Virology 499 (2016) 219–229

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses

Lisa T. Crummett [a,*,1,2], Richard J. Puxty [a,1], Claudia Weihe [a], Marcia F. Marston [b], Jennifer B.H. Martiny [a]

[a] Dept. of Ecology and Evolutionary Biology, University of California, Irvine, CA 92612, USA
[b] Dept. of Biology and Marine Biology, Roger Williams University, Bristol, RI 02809, USA

ARTICLE INFO

Article history:
Received 24 May 2016
Returned to author for revisions
16 September 2016
Accepted 17 September 2016

Keywords:
Cyanophage
Myoviruses
Auxiliary
Metabolic
Genes
Selection
Genome
Evolution
Marine

ABSTRACT

Viruses of marine cyanobacteria frequently contain auxiliary metabolic genes (AMGs) that augment host metabolism during infection, but little is known about their adaptive significance. We analyzed the distribution and genomic context of 33 AMGs across 60 cyanomyovirus genomes. Similarity in AMG content among cyanomyoviruses was only weakly correlated with phylogenetic relatedness; however, AMG content was generally conserved within the same operational taxonomic unit (OTU). A virus' AMG repertoire was also correlated with its isolation host and environment (coastal versus open ocean). A new analytical method based on shared co-linear blocks revealed that variation in the genomic location of an AMG was negatively correlated with its frequency across the genomes. We propose that rare AMGs are more frequently gained or lost as a result of fluctuating selection pressures, whereas common AMGs are associated with stable selection pressures. Finally, we describe a unique cyanomyovirus (S-CAM7) that lacks many AMGs including the photosynthesis gene *psbA*.

1. Introduction

Marine bacteriophages play a key role in ocean carbon and nutrient cycling via lysis of host cells (Breitbart et al., 2007; Fuhrman, 1999; Suttle, 2005a, 2005b), and their sheer abundance means they themselves may be an important reservoir of dissolved organic phosphorous (Bratbak et al., 1994; Jover et al., 2014). Bacteriophages also influence ocean biogeochemistry by modulating host cell metabolism during infection. These metabolic changes include expression of genes that are carried by phages but originated in bacterial cells (Anantharaman et al., 2014; Hagay et al., 2014; Mann et al., 2003; Sharon et al., 2011; Thompson et al., 2011). Such genes are referred to as auxiliary metabolic genes (AMGs) (Breitbart et al., 2007). Bacteriophages that infect the abundant marine cyanobacteria *Synechococcus* and *Prochlorococcus* (cyanophages) carry AMGs that have been acquired from their immediate host as well as more distantly-related bacteria (Ignacio-Espinoza and Sullivan, 2012; Kelly et al., 2013; Lindell et al., 2004; Millard et al., 2009; Sharon et al., 2009; Sullivan et al., 2010). Analysis of metagenomic samples has shown that AMGs are abundant, diverse, and widespread in the oceans (Williamson et al., 2008).

Cyanophage AMGs are associated with a variety of functions including photosynthesis (Mann et al., 2003; Philosof et al., 2011; Sharon et al., 2011), carbon metabolism (Thompson et al., 2011), nucleic acid synthesis and metabolism (Dwivedi et al., 2013; Hagay et al., 2014), and stress tolerance (He et al., 2001; Kelly et al., 2013; Lindell et al., 2005, 2004). AMGs linked to photosynthesis, such as *psbA* (photosystem II D1 protein), *psbD* (photosystem II D2 protein), and *hli* (highlight-inducible protein) genes, are expressed in cyanobacteria during phage infection (Clokie and Mann, 2006; Lindell et al., 2005) and thus, may make more energy available for phage reproduction (Clokie and Mann, 2006; Hellweger, 2009; Lindell et al., 2005; Millard et al., 2009). Generally, however, the adaptive significance of most cyanophage AMGs remains unclear.

