

BioSpace25 - Biodiversity insight from Space  
10 - 14 February 2025 | ESA-ESRIN | Frascati - Italy

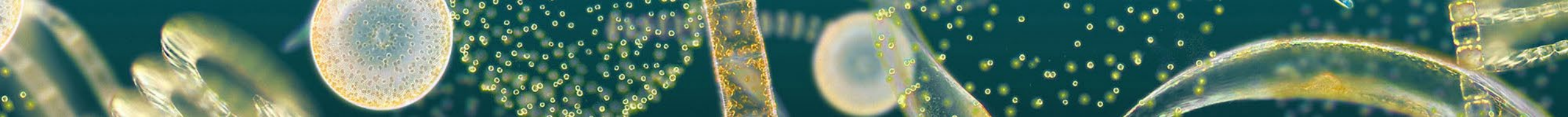


# Linking satellites to genes to estimate phytoplankton community structure from space

Roy El Hourany<sup>1</sup>, Juan Pierella Karlusich<sup>2,3</sup>, Pedro Junger<sup>2</sup>, Lucie Zinger<sup>3</sup>, Hubert Loisel<sup>1</sup>, Marina Levy<sup>4</sup>, and Chris Bowler<sup>3</sup>

<sup>1</sup>LOG, <sup>2</sup>FAS MIT, <sup>3</sup>IBENS, <sup>4</sup>LOCEAN





# Phytoplankton Community structure

## Taxonomic groups

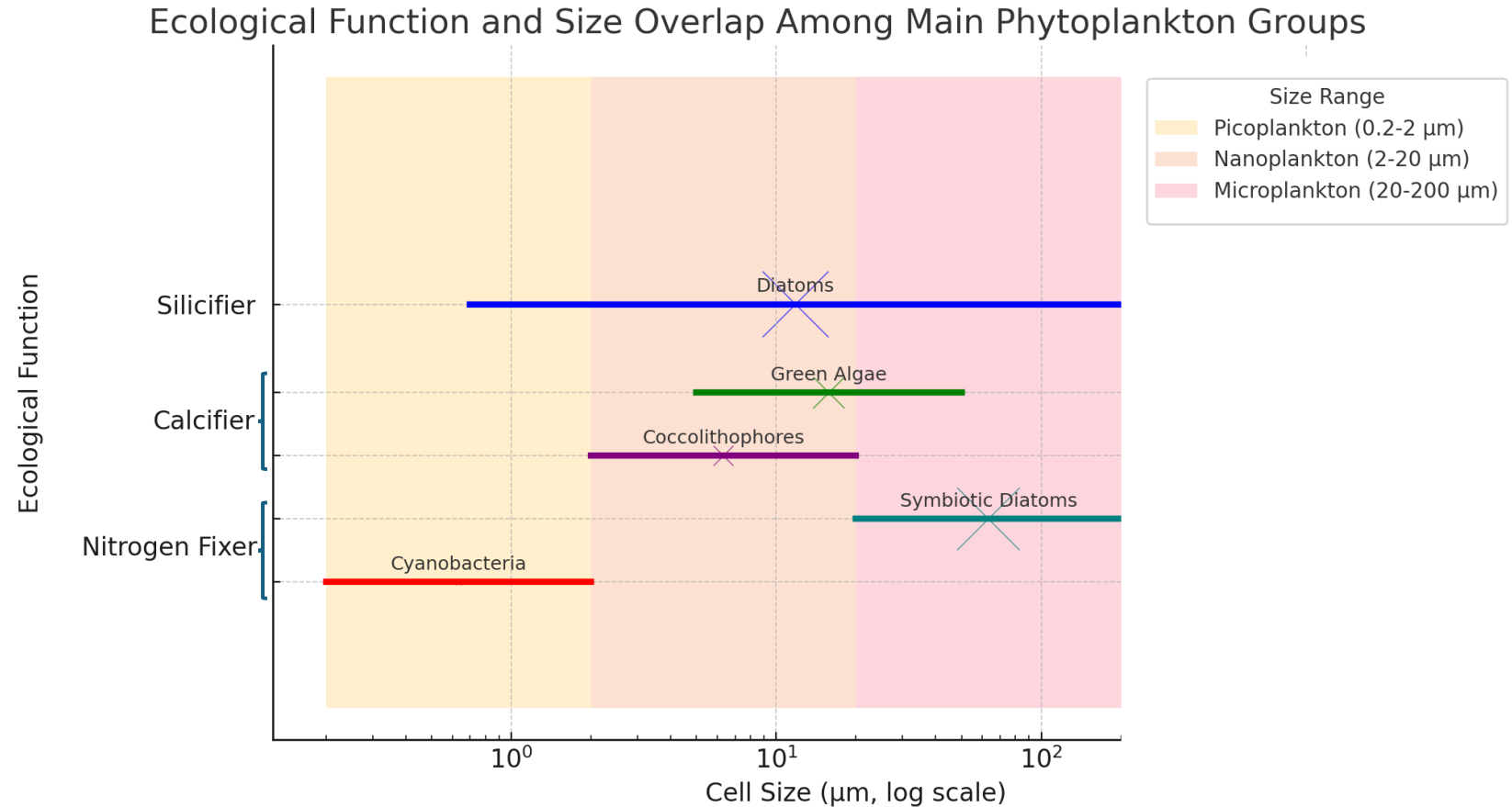
- Diatoms
- Dinoflagellates
- Green Algae
- Haptophytes
- Cyanobacteria...

