Chapter 13

Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Dual-PCR Method

Sarah J. Bourlat, Quiterie Haenel, Jennie Finnman, and Matthieu Leray

Abstract

This protocol details the preparation of multiplexed amplicon libraries for metabarcoding (amplicon-based) studies of microscopic marine eukaryotes. Metabarcoding studies, based on the amplification of a taxonomically informative marker from a collection of organisms or an environmental sample, can be performed to analyze biodiversity patterns or predator–prey interactions. For Metazoa, we use the mitochondrial cytochrome oxidase 1 (CO1) or the small ribosomal subunit (SSU) markers. Here, we describe a strategy for the preparation of multiplexed Illumina MiSeq libraries using a dual-PCR approach for the addition of index and adaptor sequences.

Key words Metabarcoding, Biodiversity, Microscopic eukaryotes, Illumina MiSeq

1 Introduction

The development of high-throughput sequencing technologies offers the possibility to recover DNA information from whole community samples, a technique routinely used for prokaryotes, based on the 16S ribosomal RNA gene [1]. Advances in DNA sequencing and analytical techniques now allow biodiversity assessments from eukaryotes and from various types of environmental samples. Recent studies have shown it possible to study bulk environmental samples dominated by microscopic eukaryotes, such as marine sediments [2, 3], analyze the contents of fish stomachs [4], carry out environmental status assessments for benthic macroinvertebrates [5], or reveal patterns of marine benthic diversity on autonomous reef monitoring structures [6]. The vast majority of metabarcoding studies have so far employed Roche 454 for sequencing due to its long read lengths compared to other technologies [7], but Illumina MiSeq technology is now getting more

Sarah J. Bourlat (ed.), *Marine Genomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1452,
DOI 10.1007/978-1-4939-3774-5_13,

