MICEMICO), M. Théry by the INCA (PLBIO-2011-141).

References
1. Keller, P.J., Schmidt, A.D., Wittbrodt, J., and Stelzer, E.H.K. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science *322*, 1065–1069.
2. Rembold, M., Loosli, F., Adams, R.J., and Wittbrodt, J. (2006). Individual cell migration serves as the driving force for optic vesicle evagination. Science *313*, 1130–1134.
3. Faure-André, G., Vargas, P., Yuseff, M.-I., Heuzé, M., Diaz, J., Lankar, D., Steri, V., Manry, J., Hugues, S., and Vascotto, F., *et al.* (2008). Regulation of dendritic cell migration by CD74, the MHC class II-associated invariant chain. Science *322*, 1705–1710.
4. Gligorijevic, B., Wyckoff, J., Yamaguchi, H., Wang, Y., Roussos, E.T., and Condeelis, J. (2012). N-WASP-mediated invadopodium formation is involved in intravasation and lung metastasis of mammary tumors. J. Cell Sci. *125*, 724–734.
5. Pouthas, F., Girard, P., Lecaudey, V., Ly, T.B.N., Gilmour, D., Boulin, C., Pepperkok, R., and Reynaud, E.G. (2008). In migrating cells, the Golgi complex and the position of the centrosome depend on geometrical constraints of the substratum. J. Cell Sci. *121*, 2406–2414.
6. Doyle, A.D., Wang, F.W., Matsumoto, K., and Yamada, K.M. (2009). One-dimensional topography underlies three-dimensional fibrillar cell migration. J. Cell Biol. *184*, 481–490.
7. Codling, E.A, Plank, M.J., and Benhamou, S. (2008). Random walk models in biology. J. R. Soc. Interface *5*, 813–834.

[1]Institut Curie, CNRS, UMR144, 75248, Paris, France. [2]Institut de Recherches en Sciences et Technologies pour le Vivant, CEA, UJF, CNRS, INRA, 38054, Grenoble, France. [3]Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, CA 94158, USA. [4]Harvard Medical School, Systems Biology, Boston, MA 02115, USA. [5]Kings College London, SE1 1UL London, UK. [6]BioQuant, Heidelberg University, Germany. [7]Institute of Molecular and Cell Biology, A*Star, Proteos, 138673, Singapore. [8]World Wide, listed in Supplemental Information. [9]Institut Curie, CNRS, UMR168, 75248, Paris, France. [10]Institut Curie, INSERM, U639, 75248, Paris, France.
*E-mail: matthieu.piel@curie.fr, manuel.thery@cea.fr

Global distribution of a wild alga revealed by targeted metagenomics

Alexandra Z. Worden[1,*],
Jan Janouskovec[2], Darcy McRose[1],
Augustin Engman[1], Rory M. Welsh[1],
Stephanie Malfatti[3],
Susannah G. Tringe[3],
and Patrick J. Keeling[2]

Eukaryotic phytoplankton play key roles in atmospheric CO_2 uptake and sequestration in marine environments [1,2]. Community shifts attributed to climate change have already been reported in the Arctic ocean, where tiny, photosynthetic picoeukaryotes (≤3 μm diameter) have increased, while larger taxa have decreased [3]. Unfortunately, for vast regions of the world's oceans, little is known about distributions of different genera and levels of genetic variation between ocean basins. This lack of baseline information makes it impossible to assess the impacts of environmental change on phytoplankton diversity, and global carbon cycling. A major knowledge impediment is that these organisms are highly diverse, and most remain uncultured [2]. Metagenomics avoids the culturing step and provides insights into genes present in the environment without some of the biases associated with conventional molecular survey methods. However, connecting metagenomic sequences to the organisms containing them is challenging. For many unicellular eukaryotes the reference genomes needed to make this connection are not available. We circumvented this problem using at-sea fluorescence activated cell sorting (FACS) to separate abundant natural populations of photosynthetic eukaryotes and sequence their DNA, generating reference genome information while eliminating the need for culturing [2]. Here, we present the complete chloroplast genome from an Atlantic picoeukaryote population and discoveries it enabled on the evolution, distribution, and potential carbon sequestration role of a tiny, wild alga.