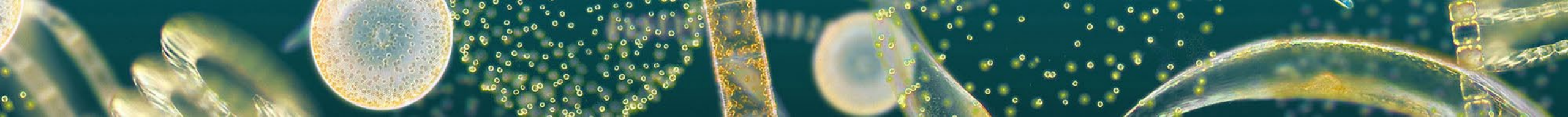
## Ecosystem functions

- Silicifiers
- Calcifiers
- N<sub>2</sub> fixer etc...

## Size ranges

- Micro-phytoplankton 20 - 200  $\mu\text{m}$
- Nano-phytoplankton 2 - 20  $\mu\text{m}$
- Pico-phytoplankton 0.2 - 2  $\mu\text{m}$

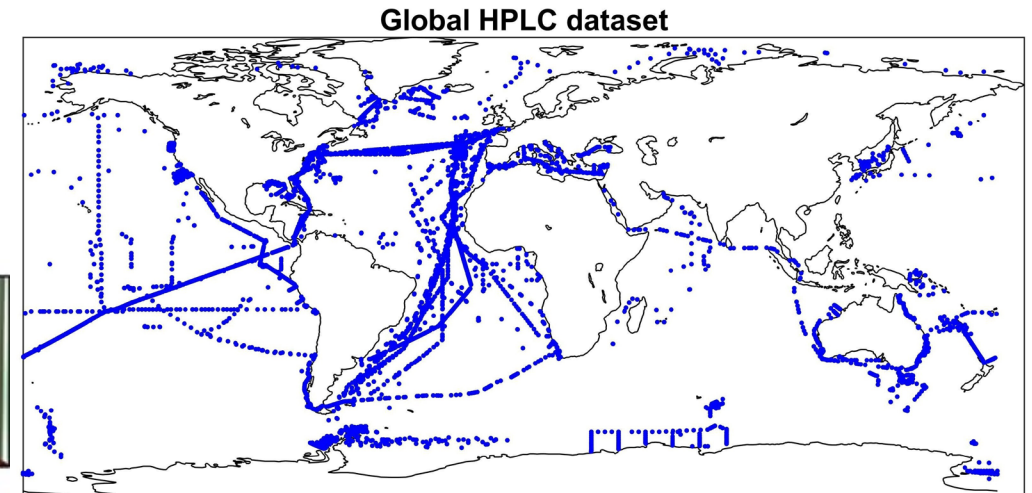
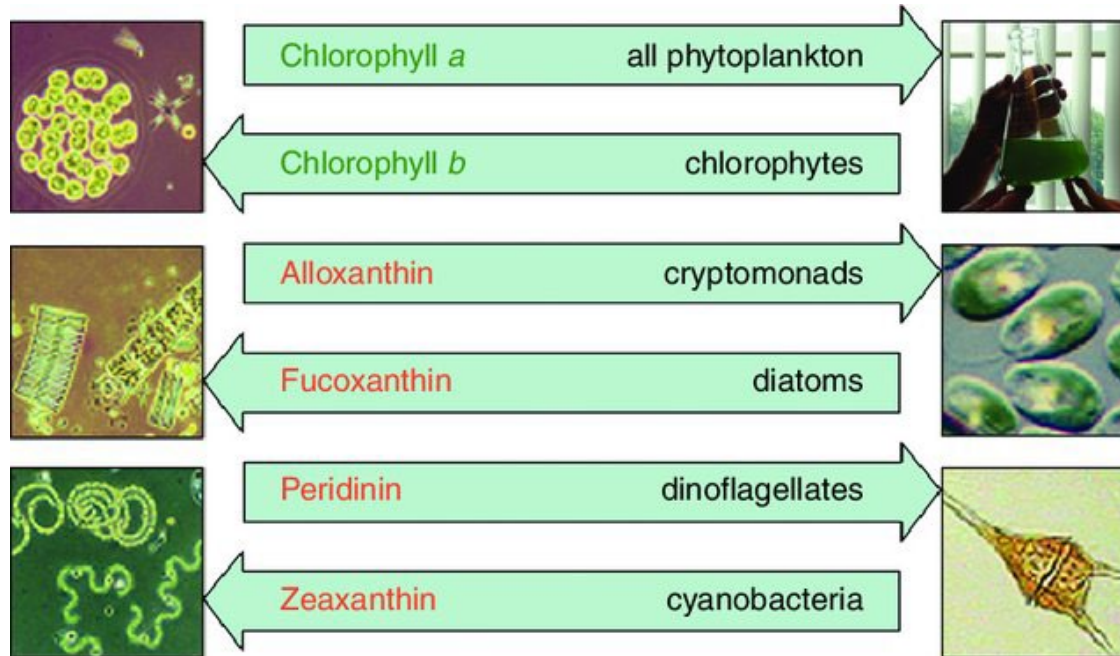




