encoded putative proteins with unknown

functions, a common characteristic of phage genomes (Fig. 3). The most abundant one, VGS 798 (0.17% of recruited reads), infects an unknown host and except for the predicted virulence gene, all the remaining ORFs encoded putative proteins of unknown function. VGS 194063, the second most abundant, encoded phage structural and replication proteins, and two virulence factors: *csgG* (Curli production assembly/transport component) and UDP-glucose epimerase (GALE). They are



**Fig. 2** Predicted hosts of virulence-encoding viruses. Relative abundance ( $\text{Log}_{10}$ ) of viral genomes grouped by predicted host. **a** Abundance of genomes encoding bacterial virulence genes and **b** abundance of all viral genomes in the coral reef communities. In both cases, most viruses infect unidentified hosts, followed by Proteobacteria and Cyanobacteria



**Fig. 3** Genomes of predicted viruses encoding bacterial virulence genes. Arrows indicate Open Reading Frames (ORFs) predicted from nucleotide sequences. Bacterial virulence genes are in red, with their specific gene annotation. Gray arrows indicate putative ORFs with unknown function, light blue indicates genes of unknown function identified as phage genes, dark blue indicates phage structural genes, purple indicates an integrase or transposase, and light pink indicates auxiliary metabolic genes. Individual scale bars are provided for each genome

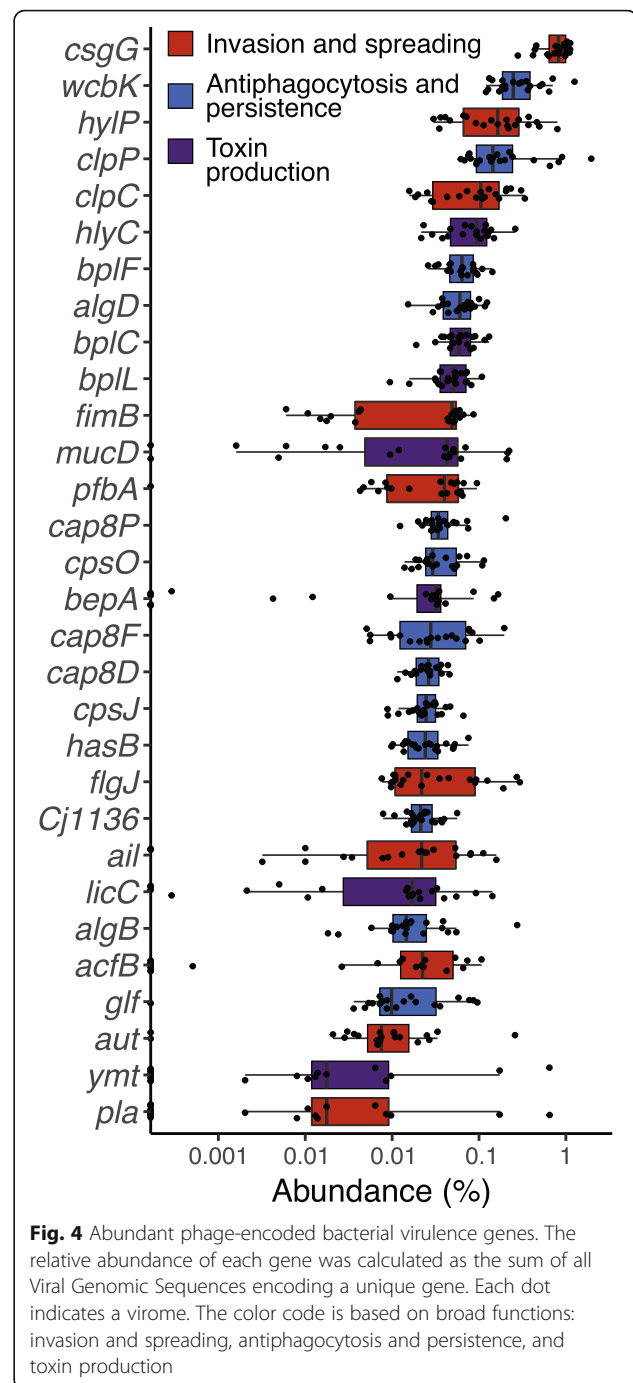
followed by Cyanophage VGS 157628, which had a genome 190 Kbp-long, encoded multiple T4-like structural and replication proteins and the genes GALE and *wcbK* (GDP-mannose 4,6-dehydratase). Three Proteobacteria-infecting phage genomes are shown in Fig. 3, two of which encoded *hig* genes, involved in a toxin-antitoxin system used by phages to regulate bacterial protein translation modulating virulence [29]. These proteobacterial phages also encoded virulence genes directly involved in attachment and invasion of eukaryotic hosts: *pla* (Plasminogen activator), *bepA* (Protein adenyltransferase) and *ail* (attachment and invasion locus).

When summing the abundance of all phage genomic sequences encoding a unique virulence gene, the most abundant genes were involved in eukaryotic host attachment, invasion, immune system evasion, and toxin production (Fig. 4). The most abundant genes were *csgG* (Curli production assembly/transport component, involved in host invasion), *wcbK* (GDP-mannose 4,6-dehydratase, involved in immune evasion), *hylP* (hyaluronidase, involved in spreading through animal tissue), *clpP* and *clpB* (proteases involved in immune system evasion), *hlyC* (hemolysin C, a toxin), and *bplF*, *C* and *L* (Lipopolysaccharide biosynthesis protein, involved in antiphagocytosis), among others. The abundances of the top 30 virulence genes, as calculated by the sum of abundances of all viral genomes encoding a unique gene are provided in Additional file 1: Table S3).

