Jan 09, 2020

Sequencing library preparation

In 1 collection

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.qwidxce

Adriana Alberti
CEA, Genoscope, France

ABSTRACT

This protocol describes the sequencing library preparation for the *Tara* Oceans expedition and is part of Viral to metazoan marine plankton nucleotide sequences from the *Tara* Oceans expedition.

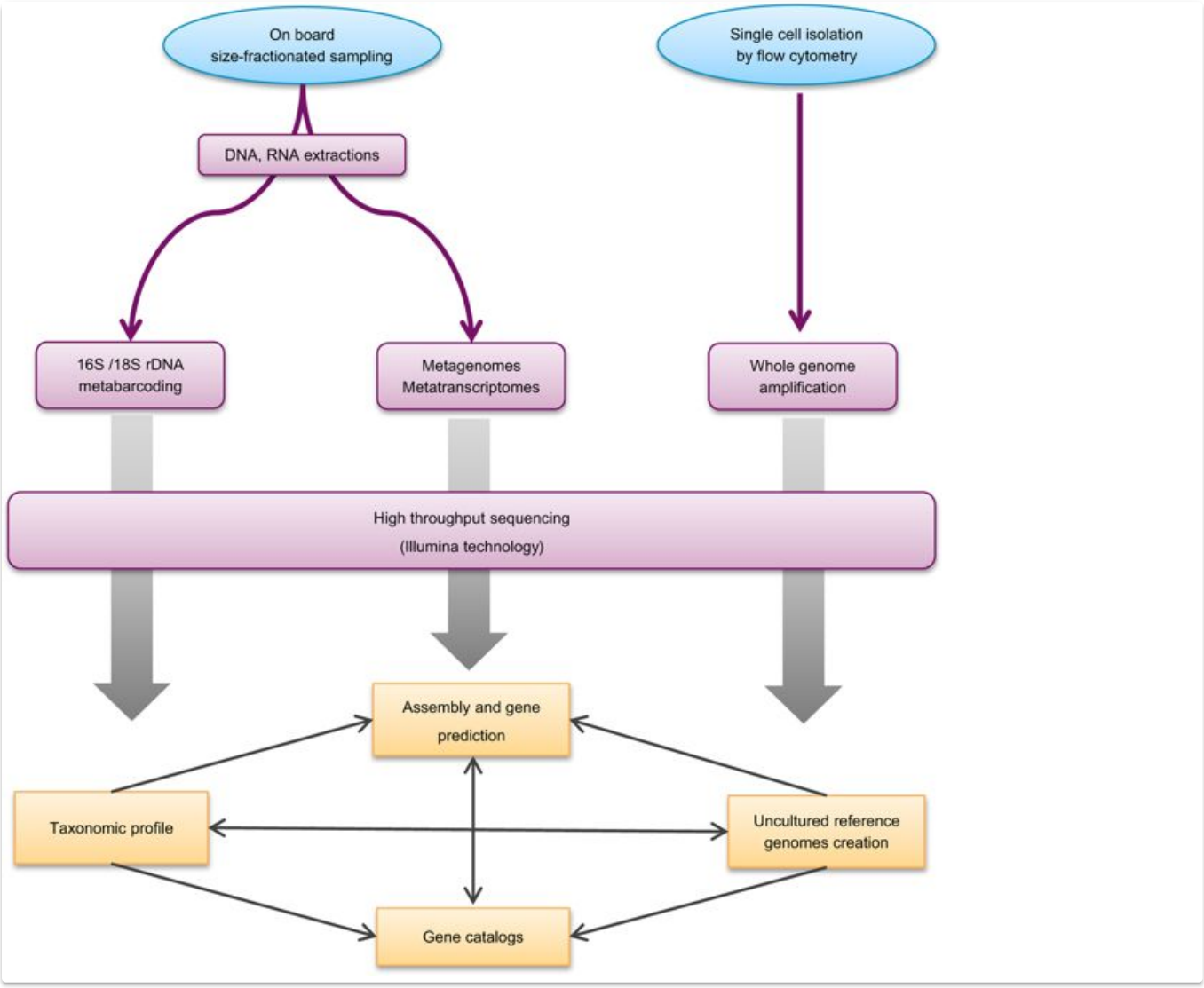

Figure 1: Overview of -omics analysis strategy applied on Tara Oceans samples.