# Phytoplankton community structure from space:

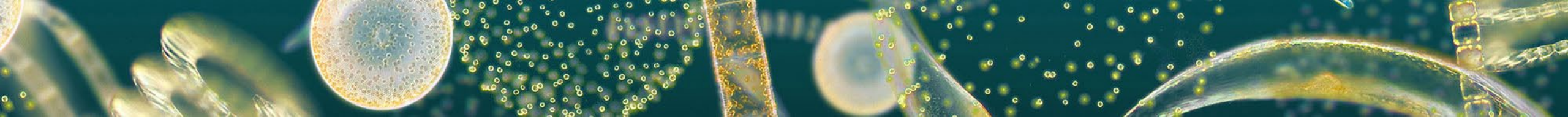
## Integration of diagnostic pigment concept in ocean color studies

Over 17000 obs compiled HPLC dataset of several pigments measured at the surface (<30m), (El Hourany et al., 2019, 2024)

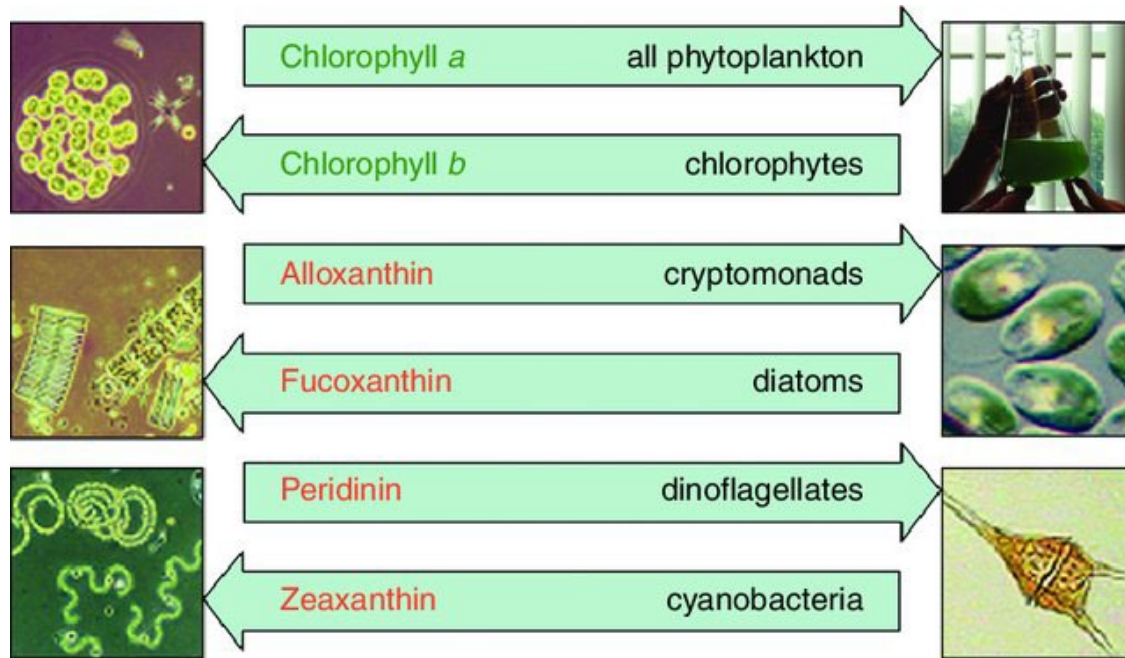


## A methodology covering Prok to Euk kingdoms

Examples of the major phytoplankton groups and their diagnostic photopigments. From Paerl, H.W., Valdes, L.M., Pinckney, J.L., Piehler, M.F., Dyble, J., Moisaner, P.H., 2003. Phytoplankton photopigments as indicators of estuarine and coastal eutrophication. *BioScience* 53 (10), 953-964.



# Biased information on phytoplankton groups due to overlapping pigments or to the physiological state



Examples of the major phytoplankton groups and their diagnostic photopigments. From Paerl, H.W., Valdes, L.M., Pinckney, J.L., Piehler, M.F., Dyble, J., Moisaner, P.H., 2003. Phytoplankton photopigments as indicators of estuarine and coastal eutrophication. *BioScience* 53 (10), 953-964.

Pigments	Abbreviation	Taxonomic or biogeochemical significance
Chlorophyll-a	Chla	All—except Prochlorophytes
Divinyl Chlorophyll-a	DVChla	<b>Prochlorophytes</b>
Phaeophorbide a	Phda	Grazor fecal pellets, cellular senescence

## Alternative methodology ?

Chlorophyll-b	Chlb	<b>Chlorophytes, Prasinophytes</b>
Divinyl Chlorophyll-b	DVChlb	<b>Prochlorophytes</b>
Total Chlorophyll-b	TChlb	<b>Chlorophytes, Prochlorophytes</b>
Peridinin	Peri	<b>Dinoflagellates</b>
19'-Butanoyloxyfucoxanthin	19'BF	<b>Pelagophytes, prymnesiophytes</b>
Fucoxanthin	Fuco	<b>Diatoms, Prymnesiophytes and some Dinoflagellates</b>
19'-hexanoyloxyfucoxanthin	19'HF	<b>Prymnesiophytes</b>
Zeaxanthin	Zea	<b>Cyanobacteria, Prochlorophytes</b>
Alloxanthin	Allo	Cryptophytes
Neoxanthin	Neo	<b>Chlorophytes, prasinophytes</b>
Violaxanthin	Viola	<b>Chlorophytes, prasinophytes, chrysophytes</b>
Prasinoxanthin	Pras	<b>Prasinophytes</b>