#### Drivers of virulence gene profiles

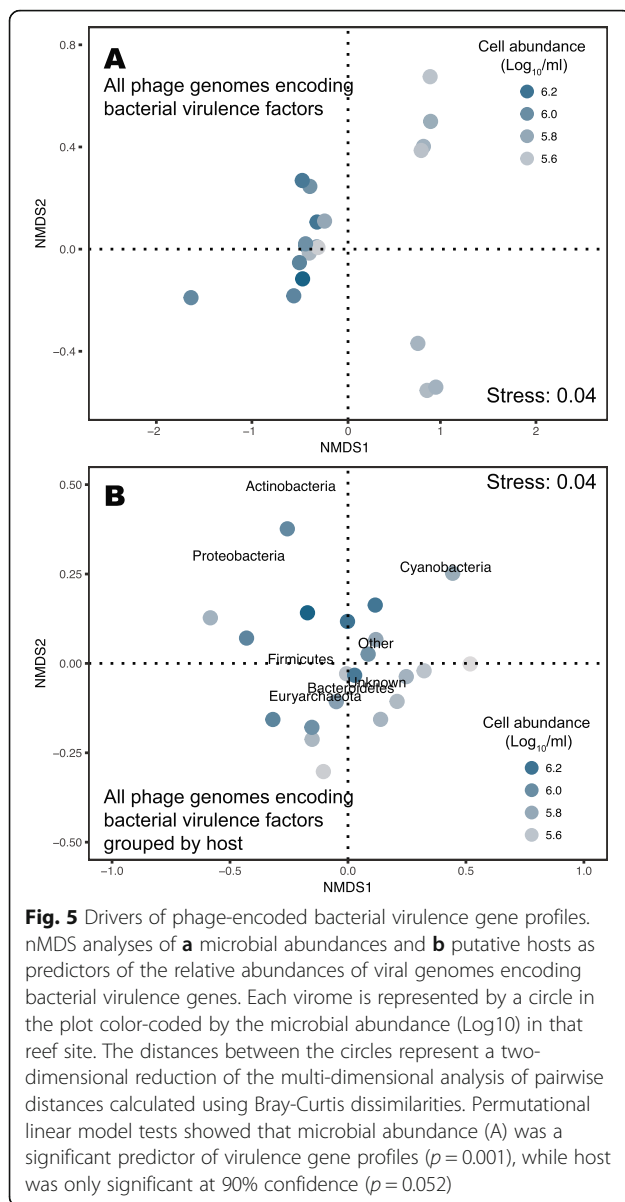
The abundances of viral genomes encoding virulence genes were significantly predicted by environmental microbial abundances (Fig. 5a; permutational linear model  $p = 0.001$ , pseudo- $F_{1,19} = 4.48$  using  $\log_{10}$  of cell abundance as predictor variable). A second nMDS analysis using the relative abundance of each virulence gene (calculated the sum of all viral genomes encoding that given gene) and cell abundance as predictor showed the same pattern, with virulence gene profile being significantly predicted by cell abundance (Additional file 1: Figure S4, permutational linear model  $p = 0.001$ , pseudo- $F_{1,19} = 4.23$  using  $\log_{10}$  of cell abundance as predictor variable). Viral genomes were then grouped according to host phylum and host annotation was tested as a predictor of the relative abundances of genomes encoding bacterial virulence genes. This analysis showed that host profile was a weak predictor of virulence gene profiles (Fig. 5b, permutational linear model  $p = 0.052$ , pseudo- $F_{1,19} = 3.14$ ).

A permutational random forest statistical learning approach determined which virulence gene-encoding genomes were best at predicting the differences across the cell abundance gradient. The random forest analysis showed that the abundance of virulence-encoding genomes explained 39.2% of the variance in cell abundances across viromes. The genomes that displayed high importance on the random forest (increase in mean square error



**Fig. 4** Abundant phage-encoded bacterial virulence genes. The relative abundance of each gene was calculated as the sum of all Viral Genomic Sequences encoding a unique gene. Each dot indicates a virome. The color code is based on broad functions: invasion and spreading, antiphagocytosis and persistence, and toxin production

and  $p$ -values below 0.05 in the permutation) were selected (Fig. 6 and Additional file 1: Figure S5). At high cell abundances, 8 genomes encoding genes involved in two broad functions were enriched: invasion and immune system evasion. The specific genes enriched were *tsr* (chemotaxis and invasion), *fimB* (regulating fimbria assembly for attachment), *ail* (attachment and invasion), and *clp*, *bsc*, *alg* and *muc*, involved in antiphagocytosis. All the eight virulence-encoding VGS enriched at high cell abundance



were predicted to infect Proteobacteria, and five encoded an integrase or transposase.

At low microbial abundances, the 12 viral genomes with highest importance in the random forest analysis had lower relative abundances compared to the ones at high microbial abundances (Fig. 6). Ten of these genomes were predicted to infect unknown hosts, one was predicted to infect Proteobacteria and one to infect Flavobacteria. None of these encoded an integrase or transposase. When the gene abundance (as the sum of all phages encoding a unique gene) was tested by the same random forest analysis to predict cell abundance, only 5.06% of the variance was explained (Additional file 1: Figure S6).

