# **Cell Genomics**



### **Preview**

# Metagenome assembled genomes are for eukaryotes too

Ramon Massana<sup>1,\*</sup> and David López-Escardó<sup>1</sup>

<sup>1</sup>Institut de Ciències del Mar (ICM-CSIC), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain \*Correspondence: ramonm@icm.csic.es https://doi.org/10.1016/j.xgen.2022.100130

Environmental community sequencing is suitable for producing metagenome assembled genomes (MAGs) of prokaryotes, but there is the perception that it cannot work for eukaryotes. In this issue of *Cell Genomics*, Delmont et al<sup>1</sup> process a massive sequencing dataset from marine plankton to produce 683 eukaryotic MAGs; the study also includes useful functional information.

Pioneers in microbial ecology started their journey by studying the abundance, activity, and diversity of microbes in natural samples, but many already had in mind the potential of metagenomics to investigate the functional and metabolic capacity of microbial communities. Initially, this was limited by technical capacities, and only few genes were captured, providing new insights of a tiny section of the microbial community. However, advances in sequencing, computer power, and bioinformatic tools have been able to break this barrier and today allow reconstruction of species-level prokaryotic genomes from microbial communities, the socalled metagenome assembled genomes (MAGs). Thanks to MAGs, the prokaryotic Tree of Life has been populated and expanded with uncultured species, providing new and relevant functional and evolutionary insights, with the latest studies reporting thousands of new genomes.<sup>2</sup> However, there was the perception that this approach could not work for eukaryotes, given their larger genomes, split in different chromosomes, and the many non-coding and repetitive regions, which add a level of complexity to current bioinformatic tools. Only a few examples of eukaryotic MAGs have been reported, often derived from relatively low-diverse communities.<sup>3,4</sup> In this issue of Cell Genomics, Delmont et al.<sup>1</sup> question this perception by presenting 683 new MAGs from planktonic marine eukaryotes. The main reason behind their success seems to be the huge amount of sequencing data (798 metagenomic samples, a total of 280 billion reads), the use of an exhaustive dataset that includes size-fractionated plankton communities taken from around the globe during the Tara Oceans

expedition, and allowing exclusion of part of the prokaryotic signal. In the study, the authors also include 30 single-cell amplified genomes (SAGs) from microbial eukaryotes to construct a solid genomic repository of dominant marine eukaryotes.

To build eukaryotic MAGs, Delmont et al.<sup>1</sup> performed metagenome assembly and binning of MAGs on samples taken from the same geographic region. The 11 prepared co-assemblies vielded 837 MAGs that, after removing redundancy, resulted in 683 new eukaryotic genomes of at least 10 Mb in size. The authors meticulously removed spurious contamination in MAGs by manual inspection with the visualizing platform anvi'o. In the few cases that MAGs affiliated to a species that has been cultured and genome sequenced, the results were verv satisfactory, as both genomic datasets were virtually identical. The averaged statistics of the whole dataset were 35.4 Mb in genome size, about 14,000 genes per MAG, and a BUSCO completeness of 40%. Following well-established criteria<sup>5</sup> of considering a decent MAG those with completeness above 50%, the collection is reduced to 240 MAGs; thus, still a substantial amount of data remains. In fact, this collection of Tara Oceans MAGs plus the 30 SAGs is so far the best representation of genomic diversity of marine eukaryotes, recruiting, on average, 11% of the metagenomic reads in the samples, outperforming any other reference genomic database. Therefore, although with some gaps, these MAGs represents a new, timely, and useful resource to interpret marine eukaryotic diversity.

MAGs did not contain the 18S rDNA gene because of the binning process.

So, their taxonomic annotation was achieved by conducting new phylogenies with DNA-dependent RNA polymerase genes retrieved from reference genomes and MAGs. Very interestingly, the taxonomic classification of MAGs corresponded partially to what is known as the dominant groups in marine ecosystems,<sup>6</sup> such as copepods within metazoans and stramenopiles, haptophytes, and chlorophytes within unicellular eukaryotes (Figure 1). However, some important protist groups were absentmost remarkably, dinoflagellates, MALVs (marine alveolates), and radiolaria and, to a lower extent, the excavates diplonemea and kinetoplastids. This can be explained by several reasons compromising the assembly of these particular taxa, such as larger and more complex genomes within these groups and the fact that many of them tend to be hyperdiverse in natural samples, lacking a dominant genotype.<sup>7</sup> Finally, Delmont et al.<sup>1</sup> also recovered three related MAGs that could putatively belong to a new high-rank eukaryotic group.

The BUSCO completeness was also very dependent on the taxonomic groups. At one end, some groups of picoeukaryotes, such as Mamiellophyceae, Chloropicophyceae, and Bolidophyceae, have very large scores with close to complete genomes, probably because they tend to present smaller and less complex genomes with few non-coding regions. Other groups also belonging to picoplankton and nanoplankton, such as diatoms, chrysophyceae, MASTs (marine stramenopiles), Cryptophyta, and Pelagophyceae, have completeness between 40% and 65%. Groups with very poor genome coverage include ciliates,





Cell Genomics Preview



Figure 1. BUSCO completeness of MAGs within taxonomic groups ordered by functional clustering Ciliate MAGs (30) and MAGs with a functional profile different from its taxa (8) are not included. Lineages from the same supergroup are marked with the same color in the violin plots: Opisthokonta in orange, Stramenopile in yellow, Archaeplastida in green, Alveolates in blue, and the remaining groups in gray. Left column colors correspond to the five functional groups described by the authors in the article.

myzozoa (probably composed by MALV species), and the opisthokonts metazoa and chaonoflagellates, for which less than 18% of the MAGs pass the 50% completeness threshold.

