Jan 09, 2020

Viral to metazoan marine plankton nucleotide sequences from the *Tara* Oceans expedition

Adriana Alberti[1], Julie Poulain[1], Stefan Engelen[1], Karine Labadie[1], Sarah Romac[2,3], Isabel Ferrera[4], Guillaume Albini[1], Jean-Marc Aury[1], Caroline Belser[1], Alexis Bertrand[1], Corinne Cruaud[1], Corinne Da Silva[1], Carole Dossat[1], Frédéric Gavory[1], Shahinaz Gas[1], Julie Guy[1], Maud Haquelle[1], E'krame Jacoby[1], Olivier Jaillon[1,5,6], Arnaud Lemainque[1], Eric Pelletier[1], Gaëlle Samson[1], Marc Wessner[1], Genoscope Technical Team[1], Silvia G. Acinas[4], Marta Royo-Llonch[4], Francisco M. Cornejo-Castillo[4], Ramiro Logares[4], Beatriz Fernández-Gómez[4,7,8], Chris Bowler[9], Guy Cochrane[10], Clara Amid[10], Petra Ten Hoopen[10], Colomban De Vargas[2,3], Nigel Grimsley[11,12], Elodie Desgranges[11,12], Stefanie Kandels-Lewis[13,14], Hiroyuki Ogata[15], Nicole Poulton[16], Michael E. Sieracki[16,17], Ramunas Stepanauskas[16], Matthew B. Sullivan[18,19], Jennifer R. Brum[19,20], Melissa B. Duhaime[21], Bonnie T. Poulos[22], Bonnie L. Hurwitz[23], Stéphane Pesant[24,25], Eric Karsenti[9,13,26], Patrick Wincker[1,5,6]

[1]CEA - Institut de Biologie François Jacob, Genoscope, Evry, France, [2]CNRS, UMR 7144, Station Biologique de Roscoff, France, [3]Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Station Biologique de Roscoff, France, [4]Departament de Biologia Marina i Oceanografia, Institute of Marine Sciences (ICM), CSIC, Barcelona, Spain, [5]CNRS, UMR 8030, Evry , France, [6]Université d'Evry, UMR 8030, Evry, France, [7]FONDAP Center for Genome Regulation, Santiago, Chile, [8]Laboratorio de Bioinformática y Expresión Génica, Instituto de Nutrición y Tecnología de los Alimentos (INTA), Universidad de Chile, El Libano Macul, Santiago, Chile, [9]Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR 8197, INSERM U1024, Paris, France, [10]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genomes Campus, Hinxton, Cambridge , UK, [11]CNRS UMR 7232, BIOM, Banyuls-sur-Mer, France, [12]Sorbonne Universités Paris 06, OOB UPMC, Banyuls-sur-Mer , France, [13]Directors' Research European Molecular Biology Laboratory, Heidelberg, Germany, [14]Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany, [15]Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, Japan, [16]Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA, [17]National Science Foundation, Arlington, Virginia, USA, [18]Departments of Microbiology and Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, Ohio, USA, [19]Department of Microbiology, The Ohio State University, Columbus, Ohio, USA, [20]Present address: Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana, USA, [21]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA, [22]University of Arizona, Tucson, Arizona, USA, [23]Department of Agricultural and Biosystems Engineering, University of Arizona, Tucson, Arizona, USA, [24]MARUM, Center for Marine Environmental Sciences, University of Bremen, Germany, [25]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Germany, [26]Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire d'oceanographie de Villefranche (LOV), Observatoire Océanologique, Villefranche-sur-mer, France

1 *Works for me* dx.doi.org/10.17504/protocols.io.qv6dw9e

Adriana Alberti
CEA, Genoscope, France

ABSTRACT

The protocols in this collection are from the Alberti A., et al manuscript (Alberti A. 2017, Scientific Data).

These protocols provide detailed procedures applied for genomic data generation, from nucleic acids extraction to sequence production, and we describe registries of genomics datasets available at the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena). This collection complements other efforts to provide a full description of experiments and open science resources generated from the *Tara* Oceans project, further extending their value for the study of the world's planktonic ecosystems.'

From the Methods section:

'The generation of information-rich data from marine plankton samples presents unique challenges that are inherent to the particular sampling conditions at sea and the wide spectrum of organisms included in that environment. All processing steps, including biomass collection, sample preservation, nucleic acids extractions and sequencing library preparation, are critical and require specific protocols and robust methods in order to ensure comparability of results and limit potential biases.

Our methods were either developed specifically for *Tara* Oceans samples or carefully selected among existing ones in order to meet the requirements of our sequencing strategy and to produce optimized datasets for downstream bioinformatics analyses, as for example the production of overlapping reads from metagenomics libraries to facilitate assembly. They are presented in five sub-sections, starting with a brief description of how samples were handled between the research vessel and the processing laboratories (protocol 1). Protocol

2 reports on DNA and RNA extractions procedures for -omics analyses, including the generation of amplified genomic DNA from uncultured isolated unicellular eukaryotes. The generation of 18S and 16S rRNA amplicons from DNA of specific size-fractions is described in protocol 3 and Illumina libraries preparation in protocol 4. Sequencing procedures and post-sequencing data processing are described in protocol 5. For details on the onboard sampling protocols, see Pesant *et al*. '

EXTERNAL LINK

https://www.nature.com/articles/sdata201793

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Alberti, A. (2017). Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific Data***4**, 170093 (2017)
doi: 10.1038/sdata.2017.93

Files

- Handling of genomics samples
  by Adriana Alberti,
  CEA, Genoscope, France
- Nucleic acids preparations
  by Adriana Alberti,
  CEA, Genoscope, France
- DNA and RNA backups
  by Adriana Alberti,
  CEA, Genoscope, France
- 18S and 16S rRNA genes amplicon generation for eukaryotic and prokaryotic metabarcoding
  by Adriana Alberti,
  CEA, Genoscope, France
- Sequencing library preparation
  by Adriana Alberti,
  CEA, Genoscope, France
- Sequencing and data quality control
  by Adriana Alberti,
  CEA, Genoscope, France