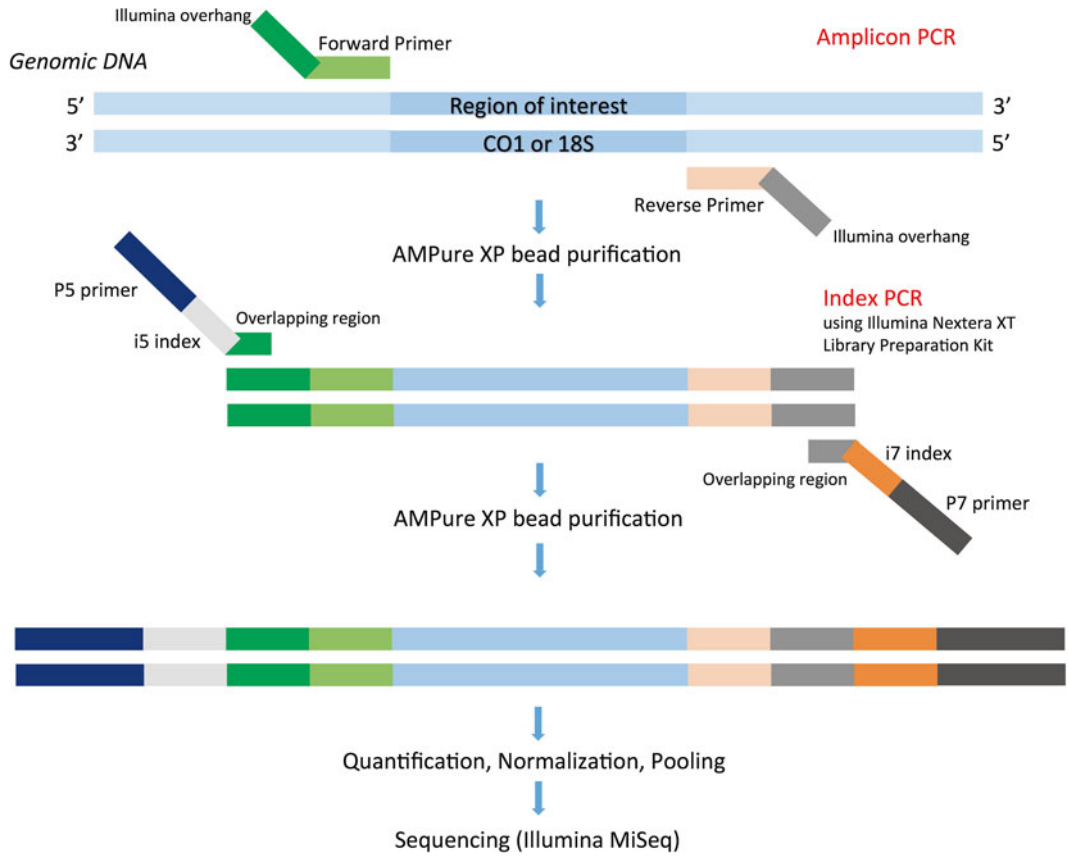

Fig. 1 Schema for Illumina MiSeq library preparation using dual-PCR. The first PCR step uses amplicon-specific primers including Illumina adapter overhangs (amplicon PCR) and the second cycle limited PCR allows the incorporation of Illumina index adapters i5 and i7 (index PCR). Bead purifications are carried out after each step. Quantification, normalization, and pooling are carried out before sequencing on Illumina MiSeq (Figure adapted from [8])

competitive in terms of read length, reaching up to 2 × 300bp with MiSeq Reagent Kit v3. Marker sequences amplified can have up to 500 bp, including a 100 bp overlap for paired end reads.

This method consists in dual PCR amplification, the first PCR using amplicon-specific primers including an Illumina adapter overhang (referred to here as the amplicon PCR) and a second cycle limited PCR for the incorporation of Illumina index adapters (referred to here as the index PCR) (Fig 1, adapted from [8]).

2 Materials

2.1 DNA Extraction

1. DNA extraction kit. For sediment samples, use the PowerSoil or for larger volumes PowerMax soil DNA isolation kit (MoBio). For tissue samples, use UltraClean tissue and cells DNA isolation (MoBio) or an equivalent kit.

2.2 Amplicon PCR

1. 96-well PCR plates and microseal film.
2. Pfu proofreading polymerase.
3. dNTP mix containing sodium salts of dATP, dCTP, dGTP, and dTTP, each at 10 mM (total concentration 40 mM).
4. Nuclease-free water.
5. Thermocycler.
6. Gel electrophoresis apparatus and reagents.
7. 100 bp DNA ladder.
8. Locus-specific primers with Illumina overhang. Illumina overhang adapter sequences derived from the 16S demonstrated protocol (Illumina) [8] were appended to locus-specific primers for compatibility with Illumina index and sequencing adapters [8]. The universal CO1 primers available for the Metazoa amplify a 658 bp region [9], which is too long for most NGS applications. For the CO1 gene, Illumina overhangs were appended to each of the primer sequences proposed by Leray et al. to amplify a 313 bp fragment, termed the 'mini-barcode' (mlCOIintF-dgHCO2198) [10]. For the reverse CO1 primer, a variation of the primer proposed by Lobo et al. was used, which was shown to enhance amplification of the CO1 region in a wide range of invertebrates [11]. For the SSU region, Illumina overhang adapter sequences were appended to primers modified from Fonseca et al. and yielding a 364 bp fragment (SSU_FO4—SSU_R22) [3]. These primers target a homologous region of the gene and flank a region that is highly divergent, corresponding to the V1–V2 region of the 18S gene. *See* Chapter 12 by Fonseca and Lallias for further details on the SSU_FO4 and SSU_R22 primers.

2.3 Bead Purification

1. Magnetic beads for DNA purification, such as AMPure XP (Agencourt), or equivalent. These are also called SPRI beads for solid-phase reversible immobilization.
2. Magnetic 96-well plate stand.
3. Freshly prepared 80% ethanol.
4. Nuclease-free water.