EXTERNAL LINK

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

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

ATTACHMENTS

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

GUIDELINES

All library preparations were performed at Genoscope.

1. Metagenomic library preparation from size fractionated filters DNA (Method ID: MetaG)

When the *Tara* Oceans project started in 2009, Illumina offered a high throughput system that could enable to gain biological insights for complex samples. The counterpart was to obtain maximum read lengths of 100 bp. As short read lengths may be challenging for *de novo* assemblies, library preparation protocols for complex metagenomics samples were improved in order to generate much longer

reads by overlapping and merging read pairs before assembly. For this purpose, a size selection step was added at the end of library preparation obtaining narrowly sized libraries around 300 bp. This corresponded to an insert fragment at around 180 bp, allowing ~20 bp paired read overlaps.

Depending on extracted DNA yields, libraries were prepared manually or in a semi-automatic manner. Genomic DNA was first sheared to a mean target size of 300 bp using a Covaris E210 instrument (Covaris, Woburn, MA). DNA inputs in fragmentation step were 30–100 ng in the case of a downstream manual preparation, or 250 ng for semi-automatized protocol, more demanding in DNA quantity. Size profiles of sheared materials were visualized on an Agilent Bioanalyzer DNA High Sensitivity chip.

In the manual protocol, the resulting fragmented DNA was end-repaired, A-tailed at the 3'end, and ligated to Illumina compatible adapters using the NEBNext DNA Sample Prep Master Mix Set 1 (New England Biolabs). Ligation products were subsequently cleaned up using 1x AMPure XP beads.

For >250 ng input gDNA, end repair, A-tailing, adaptors ligation and a 200–400 bp size selection were performed using the SPRIWorks Library Preparation System and SPRI TE instrument (Beckmann Coulter Genomics), according to the manufacturer protocol. This allowed to process rapidly and with few hands-on time, up to 10 samples in parallel.

Ligation products were then enriched by performing 12 cycles of amplification (98 °C for 30 s, 12 cycles of 10 s at 98 °C, 30 s at 60 °C, 30 s at 72 °C and 72 °C for 5 min) using Platinum Pfx Taq Polymerase (Thermo Fisher Scientific) and P5 and P7 primers. Amplified products were purified using AMPure XP beads (1 volume) and samples were run on a 3% agarose gel in order to size-select gel slices around 300 bp. The excised band (280–310 bp) was finally purified using the Nucleospin Extract II DNA purification kit (Macherey-Nagel).

Later on, further optimizations to the original manual protocol were applied to the processing of samples collected during the Tara Oceans Polar Circle campaign (stations 155–210). In particular, after gDNA shearing, libraries were prepared using the NEBNext DNA Sample Prep Master Mix kit with a 'on beads' protocol that achieves higher library yields. Performing each reaction step on the same AMPure XP beads used for first purification after end repair minimizes sample losses during the successive clean up steps. Ligation was performed with adapted concentrations of Nextflex DNA barcodes (Bioo Scientific, Austin, TX,) and cleaned up by two rounds of AMPure XP beads purifications.

For higher sample inputs (250 ng), library preparation benefitted of high throughput automatized instruments. End-repair, A-tailing and ligation were made by a liquid handler, the Biomek FX Laboratory Automation Workstation (Beckmann Coulter Genomics), able to perform up to 96 reactions in parallel in half a day. Library amplification was performed using Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA) (98 °C for 45 s, 12 cycles of 15 s at 98 °C, 30 s at 60 °C, 30 s at 72 °C and 72 °C for 1 min) instead of Platinum Pfx Taq Polymerase. Amplified library was purified and size-selected as described above.

2. Library preparation from viral samples (Method ID: MetaG_virus)

Due to very low DNA extraction yields obtained from concentrated viral samples (usually only a few nanograms), library preparation protocol was adapted in order to improve its efficiency starting from very low input DNA. Following an extensive study of the impact of DNA amount, amplification and library preparation protocol, the method developed at Genoscope and described in detail at https://www.protocols.io/groups/sullivan-lab was chosen for preparation of all viral metagenomics libraries from *Tara* Oceans stations.