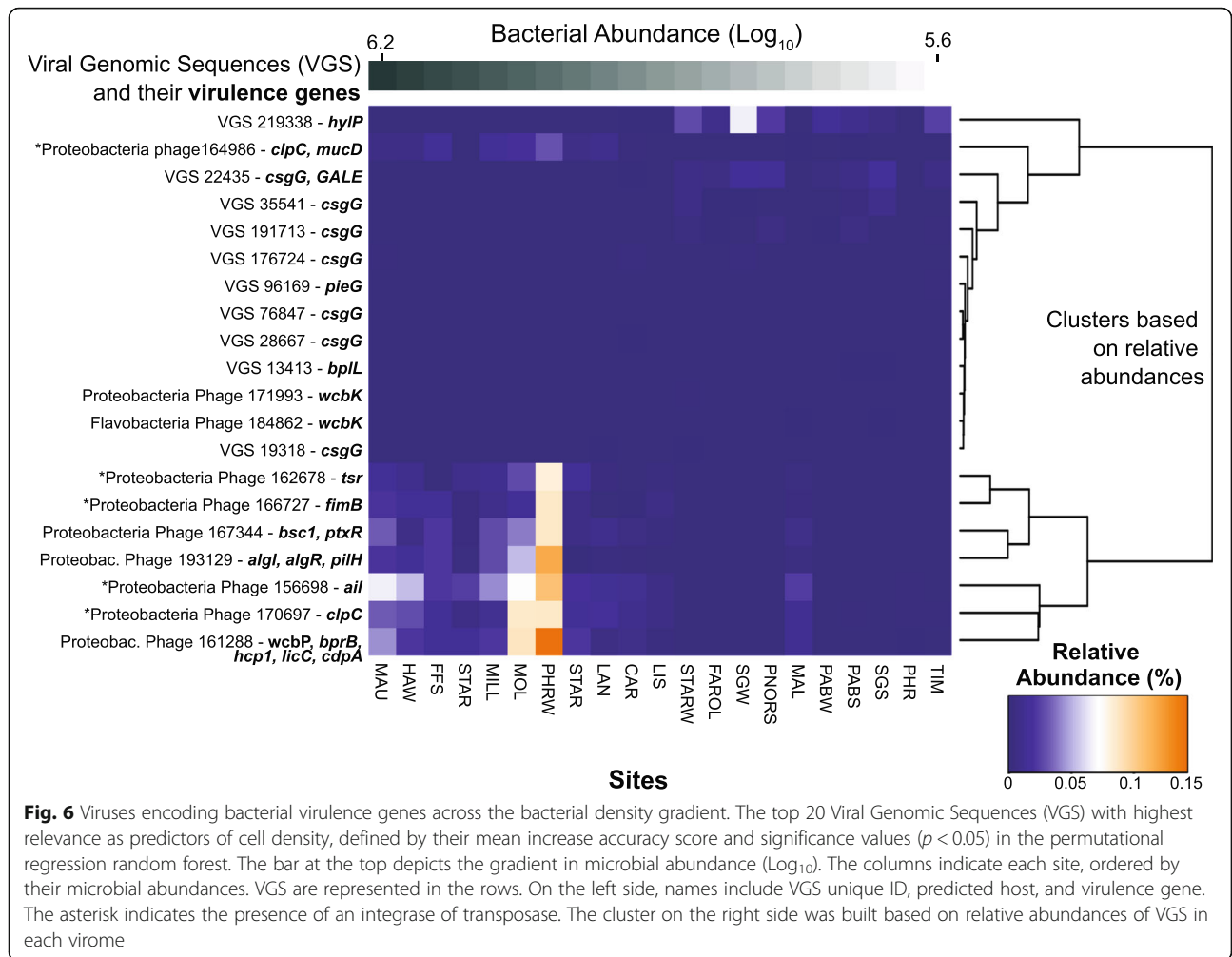
## Discussion

### Drivers of phage-encoded virulence gene profiles

Here we tested the hypothesis that in coral reefs, the distribution of phage genes with homology to bacterial virulence factors is associated with microbial densities and host taxonomy. This association is predicted to result from an increased frequency of viral infection and selection of genes that bring competitive advantages to the bacterial host. The results corroborated the first predictions of this hypothesis (the relationship between phage-encoded virulence and microbial density), but did not support the second prediction (relationship between bacterial host and phage virulence genes). The decoupling between functional genes and taxonomy is a common feature of microbial communities and has been previously observed in coral reef microbiomes [25].

The significant relationships between microbial density and the abundance profiles of the whole viral community (Additional file 1: Figure S2) and the fraction of the community encoding virulence factors (Fig. 5a) indicated that host availability is a major driver of phage community structure. These results were consistent with previous observations of viral and bacterial community structure being associated with bacterial densities [19, 23, 27]. The decrease in diversity and richness of virulence-encoding phage genomes with increasing microbial density (Additional file 1: Figure S3) supports the idea of increased abundance of opportunistic strains at high densities [23, 27]. If the acquisition of a virulence gene by a bacterium during lysogeny increases fitness, it would also increase the abundance of this strain in the environment. In this case, high microbial density is an outcome of the gene acquisition, closing a positive feedback loop of microbial biomass accumulation [20, 21].

