



Master project 2024-2025

Personal Information

Supervisor	Miguel Méndez Sandín
Email	miguel.mendez@ibe.upf-csic.es
Institution	Institute of Evolutionary Biology (CSIC-UPF)
Website	https://www.beaplab.org/
Group	Biology and Ecology of Abundant Protist Lab

Project

Computational systems biology

Project Title:

Do new environmental sequences truly represent new species?

Keywords:

Metabarcoding, Intragenomic variability, Cryptic diversity, Biogeography, Ecology

Summary:

Microbial eukaryotes represent the backbone of the eukaryotic tree of life, where their multicellular counterparts evolved several times independently. While the diversity of animals, plants and fungi is well studied, that of microbial eukaryotes lag far behind despite they represent the vast bulk of eukaryotic diversity. In recent years, an enormous genetic diversity of microbial eukaryotes has been unveiled by the approach of metabarcoding (Burki et al., 2021) which consist in the high-throughput sequencing of genetic markers from environmental samples. The large number of reads generated by this approach is normally classified into clusters of diversity to reduce amplification or sequencing errors, generally so-called metabarcodes. And yet, a global oceanic survey found that 86% of the metabarcodes do not match any reference sequence assigned to a given species (de Vargas et al., 2015). Since most of such novel diversity is highly similar to characterized species, it has been regarded as intragenomic variability (Caron and Hu, 2019) and even shown in specific taxa; as in Radiolaria (Decelle et al., 2014; Sandin et al., 2022), Foraminifera (Greco et al., 2023), or ciliates (Bachy et al., 2013; Gong et al., 2013; Santoferrara et al., 2016). Recent global scale comparative analyses have shown that several metabarcodes with distinct biogeographic patterns can hide under the same morphological species, suggesting that genetic diversification is decoupled from morphological diversification (Morard et al., 2024). A subsequent question is thus, are such intragenomic biogeographical patterns common across eukaryotes? And if it so, could differential expression of intragenomic variability be a consequence of ecological speciation? In this project we will mine the most comprehensive global metabarcoding dataset (EukBank; Berney et al., 2023) to explore whether highly similar metabarcodes co-occur or have distinct biogeographical patterns. To do so, the specific objectives are to (Objective 1) characterize the worldwide biogeography and abundance of described examples of intragenomic variability and their related metabarcodes, and to (Objective 2) extend such search of highly similar metabarcodes to all eukaryotes. The master student will benefit from the expertise and facilities available at the Biology and Ecology of Abundant Protists Lab (beaplab.org/), within the Institute of Evolutionary Biology (ibe.upf-csic.es/). Promising preliminary data and resources are already available. We anticipate that the outcome of the work proposed here will be submitted for publication in a peer-reviewed scientific journal, such as Environmental Microbiology or The ISME Journal.

References:

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *ISME J.* 7(2):244-55. doi: 10.1038/ismej.2012.106
Berney C, Mahé F, Henry N, Lara E, de Vargas C, EukBank consortium (2023) EukBank 18S V4 dataset [Data set]. Zenodo. doi: 10.5281/zenodo.7804946
Burki F, Sandin MM, Jamy M (2021) Diversity and ecology of protists revealed by metabarcoding. *Curr Biol.* 31(19):R1267-R1280. doi: 10.1016/j.cub.2021.07.066
Caron DA, Hu SK (2019) Are We Overestimating Protistan Diversity in Nature? *Trends Microbiol.* 27(3):197-205. doi: 10.1016/j.tim.2018.10.009
Decelle J, Romac S, Sasaki E, Not F, Mahé F (2014) Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. *PLoS One.* 9(8):e104297. doi: 10.1371/journal.pone.0104297
de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N,

Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury JM, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S; Tara Oceans Coordinators, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E. (2015) Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science*. 348(6237):1261605. doi: 10.1126/science.1261605 Gong J, Dong J, Liu X, Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist*. 164(3):369-79. doi: 10.1016/j.protis.2012.11.006 Greco M, Morard R, Darling K, Kucera M (2023) Macroevolutionary patterns in intragenomic rDNA variability among planktonic foraminifera. *PeerJ*. 11:e15255. doi: 10.7717/peerj.15255 Morard R, Darling KF, Weiner AKM, Hassenrück C, Vanni C, Cordier T, Henry N, Greco M, Vollmar NM, Milivojevic T, Rahman SN, Siccha M, Meilland J, Jonkers L, Quillévéré F, Escarguel G, Douady CJ, de Garidel-Thoron T, de Vargas C, Kucera M (2024) The global genetic diversity of planktonic foraminifera reveals the structure of cryptic speciation in plankton. *Biol Rev Camb Philos Soc*. doi: 10.1111/brv.13065 Sandin MM, Romac S, Not F (2022) Intra-genomic rDNA gene variability of Nassellaria and Spumellaria (Rhizaria, Radiolaria) assessed by Sanger, MinION and Illumina sequencing. *Environ Microbiol*. 24(7):2979-2993. doi: 10.1111/1462-2920.16081 Santoferrara LF, Grattepanche JD, Katz LA, McManus GB (2016) Patterns and processes in microbial biogeography: do molecules and morphologies give the same answers? *ISME J*. 10(7):1779-90. doi: 10.1038/ismej.2015.224

Expected skills:

A programming language (preferably R, bash and/or python), and basic understanding on community ecology and/or numerical ecology.

Possibility of funding:

No

Possible continuity with PhD:

No

Comments:

Possibility of remote/hybrid work