Briefly, 10–15 ng DNA were fragmented to a 150–600 bp size range using the E210 Covaris instrument. End repair, A-tailing and ligation with adjusted concentrations of homemade adaptors were performed using the NEBNext DNA Sample Preparation Reagent Set 1 (New England Biolabs). After two consecutive 1x AMPure XP clean ups, the ligated product was amplified by 12 cycles PCR using Platinium Pfx DNA polymerase followed by 0.6x AMPure XP purification.

For samples collected during the *Tara* Oceans Polar Circle campaign, similar optimizations were applied as described for metagenomics libraries. Manual 'on beads' protocol was used on lower inputs (1.3–20 ng). The ligated product was amplified by 12 to 18 cycles PCR using Kapa Hifi HotStart NGS library Amplification kit and purified by 0.6x AMPure XP clean up.

3. Library preparation from SAGs (Method ID: MetaG_SAGs)

A fixed volume (7.5 µl of single cell-amplified DNA) was used as input for DNA shearing. Then, the same library preparation protocol used for viral libraries was applied without significant modifications.

4. Metatranscriptomic libraries

Different cDNA synthesis protocols were applied according to the fractions from which RNA originated. A first problem to be solved was to limit the generation of rRNA reads coming from this predominant RNA fraction. In the case of RNA issued from fractions enriched in

protists and metazoans (0.8–5 µm (or 0.8–3 µm), 0.8–2,000 µm, 3–2,000 µm, 5–20 µm (or 3–20 µm), 20–180 µm and 180–2,000 µm membrane filters), methods including a poly(A)$^+$ RNA selection step were chosen. Whereas this approach is very efficient in lowering the number of rRNA reads, it does not allow to retrotranscribe mRNAs from prokaryotic species, thus leading to eukaryote-only metatranscriptomes.

In contrast, cDNA synthesis from prokaryote- and virus-enriched fractions RNAs (0.2–1.6 µm and 0.2–3 µm) was performed by a random priming approach, preceded by a prokaryotic rRNA depletion step. This method allows cDNA synthesis from both eukaryotic and prokaryotic mRNA and organellar transcripts but also from residual, poorly-depleted eukaryotic rRNA resulting in high percentage of rRNA reads when small protists are abundant.

cDNA synthesis and library preparation from eukaryote-enriched fractions

The quantity of extracted total RNA was an additional factor, which conditioned the choice of cDNA synthesis method. When at least 2 µg total RNA were available, cDNA synthesis was carried out using the TruSeq mRNA Sample preparation kit (Illumina, San Diego, CA) (Method ID: TS_RNA). Briefly, poly(A)$^+$ RNA was selected with oligo(dT) beads, chemically fragmented and converted into single-stranded cDNA using random hexamer priming. Then, the second strand was generated to create double-stranded cDNA. Next, library preparation was performed according to the protocol described for viral metagenomics libraries by omitting cDNA shearing and performing a post-PCR 1x AMPure XP purification.

In 2012, Illumina released a new version of the kit, the TruSeq Stranded mRNA kit, which allows retaining strand information of RNA transcripts (sequence reads occur in the same orientation as anti-sense RNA). Strand specificity is achieved by quenching the second strand during final amplification thanks to incorporation of dUTP instead of dTTP during second strand synthesis. As strand orientation provides additional valuable information for downstream RNAseq data analysis, this method (Method ID: TS_strand) was applied for processing RNA from samples collected during the Polar Circle campaign. The minimal RNA input used for this library was 1 µg total RNA. After second strand synthesis, ready-to-sequence Illumina library was generated following the manufacturer's instructions using the reagents included in the kit.

RNA extractions yielding insufficient quantities for TruSeq preparations were processed using the SMARTer Ultra Low RNA Kit (Clontech, Mountain View, CA) (Method ID: SMART_dT). This method, successfully used for eukaryotic single cell transcriptomic studies, converts poly(A)$^+$ RNA to full-length cDNA using a modified oligo(dT) primer combined with SMART (*S*witching *M*echanism *a*t the 5' end of *R*NA *T*emplate) technology. Fifty nanograms or less total RNA were used for cDNA synthesis, followed by 12 cycles of PCR preamplification of cDNA. Before Illumina library preparation, 5–50 ng double stranded cDNA were fragmented to a 150–600 bp size range using the E210 Covaris instrument. Then, sheared cDNA were used for Illumina library preparation following the protocol described for viral metagenomes libraries, except for the post amplification AMPure XP purification performed at a ratio 1:1.