Phages infecting Proteobacteria were the most abundant among viral genomes for which putative hosts were identified (Fig. 2). Proteobacteria, mainly belonging to the genus *Vibrio*, are common marine pathogens found in high abundances in microbialized reefs, stressed corals, and other animals [25, 28, 30]. Lysogenic conversion was proposed as a virulence mechanism in the coral pathogen *Vibrio corallilyticus*, based on sequence similarity between *V. corallilyticus* prophages and virulence-encoding *V. cholerae* phages [16]. The results described here support the role of lysogenic conversion in coral reef *Vibrio* and extend that to other bacterial groups, suggesting that the lysogenic conversion mechanism is widespread among marine pathogens. Another possible explanation is that these genes are participating in the mediation of commensal or even mutualistic relationships, as marine *Vibrio* can establish diverse symbiotic interactions with eukaryotes [31]. Most virulence-encoding viruses described here infected unknown hosts



(Fig. 2), limiting further interpretation of the host-related results, despite the best available tools being applied for host inference [3, 4, 32, 33]. Other biases derived from sample preparation methods could also interfere with these analyses.

### Phage-encoded virulence genes and genomic islands

The most abundant phage-encoded bacterial virulence genes and those enriched at high bacterial densities encoded proteins that are expressed on the bacterial cell surface during phage infection and have functions of invasion, spreading, and immune system evasion (Figs. 4 and 6). The lateral acquisition of these genes and traits is the first step for a bacterial strain that is originally free-living to explore a new niche by associating with an animal host [34, 35]. Exploring this new niche requires successful competition with resident microbiome associated with that animal, and evasion from the animal immune system [36]. Toxins and immune evasion genes perform this function, while other unidentified genes may play roles in bacteria-bacteria competition and bacteria-animal communication.

Many of the genes identified here are located in genomic islands or flanked by transposons in reference bacterial genomes. Some examples are: *hlyC*, encoding the toxin hemolysin and found in genomic islands of pathogenic *E. coli* predicted to originate from defective prophages: 10 to 200 kb regions containing an integrase gene, flanked by tRNA genes, and with GC content that significantly deviates from the host genome [37]; Homologs of Clp proteases, some of the most abundant genes in this dataset, are common in bacterial genome and can have different functions, being exchanged between strains through homologous flanking regions. The viral version of this gene is involved in both virion assembly and regulation of the expression of proteins mediating bacterial evasion of immune cells [38, 39]; the genes *csg* and *fim*, involved in the synthesis of two types of fimbria, are enriched at high cell densities and found in genomic islands of bacterial genomes with evidence of horizontal transfer [40]. Fimbria mediate bacterial recognition and invasion of animal hosts, being common in *Pseudovibrio* spp. genomes infecting sponges, corals, flatworms, and tunicates [40].

Virulence genes, specially secreted proteins (toxins) and secretion systems are commonly flanked by mobile genetic elements in bacterial genomes, and transferred along addition systems, which kill cells in the absence of the mobile genetic element [41]. These genomic islands often originate from integrated phages [42]. Bacterial mutation rates and the selective advantages brought by the virulence gene to the bacteria lead to the complete inactivation of the integrated phage's ability to form new particles and lyse the host, while maintaining the virulence gene functionality [43].

### Lysogenic conversion in coral reefs

In coral reefs, high rates of lysogenic conversion during microbialization are predicted to contribute to the rise of pathogens [23, 44]. The functional analyses of virulence genes in phage genomes showed that the pathogenic functions encoded by phages enable bacteria to recognize and invade animals. These results suggest that phages are agents of diseases in microbialized systems. The abundant *ail* gene described here (also known as *lom* – lambda outer membrane) is carried by phages infecting *Enterobacteria* spp. and expressed on the outer membrane, triggering attachment to animal epithelial cells and initiation of infection in mice [45]. It is possible that this gene has a similar function in the marine environment, enabling *Enterobacteria* to attach to animal epithelium and explaining the implication of *Enterobacteria* spp. in white-plague disease in the coral species *Orbicella faveolata* and *Mussismilia braziliensis* [46–48]. These predictions provide a platform in the search for potential phage-mediated mechanisms causing marine diseases of unknown pathogen, such as black band disease, white plague, white pox, and stony coral tissue loss [47, 49, 50].

Many genes identified in this in silico analysis as virulence factors may have other functions in vivo. Secreted proteins and secretions systems have been associated with bacteria-bacteria communication and cooperation [41]. Many canonical virulence genes may also enable bacteria to escape protist grazing by modifying bacterial surface, inhibiting their degradation and increasing long-term persistence in association with the eukaryote [51, 52]. The gene *wcbK*, for instance, has immune system evasion roles in Proteobacteria through the modification of cell surface [53, 54], however its presence in Cyanophage genomes may indicate a function in protection against grazing [55, 56]. Here, many of the Cyanophage genomes containing virulence genes did not encode an integrase, preventing the identification of these as lytic or temperate (Fig. 1). However, even in the absence of integration, the expression of phage genes during long-term associations (such as pseudo-lysogeny or simply delayed lysis) may

modify bacterial phenotype before death by lysis [57, 58]. Phage T4, for instance, can establish pseudolysogenic infections in *E. coli* through the expression of *rI* [59]. If that is the case for the Cyanophages found here, lysis delay combined with grazing protection could increase short-term bacterial survival and long term-coexistence of the bacteria and the phage. The combined effects of high virulence and escape from viral and protist predation are predicted to exacerbate coral reef decline and microbialization by increasing bacterial fitness.