2.4 Index PCR

1. Nextera XT index Kit, 96 indices, 384 samples for a 96-well plate (Illumina).
2. KAPA HiFi HotStart ready mix or equivalent high-fidelity DNA polymerase mix containing buffer, $MgCl_2$, dNTPs, and polymerase.

2.5 Normalization and Pooling

1. Tapestation (Agilent 2200) with high Sensitivity D1000 ScreenTape.
2. Qubit® 2.0 Fluorometer and Qubit™ dsDNA HS assay kit.
3. Tris-Cl 10 nM buffer, pH 8.5 with 0.1% Tween 20.

2.6 Library Preparation and MiSeq Sequencing

1. MiSeq Instrument (Illumina).
2. MiSeq Reagent Kit v3 (600 cycles).
   - Box 1—MiSeq Reagent Cartridge and Hybridization Buffer (HT1).
   - Box 2—Flow Cell and PR2 Bottle.
3. Heat Block.
4. NaOH stock (1.0 N).
5. Microcentrifuge tubes.
6. PhiX (10 nM).
7. Tween 20.
8. Laboratory grade water.
9. 70% EtOH.
10. Whatman, lens cleaning tissue.

3 Methods

3.1 DNA Extraction

For total DNA extraction from sediment samples, use MoBio's PowerSoil or PowerMax DNA isolation kit, depending on the volume of sediment processed. For DNA extraction from tissue samples, use MoBio's UltraClean tissue and cells DNA isolation kit or an equivalent kit according to the manufacturer's instructions.

3.2 Amplicon PCR

Here, we use locus-specific primers with an Illumina overhang as described in Subheading 2.2. To minimize PCR errors, use Pfu proofreading polymerase. Also, we recommend running three PCR replicates to minimize biases.

1. For a 50 μl reaction volume, use 5 μl Pfu polymerase buffer (10×), 1 μl dNTP mix (final concentration of each dNTP 200 μM), 0.5 μl of each primer at 50 pm/μl, 2 μl DNA template (~10 ng), 0.5 μl Pfu DNA polymerase, and 40.5 μl of nuclease-free water (*see* **Note 1**).
2. Run three PCR replicates (e.g., three independent PCRs for the same sample) using the following cycling conditions: 2 min at 95 °C (1×); 1 min at 95 °C, 45 s at *X*°C, 1 min at 72 °C (25×); 5 min at 72 °C (1×); hold at 4 °C. Replace *X* with the adequate annealing temperature for your particular primers (*see* **Note 2**).
3. Run a 1–2% agarose gel to check the size of the amplicons using a DNA ladder. % Agarose depends on the size of your fragment.
4. If any additional bands appear that are not the size of the desired product, increase the annealing temperature of the PCR or perform additional purification steps (*see* **Note 3**).

5. Pool PCR replicates.

3.3 Bead Purification

Purify amplicons using magnetic beads and a magnetic stand. Size selection can be achieved using different ratios of beads to sample (*see* **Note 3**). A ratio of bead to sample of 0.8:1 will efficiently purify the amplicons away from primers and primer dimers and allow selection of fragments larger than 200 bp [12].

1. Vortex the beads before use. Add 40 μl beads to 50 μl of PCR product to obtain a ratio of 0.8. Pipette up and down ten times. Incubate a room temperature without shaking for 5 min.
2. Place the plate on the magnetic stand until the supernatant has cleared, at least 3 min.
3. Remove the supernatant with a multichannel pipette if you are using a 96-well plate, making sure not to disturb the beads.
4. With the samples on the magnetic stand, wash the beads by adding 200 μl of freshly prepared 80% ethanol and incubate for 30 s. Carefully remove the supernatant without disturbing the beads.
5. Repeat washing **step 4**.
6. Remove all residual ethanol using a pipette and air dry, leaving the samples on the magnetic stand.
7. Remove the plate from the stand and add 40 μl of nuclease-free water for elution, gently pipetting up and down ten times to resuspend the beads. Incubate the plate at room temperature for 5 min.
8. Place the plate back on the magnetic stand at least 5 min or until the supernatant has cleared. *See* also **Note 4** about bead carryover.
9. Carefully transfer the supernatant to a new plate.
10. Seal the plate and freeze the samples at this point, or proceed with the index PCR.

