

# Impact of sequencing platform, target amplicon and OTU-clustering on DNA metabarcoding of mock communities of marine macrobenthos

B. R. Leite<sup>A</sup>, C. Hollatz<sup>A</sup>, J. Lobo<sup>AB</sup>, H. Froufe<sup>C</sup>, C. Egas<sup>C</sup> & F. O. Costa<sup>A</sup>

<sup>A</sup>CBMA—Centre of Molecular and Environmental Biology, Biology Department, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.

<sup>B</sup>MARE—Marine and Environmental Sciences Centre, Department of Environmental Sciences and Engineering, Faculty of Science and Technology, Lisbon New University, 2829-516 Monte de Caparica, Portugal.

<sup>C</sup>Genoiseq, UC-Biotech, Center for Neurosciences and Cell Biology, Biocant Park, 3060-197, Cantanhede, Portugal.



## Abstract

In a previous study<sup>(1)</sup>, we tested 5 COI primer-pairs in two mock communities (MC1 and MC2), each containing 21 macrobenthic species in different proportions, using a Roche-454 platform. Here we compare the detection success obtained for the same communities with an Illumina-MiSeq platform using a) one of the 5 COI primer-pairs (ArF2/ArR5) and b) a primer-pair targeting an alternative marker: 18S rDNA. Furthermore, in both platforms we tested the impact of species assignment tools using two different approaches: read-based assignment and OTU-based assignment. Compared to 454, in MiSeq platform the detection success increased in both communities. Using the V4 region, species-level resolution was only attained for *Lepidochitona cinerea*. Moreover, some taxa were detected solely by V4, demonstrating a tendency to detect preferentially other taxa than the target macrobenthic species. Compared to the individual read-based assignments, OTU-based assignments resulted both in a lower detection success of the target species, together with an excess of putative taxonomic units.

## Methods

### Experimental contrived communities (MC1 and MC2)

#### 1. Impact of species assignment tools

##### COI: Roche-454 platform

ArF2/LoboR<sup>(2)</sup> invF/LoboR<sup>(1,2)</sup> jgLCO1490/jgHCO2198<sup>(3)</sup>  
mICOLintF/LoboR<sup>(4)</sup> ArF2/ArR5<sup>(5)</sup>

Read-based assignment

Previous study<sup>(1)</sup>

OTU-based assignment optimized

through "elbow"<sup>(6)</sup> analyses

Taxonomic assignment threshold: BOLD  $\geq$  97% sequence similarity

#### 2. Comparison of species detection success among target region and sequencing platform

##### Roche-454 and Illumina-MiSeq platforms

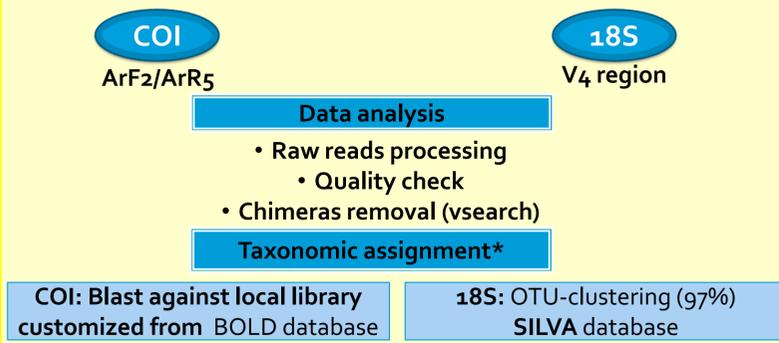


Figure 1. Overview of the experimental design.

## Roche-454 vs Illumina-MiSeq

- ✓ Application of the Illumina-MiSeq platform resulted in higher species detection success compared to Roche-454, in both mock communities.