### Conclusion

Our study shows that coral reef microbialization is associated with the rise in phage-encoded virulence genes that enable bacterial recognition and invasion of metazoans. The virulence genes encoded by phage genomes were distinct between high and low microbial abundance reefs, suggesting that phage-encoded genes contribute to bacterial fitness at high microbial densities. Predicted bacterial host was not a significant variable driving the virulence gene profiles, indicating that total bacterial density is a stronger predictor of phage-mediated pathogenicity. These results point towards lysogenic conversion as a cause of polymicrobial diseases contributing to coral reef decline.

### Methods

#### Viral metagenomic dataset

Viral metagenomes (viromes) from coral reefs analyzed here were previously published in Knowles et al. 2016 [23]. Briefly, samples were collected from reef boundary layers across the Abrolhos Bank in the Southwest Atlantic (7 viromes), the Line Islands archipelago in the central Pacific (6 viromes) and the Hawaiian archipelago (8 viromes). Samples were taken from the benthos-water interface, within 30 cm above the benthic surface [60]. Before library preparation for sequencing, viral DNA was tested for the presence of bacterial 16S by PCR, and only samples that tested negative were analyzed. A description of the 21 viromes analyzed here including site geographical coordinates is provided in Additional file 1: Table S1. Viral and microbial abundances from each sample were determined by epifluorescence microscopy and flow cytometry [23].

#### Viral genome prediction

A database of viral genomic sequences (VGS) was compiled including i) viral genomes from RefSeq and ii) uncultured marine viral genomes assembled from multiple studies (including the coral reef viromes analyzed here) and amounted to a total of 195,698 sequences [3, 4, 33, 62–66]. The selection of metagenome-assembled viral genomic sequences (VGS) and exclusion of potential



bacterial contamination was performed as follows: contigs derived from the assemblies were combined, and those shorter than 2500 bp were removed to decrease the chances of false-positives. Coding DNA sequences were identified with Prodigal within Prokka and protein sequences were queried against the NCBI-nr database for annotation using Diamond, setting a minimum e-value of  $10^{-5}$ . A database of known phage genomes was built by merging genomes obtained from three sources: (1) the NCBI RefSeq database; (2) the complete marine phage genomes obtained from fosmid libraries [66] and (3) prophages identified in bacterial genomes with Vir-Sorter [67]. The database was made non-redundant by clustering the genomes with BLASTn with a 95% identity and a 40% coverage cutoff. The Dice coefficient score was used to estimate the distances between the contigs and the database of known viral genomes to organize them into a phylogenomic framework [66]. Only contigs that had at least one detectable homolog among known viruses as determined by tBLASTx searches were used for further analysis. Reads from the coral reef viral metagenomes were queried against the viral genome database using the very-sensitive-local mode of Bowtie2 [68]. This resulted in 28,483 contigs recruiting at least 10 reads from a given sample. These contigs were selected for visualization with Anvi'o [69]. The top 14,000 contigs in abundance were analyzed with Phanotate [70] to identify ORFs that were compared through BLASTp with a viral RefSeq integrase database.

Microbial hosts infected by VGSs were predicted using multiple bioinformatics methods introduced in Coutinho et al. 2017. Briefly, host prediction was based on: 1) homology matches against bacterial and archaeal genomes; 2) homology matches of the VGSs against the TARA oceans microbial contigs [71]; 3) CRISPR spacers within the microbial genomes [72]; and 4) transfer RNAs identified in VGSs.

### Virulence gene assignment

Coding sequences from VGS were compared to the virulence factors of pathogenic bacteria database from the Pathosystems Resource Integration Center, PATRIC, which contains genes with virulence functions demonstrated in vivo [73]. To avoid false-positives, the PATRIC database sequences were clustered in orthologous groups using the OrthoMCL algorithm within ght GetHomologues pipeline (default parameters, inflation factor = 1.5) [74, 75] and genes encoding ABC transporters, integrases, and chaperones were removed, as these proteins are common in the marine environment and are involved in functions unrelated to virulence. The final curated database is provided the FigShare repository (<https://doi.org/10.6084/m9.figshare.8232935>). Viral proteins predicted from contig ORFs were queried

against the curated database using BLASTp with an e-value cut-off of 0.00001 and 40% identity across 20 amino acids. The statistical analyses were performed with either the relative abundances of VGS which encode virulence factors or the abundance of virulence genes, calculated as the sum of the abundances of all VGS in which unique genes were encoded.

### Statistical analyses

Non-metric Multi-Dimensional Scaling (nMDS) followed by permutational multivariate analysis of variance using distance matrices (Adonis) with Bray-Curtis dissimilarity and 999 permutations were performed to test the relationship between bacterial density and host taxonomy. Pairwise distances between samples was calculated from i) the relative abundances of VGS, ii) relative abundance of virulence-encoding VGS, iii) virulence genes as the sum all VGS encoding each gene, and iv) VGS grouped by the host they infect. The analyses were performed using the package Vegan in R [76]. A dimension-reduction analysis was performed using supervised regression random forests in the R package rfPermute with 1000 trees [77].

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-020-6523-2>.