Comparative genomics provides several approaches for generating hypotheses about the adaptive significance, if any, of AMGs in natural cyanophage communities. First, the prevalence of

* Corresponding author at: Dept. of Ecology and Evolutionary Biology, University of California, Irvine, CA 92612, USA.
E-mail addresses: lcrummett@soka.edu (L.T. Crummett), rpuxty@uci.edu (R.J. Puxty), cweihe@uci.edu (C. Weihe), mmarston@rwu.edu (M.F. Marston), jmartiny@uci.edu (J.B.H. Martiny).
[1] Contributed equally to the work.
[2] Current address: Soka University of America, 1 University Drive, Aliso Viejo, CA 92656, USA.

http://dx.doi.org/10.1016/j.virol.2016.09.016
0042-6822/

specific AMGs varies widely among isolate genomes, and these patterns may provide insights into their role during viral infection. For instance, *psbA*, *cobS* (cobalamin synthetase), and *mazG* (pyrophosphatase) genes have been found in all known cyanophages in the family Myoviridae (cyanomyoviruses) and are therefore part of the lineage's core genome. In contrast, two genes encoding electron transporters involved in photosynthesis, *petF* (ferredoxin) and *ptox* (plastoquinol terminal oxidase), are sporadically distributed among these viruses (Sullivan et al., 2010) and are therefore part of the flexible genome, analogous to that observed in lineages of bacteria and archaea (Polz et al., 2013; Tettelin et al., 2008). AMGs that are common among cyanophages may encode metabolic functions that are essential under the range of conditions they experience, whereas less common AMGs may be adaptive for only a subset experiencing a particular set of conditions (microhabitat or host-type) (Cordero and Polz, 2014).

Second, environmental correlations between AMG prevalence among genomes and abiotic parameters at the site of isolation provide clues to the adaptive benefit of particular AMGs (Kelly et al., 2013; Williamson et al., 2008). In ocean metagenomes, the relative abundance of some AMGs was positively correlated with temperature (Williamson et al., 2008). Among cyanophage isolates, the relative abundances of *pstS* (phosphate-binding protein) and *phoA* (alkaline phosphatase) genes were higher in those originating from regions with lower phosphate concentrations (Kelly et al., 2013), suggesting that the genes are associated with phosphate stress. Thus, environmental variables like temperature and nutrient availability may select for AMG content (the number and identity of AMGs) in cyanophages.

Finally, the genomic context of AMGs may shed light on their evolution. AMGs that confer an advantage in all circumstances may be located in highly conserved genomic regions. Alternatively, AMGs that are only adaptive under specific conditions, and therefore subject to higher rates of gain and loss, may be found in highly variable genomic regions. Indeed, previous work on cyanomyoviruses revealed a hypothesized mobile gene cassette containing four carbon metabolism AMGs (*ptox*, *petE*, *zwf*, and *gnd*, encoding plastoquinol terminal oxidase, plastocyanin, glucose 6-phosphate dehydrogenase, and 6-phosphogluconate dehydrogenase, respectively). These genes are sporadically distributed across the genomes and located in a hypervariable region between *g16* and *g17* (Millard et al., 2009; Sullivan et al., 2010). Such hypervariable regions have been identified in other T4-like phages. The regions are composed primarily of genes of unknown function, but also contain genes implicated in phage adaptation to the host (Comeau et al., 2007).

To investigate the potential adaptive role and evolution of AMGs in marine cyanomyoviruses, we examined patterns of AMG content and genomic context across 60 genomes isolated from various regions and hosts, 25 of which were newly sequenced. In contrast to previous analyses, we sequenced genomes that were both closely related and genetically diverged, allowing us to examine the degree of AMG conservation at both fine and broad phylogenetic resolution. To date, genomic comparison of AMGs has mainly focused on distantly-related viral isolates from various locations, but slight differences in AMG content between isolates with nearly identical genomes may reveal incremental adaptation to local environmental conditions and/or hosts (Petrov et al., 2010). To target closely-related genomes, we selected several isolates within previously-defined operational taxonomic units or OTUs (Marston and Sallee 2003; Clasen et al., 2013). Previous work indicates that cyanomyovirus OTUs defined by the similarity of *g20* sequences represent discrete natural populations (Deng et al., 2014; Marston and Amrich, 2009) that display seasonal and spatial biogeographic patterns (Marston et al., 2013).