3.4 Index PCR

In this step, barcodes for dual indexing are attached to the purified amplicons containing Illumina overhangs. For indexing, we use the Nextera XT DNA index kit according to the manufacturer's instructions.

1. For a 50 μl reaction volume, use 5 μl of cleaned up PCR amplicons, 5 μl Nextera XT Index Primer i5, 5 μl Nextera XT Index Primer i7, 25 μl 2× KAPA HiFi HotStart ready mix, and 10 μl nuclease-free water.
2. Run the PCR using the following cycling conditions: 3 min at 95 °C (1×); 30 s at 95 °C, 30 s at 55 °C, 30 s at 72 °C (8×); 5 min at 72 °C (1×); hold at 4 °C.

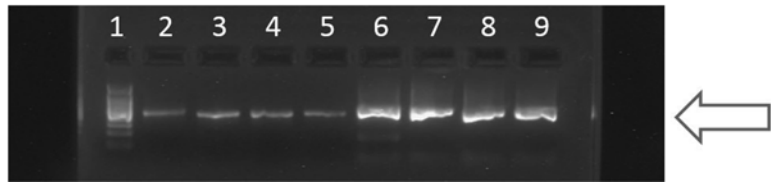

Fig. 2 PCR products run on a 1.5% agarose gel to check the size of the amplicons using a DNA ladder. Well 1: 100 bp DNA ladder; wells 2–5: samples after amplicon PCR and bead purification; wells 6–9: samples after index PCR. The *arrow* shows the desired product at 500 bp

3. Run the product on a 1.5% agarose gel to check the size of the amplicons using a DNA ladder (Fig. 2).
4. If any additional bands appear that are not the size of the desired product, additional purification steps need to be carried out (*see* **Note 3**).

3.5 Bead Purification

Purify the amplicons after the index PCR using magnetic beads and a magnetic stand. Size selection can be achieved using different ratios of beads to sample (*see* **Note 3**). A ratio bead:sample of 0.8:1 effectively purifies amplicons away from primers and primer dimers and allows selection of fragments larger than 200 bp [12]. Follow instructions in Subheading 3.3.

3.6 Library Validation

Verify the average fragment size of the individual samples with Tapestation. Ensure you have the right fragment and that you don't have any additional unwanted peaks (*see* **Notes 4** and **5**). The result should be a Tapestation trace with one main peak of the right size and the upper and lower marker peaks (Fig. 3).

3.7 Library Quantification

1. Measure the concentration of your libraries with Qubit (or some other fluorometric quantification method that uses dsDNA dyes) (*see* **Note 6**).
2. Calculate sample concentration in nM using the following formula. The assumed molecular weight of 1 bp is 660 Da.

$$\frac{(\text{Concentration in ng}/\mu\text{l}) \times 1{,}000{,}000}{(660\ \text{g}/\text{mol} \times \text{average library size})} = \text{nM}.$$

3.8 Normalization and Pooling

1. Dilute the samples to 50 nM using Tris–Cl 10 nM buffer, pH 8.5 with 0.1% Tween 20.
2. Dilute the samples again to 10 nM using Tris–Cl 10 nM buffer, pH 8.5 with 0.1% Tween 20.
3. Measure sample concentrations on Qubit to check the accuracy of dilution.
4. Correct the concentration of the samples to 10 nM.

