



Conference Abstract

Bacteria are everywhere, even in your COI data: The art of getting to know the unknown unknowns and shine light on the dark matter!

Haris Zafeiropoulos^{‡,§}, Laura Gargan^l, Christina Pavloudi[§], Evangelos Pafilis[§], Jens Carlsson^l

[‡] University of Crete, Heraklion, Greece

[§] Hellenic Centre for Marine Research (HCMR), Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Heraklion, Crete, Greece

^l Area 52 Research Group, Earth Institute/School of Biology and Environmental Science, University College Dublin, Dublin, Ireland

Corresponding author: Laura Gargan (lauramgargan@gmail.com)

Received: 24 Feb 2021 | Published: 04 Mar 2021

Citation: Zafeiropoulos H, Gargan L, Pavloudi C, Pafilis E, Carlsson J (2021) Bacteria are everywhere, even in your COI data: The art of getting to know the unknown unknowns and shine light on the dark matter! ARPHA Conference Abstracts 4: e64966. <https://doi.org/10.3897/aca.4.e64966>

Abstract

Environmental DNA (eDNA) metabarcoding has been commonly used in recent years (Jeunen et al. 2019) for the identification of the species composition of environmental samples. By making use of genetic markers anchored in conserved gene regions, universally present across the species of large taxonomy groups, eDNA metabarcoding exploits both extra- and intra-cellular DNA fragments for biodiversity assessment.

However, there is not a truly “universal” marker gene that is capable of amplifying all species across different taxa (Kress et al. 2015). The mitochondrial cytochrome C oxidase subunit I gene (COI) has many of the desirable properties of a “universal” marker and has been widely used for assessing species identity in Eukaryotes, especially metazoans (Andjar et al. 2018). However, a great number of COI Operational Taxonomic Units (OTUs) or/and Amplicon Sequence Variants (ASVs) retrieved from such studies do not match reference sequences and are often referred to as “dark matter” (Deagle et al. 2014). The aim of this study was to discover the origins and identities of these COI dark matter sequences.

We built a reference phylogenetic tree that included as many COI-sequence-related information across the tree of life as possible. An overview of the steps followed is presented in Fig. 1a. Briefly, the [Midori reference 2](#) database was used to retrieve eukaryotes sequences (183,330 species). In addition, the API of the [BOLD database](#) was used as source for the corresponding Bacteria (559 genera) and Archaea (41 genera) sequences. Consensus sequences at the family level were constructed from each of these three initial COI datasets. The COI-oriented reference phylogenetic tree of life was then built by using 1,240 consensus sequences with more than 80% of those coming from eukaryotic taxa.

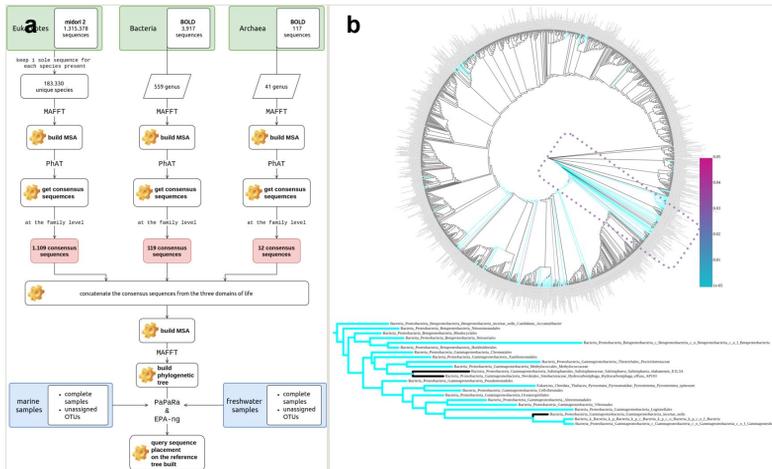


Figure 1.

Investigating COI dark matter in a nutshell.

a: Overview of the bioinformatic steps followed.

b: Upper figure: Overview of the reference phylogenetic tree built and the sequence query placements assigned for a marine sample. Black branches have no sequences assigned to them while the blue to purple colour scale. Bottom figure: Magnification of the bacterial branches of the tree where a great number of placements were assigned.

Phylogeny-based taxonomic assignment was then used to place query sequences. The a) total number of sequences, b) sequences assigned to Eukaryotes and c) unassigned subsets of OTUs, from marine and freshwater samples, retrieved during in-house metabarcoding experiments, were placed in the reference tree (Fig. 1b). It is clear that a large proportion of sequences targeting the COI region of Eukaryotes actually represents bacterial branches in the phylogenetic tree (Fig. 1b).

We conclude that COI metabarcoding studies targeting Eukaryotes may come with a great bias derived from amplification and sequencing of bacterial taxa, depending on the primer pair used. However, for the time being, publicly available bacterial COI sequences are far too few to represent the bacterial variability; thus, a reliable taxonomic identification of them is not possible. We suggest that bacterial COI sequences should be included in the reference databases used for the taxonomy assignment of OTUs/ASVs in COI-based

eukaryote metabarcoding studies to allow for bacterial sequences that were amplified to be excluded enabling researchers to exclude non-target sequences. Further, the approach presented here allows researchers to better understand the unknown unknowns and shed light on the dark matter of their metabarcoding sequence data.

Keywords

eDNA, metabarcoding, marker genes, COI, Bacteria, phylogenetic tree, taxonomy assignment, dark matter

Presenting author

Laura Gargan

Presented at

1st DNAQUA International Conference (March 9-11, 2021)

Acknowledgements

This research was supported in part through computational resources provided by IMBBC (Institute of Marine Biology, Biotechnology and Aquaculture) of the HCMR (Hellenic Centre for Marine Research). Funding for establishing the IMBBC HPC has been received by the MARBIGEN (EU Regpot) project, LifeWatchGreece RI and the CMBR (Centre for the study and sustainable exploitation of Marine Biological Resources) RI.

Funding program

This project has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Innovation (GSRI), under grant agreement No. 241 ([PREGO](#) project). It has been also supported by the "[ELIXIR-GR: Managing and Analysing Life Sciences Data \(MIS: 5002780\)](#)" project co-financed by Greece and the European Union - European Regional Development Fund.

References

- Andjar C, Arribas P, Yu DW, Vogler AP, Emerson BC (2018) Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology* 27 (20): 3968-3975. <https://doi.org/10.1111/mec.14844>

- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol Lett* 10: 20140562. <https://doi.org/10.1098/rsbl.2014.0562>
- Jeunen G, Knapp M, Spencer HG, Taylor HR, Lamare MD, Stat M (2019) Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecology and Evolution* 9: 1323-1335. <https://doi.org/10.1002/ece3.4843>
- Kress WJ, Garca-Robledo C, Uriarte M, Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution* 30 : 25-35. <https://doi.org/10.1016/j.tree.2014.10.008>