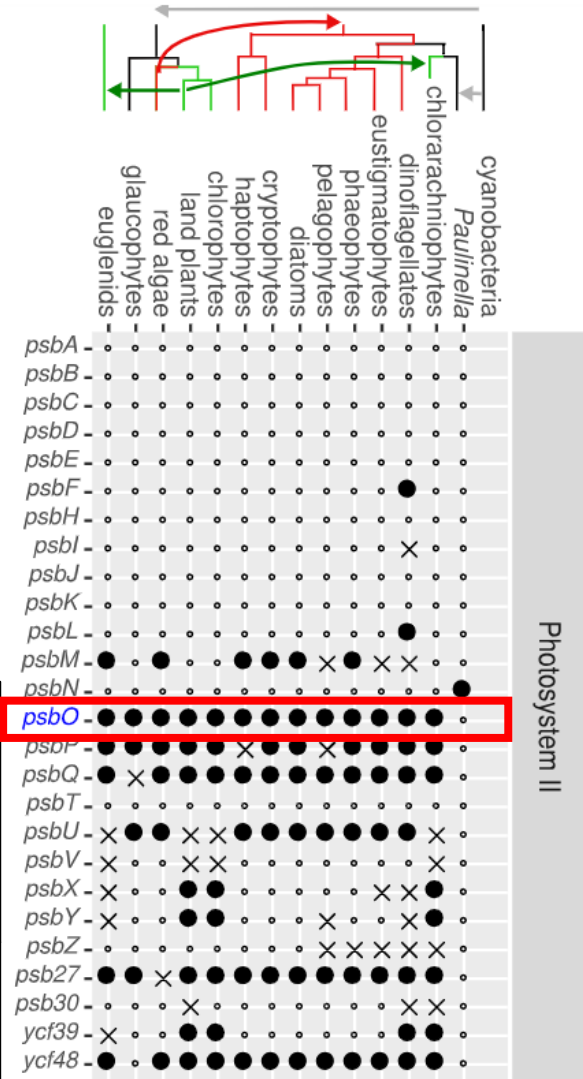
# Alternative methodology covering the entire community structure



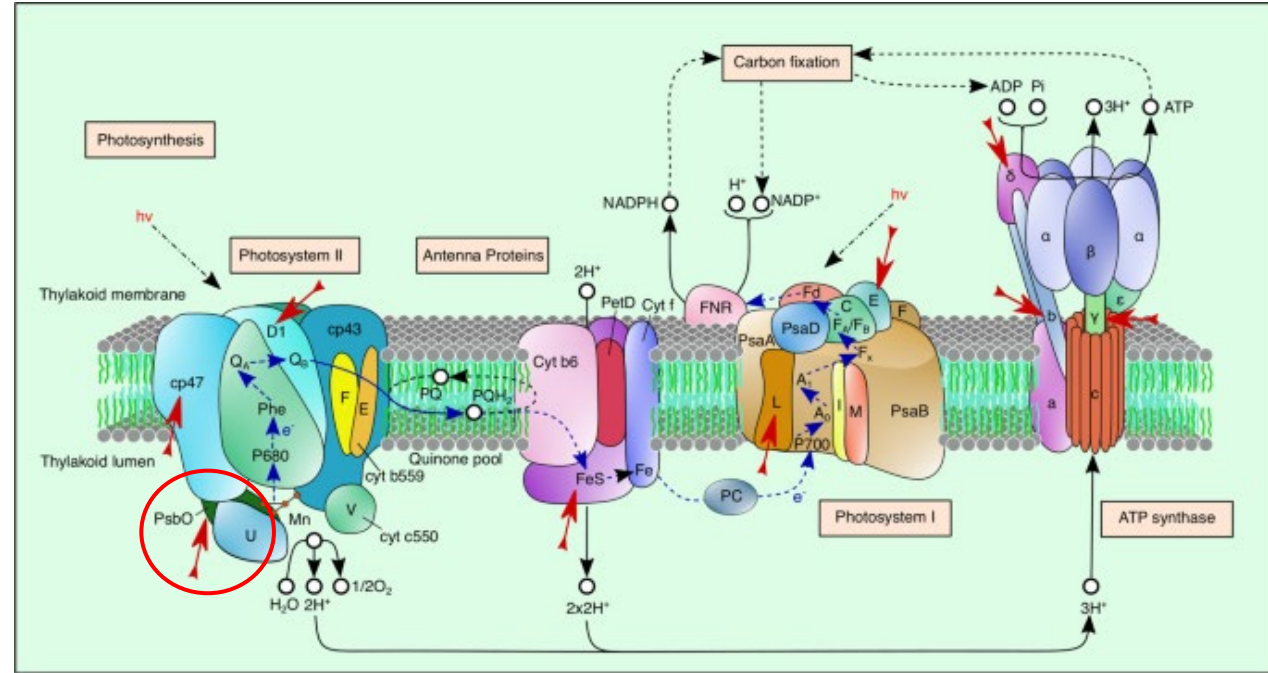
The *psbO* protein; a core subunit of photosystem II (PSII), unique to organisms that carry out oxygenic photosynthesis.