We focus here on three questions: (1) At what phylogenetic scale do we find significant differences in AMG content? For instance, do members of the same operational taxonomic unit (OTU) vary in AMG content? (2) Is AMG content correlated with environmental parameters such as geography or host type, each of which could impose a selective filter on AMG content? (3) Is the genomic context of an AMG related to its frequency among the genomes, perhaps providing insight into its adaptive significance?

2. Results and discussion

2.1. Phylogenetic conservation of AMG content

We focused on the presence of 33 different AMGs (Table S1) in 60 cyanomyovirus genomes (Table 1). Among the 60 genomes, 36 OTUs (defined as $\geq$99% *g20* nucleotide similarity), were represented, 14 of which included at least two representative genomes. The genomes of isolates within an OTU had an average nucleotide identity (ANI) value of 92.6–99.9% ($\bar{x}$=98.8%) across the entire genome and 98.4–100% ($\bar{x}$=99.8%) across seven core cyanomyoviral genes (Table 1). The 33 AMGs were chosen based on prior studies (Sullivan et al., 2010), and the ability for unambiguous annotation across diverse taxa. The total number of AMGs in a genome ranged from 15 to 26 ($\bar{x}$=17.32) (Table 1), excluding two S-CAM7 isolates that we will discuss below as outliers. The frequency of an AMG across the genomes varied greatly, from only 2 occurrences of *pyrE*, *purN*, and *purH* to 100% of the genomes for *phoH*, *cobS*, *heat shock protein*, *mazG*, and *hli* (including the S-CAM7 genomes).

The similarity in AMG content (presence/absence) between representatives of any two OTUs (n=36) was weakly positively correlated (RELATE test, rho=0.354, p=0.001) with their phylogenetic similarity based on a core gene phylogeny (Fig. S1). However, AMG content among members of the same clade was still quite variable (Fig. 1a). For instance, S-RIM44 and S-RIM32 are representatives of two closely-related OTUs (Fig. S1), and yet these isolates differ in the presence of seven AMGs.

At a finer phylogenetic scale, AMG content was conserved within an OTU for the vast majority of AMGs (30 out of 33). Still, three of the 14 OTUs that had multiple representatives showed variation in AMG content among those representatives. For instance, viral isolate S-RIM44 (W2-07-0710), collected from the coastal waters of Rhode Island in 2010, belongs to the same OTU as viral isolate Syn1, collected from Woods Hole, MA in 1990, and yet they differed in the presence of phosphoribosylaminoimidazole synthetase (*purM*) (Fig. 2a), an enzyme involved in purine biosynthesis. A comparison of the genetic context of the region containing *purM* in viral isolate S-RIM44 with viral isolate Syn1 suggests that there has either been a deletion of *purM* in Syn1 or an insertion into S-RIM44 given high conservation of nucleotide identity and gene order in the upstream and downstream regions (Fig. 2a). The *purM* homolog in S-RIM44 was found on a fragment containing a *hyp-purM-hyp-nrdC* (glutaredoxin) cluster of genes. We searched for this gene cluster in other cyanomyoviruses within this dataset to possibly identify a source for recombination, but were unsuccessful. However, a distantly related *Synechococcus* virus, Syn33, also possessed a copy of *purM* downstream of the *purC-hyp-hyp* gene cluster that was shared in viral isolates S-RIM44 and Syn1, albeit in a disparate location (Fig. 2a). Thus, the presence of *purM* in this context is not unprecedented and suggests that S-RIM44 may have acquired *purM* by homologous recombination with a divergent virus.

A second example of within-OTU variability was S-CAM9, for which two of the three isolates possessed the 6-phosphogluconate dehydrogenase (*gnd*) gene (Fig. 2b). This difference appeared to be the result of a deletion of *gnd* in S-CAM9 0908SB82. The isolate