OCEAN PLANKTON

Patterns and ecological drivers of ocean viral communities

Jennifer R. Brum,[1*†] J. Cesar Ignacio-Espinoza,[2*†] Simon Roux,[1*†] Guilhem Doulcier,[1,3] Silvia G. Acinas,[4] Adriana Alberti,[5] Samuel Chaffron,[6,7,8] Corinne Cruaud,[5] Colomban de Vargas,[9,10] Josep M. Gasol,[4] Gabriel Gorsky,[11,12] Ann C. Gregory,[13†] Lionel Guidi,[11,12] Pascal Hingamp,[14] Daniele Iudicone,[15] Fabrice Not,[9,10] Hiroyuki Ogata,[16] Stéphane Pesant,[17,18] Bonnie T. Poulos,[1] Sarah M. Schwenck,[1] Sabrina Speich,[19‡] Celine Dimier,[9,10,20] Stefanie Kandels-Lewis,[21,22] Marc Picheral,[11,12] Sarah Searson,[11,12] *Tara* Oceans Coordinators,[§] Peer Bork,[21,23] Chris Bowler,[20] Shinichi Sunagawa,[21] Patrick Wincker,[5,24,25] Eric Karsenti,[20,22‖] Matthew B. Sullivan[1,2,13†‖]

Viruses influence ecosystems by modulating microbial population size, diversity, metabolic outputs, and gene flow. Here, we use quantitative double-stranded DNA (dsDNA) viral-fraction metagenomes (viromes) and whole viral community morphological data sets from 43 *Tara* Oceans expedition samples to assess viral community patterns and structure in the upper ocean. Protein cluster cataloging defined pelagic upper-ocean viral community pan and core gene sets and suggested that this sequence space is well-sampled. Analyses of viral protein clusters, populations, and morphology revealed biogeographic patterns whereby viral communities were passively transported on oceanic currents and locally structured by environmental conditions that affect host community structure. Together, these investigations establish a global ocean dsDNA viromic data set with analyses supporting the seed-bank hypothesis to explain how oceanic viral communities maintain high local diversity.

Ocean microbes produce half of the oxygen we breathe (*1*) and drive much of the substrate and redox transformations that fuel Earth's ecosystems (*2*). However, they do so in a constantly evolving network of chemical, physical, and biotic constraints—interactions that are only beginning to be explored. Marine viruses are presumably key players in these interactions (*3*, *4*), as they affect microbial populations through lysis, reprogramming of host metabolism, and horizontal gene transfer. Here, we strive to develop an overview of ocean viral community patterns and ecological drivers.

The *Tara* Oceans expedition provided a platform for sampling ocean biota from viruses to fish larvae within a comprehensive environmental context (*5*). Prior virus-focused work from this expedition has helped optimize the double-stranded DNA (dsDNA) viromic sample-to-sequence workflow (*6*), evaluate ecological drivers of viral community structure as inferred from morphology (*7*), and map ecological patterns in the large dsDNA nucleo-cytoplasmic viruses using marker genes (*8*). Here, we explore global patterns and structure of ocean viral communities using 43 samples from 26 stations in the *Tara* Oceans expedition (see supplementary file S1) to establish dsDNA viromes from viral-fraction (<0.22 μm) concentrates and quantitative whole viral community morphological data sets from unfiltered seawater. Viruses lack shared genes that can be used for investigation of community patterns. Therefore, we used three levels of information to study such patterns: (i) protein clusters (PCs) (*9*) as a means to organize virome sequence space commonly dominated by unknown sequences (63 to 93%) (*10*), (ii) populations, using established metrics for viral contig recruitment (*11*), and (iii) morphology, using quantitative transmission electron microscopy (qTEM) (*7*).