The *psbO* is a **single-copy gene**, present in both eukaryotes and prokaryotes groups

= Proxy of cell abundance



FROM THE COVER  
**A robust approach to estimate relative phytoplankton cell abundances from metagenomes**  
 Juan José Pierella-Karlusch<sup>1,2</sup> | Eric Pelletier<sup>2,3</sup> | Lucie Zingales<sup>1,2</sup> | Fabien Lombard<sup>2,4,5</sup> | Adriana Zingone<sup>6</sup> | Sébastien Colin<sup>2,6</sup> | Josep M. Gasol<sup>7</sup> | Richard G. Dorris<sup>8</sup> | Nicolas Henry<sup>2,10</sup> | Eleonora Scalo<sup>9</sup> | Silvia G. Acinas<sup>11</sup> | Patrick Wincker<sup>2,3</sup> | Colomán de Vargas<sup>2,8</sup> | Chris Bowler<sup>1,2</sup>



Pierella et al., 2022, <https://doi.org/10.1111/1755-0998.13592>

# Alternative methodology covering the entire community structure

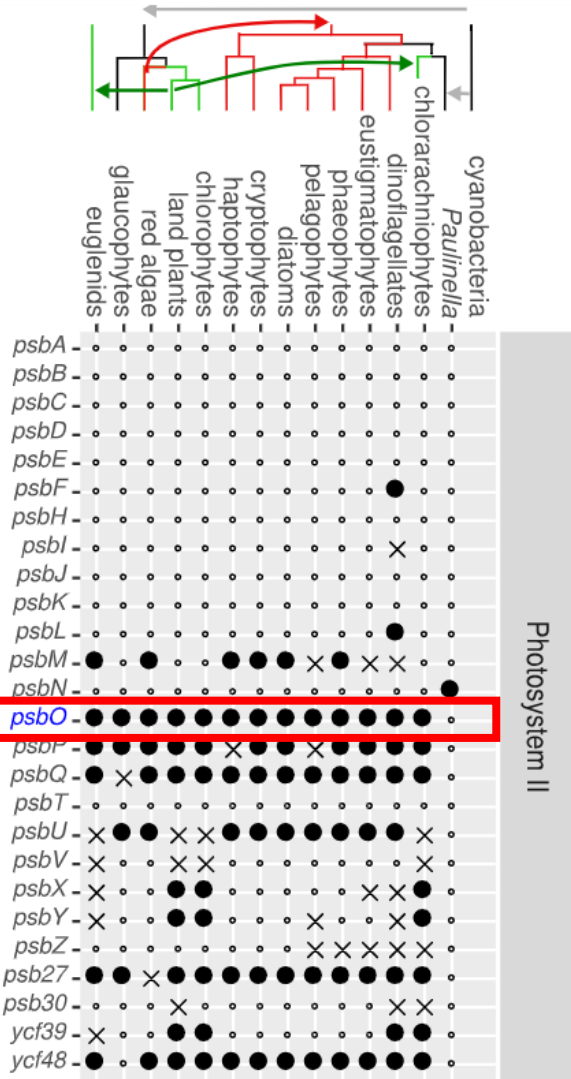


TARA OCEANS

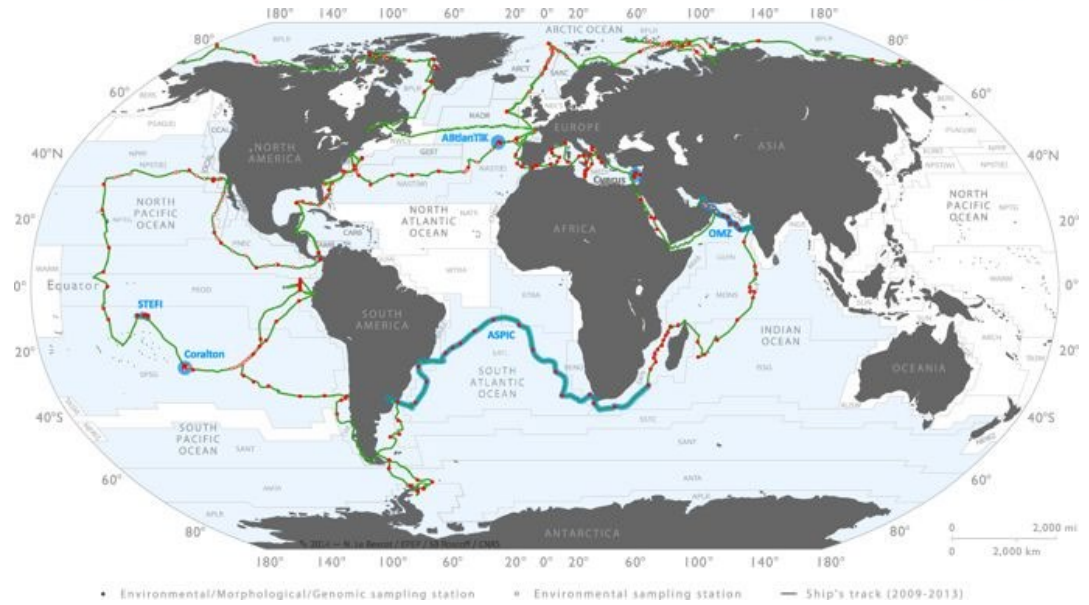
The *psbO* protein; a core subunit of photosystem II (PSII), unique to organisms that carry out oxygenic photosynthesis.