We assembled a complete chloroplast genome from a coherent picoeukaryote population sorted from the Gulf Stream Current. The sorting step reduced bioinformatic complexity to a level where high quality *de novo* sequence assembly was possible. The resulting circular plastid genome was 91,306 bp, 35% G+C and encoded 106 proteins, 27 tRNAs, an rRNA operon as well as other features (Figure S1 in Supplemental Information, published with this article online). Multiple lines of evidence demonstrate this genome is from a member of the Pelagophyceae, a recently discovered phytoplankton class [4]. Complete plastid genomes are available from two cultured Pelagophyceae, the brown-tide forming *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. Genome organization in the uncultured pelagophyte was similar to *Aureococcus*, and more divergent from *Aureoumbra* (Figure 1A), consistent with evolutionary relationships deduced from our 105 plastid-protein phylogeny (Figure S2). The uncultured pelagophyte encoded all *Aureococcus* genes plus one (*ycf45*) of *Aureoumbra*'s five additional proteins. All three Pelagophyceae encoded a 267 residue protein with multiple predicted transmembrane domains not seen in any other organisms based on tblastn and blastp against the full GenBank repository. Comparisons with the best-sampled protein-encoding pelagophyte plastid gene, *rbcL*, showed the uncultured population was most similar to *Pelagomonas calceolata* (99.0% nucleotide identity, ≤95.2% to other Pelagophyceae). The 16S rRNA gene, which is highly conserved across genera, had 100% identity to partial sequences available for *P. calceolata*. The uncultured population may therefore be *P. calceolata*, but based on extant sampling we call it 'wild *Pelagomonas*'.

With the plastid genome in hand, we addressed the distribution and ecological significance of this lineage. The complete set of coding regions from the chloroplast genome was compared with marine metagenomic samples using a cutoff of 97.0% nucleotide identity. We found that the wild *Pelagomonas*

A Wild *Pelagomonas*

B

Chlorophyll (mg m^{-3})
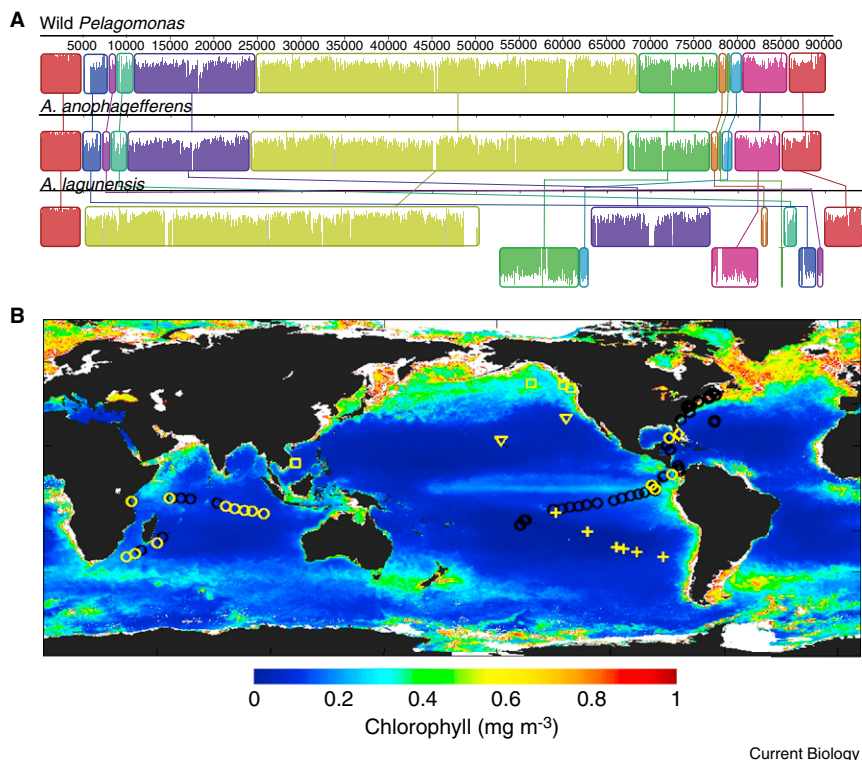
Current Biology

Figure 1. The wild *Pelagomonas* and comparisons to cultured taxa.
(A) Locally collinear genome blocks show sequence divergence between the wild *Pelagomonas* plastid genome as compared to the only sequences available from cultured Pelagophyceae, i.e. *Aureococcus* and *Aureoumbra*. Colored blocks represent homologous sections between genomes (with regions on opposing strands shown inverted). The height of colored shading is proportional to sequence identity in the respective block. (B) Locations where sequences from the wild *Pelagomonas* were detected in GOS (yellow circles), Station ALOHA (yellow triangle) and Station 67-155 (yellow triangle) metagenomes. Black circles represent GOS samples where sequences were not detected. Note that GOS sampling involved pre-filtration through a 0.8 µm pore-size filter (except a few Sargasso Sea samples) [5]. While we detected eukaryotic sequences in the survey data, the filtration procedure minimized representation of eukaryotes and likely influenced taxon recovery differentially, depending on respective cell sizes and fragility, making comparisons between eukaryotic taxa unreliable. Yellow squares represent 16S rRNA gene data previously attributed to bacteria. Our data also exhibited 99–100% identity to partial rRNA sequences (yellow cross) from the southeast Pacific generated using multiple PCR primers, one of which frequently recovered pelagophyte sequences [7]. Location of FACS sort from 75 m is shown with a yellow diamond.