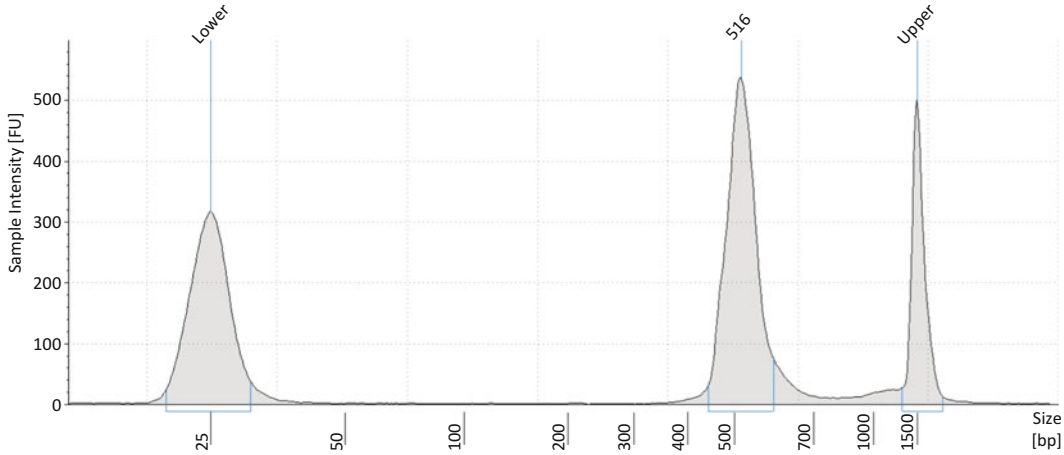

Fig. 3 Tapestation trace after bead purification and pooling

5. Pool the samples, using 2 µl of each sample. Adjust this volume if some of the samples are more or less concentrated than 10 nM.

3.9 Pool Validation

Validate the pool using Qubit and Tapestation. If an unwanted peak is observed at this stage, proceed with additional magnetic bead purification on the pooled samples. Bead purifications should be carried out until all unwanted products are removed (*see* **Note 4**).

1. Run Tapestation to see if the purification was successful.
2. Dilute the pool to 4 nM.

3.10 Library Denaturation

To prepare the DNA for cluster generation and sequencing, the libraries are denatured with NaOH, diluted with hybridization buffer (HT1), and heat denatured. The following steps are done accordingly Illuminas 16S Metagenomic Sequencing Library Preparation protocol [8].

1. Combine 5 µl pool (4 nM) with 5 µl NaOH (0.2 N) in a microcentrifuge tube.
2. Vortex and spin down.
3. Incubate at room temperature for 5 min to make the DNA single stranded.
4. Add 990 µl HT1.

This results in a 20 pM Library in 1 mM NaOH, which can be further diluted to your final concentration. Keep the denatured DNA on ice until you are ready to do the final dilution.

3.11 Dilution of Denatured Library

Dilute the denatured DNA with HT1 to the desired concentration. Illumina recommends starting the first run using a 4 pM loading concentration and to adjust that in following runs (*see* **Note 7**). For this project we diluted the library to 5 pM.

1. Mix 150 µl of your 20 pM denatured DNA with 450 µl pre-chilled HT1 to get 5 pM.
2. Invert and spin down.
3. Place the denatured and diluted DNA on ice.

3.12 Denaturation and Dilution of PhiX Control

PhiX is a balanced and diverse library that can be used as a control in sequencing runs. In this case we spiked in 15% (*see* **Note 8**).

Follow these steps to denature and dilute PhiX to the same concentration as the amplicon library (5 pM).