The *psbO* is a **single-copy gene**, present in both eukaryotes and prokaryotes groups

= Proxy of cell abundance



FROM THE COVER  
**A robust approach to estimate relative phytoplankton cell abundances from metagenomes**  
 Juan José Pierella-Karlusch<sup>1,2</sup> | Eric Pelletier<sup>2,3</sup> | Lucie Zingone<sup>1,2</sup> | Fabien Lombard<sup>2,4,5</sup> | Adriana Zingone<sup>6</sup> | Sébastien Colin<sup>7,8</sup> | Josep M. Gasol<sup>9</sup> | Richard G. Dorris<sup>10</sup> | Nicolas Henry<sup>2,10</sup> | Eleonora Scalo<sup>9</sup> | Silvia G. Acinas<sup>11</sup> | Patrick Wincker<sup>2,3</sup> | Colomán de Vargas<sup>2,8</sup> | Chris Bowler<sup>1,2</sup>



Tara Oceans Expedition 2009-2013, 145 stations

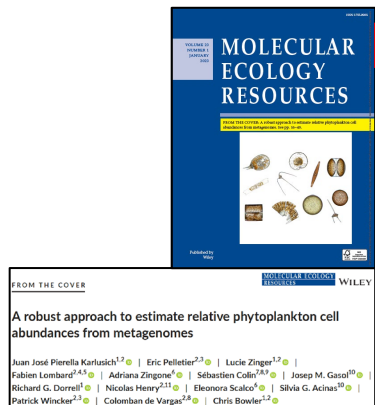
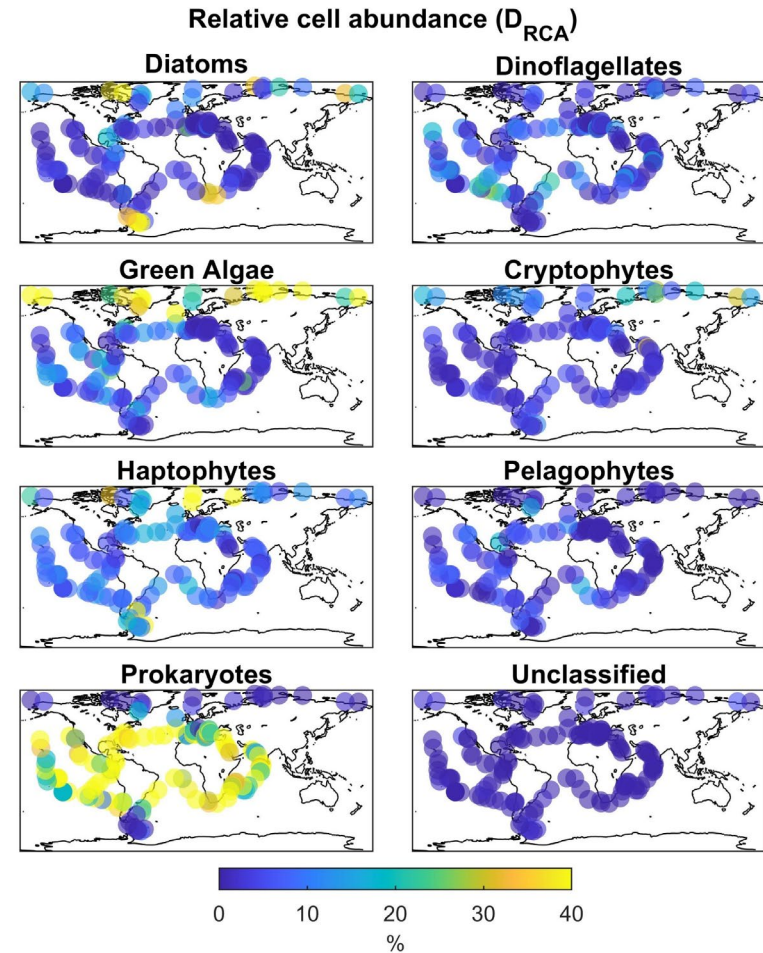
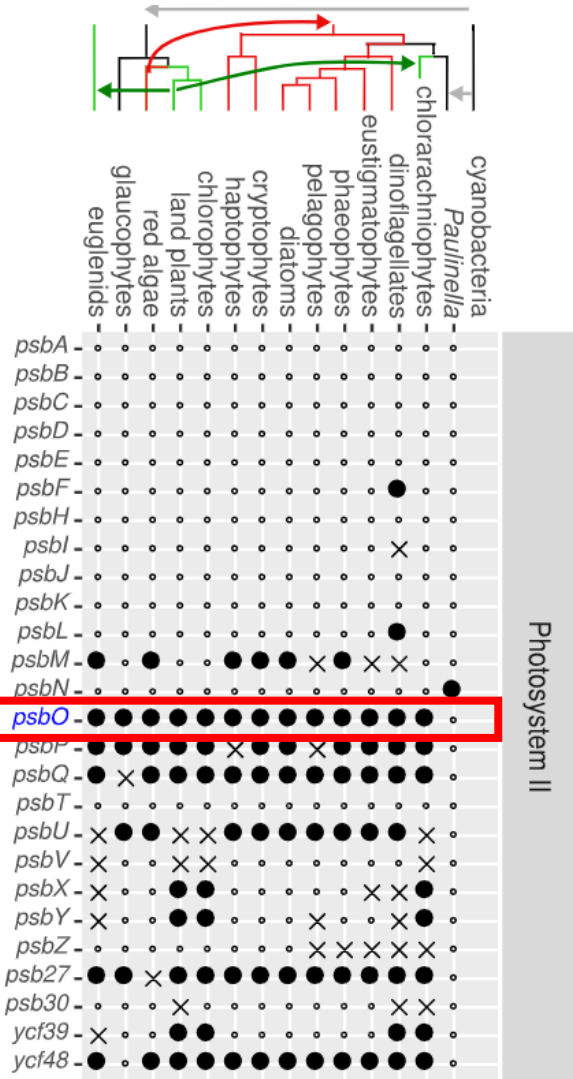
Pierella et al., 2022, <https://doi.org/10.1111/1755-0998.13592>

# Alternative methodology covering the entire community structure

The *psbO* protein; a core subunit of photosystem II (PSII), unique to organisms that carry out oxygenic photosynthesis.

