Virology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Virology

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

Dip in the gene pool: Metagenomic survey of natural coccolithovirus communities

António Pagarete [a,*], Kanthida Kusonmano [b], Kjell Petersen [b], Susan A. Kimmance [c], Joaquín Martínez Martínez [d], William H. Wilson [c,d], Jan-Hendrik Hehemann [e], Michael J. Allen [c], Ruth-Anne Sandaa [a]

[a] Department of Biology, University of Bergen, Norway
[b] Computational Biology Unit, University of Bergen, Norway
[c] Plymouth Marine Laboratory, Plymouth, UK
[d] Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA
[e] Department of Civil & Environmental Engineering, Massachusetts Institute of Technology, USA

ARTICLE INFO

Keywords:
Coccolithovirus
Metagenome
Algae virus
Emiliania huxleyi virus
EhV
Emiliania huxleyi
Giant virus
Marine metagenomics
Virome
Genomic hyper-variable region
Bloom dynamics
Viral genomic diversity

ABSTRACT

Despite the global oceanic distribution and recognized biogeochemical impact of coccolithoviruses (EhV), their diversity remains poorly understood. Here we employed a metagenomic approach to study the occurrence and progression of natural EhV community genomic variability. Analysis of EhV metagenomes from the early and late stages of an induced bloom led to three main discoveries. First, we observed resilient and specific genomic signatures in the EhV community associated with the Norwegian coast, which reinforce the existence of limitations to the capacity of dispersal and genomic exchange among EhV populations. Second, we identified a hyper-variable region (approximately 21 kbp long) in the coccolithovirus genome. Third, we observed a clear trend for EhV relative amino-acid diversity to reduce from early to late stages of the bloom. This study validated two new methodological combinations, and proved very useful in the discovery of new genomic features associated with coccolithovirus natural communities.

Introduction

Viruses that infect phytoplankton play a key role in shaping the evolution and dynamics of the oceanic micro-scale ecosystem (Fuhrman, 1999; Sandaa, 2008; Suttle, 2005). Several studies have highlighted the role of viruses as major players in high phytoplankton turnover rates, a process termed as the viral shunt (Wilhelm and Suttle, 1999). The interplay of viruses with their host communities is complex, and may assume different forms. Traditionally regarded as simple agents of mortality and catalysts for nutrient transformation (Suttle, 2005; Weinbauer and Rassoulzadegan, 2004), viruses are now also believed to play a fundamental role in controlling the biodiversity and functioning of their associated host communities (Frada et al., 2008; Thingstad, 2000; Thingstad and Lignell, 1997).

Emiliania huxleyi (Lohmann) Hay et Mohler, a single-celled phytoplankton, is the most abundant and ubiquitous coccolithophore in extant marine systems (Brown and Yoder, 1994). *E. huxleyi* is an important species with respect to past and present marine primary productivity, and in particular global carbon and sulphur cycles (Burkill et al., 2002; Westbroek et al., 1993). Poorly understood until recently, it is now clear that *E. huxleyi*-specific viruses (EhV, Coccolithoviridae) are closely involved in the control of their host's populations, a phenomenon better appreciated during the sudden crashes of vast *E. huxleyi* coastal and mid-oceanic blooms (Bratbak et al., 1993; Jacquet et al., 2002; Schroeder et al., 2003; Wilson et al., 2002).

This long-established host–virus interaction (Coolen, 2011) will have driven the genomic evolution of both virus and host systems, leading to the development of infection/resistance strategies that are now fundamental to their ecology. An evolutionary consequence of the close intracellular interaction between the *E. huxleyi* and EhV systems is the high level of promiscuity between the two genomes that has enabled a series of horizontal gene transfer (HGT) events (Read et al., 2013). Some of these genes have potential implications for the infection strategy of these viruses and/or relate to their host's defence system (Monier et al., 2009; Pagarete et al., 2009; Vardi et al., 2012; Vardi et al., 2009). At an

* Corresponding author.
E-mail address: antonio.pagarete@bio.uib.no (A. Pagarete).

http://dx.doi.org/10.1016/j.virol.2014.05.020
0042-6822/

ecological level we observe how the selection pressure imposed by these viruses is potentially linked to profound somatic consequences in *E. huxleyi*'s life cycle, namely, the alternation between diploid and haploid phases as a key mechanism to evade infection (Frada et al., 2008).

Host–virus interaction analyses have commonly reported established phenomena where there is a significant decrease in EhV major capsid protein (MCP) diversity during the progression of bloom events (Martínez Martínez et al., 2007; Schroeder et al., 2003). Put simply, from an initial high diversity, a few dominant ecotypes eventually dominate as the bloom develops. The selection pressure acting upon the host–virus pairs is not trivial, especially when considering the ecological dynamics and consequences. For instance, if relative EhV MCP diversity is significantly reduced during a single bloom event, how can diversity be maintained between blooms? Despite the omnipresence of *E. huxleyi* cells in marine samples, the truth is that for the majority of cases (meaning non-blooming situations) these cells exist in low concentrations. For example, in the Norwegian coastal area studied here *E. huxleyi* concentrations are below 30 cells ml^{-1} for at least 6 months every year (unpublished data). Consequently, the probability of an EhV virion finding a suitable *E. huxleyi* host outside the bloom windows is significantly decreased. This should in theory favour geographical isolation of EhV populations. Indeed, comparison of two EhVs isolated from the west Norwegian coast with thirteen isolates from the English Channel (E.C., ~1000 km distance) has revealed the presence and absence of many genes that were found only in one of these locations (Pagarete et al., 2012). Conversely, cross-infectivity experiments with EhV isolates from the North Sea and a collection of geographically diversified *E. huxleyi* strains has shown no increased capacity of EhV strains to successfully infect either closely or distantly isolated host strains (Allen et al., 2007; Pagarete, 2010). Hence, the question on the capacity and relevance of dispersal and gene exchange of these oceanic viruses remains unanswered.