**Additional file 1: Figure S1.** Rank-abundance curve. **Figure S2.** nMDS analysis of the relative abundances of all Viral Genomic Sequences in each virome. **Figure S3.** Relationships between viral community diversity and microbial abundance. **Figure S4.** nMDS analysis of the relative abundances of virulence genes (calculated as the sum of Viral Genomic Sequences encoding that gene in a given sample). **Figure S5.** Viral Genomic Sequences with highest importance in the random forest analysis of the relative abundance of virulence-encoding VGS predicted by the cell abundance gradient. **Figure S6.** Virulence genes with highest importance in the random forest analysis of the relative abundance of virulence genes (as the sum of all phages encoding a given gene) predicted by the cell abundance gradient. **Table S1.** Coral reef virome sampling sites and diversity. **Table S2.** Top 30 most abundant viral genomes encoding virulence genes. **Table S3.** Top 30 most abundant virulence genes across all sites.

### Acknowledgements

We thank Esther Rubio-Portillo, Helena Vilella, Adam Barno and the SDSU bioinformatics breakfast group for thoughtful discussions on phage-encoded virulence genes and manuscript review.

### Authors' contributions

Study Design: CBS, FHC. Data Collection: CBS, FHC, GSC, RBFF, EAD, RAE, CCT. Analyses: CBS, FHC, SB, MPD, AL, FLR, FT. Manuscript Writing: CBS wrote the manuscript and all authors contributed to revisions. All authors read and approved the final manuscript.

### Funding

This research was funded by the Science Without Borders Program - CNPq (Fellowship 234702 to CBS), Gordon and Betty Moore Foundation (Grant 3781 to FLR), and Spruance Foundation. Computational resources for bioinformatic analyses were funded by NSF (Grant CNS-1305112 to RE). EAD was supported by NSF (Grants 1323809 and 1330800). FT was funded by CNPq, CAPES, and FAPERJ. Funders had no role in study design, data collection, and interpretation.

**Availability of data and materials**

All viromes analyzed here are available on the FigShare repository (<https://doi.org/10.6084/m9.figshare.4290056.v1> and <https://doi.org/10.6084/m9.figshare.4290044.v1>). The curated virulence factors database is available on the FigShare repository (<https://doi.org/10.6084/m9.figshare.8232935>).

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Department of Biology, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA. <sup>2</sup>Viral Information Institute, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA. <sup>3</sup>Department of Biology, University of Miami, 1301 Memorial Dr., Coral Gables, FL 33146, USA. <sup>4</sup>Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, Apartado 18, 03550 San Juan de Alicante, Spain. <sup>5</sup>Sydney Institute of Marine Science, 19 Chowder Bay Rd, Mosman, NSW 2088, Australia. <sup>6</sup>Centro de Biologia Marinha, Universidade de São Paulo, Rodovia Manoel Hypólito do Rego, Km 131,50, São Sebastião, SP 11600-000, Brazil. <sup>7</sup>Instituto de Biologia, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho, 373, Rio de Janeiro, RJ 21941- 599, Brazil. <sup>8</sup>Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA. <sup>9</sup>Computational Science Research Center, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA. <sup>10</sup>SAGE/COPPE, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho, 373, Rio de Janeiro, RJ 21941- 599, Brazil.