The *psbO* is a **single-copy gene**, present in both eukaryotes and prokaryotes groups

= Proxy of cell abundance



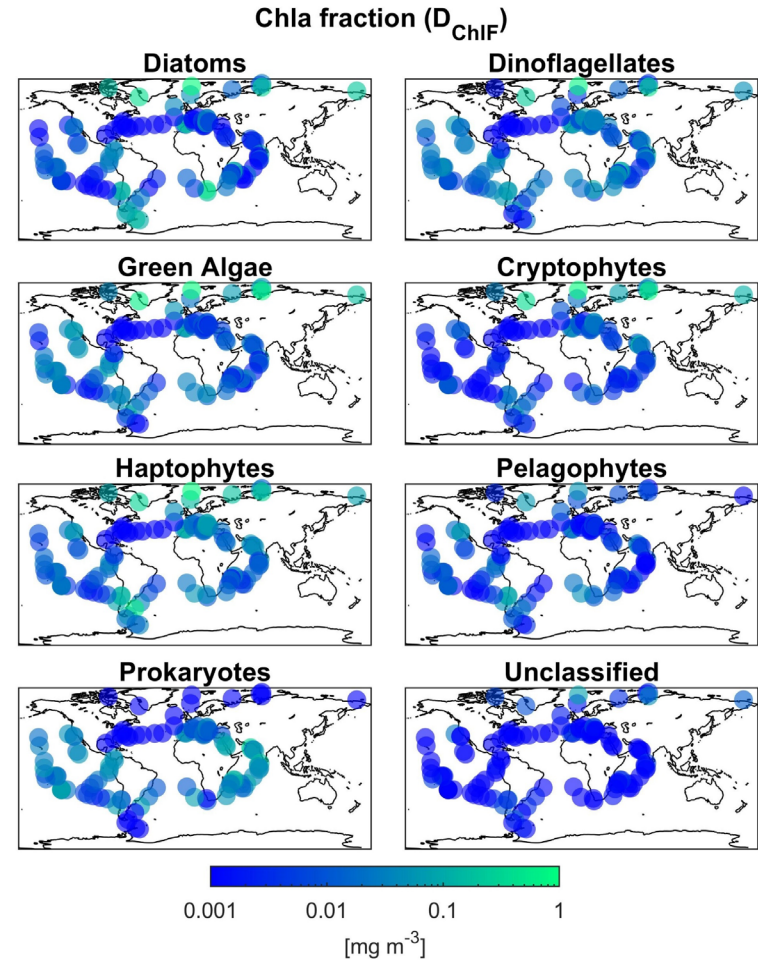
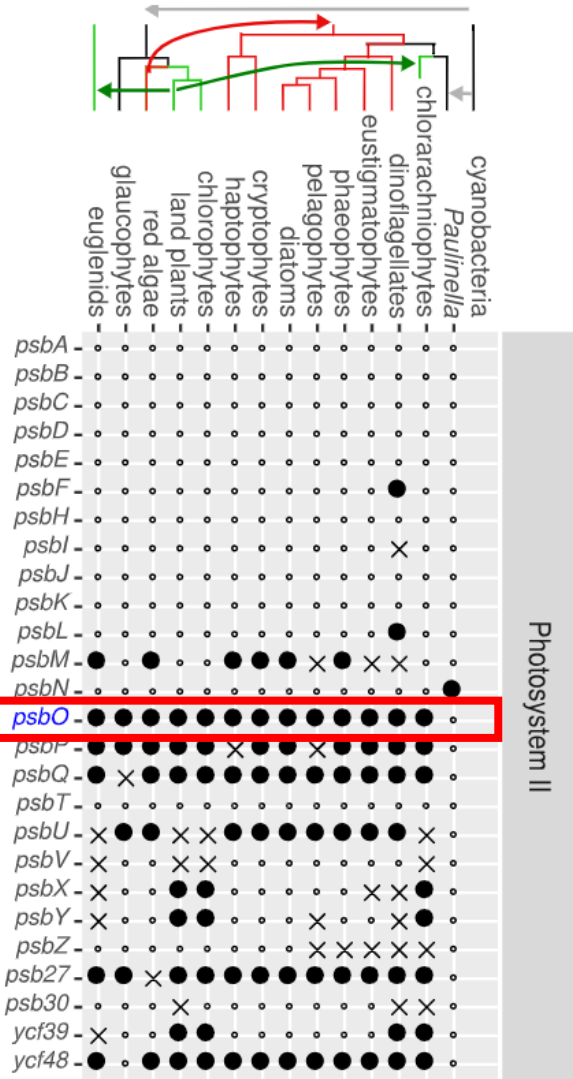
El Hourany, R., 2024, Zenodo. <https://doi.org/10.5281/zenodo.10361485>

# Alternative methodology covering the entire community structure

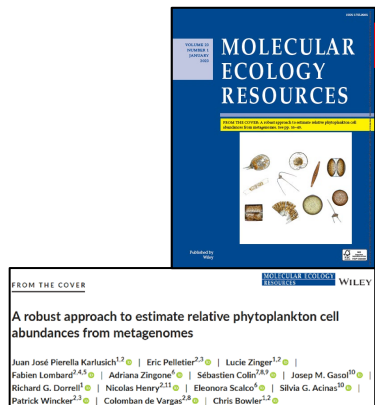
The *psbO* protein; a core subunit of photosystem II (PSII), unique to organisms that carry out oxygenic photosynthesis.