was broadly distributed in surface metagenomic data from the Global Ocean Survey (GOS) [5] (Figure 1B), an expedition that set out to circumnavigate the globe. By contrast, neither *Aureoumbra* nor the model pelagophyte *Aureococcus* was detected in these data [5] using the same search criteria. Surface and deep chlorophyll maximum (DCM) metagenomic data from the central North Pacific (NP) Gyre also contained sequences from the wild *Pelagomonas*. Likewise, sequences from the wild *Pelagomonas* were present at the DCM in the eastern NP Gyre (Figure 1B), at a depth similar to the Gulf Stream DCM (75 m) where the sort was performed. Our analyses

suggest future studies should incorporate balanced sampling of surface and deep sunlit waters, where eukaryotic phytoplankton are most abundant [2,6]

Notably, the wild *Pelagomonas* was found in the majority of GOS Indian Ocean samples, a region where data on pelagophytes conflict. Pigment analyses initially indicated pelagophytes were abundant in the Indian Ocean, but few pelagophyte 18S rDNA sequences were retrieved there [6]. Pigment data are often considered indeterminate because interpretation can be confounded by the presence of similar marker pigments in multiple algal groups [2,6]. However, PCR-clone library studies

also produce conflicting information on phytoplankton community composition due to primer biases and other issues [7]. Fluorescence *in situ* hybridization (FISH) studies appear to corroborate pigment-based observations. For example, FISH results show pelagophytes as a whole contribute significantly to picophytoplankton biomass in the tropical Atlantic Ocean at 5 and 20 m (deeper samples were not analyzed) [8]. Thus, the paucity of pelagophyte 18S rDNA clone library sequences in the Indian Ocean [6] may reflect biases circumvented by our metagenomic approach. Pigment analyses do not have the power to resolve which genera are present, although they indicate that in general Pelagophyceae are important marine primary producers [9] and their contributions appear to be increasing, at least in the North Atlantic [10]. The chloroplast genome sequenced here provides unprecedented taxon resolution and shows that cells with high genetic identity to the FACS-sorted population are distributed across several oceans, suggesting considerable ecological importance.

How important might this wild population be to export of CO$_2$ to the deep ocean? We discovered sequences highly similar to the wild *Pelagomonas* in subarctic Pacific Ocean surface (e.g., accession HQ671892, 99.9% identity) and deep sea sediment (e.g., DQ513100, 99.8% identity) data previously categorized as bacterial. Presence in sediments indicates that the wild *Pelagomonas* may be performing a key ecosystem service by transporting atmospheric carbon it sequesters by photosynthesis to the deep ocean. The data provided here will enable focused studies on the role of pelagophytes in carbon cycling. Our results demonstrate the value of targeted population metagenomics for discovering ecologically relevant taxa and the importance of the wild pelagophyte across multiple oceans.

Accession Numbers
The plastid genome and annotation have been deposited under GenBank accession JX297813.

Supplemental Information
Supplemental information includes two figures and experimental procedures and

can be found with this article online at http://dx.doi.org/10.1016/j.cub.2012.07.054.

References

1. Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. Science *281*, 237–240.
2. Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messié, M., Tringe, S.G., Woyke, T., Welsh, R., Ishoey, T., Lee, J.H., *et al*. (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. Proc. Natl. Acad. Sci. USA *107*, 14679–14684.
3. Li, W.K.W., McLaughlin, F.A., Lovejoy, C., and Carmack, E.C. (2009). Smallest algae thrive as the Arctic Ocean freshens. Science *326*, 539.
4. Andersen, R.A., Sounders, G.W., Paskind, M.P., and Sexton, J.P. (1993). Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the discription of a new algal class, the Pelagophyceae classis nov. J. Phycol. *29*, 701–715.
5. Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., *et al*. (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. *5*, e77.
6. Not, F., Latasa, M., Scharek, R., Viprey, M., Karleskind, P., Balague, V., Ontoria-Oviedo, I., Cumino, A., Goetze, E., Vaulot, D., *et al*. (2008). Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. Deep Sea Res. Part I Oceanogr. Res. Pap. *55*, 1456–1473.
7. Shi, X.L., Lepere, C., Scanlan, D.J., and Vaulot, D. (2011). Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. PLoS ONE *6*, e18979.
8. Jardillier, L., Zubkov, M.V., Pearman, J., and Scanlan, D.J. (2010). Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J. *4*, 1180–1192.
9. DiTullio, G.R., Geesey, M.E., Jones, D.R., Daly, K.L., Campbell, L., and Smith, W.O. (2003). Phytoplankton assemblage structure and primary productivity along 170°W in the South Pacific Ocean. Mar. Ecol. Prog. Ser. *255*, 55–80.
10. Lomas, M.W., Steinberg, D.K., Dickey, T., Carlson, C.A., Nelson, N.B., Condon, R.H., and Bates, N.R. (2010). Increased ocean carbon export in the Sargasso Sea linked to climate variability is countered by its enhanced mesopelagic attenuation. Biogeosciences *7*, 57–70.

[1]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95060, USA. [2]Department of Botany, University of British Columbia, Vancouver BC V6T 1Z4, Canada. [3]DOE Joint Genome Institute, Walnut Creek, CA 94598, USA.
[*]E-mail: azworden@mbari.org