1. Combine 2 µl PhiX library (10 nM) with 3 µl Tris (10 mM, pH 8.5) in a microcentrifuge tube to dilute PhiX to 4 nM.
2. Vortex and spin down.
3. Add 5 µl NaOH (0.2 N) to the tube with 5 µl PhiX library (4 nM) to get a 2 nM PhiX library.
4. Vortex and spin down. Incubate at room temperature for 5 min to make the DNA single stranded.
5. Add 990 µl HT1 to dilute PhiX to 20 pM.
6. Mix 150 µl of your 20 pM denatured PhiX with 450 µl pre-chilled HT1 to dilute PhiX to the same loading concentration as your amplicon library (5 pM).
7. Invert and spin down.
8. Place the denatured and diluted PhiX on ice.

3.13 Combining the Library and PhiX Control

This step should be performed directly before loading the library into the MiSeq reagent cartridge.

1. Combine 90 µl PhiX (5 pM) with 510 µl of your library (5 pM) to get a 15% PhiX spike in.
2. Incubate the combined library and PhiX in a heat block at 96 °C for 2 min.
3. Invert the tube to mix and place it in an ice water bath for 5 min.

3.14 MiSeq Sequencing

The MiSeq Control Software (MCS) guides you through the steps to load the flow cell and reagents. MCS also provides you with an overview of the quality statistics during the run. MCS controls the flow cell stage, fluidics system, and flow cell temperatures, and captures images of clusters on the flow cell during the run. The following steps are done accordingly Illuminas MiSeq System User Guide [14].

1. Prepare the prefilled sequencing reagent cartridge for use (MiSeq Reagent Kit v3, 600c), by thawing it in a water bath with ultrapure water.
2. Wash and dry the flow cell.
3. Load the flow cell.
4. Load the PR2 bottle and empty the waste bottle.

5. Pierce the foil seal in the cartridge where it says "load sample" with a pipette tip and add 600 μl of your denatured and diluted sample.
6. Load the cartridge.
7. After loading the flow cell, buffer and reagent cartridge, MiSeq will search for the correct sample sheet based on the barcode on the cartridge. Review the run parameters in your sample sheet, created with Illumina Experiment Manager (IEM) (*see* **Note 9**). MCS performs the run according to parameters specified in the sample sheet.
8. Review prerun check results.
9. Select 'Start Run' to start sequencing (2 × 300 bp run).
10. Monitor the run from the MCS.
11. Perform a postrun wash when the sequencing run is ready.

4 Notes

1. A smaller reaction volume can also be used if necessary as the triplicate reactions will be later pooled.
2. Cycling conditions should be modified according to the target gene. For example, 48 °C annealing temperature will be best suited if the PCR is carried out with the mlCOIintF-dgHCO2198 primer set. Primer annealing temperature can be calculated as about 5 °C lower than the melting temperature (Tm) of your primers. You can also use a web-based Tm calculator tool. If no product is visible on the gel, lower annealing temperatures can be chosen but this can result in unspecific binding to sequences other than the intended target, visible as additional bands on the gel. Optimal annealing temperatures should result in the highest product yield for the intended amplicon. Regarding number of PCR cycles used, we recommend using the lowest possible number to prevent the introduction of PCR errors. Here we use 25 cycles for the amplicon PCR, but in cases where no PCR product is visible on the gel with 25 cycles, this can be increased to 30 or 35 cycles.
3. Size selection can be carried out by gel purification. An alternative better suited to high-throughput sequencing and low DNA concentrations is to use magnetic beads, as these will give better DNA recovery. Depending on the ratio of beads to sample, different size fragments can be purified [12]. In addition, selection of fragments to the left and right side of the desired fragment range can be carried out. Left side selection is done by binding the larger fragments to the right of the desired range to the beads and eluting the smaller fragments. For right size selection, the larger fragments to the right of the desired range are bound to the beads, and the supernatant containing the smaller fragments is removed to a fresh tube. For more details on this procedure *see* [13].
4. If an unwanted product is seen at 1000 bp, it can be due to bead carryover. To ensure that all magnetic beads are removed

from the sample, an additional purification step can be carried out, by placing the samples on the magnetic stand for 15 min and transferring the supernatant to a new tube.