Roche-454: 43% of species detected in both MC

Illumina-MiSeq: 52% in MC1 and 63% in MC2 of species detected

## COI vs 18S

- ✓ Using 18S rDNA, 22 and 63 OTUs were obtained that could be assigned to taxa present in the mock communities.
  - ✓ However, 18S species-level resolution was only attained for *Lepidochitona cinerea*.
  - ✓ Compared to 18S, COI performance demonstrated higher species discrimination ability
- COI: 11 species detected in MC1 and 13 in MC2

## Significance

Using communities of known composition we demonstrated that HTS-platforms with deeper sequencing capacity can improve species detection success. Furthermore, this test also illustrated the influence of target genomic regions on the ability of taxa detection. Additionally, we observed that OTU-clustering procedures, even when optimized, might still result in considerable overestimation of the taxonomic richness.

To avoid potential operational artifacts, circumvent OTU-clustering, and improve the performance of HTS-based macrobenthos monitoring, more effort should be allocated to the completion of reference libraries.

### References

- Hollatz C et al. (2017). Genome 60, 260-271. doi: 10.1139/gen-2015-0220.
- Lobo J et al. (2013). BMC Ecology 13, 34. doi:10.1186/1472-6785-13-34.
- Geller J et al. (2013). Molecular Ecology Resources 13, 851-861. doi:10.1111/1755-0998.12138.
- Leray M et al. (2013). Frontiers in Zoology 10, 34. doi:10.1186/1742-9994-10-34.
- Gibson J et al. (2014). Proceedings of the National Academy of Science U.S.A. 111, 8007-8012. doi:10.1073/pnas.1406468111.
- Young RG et al. (2017). Genome 60, 169-182. doi:10.1139/gen-2015-0209.

- ✓ Differences in success rates of species detection, between COI and 18S, can be attributed mainly to 2 reasons: differences in taxonomic coverage of the reference databases and potential primer bias in PCR amplification, promoting preferential amplification of certain taxa for each primer/target region.

## Read-based assignment vs OTU-based assignment

- ✓ Globally the species detection success with read-based assignment remained higher than OTU-based assignment, except with mICOLintF/LoboR in MC2 where the success was identical.

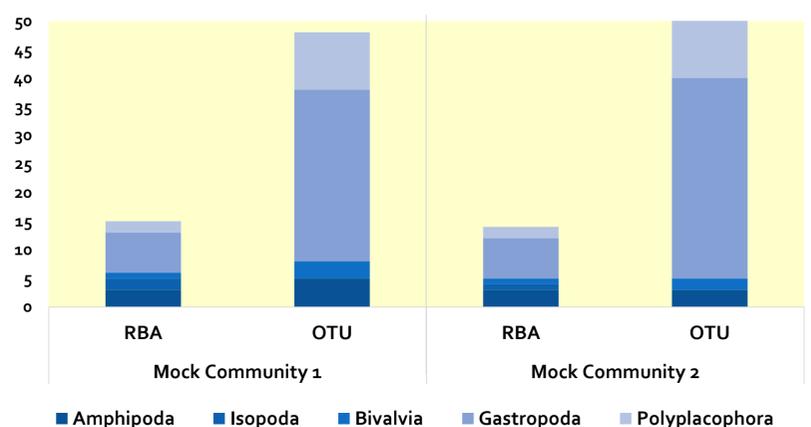


Figure 2. Comparison of species detected by Read-based assignment with OTU-based assignment for 5 taxonomic groups present in mock communities.

- ✓ OTU clustering resulted in an excess of retrieved taxonomic units in all primer-pairs (range 29-169), compared to known composition of the mock communities. For example, multiple OTUs were produced for the same species in various cases and primers (e.g. 4 OTUs for *Patella aspera* in mICOLintF/LoboR in MC2).
- ✓ Application of the OTU-based taxonomic assignment resulted in an "artificial" community composition.
- ✓ Hence, if this was applied in a "blind" assessment of a natural community, the detection of higher taxonomic ranks through the use of OTU-clustering can result in an overestimation of the taxonomic richness.

### Acknowledgments

The present study was financed by FEDER through POFC-COMPETE, in the scope the project FCOMP-01-0124-FEDER-015429 granted by "Fundação para a Ciência e a Tecnologia" (FCT), Portugal. Work at CBMA was supported by FCT I.P. through the strategic funding UID/BIA/04050/2013. Barbara Leite was supported by a PhD fellowship (PD/BD/127994/2016) from FCT.

