

RESEARCH HIGHLIGHT

Genomes of uncultured eukaryotes: sorting FACS from fiction

Alexandra Z Worden^{1*}, Christopher Dupont² and Andrew E Allen²

Abstract

A recent study explores the genome content of uncultured unicellular marine eukaryotes and provides insights about interactions between uncultured eukaryotes and other biological entities.

Uncultured unicellular eukaryotes have critical roles in global CO₂ fixation in the oceans. The dilemma is that as earth systems undergo climate change, the responses of these elusive organisms and other uncultured taxa are nearly impossible to study or predict. Researchers are now investigating uncultured microbes by sequencing their genomes directly from the environment. The approach itself sounds straightforward: use fluorescently activated cell sorting (FACS) to separate populations, amplify their DNA by multiple displacement amplification (MDA), and sequence individual genes or genome fragments. But artifacts can be introduced at each step in the process and careful consideration is required during the data interpretation phase.

For many years, oceanographers inferred from pigment distributions that a group of unicellular eukaryotic phytoplankton belonging to the prymnesiophytes were important - and that these prymnesiophytes were tiny (picoplankton, up to 2 to 3 μm in diameter). Yet no such prymnesiophyte existed in culture - at least none that matched the 18S rDNA sequences commonly retrieved from the ocean by PCR. Two recent publications showed that uncultured pico-prymnesiophytes are responsible for 25 ± 9% of the overall primary production (photosynthetic uptake of CO₂ into plant-like biomass) in the North Atlantic [1,2]. Moreover, pico-prymnesiophytes form a significant portion of picoplanktonic photosynthetic

biomass in biogeographical provinces stretching from the tropics to high latitude seas [1,2].

In 2010 we sequenced a targeted metagenome from a wild pico-prymnesiophyte population using their natural photosynthetic pigments and size characteristics, at-sea FACS and MDA [2], a non-PCR based DNA amplification technique that uses random hexamer primers and the bacteriophage-derived Φ29 polymerase. The resulting partial genome assemblies revealed densely packed genomes with sparse intergenic regions and novel features for phytoplankton. Since then, a second study has explored the 18S rDNA diversity of discrete eukaryotic phytoplankton populations, again using FACS, MDA and then PCR and 18S rRNA gene clone library construction [3].

In addition to phytoplankton such as pico-prymnesiophytes, there are rare eukaryotes in seawater. An intriguing uncultivated group of such eukaryotes, the biliphytes, has been an elusive target because of their sparseness in marine samples and difficulties in attaining statistically supported data. However, a new study [4] reports sequencing of biliphytes using FACS and MDA. Biliphytes were initially thought to potentially represent a unique group of red algae; however, more comprehensive phylogenetic analysis placed them elsewhere in the eukaryotic tree of life [5]. Moreover, microscopy work indicated they contained orange, phycobilin-like fluorescence (a photosynthetic pigment found in cyanobacteria and some eukaryotic algae) and were picoplanktonic in size, and hence they were named 'picobiliphytes' [5]. The same study also suggested they contained a remnant nucleus (a nucleomorph) from a eukaryote engulfed in an ancient secondary endosymbiotic event [5]. This was very exciting because only two lineages are known to have nucleomorphs, making it difficult to trace evolutionary relationships between different ancestral host eukaryotes because of mixing of genes from hosts and the photosynthetic eukaryotes that they engulfed long ago. A subsequent study [6] more tentatively suggested that they were photosynthetic, and placed the group in a similar (although not statistically supported) phylogenetic position, but found no evidence for a nucleomorph. The uncultured cells were more abundant in this study and

*Correspondence: azworden@mbari.org

¹Monterey Bay Aquarium Research Institute (MBARI), 7700 Sandholdt Road, Moss Landing, CA 95039, USA

Full list of author information is available at the end of the article

appeared larger, about 3.5 to 4 μm diameter, so the group was renamed 'biliphytes'.

The recent FACS/MDA study of uncultured marine eukaryotes looked at single biliphyte cells [4]. Biliphyte genome fragments were sequenced along with those of co-associated entities. The study [4] found no nucleomorph-like genes, supporting inferences from microscopy results [2,6] that no nucleomorph was present. The results also supported inferences [7] that biliphytes may not be photosynthetic but perhaps facultative mixotrophs or phagotrophs, whereby transient detection of orange fluorescence could represent ingested prey items (for example, *Synechococcus*) [7].

Assessing characteristics that might be absent from a genome using partial genome sequences from single cells or populations hinges on the relationship between genome recovery and arguments about absence. These arguments can be tested using the Bernoulli distribution, a probability distribution of the number of successes from multiple independent yes/no experiments, each with the same probability of success, but only after the critical task of estimating genome recovery has been accomplished. There are inherent biases in MDA reactions that lead to insufficient coverage of entire genomes, and this is confounded by the possibility that a single FACS gating event can include more than one organism. Bacteria and viruses often reside in close extracellular association with eukaryotic cells [8]. Diverse uncultured fungi have recently been discovered; these attach to diatoms and presumably other microbes too [9]. MDA itself can also introduce artificial contaminants.

We [2] and Yoon *et al.* [4] both highlight the confounding influence of natural or artificial genomic contaminants in FACS/MDA-derived data. In [4], in addition to biliphyte sequences, assemblies were generated from viruses and a bacterium hypothesized to be an ingested prey, but alternatively may have been attached to the cell surface of the sorted biliphyte cell. Each contaminating genome fragment, regardless of derivation, increases the apparent total genome pool that is sampled, reducing the probability of sampling the targeted taxon. The chances of Yoon *et al.* [4] not recovering 150 genes encoded by the plastid (that one would expect to find if the organism was photosynthetic) in 6,000 independent sampling events from a pool of 12,000 genes is unlikely, but not implausible given MDA biases. The chances of not detecting specific genes increase if the gene pool is larger; biliphytes may have a larger gene pool given that a comparison with complete genomes from smaller eukaryotes was used to generate this estimate.