5. It is not unusual to detect a peak around ~120–130 bp when validating your library. It is important to try to remove this peak (which could be adapter dimer). Adapter dimers will also bind to the flow cell, cluster very efficiently, and represent a high proportion of the total sequencing yield. This will also have an impact on the overall quality of the run, which tends to drop after the adapter dimer. You should be able to successfully remove adapter dimer by carrying out another round of magnetic bead cleanup. If your unwanted peak is much higher in proportion to your library peak, or if you cannot get rid of it by bead purification, a more aggressive selective approach by gel extraction is recommended. Keep in mind that the yield may be compromised when using this method.
6. Because it is very important to get the quantification as correct as possible, it is strongly advised to use a fluorescent dye method such as Qubit or PicoGreen instead of NanoDrop, which is based on UV absorbance. Instruments that use UV absorbance cannot distinguish between DNA, RNA, degraded nucleic acids, and other contaminants. Quantification with fluorescent dyes only detects the molecule of interest, and hence gives the most accurate values.
7. We expect amplicons to have low diversity, this is not an issue in general but with low diversity libraries it is advisable to keep cluster density between 600 and 800 K/mm^2 (or even lower). It is important to keep this in mind, so you don't overcluster the flowcell, which can lead to run failure. It is better to start with a low loading concentration and adjust it in following runs.
8. Low diversity libraries, such as amplicon libraries, where a large number of the reads have the same sequence, require a PhiX spike in to create a more diverse set of clusters. Illumina recommends a 1% PhiX spike in to all libraries. For low-diversity libraries, the percentage of PhiX depends on the diversity of the library and requires optimization. Between 10 and 20% should be enough.
9. Illumina Experiment Manager (IEM) is an application that helps you with the sample sheet setup of your run parameters for your MiSeq run.

Acknowledgements

This work was supported by Swedish research council grant C0344601 to S.J.B. Financial support to M.L. was provided by the Sant Chair and the Smithsonian Tennenbaum Marine Observatories Network, for which this Contribution No. 4. We would like to thank the Genomics Core Facility platform, at the Sahlgrenska Academy, University of Gothenburg.

References

1. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108(Suppl 1):4516–4522. doi:10.1073/pnas.1000080107
2. Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK (2012) Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol Evol 27:233–234
3. Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lambshead PJD, Thomas WK, Creer S (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nat Commun 1:98. doi:10.1038/Ncomms1095
4. Leray M, Agudelo N, Mills SC, Meyer CP (2013) Effectiveness of annealing blocking primers versus restriction enzymes for characterization of generalist diets: unexpected prey revealed in the gut contents of two coral reef fish species. PLoS One 8(4), e58076. doi:10.1371/journal.pone.0058076
5. Aylagas E, Borja A, Rodriguez-Ezpeleta N (2014) Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). PLoS One 9(3):e90529. doi:10.1371/journal.pone.0090529
6. Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proc Natl Acad Sci U S A 112(7):2076–2081. doi:10.1073/Pnas.1424997112
7. Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21(8):1794–1805. doi:10.1111/J.1365-294x.2012.05538.X
8. Illumina (2013) 16S metagenomic sequencing library preparation guide, Part # 15044223 Rev. B. http://www.support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf
9. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3(5):294–299
10. Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool 10:Unsp 34. doi:10.1186/1742-9994-10-34
11. Lobo J, Costa PM, Teixeira MAL, Ferreira MSG, Costa MH, Costa FO (2013) Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecol 13:Artn 34. doi:10.1186/1472-6785-13-34
12. CoreGenomics How do SPRI beads work? http://core-genomics.blogspot.co.uk/2012/04/how-do-spri-beads-work.html
13. BeckmanCoulter. http://www.beckmancoulter.com/wsrportal/bibliography?docname=SPRIselect.pdf
14. MiSeq system user guide, part # 15027617 Rev O http://www.support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-system-guide-15027617-o.pdf