The *Tara* Oceans Viromes (TOV) data set

The 43 *Tara* Oceans Viromes (TOV) data set comprises 2.16 billion ~101-base pair (bp) paired-end Illumina reads (file S1), which largely represent epipelagic ocean viral communities from the surface (ENVO:00002042) and deep chlorophyll maximum (DCM; ENVO:01000326) throughout seven oceans and seas; only 1 of 43 viromes is from mesopelagic waters, Environment Ontology feature ENVO:00000213 (file S1). The TOV data set offers deeper sampling of surface ocean viral communities but underrepresents the deep ocean relative to the Pacific Ocean Viromes data set (POV) (*10*), which includes 16 viromes from aphotic zone waters. In all viromes, sampling and processing affects which viruses are represented (*6*, *12*–*14*). We filtered TOV seawater samples through 0.22-μm-pore–sized filters and then concentrated viruses in the filtrate using iron chloride flocculation (*15*). These steps would have removed most cells but also would have excluded any viruses larger than 0.22 μm. We then purified the resulting TOV viral concentrates using deoxyribonuclease (DNase) treatment, which is as effective as density gradients for purifying ocean viral concentrates (*14*). This DNAse-only step is unlikely to affect viral representation in the viromes but reduces nonviral DNA contamination. Finally, we extracted DNA from the samples and prepared sequence libraries using linker amplification (*13*). These steps preserve quantitative representation of dsDNA viruses in the resulting viromes (*12*, *13*), but the ligation step excludes RNA viruses and is biased against single-stranded DNA (ssDNA) viruses (*12*).

We additionally applied quantitative transmission electron microscopy (qTEM) (*7*) to paired whole seawater samples to evaluate patterns in whole viral communities. This method simultaneously considers ssDNA, dsDNA, and RNA viruses, although without knowledge of their relative abundances because particle morphology does not identify nucleic acid type. In the oceans, total virus abundance estimates based on TEM analyses, which include all viral particles, are similar to estimates based on fluorescent staining, which inefficiently stains ssDNA and RNA viruses (*16*–*24*). This suggests that most

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. [2]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA. [3]Environmental and Evolutionary Genomics Section, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS, UMR8197, INSERM U1024, 75230 Paris, France. [4]Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM)-CSIC, Pg. Marítim de la Barceloneta 37-49, Barcelona, E08003, Spain. [5]Genoscope, Commissariat à l'Energie Atomique (CEA)—Institut de Génomique, 2 rue Gaston Crémieux, 91057 Evry, France. [6]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [7]Center for the Biology of Disease, VIB KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [8]Department of Applied Biological Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [9]CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [10]Sorbonne Universités, Université Pierre et Marie Curie, Université Paris 06, and UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [11]CNRS, UMR 7093, Laboratoire d'océanographie de Villefranche (LOV), Observatoire Océanologique, 06230 Villefranche-sur-mer, France. [12]Sorbonne Universités, Uiversité Pierre et Marie Curie, Université Paris 06, UMR 7093, Laboratoire d'océanographie de Villefranche (LOV), Observatoire Océanologique, 06230 Villefranche-sur-mer, France. [13]Soil, Water, and Environmental Science, University of Arizona, Tucson, AZ 85721, USA. [14]Aix Marseille Université, CNRS IGS UMR 7256, 13288 Marseille, France. [15]Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy. [16]Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0001, Japan. [17]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, 28359 Bremen, Germany. [18]MARUM, Center for Marine Environmental Sciences, University of Bremen, 28359 Bremen, Germany. [19]Laboratoire de Physique des Océans, Institut Universitaire Européen de la Mer, Université de Bretagne Occidentale (UBO-IUEM), Place Copernic, 29820 Plouzané, France. [20]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), and INSERM U1024, and CNRS UMR 8197, Paris, 75005, France. [21]Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [22]Directors' Research, European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany. [23]Max-Delbrück-Centre for Molecular Medicine, 13092 Berlin, Germany. [24]CNRS, UMR 8030, CP5706, 91057 Evry, France. [25]Université d'Evry, UMR 8030, CP5706, 91057 Evry, France. *These authors contributed equally to this work. †Present address: Department of Microbiology, Ohio State University, Columbus, OH 43210, USA. ‡Present address: Department of Geosciences, Laboratoire de Météorologie Dynamique (LMD), Ecole Normale Supérieure, 24 rue Lhomond 75231 Paris, Cedex 05, France. §*Tara* Oceans coordinators and affiliations are listed after the Acknowledgments. ‖Corresponding authors. E-mail: mbsulli@gmail.com (M.B.S.); karsenti@embl.de (E.K.)