Received: 19 August 2019 Accepted: 21 January 2020

Published online: 05 February 2020

**References**

- Cobián Güemes AG, Youle M, Cantú VA, Felts B, Nulton J, Rohwer F. Viruses as winners in the game of life. *Annual Review of Virology*. 2016;3:197–214.
- Mavrich TN, Hatfull GF. Bacteriophage evolution differs by host, lifestyle and genome. *Nat Microbiol*. 2017;2:17112.
- Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CP, et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun*. 2017;8:15955.
- Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature*. 2016;536:425–30.
- Belshaw R, Pybus OG, Rambaut A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res*. 2007;17:1496–504.
- Chelikani V, Ranjan T, Kondabagil K. Revisiting the genome packaging in viruses with lessons from the "Giants.". *Virology*. 2014;466–467:15–26. <https://doi.org/10.1016/j.virol.2014.06.022>.
- Duffy S, Shackelton L a, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*. 2008;9:267–76.
- Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc Natl Acad Sci*. 2014;111:10714–9. <https://doi.org/10.1073/pnas.1319781111>.
- Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, et al. The marine viromes of four oceanic regions. *PLoS Biol*. 2006;4:2121–31. <https://doi.org/10.1371/journal.pbio.0040368>.
- Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann M-L, Brüßow H. Phage as agents of lateral gene transfer. *Curr Opin Microbiol*. 2003;6:417–24. [https://doi.org/10.1016/S1369-5274\(03\)00086-9](https://doi.org/10.1016/S1369-5274(03)00086-9).
- Pál C, Papp B, Lercher MJ. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet*. 2005;37:1372–5.
- Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Micro*. 2015;13:641–50. <https://doi.org/10.1038/nrmicro3527>.
- Brüssow H, Canchaya C, Hardt W, Bru H. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*. 2004;68:560–602.
- Munro J, Oakey J, Bromage E, Owens L. Experimental bacteriophage-mediated virulence in strains of *Vibrio harveyi*. *Dis Aquat Org*. 2003;54:187–94.
- Rivera-Chávez F, Mekalanos JJ. Cholera toxin promotes pathogen acquisition of host-derived nutrients. *Nature*. 2019;572(7768):244–8.
- Weynberg KD, Voolstra CR, Neave MJ, Buerger P, van Oppen MJH. From cholera to corals: viruses as drivers of virulence in a major coral bacterial pathogen. *Sci Rep*. 2015;5:17889.
- Santos EDO NA Jr, Dias GM, Mazotto AM, Vermelho A, Vora GJ, et al. Genomic and proteomic analyses of the coral pathogen *Vibrio coralliilyticus* reveal a diverse virulence repertoire. *ISME J*. 2011;5:1471–83.
- McDole T, Nulton J, Barott K, Felts B, Hand C, Hatay M, et al. Assessing coral reefs on a Pacific-wide scale using the microbialization score. *PLoS One*. 2012;7:e432331–10. <https://doi.org/10.1371/journal.pone.0043233>.
- Haas AF, MFMFM F, LWLW K, CECE N, Wei Y, Knowles B, et al. Global microbialization of coral reefs. *Nature Microbiology*. 2016;1:16042.
- Silveira C, Cavalcanti G, Walter J, Silva-Lima A, Dinsdale E, Bourne D, et al. Microbial processes driving coral reef organic carbon flow. *FEMS Microbiol Rev*. 2017;41:575–95.
- Silveira CB, Luque A, Roach TNF, Villela H, Barno A, Green K, et al. Biophysical and physiological processes causing oxygen loss from coral reefs. *eLife*. 2019;8:e49114.
- Pleška M, Lang M, Refardt D, Levin BR, Guet CC. Phage–host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nature Ecol Evol*. 2018;2:359.
- Knowles B, Silveira CCB, Bailey BBA, Barott K, Cantu VAV, Cobian-Guëmes AG, et al. Lytic to temperate switching of viral communities. *Nature*. 2016;531:466–70.
- Silveira CB, Silva-Lima AW, Francini-Filho RB, Marques JSM, Almeida MG, Thompson CC, et al. Microbial and sponge loops modify fish production in phase-shifting coral reefs. *Environ Microbiol*. 2015;17:3832–46.
- Kelly LW, Williams GJ, Barott KL, Carlson CA, Dinsdale EA, Edwards RA, et al. Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci*. 2014;111:10227–32. <https://doi.org/10.1073/pnas.1403319111>.
- Cárdenas A, Neave MJ, Haroon MF, Pogoreutz C, Räderer N, Wild C, et al. Excess labile carbon promotes the expression of virulence factors in coral reef bacterioplankton. *ISME J*. 2018;12:59–76.
- Dinsdale EA, Pantos O, Smruga S, Edwards RA, Angly F, Wegley L, et al. Microbial ecology of four coral atolls in the northern Line Islands. *PLoS One*. 2008;3:e15841–17.
- Castillo D, Kau K, Hussain F, Kalatzis P, Rørbo N, Polz MF, et al. Widespread distribution of prophage-encoded virulence factors in marine *Vibrio* communities. 2018; March:2–10.
- Lobato-Márquez D, Díaz-Orejas R, García-del PF. Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev*. 2016;40:592–609.
- Bruce T, Meirelles PM, Garcia G, Paranhos R, Rezende CE, de Moura RL, et al. Atherosclerosis bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. *PLoS One*. 2012;7:e366871–13. <https://doi.org/10.1371/journal.pone.0036687>.
- Tonon LAC, Silva BSO, Moreira APB, Valle C, Alves N, Cavalcanti G, et al. Diversity and ecological structure of vibrios in benthic and pelagic habitats along a latitudinal gradient in the Southwest Atlantic Ocean. *Peer J*. 2015;3:e741. <https://doi.org/10.7717/peerj.741>.
- Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiol Rev*. 2016;40:258–72.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and biogeochemical impacts of uncultivated globally abundant ocean viruses. *Nature*. 2016;053090. <https://doi.org/10.1101/053090>.
- Toft C, Andersson SGE. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet*. 2010;11:465.
- Ochman H, Lawrence JG, Groisman E a. lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000;405:299–304.
- Adair KL, Douglas AE. Making a microbiome: the many determinants of host-associated microbial community composition. *Curr Opin Microbiol*. 2017;35:23–9.