A host–virus system typically evolves under the guise of an arms race between two distinct genomes, the one of the host and the other of the virus (Stern and Sorek, 2011). Yet, in that arms race some genes, intra-gene regions, or even genomic regions will face different selective pressures. In some cases peptide structure and function, along with molecule-to-molecule interaction constraints, can lead to a tendency for conservation of a particular amino-acid sequence. In other cases though, e.g., peptides related to host–virus recognition mechanisms or host immune system proteins, amino-acid variability and thus increased diversity may prove to be more advantageous (Villarreal, 2005). To date we have a poor understanding of how selection pressure influences giant EhV genomes, and consequently, the amino-acid composition of its associated proteins. It is currently unknown if selection is being homogeneously exerted on the whole of the EhV genome, or if in turn distinct conservation rates can be found for specific EhV genes or genomic regions. A recent study on Mimivirus, another giant virus, clearly showed a tendency for those viruses to endure significant and rapid genome reductions (through gene loss) after only 150 infection rounds (Boyer et al., 2011).

With these questions and issues in mind we used a new approach to study EhV metagenomic diversity within natural populations. Traditionally, studies of viral genomic diversity use sequence data from available isolates, but the large size of EhV genomes makes it virtually impossible to isolate and sequence enough viral strains to comprehensively represent the EhV genetic diversity naturally existing. Hence we employed a new combination of DNA separation methodologies (based on single band sequencing from either pulsed field gel electrophoresis (PFGE) or CsCl gradient) with next generation sequencing (454 or Illumina technologies) to study genomic variability within a natural EhV community during an *E. huxleyi* bloom. Here we present two EhV metagenomes from the early and late stages of an induced *E. huxleyi* bloom with the aim of answering 3 specific questions: (1) what is the genomic resemblance of natural EhV populations to currently isolated EhV strains, (2) how is diversity and conservation distributed along the EhV metagenome, and (3) what is the progression of EhV metagenomic diversity during a bloom?

Building upon the answers to these questions, we then focus our analysis on two particular genes (ehv060 and ehv452) to demonstrate the potential, but also the intricacies, of the metagenomic approach presented for the identification of selective constraints and infection mechanisms acting in this virus–host system. The ehv060 gene encodes two domains with putative glycan binding function: a carbohydrate binding module (CBM) specific for sialic-acid residues in host glycans and a C-type lectin-like domain. Both domains can be involved in glycan interactions, mediating viral attachment (Bowden et al., 2011; Jolly and Sattentau, 2013). Notably the ehv060 protein is present in the EhV virion (Allen et al., 2008). The ehv452 gene encodes a high mobility group (HMG) protein. HMG proteins are involved with chromatin structure, usually endowing the chromosome with nuclease sensitivity, and they also recruit transcription factors to bind to enhancers (Štros, 2010). The unexpected high levels of amino-acid diversity registered for these proteins justified their analysis in this manner.

Results

General bloom/infection dynamics

Initial *E. huxleyi* abundance at the start of the experiment (Day 0) was approximately 2.1×10^2 cells ml^{-1}. Coccolithophore concentrations inside the mesocosm enclosure started increasing exponentially from day 6, reaching a maximum number of 6.1×10^4 cells ml^{-1} on day 15, followed by sharp decline (Fig. S1). The decline in *E. huxleyi* numbers coincided with the appearance and exponential increase of coccolithoviruses from day 11 onwards. A maximum concentration of 2.8×10^7 coccolithoviruses ml^{-1} was recorded on day 15. When samples were collected for metagenomic analysis, on days 11 and 15, EhV concentrations were 8.1×10^5 and 2.8×10^7 coccolithoviruses ml^{-1}, respectively. For an in-depth description and discussion of community dynamics during the mesocosm experiment refer to these references (Kimmance et al., 2014; Pagarete et al., 2009, 2011; Vardi et al., 2012).

Characteristics of the two metagenomes

Sample S11 was sequenced using 454 technology, generating 166,940 reads of 256 bp (on average) equivalent to nearly 0.043 Gb of sequence. Sample S15 was sequenced using Illumina technology, generating 11,576,462 paired-end reads of 51 bp (on average) equivalent to nearly 1.2 Gb of sequence. In both metagenomic datasets a significant percentage of the reads was identified as EhV sequences (approx. 18% and 68% for S11 and S15, respectively). Average level of sequence depth differed between the two metagenomes by an order of magnitude. Consequently, the conservative analysis of gene identification was carried out independently for each metagenome, with different thresholds of minimum average read depth and minimum DB coverage adopted for each metagenome (Table 1). In both metagenomes the identified proteins were homogeneously scattered around the EhV genome, with no obvious sign of sequencing bias toward specific genomic regions (Fig. 1). The Simpson index was chosen as an indicator of amino-acid diversity, after tests performed on a battery of diversity indexes, due to its status as the index least affected by

Please cite this article as: Pagarete, A., et al., Dip in the gene pool: Metagenomic survey of natural coccolithovirus communities. Virology (2014), http://dx.doi.org/10.1016/j.virol.2014.05.020