A hurdle for future efforts is implementation of bioinformatic methods for separating a heterogeneous genome population into its individual constituents. Genome sorting is further hindered by the chimeric

nature of many eukaryotic genomes, which contain phylogenetic signals for other lineages and even bacterial phyla [10]. Rigorous approaches are required to confidently classify data into genetic material from target cells versus that from other co-sorted entities. Phylogenomic filters can help identify *bona fide* scaffolds assembled from target taxa reads to conservatively restrict comparative analyses. For example, Yoon *et al.* reduced their data set using blastX to select 7.9 Mbp of contigs of eukaryotic, and possibly biliphyte, origin from approximately 28 Mbp of assembly derived from just over 3 Gbp of raw sequence. Further, they reported globally on the taxonomic content of open reading frames within contigs by using BLASTx combined with phylogenomic profiling; their analyses returned 5,231 phylogenetic trees. A classifiable majority of the putative picobiliphyte proteins were phylogenetically most similar to either Metazoa, Viridiplantae or Stramenopiles [4]. An alternative approach is to include contig-level phylogenetic classification and analyses of expected and recovered gene family distributions, from which genomic properties such as gene size and density can be inferred [2].

An ongoing difficulty is the paucity of appropriate reference genomes for the phylogenomic filtering database. For studies of genome sequences from cultured eukaryotes, confident phylogenetic classification of genes from distant lineages (for example, genes of bacterial or viral origin) [10] is derived from genomic assembly - which is facilitated by sufficient sequence coverage for each nucleotide position. This type of coverage and corresponding assembly has been difficult to achieve for MDA material generated from eukaryotic nuclear DNA. This problem should be solved soon given initiatives by several single-cell genomics groups.

Perhaps most exciting in the latest publication on sequencing of uncultured eukaryotes [4] was retrieval of genomes from viruses that were either attached to or infecting biliphytes. These had high coverage, assembled well and represented viruses for which no data were available. Viruses are enormously abundant in marine environments and have important roles in shaping the population dynamics of prokaryotic and eukaryotic microbes - their genomes can contain genes that seem to be 'stolen' from their hosts and can reveal especially important adaptations and environmental pressures.

Genomics of uncultured eukaryotic microbes [2,4] is providing new information on their biology and interactions with other microbes. The findings of Yoon *et al.* [4] also highlight that diversity of organism-organism interactions might drive the low cohesiveness between the three biliphyte cells investigated. The future holds much excitement - new discoveries will come from scaling up and refinement of current approaches for sorting facts about uncultured eukaryotes from fiction.

Acknowledgements

We are grateful to Mary Ann Moran for commenting on the manuscript.

Author details

¹Monterey Bay Aquarium Research Institute (MBARI), 7700 Sandholdt Road, Moss Landing, CA 95039, USA. ²Microbial and Environmental Genomics, J Craig Venter Institute, 10355 Science Center Drive, San Diego, CA 92121, USA.

Published: 27 June 2011

References

- Jardillier L, Zubkov MV, Pearman J, Scanlan DJ: **Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean.** *ISME J* 2010, **4**:1180-1192.
- Cuvelier ML, Allen AE, Monier A, McCrow JP, Messié M, Tringe SG, Woyke T, Welsh RM, Ishoey T, Lee JH, Binder BJ, DuPont CL, Latasa M, Guigand C, Buck KR, Hilton J, Thiagarajan M, Caler E, Read B, Lasken RS, Chavez FP, Worden AZ: **Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton.** *Proc Natl Acad Sci U S A* 2010, **107**:14679-14684.
- Lepere C, Demura M, Kawachi M, Romac S, Probert I, Vaulot D: **Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations.** *FEMS Microbiol Ecol* 2011, **76**:513-523.
- Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S, Bhattacharya D: **Single-cell genomics reveals organismal interactions in uncultivated marine protists.** *Science* 2011, **332**:714-717.
- Not F, Valentin K, Romari K, Lovejoy C, Massana R, Töbe K, Vaulot D, Medlin LK: **Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes.** *Science* 2007, **315**:253-255.
- Cuvelier ML, Ortiz A, Kim E, Moehlig H, Richardson DE, Heidelberg JF, Archibald JM, Worden AZ: **Widespread distribution of a unique marine protistan lineage.** *Environ Microbiol* 2008, **10**:1621-1634.
- Kim E, Harrison JW, Sudek S, Jones MD, Wilcox HM, Richards TA, Worden AZ, Archibald JM: **Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life.** *Proc Natl Acad Sci U S A* 2011, **108**:1496-1500.
- Azam F, Worden AZ: **Oceanography. Microbes, molecules, and marine ecosystems.** *Science* 2004, **303**:1622-1624.
- Jones MD, Forn I, Gadelha C, Egan MJ, Bass D, Massana R, Richards TA: **Discovery of novel intermediate forms redefines the fungal tree of life.** *Nature* 2011, **474**:200-203.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, *et al.*: **The Phaeodactylum genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**:239-244.

doi:10.1186/gb-2011-12-6-117

Cite this article as: Worden AZ, *et al.*: Genomes of uncultured eukaryotes: sorting FACS from fiction. *Genome Biology* 2011, **12**:117.