37. Schmidt H, Hensel M. Pathogenicity Islands in Bacterial Pathogenesis. *Clin Microbiol Rev.* 2004;17:14–56.
38. Michel A, Agerer F, Hauck CR, Herrmann M, Ullrich J, Ohlsen K, et al. Global Regulatory Impact of ClpP Protease of *Staphylococcus aureus* on Regulons Involved in Virulence , Oxidative Stress Response , Autolysis , and DNA Repair †. *J Bacteriol.* 2006;188:5783–96.
39. Zheng X, Zheng H, Lan R, Ye C, Wang Y, Zhang J. Identification of Genes and Genomic Islands Correlated with High Pathogenicity in *Streptococcus suis* Using Whole Genome Tiling Microarrays. *PLoS One.* 2011;6(3):e17987.
40. Alex A, Antunes A. Genus-wide comparison of *Pseudovibrio* bacterial genomes reveal diverse adaptations to different marine invertebrate hosts. *PLoS ONE.* 2018;13(5):e0194368.
41. Nogueira T, Rankin DJ, Touchon M, Taddei F, Brown SP, Rocha EPC. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr Biol.* 2009;19:1683–91.
42. Juhas M, Van Der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands : tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev.* 2009;33:376–93.
43. Bobay L, Touchon M, Rocha EPC. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci.* 2014;111:12127–32. <https://doi.org/10.1073/pnas.1405336111>.
44. Silveira CB, Rohwer FL. Piggyback-the-winner in host-associated microbial communities. *Npj Biofilms Microbiomes.* 2016;2(1):1–5.
45. Pacheco SV, González OG, Contreras GLP. The *lom* gene of bacteriophage  $\lambda$  is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. *FEMS Microbiol Lett.* 1997;156:129–32.
46. Daniels CA, Baumgarten S, Yum LK, Michell CT, Bayer T, Arif C, et al. Metatranscriptome analysis of the reef-building coral *Orbicella faveolata* indicates holobiont response to coral disease. *Front Marine Sci.* 2015;2. <https://doi.org/10.3389/fmars.2015.00062>.
47. Garcia GD, Gregoracci GB, Santos EO, Meirelles PM, Silva GGZ, Edwards R, et al. Metagenomic analysis of healthy and white plague-affected *Mussismilia braziliensis* corals. *Microb Ecol.* 2013;65:1076–86. <https://doi.org/10.1007/s00248-012-0161-4>.
48. Silveira CB, Gregoracci GB, Coutinho FH, Soares AC, Silva GGZ, Haggerty JM, et al. Bacterial community associated with the reef coral *Mussismilia braziliensis*'s boundary layer over a diel cycle. *Front Microbiol.* 2017;8:e784.
49. Bourne DG, Garren M, Work TM, Rosenberg E, Smith GW, Harvell CD. Microbial disease and the coral holobiont. *Trends Microbiol.* 2009;17:554–62. <https://doi.org/10.1016/j.tim.2009.09.004>.
50. Sutherland KP, Berry B, Park A, Kemp DW, Kemp KM, Lipp EK, et al. Shifting white pox aetiologies affecting *Acropora palmata* in the Florida Keys , 1994–2014. 2016.
51. Brüssow H. Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. *Mol Microbiol.* 2007;65:583–9.
52. Taylor-Mulneix DL, Bendor L, Linz B, Rivera I, Ryman VE, Dewan KK, et al. *Bordetella bronchiseptica* exploits the complex life cycle of *Dictyostelium discoideum* as an amplifying transmission vector. *PLoS Biol.* 2017;15:e2000420.
53. Bacon DJ, Szymanski CM, Burr DH, Silver RP, Alm RA, Guerry P. A phase-variable capsule is involved in virulence of campylobacter jejuni 81-176. *Mol Microbiol.* 2001;40:769–77.
54. Reckseidler SL, DeShazer D, Sokol PA, Woods DE. Detection of bacterial virulence genes by subtractive hybridization: identification of capsular polysaccharide of *Burkholderia pseudomallei* as a major virulence determinant. *Infect Immun.* 2001;69:34–44.
55. Steinberg KM, Levin BR. Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. *Proc R Soc B Biol Sci.* 2007;274:1921–9.
56. Yang Z, Kong F, Shi X, Zhang M, Xing P, Cao H. Changes in the morphology and polysaccharide content of *Microcystis aeruginosa* (cyanobacteria) during flagellate grazing. *J Phycol.* 2008;44:716–20.
57. Cenens W, Makumi A, Mebrhatu MT, Lavigne R, Aertsen A. Phage–host interactions during pseudolysogeny. *Bacteriophage.* 2013;3:e25029.
58. Cenens W, Makumi A, Govers SK, Lavigne R, Aertsen A. Viral transmission dynamics at single-cell resolution reveal transiently immune subpopulations caused by a carrier state association. *PLoS Genet.* 2015;11:1–19.
59. Los M, Wegryn G, Neubauer P. A role for bacteriophage T4 rI gene function in the control of phage development during pseudolysogeny and in slowly growing host cells. *Res Microbiol.* 2003;154:547–52.
60. Barott KL, Rohwer FL. Unseen players shape benthic competition on coral reefs. *Trends Microbiol.* 2012;20:621–8.
61. Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 2008;9:386. <https://doi.org/10.1186/1471-2105-9-386>.
62. López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodríguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genet.* 2017;13:e1007018.
63. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, et al. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere.* 2017;2:e00359–16.
64. Philofof A, Yutin N, Flores-Uribe J, Sharon I, Koonin EV, Béjà O. Novel abundant oceanic viruses of uncultured marine group II Euryarchaeota. *Curr Biol.* 2017;27:1362–8.
65. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *Elife.* 2015; 4:e08490.
66. Mizuno CM, Rodríguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. *PLoS Genet.* 2013;9:e1003987.
67. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ.* 2015;3:e985. <https://doi.org/10.7717/peerj.985>.
68. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9:357–9.
69. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ.* 2015; 3:e1319.
70. McNair K, Zhou C, Dinsdale EA, Souza B, Edwards RA. PHANOTATE: A novel approach to gene identification in phage genomes. *Bioinformatics.* 2019; 35(22):4537–42.
71. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science.* 2015;348: 1261359.
72. Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRtarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* 2013;10: 817–27.
73. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 2013;42:D581–91.
74. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178–89.
75. Contreras-Moreira B, Vinuesa P. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol.* 2013;79:7696–701.
76. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
77. Liaw A, Wiener M. Classification and regression by randomForest. *R news.* 2002;2:18–22.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