cDNA synthesis and library preparation from prokaryote- and virus-enriched fractions

As for eukaryotic RNA, the extraction yields from 0.2–1.6 µm and 0.2–3 µm filters were a concern and motivated a preliminary study of different 'low input' cDNA synthesis methods adapted to prokaryotic mRNA. On the basis of the results presented in this paper, we chose to perform bacterial rRNA depletion followed by cDNA synthesis with SMARTer Stranded RNA-Seq Kit (Clontech). This method is a more recent release from Clontech than the SMARTer Low Input library kit. Differently from this oligo(dT)-based method, the SMARTer Stranded kit is based on initial chemical RNA fragmentation followed by a first cDNA strand synthesis by random priming and SMART template switching technology. Then, single-stranded cDNA is directly amplified with oligonucleotides which contain Illumina adaptors and indexes sequences to obtain a ready-to-sequence library. Finally, differently from oligo(dT) method, this one preserves the coding strand information which can be deduced after paired end sequencing of library fragments.

Bacterial rRNA depletion was carried out using Ribo-Zero Magnetic Kit for Bacteria (Epicentre Biotechnologies). Different total RNA inputs were depleted, varying between undetectable quantities by Qubit measurement up to 4 µg. Therefore, Ribo-Zero depletion protocol was modified to be adapted to low RNA input amounts according to Alberti*et al.*. Except for these modifications, depletion was performed according to the manufacturer instructions. Depleted RNA were concentrated to 10 µl total volume with RNA Clean and Concentrator-5 kit (ZymoResearch) following the procedure described for retention of >17 nt RNA fragments. Then, when total RNA input was > or equal to 1 µg, depleted RNA amount was checked by Qubit RNA HS Assay quantification and 40 ng, or less, were used to synthesize cDNA with SMARTer Stranded RNA-Seq Kit (Method ID: RiboZero_SMART_Strand). Otherwise, 7 µl were used for cDNA synthesis. Single stranded cDNA was purified by two rounds of purification with 1x AMPure XP beads. The purified product was amplified by 18 cycles PCR with SeqAmp DNA polymerase and the Illumina Index Primer set, both provided in the kit. Final library was purified with 1x AMPure XP beads.

5. Library preparation from V9-18S rRNA amplicons (Method ID: MetaBar_18S)

In order to evaluate the eukaryotic biodiversity of samples, libraries were prepared from amplicons generated by the amplification of the V9 region of the 18S rRNA gene. As the amplicon size, visualized on an Agilent Bioanalyzer, was around 160 bp (majority peak), no fragmentation was needed before library preparation. Amplicons (100 ng) generated from *Tara* Oceans samples were end-repaired, A-tailed and ligated with Illumina adaptors using the SPRIWorks Library Preparation System and SPRI TE instrument, without any size selection. Ligated products were amplified using Platinum *Pfx* Taq Polymerase and cleaned up on magnetic beads as described above for metagenomic libraries except that gel size selection was skipped.

Tara Oceans Polar Circle amplicons were treated as described in metagenomic libraries section for samples issued from the same campaign.

6. Library preparation from V4-V516S rRNA amplicons (Method ID: MetaBar_16S)

Tags generated from amplification of V4 and V5 hypervariable regions of 16S rRNA genes were used for preparation of sequencing libraries by high throughput automatized instruments. One hundred ng amplicons were directly end-repaired, A-tailed and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation. Then, library amplification was performed using Kapa Hifi HotStart NGS library Amplification kit with the same cycling conditions applied for metagenomics libraries. After AMPure XP purification (1 volume) and quantification by Qubit fluorometric measurement (HS assay), equimolar pools of amplified products were run on a 2% agarose gel to select 500–650 bp gel slices (amplicon size increased by Illumina adapters). This sizing step allowed isolating the prokaryotic 16S amplicon from non-specific amplification products. The library was finally purified using the Nucleospin Extract II DNA purification kit.

SAFETY WARNINGS

Please refer to the SDS (Safety Data Sheet) for warnings and hazard information.