After obtaining this genomic resource from dominant and ecologically relevant marine eukaryotes, Delmont et al.<sup>1</sup> used it to compare, for the first time, the functional grouping of marine species based on their genomic data. The clustering of all MAGs based on orthologous genes indicated five main functional groups of marine species, one formed by the small animals and the other four including all protists. At a low scale, virtually all species from the same group appeared together, implying that they have a similar functional performance and very strong vertical evolutionary constraints on their genomes. At a larger scale, two functional clusters still fit vertical evolution, such as cluster A that includes all haptophyte algae and cluster B that presents different groups from the supergroup teloneomids, stramenopiles, alveolates, and rhizaria. More interesting patterns emerge in the remaining clusters, which put together divergent taxonomic groups, and imply a sort of functional convergence for species

cohabiting the oceanic realm for millions of years. Cluster C unites two taxonomically distant groups of algae, diatoms and archaeplastida, both of which seem to lack phagotrophic capacity (but see Bock et al.<sup>8</sup>). Finally, cluster D represents the most intriguing functional group, including acanthoecids, Picozoa, and "oomycota." Based on the very low abundance of oomycota in the marine plankton and on its phylogenetic position, we believe that this category includes MAST-1 taxa, which are the largest MAST cells known and for which genomes obtained from single cells have

## **Cell Genomics**

**Preview** 

been recently generated.<sup>9</sup> So, cluster D is unifying non-photosynthetic large heterotrophic eukaryotes, in which the authors found an enrichment of functions associated with carbohydrate transport and metabolism, perhaps related to different prey preferences or metabolic performances.

This study precedes a preprint that analyzes the same metagenomes<sup>10</sup> and provides an excellent opportunity to compare and evaluate how similar eukaryotic MAGs generated by different pipelines are. This work represents an important step forward in the knowledge of eukaryotic genomic diversity, and we can now openly say that MAGs are for eukaryotes, too, paraphrasing a previous perspective claiming the role of protists in microbial ecology.<sup>11</sup> We envision a future when eukaryotic MAGs will be better extracted from metagenomic datasets, perhaps by importing new sequencing approaches or better bioinformatic tools, providing exhaustive genome characterizations of dominant species in nature, and saturating new phylum discovery as is occurring with prokaryotes. These genomes can then be used to study species biogeography, population genomics, and fine evolutionary events when applied to long-term monitoring series. In addition, when combined with metatranscriptomics, MAGs will provide the reference for retrieving the species transcriptomic profile, giving a timely report of the genes they use to thrive in the environment.

New times are ahead for microbial ecologists, with new tools, datasets, and resources for a better characterization of microbial roles in the biogeochemical cycles.

### **DECLARATION OF INTERESTS**

The authors declare no competing interests

### REFERENCES

- Delmont, T.O., Gaia, M., Hinsinger, D.D., Fremont, P., Vanni, C., Fernandez Guerra, A., Eren, A.M., Kourlaiev, A., d'Agata, L., Clayssen, Q., et al. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genomics 2.
- Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.-M., Huntemann, M., et al.; IMG/M Data Consortium (2021). A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 39, 499–509. https://doi.org/10. 1038/s41587-020-0718-6.
- Joli, N., Monier, A., Logares, R., and Lovejoy, C. (2017). Seasonal patterns in Arctic prasinophytes and inferred ecology of *Bathycoccus* unveiled in an Arctic winter metagenome. ISME J. *11*, 1372–1385. https://doi.org/10. 1038/ismej.2017.7.
- West, P.T., Probst, A.J., Grigoriev, I.V., Thomas, B.C., and Banfield, J.F. (2018). Genome-reconstruction for eukaryotes from complex natural microbial communities. Genome Res. 28, 569–580. https://doi.org/ 10.1101/gr.228429.117.
- Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., and Tyson, G.W. (2017).



Recovery of nearly 8,000 metagenomeassembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533–1542. https://doi.org/10.1038/s41564-017-0012-7.

- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science 348, 1261605. https://doi.org/10.1126/science.1261605.
- Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., Bowler, C., Lukes, J., and Horák, A. (2016). Extreme diversity of diplonemid eukaryotes in the ocean. Curr. Biol. *26*, 3060–3065. https://doi.org/10. 1016/j.cub.2016.09.031.
- Bock, N.A., Charvet, S., Burns, J., Gyaltshen, Y., Rozenberg, A., Duhamel, S., and Kim, E. (2021). Experimental identification and in silico prediction of bacterivory in green algae. ISME J. 15, 1987–2000. https://doi. org/10.1038/s41396-021-00899-w.
- Labarre, A., López-Escardó, D., Latorre, F., Leonard, G., Bucchini, F., Obiol, A., Cruaud, C., Sieracki, M.E., Jaillon, O., Wincker, P., et al. (2021). Comparative genomics reveals new functional insights in uncultured MAST species. ISME J. 15, 1767–1781. https://doi. org/10.1038/s41396-020-00885-8.
- Alexander, H., Hu, S.K., Krinos, A.I., Pachiadaki, M., Tully, B.J., Neely, C.J., and Reiter, T. (2021). Eukaryotic genomes from a global metagenomic dataset illuminate trophic modes and biogeography of ocean plankton. Preprint at bioRxiv. https://doi.org/10.1101/ 2021.07.25.453713.
- Caron, D.A., Worden, A.Z., Countway, P.D., Demir, E., and Heidelberg, K.B. (2009). Protists are microbes too: a perspective. ISME J. 3, 4–12. https://doi.org/10.1038/ismej.2008.101.