The *psbO* is a **single-copy gene**, present in both eukaryotes and prokaryotes groups

= Proxy of cell abundance

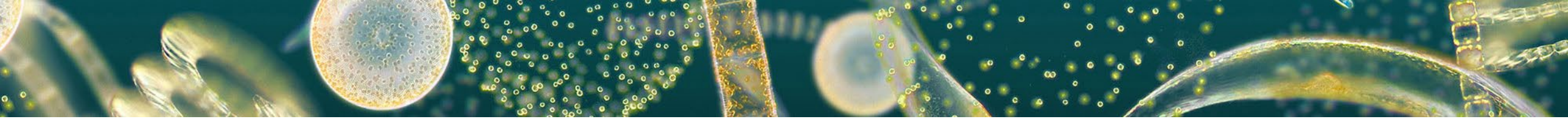


\*integrating size fractions (4 sizes)



El Hourany, R., 2024, Zenodo. <https://doi.org/10.5281/zenodo.10361485>



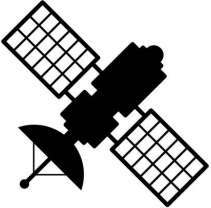


# Statistical modeling of the phytoplankton community structure using satellite and *psbO*

**Phytoplankton Groups**  
Diatom, Dino, Crypto, Hapto,  
Pelago, Green Algae, Cyano

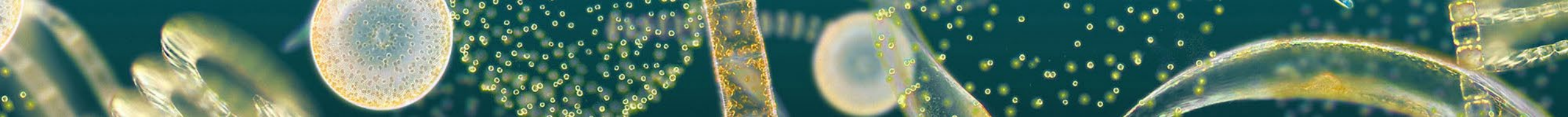


**Satellite matchups (at daily, 4km resolution)**  
(Reflectance at several wavelengths, Chlorophyll-a,  
sea surface temperature, PAR, Fluorescence,  
attenuation coefficient)



145 obs

Seven Phytoplankton groups  
expressed as Chla fraction and  
relative cell abundance

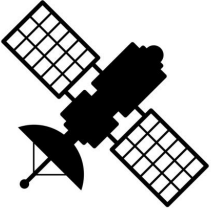


# Statistical modeling of the phytoplankton community structure using satellite and *psbO*

**Phytoplankton Groups**  
Diatom, Dino, Crypto, Hapto,  
Pelago, Green Algae, Cyano



**Satellite matchups (at daily, 4km resolution)**  
(Reflectance at several wavelengths, Chlorophyll-a,  
sea surface temperature, PAR, Fluorescence,  
attenuation coefficient)

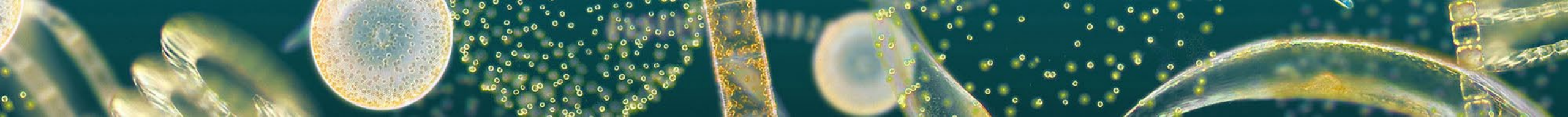


145 obs

Seven Phytoplankton groups  
expressed as Chla fraction and  
relative cell abundance

### Challenges

Low number of observations / matchups  
Multivariate target with coherence to respect

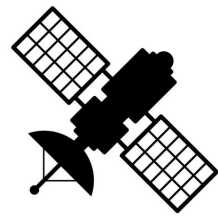


# Statistical modeling of the phytoplankton community structure using satellite and *psbO*

**Phytoplankton Groups**  
Diatom, Dino, Crypto, Hapto, Pelago, Green Algae, Cyano



**Satellite matchups (at daily, 4km resolution)**  
(Reflectance at several wavelengths, Chlorophyll-a, sea surface temperature, PAR, Fluorescence, attenuation coefficient)

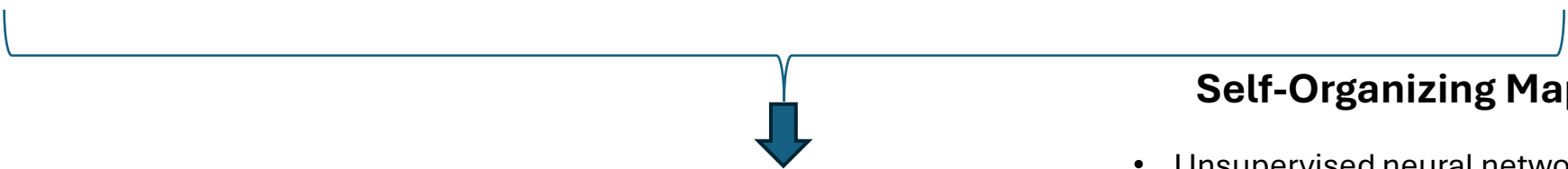


145 obs

Seven Phytoplankton groups expressed as Chla fraction and relative cell abundance

### Challenges

Low number of observations / matchups  
Multivariate target with coherence to respect

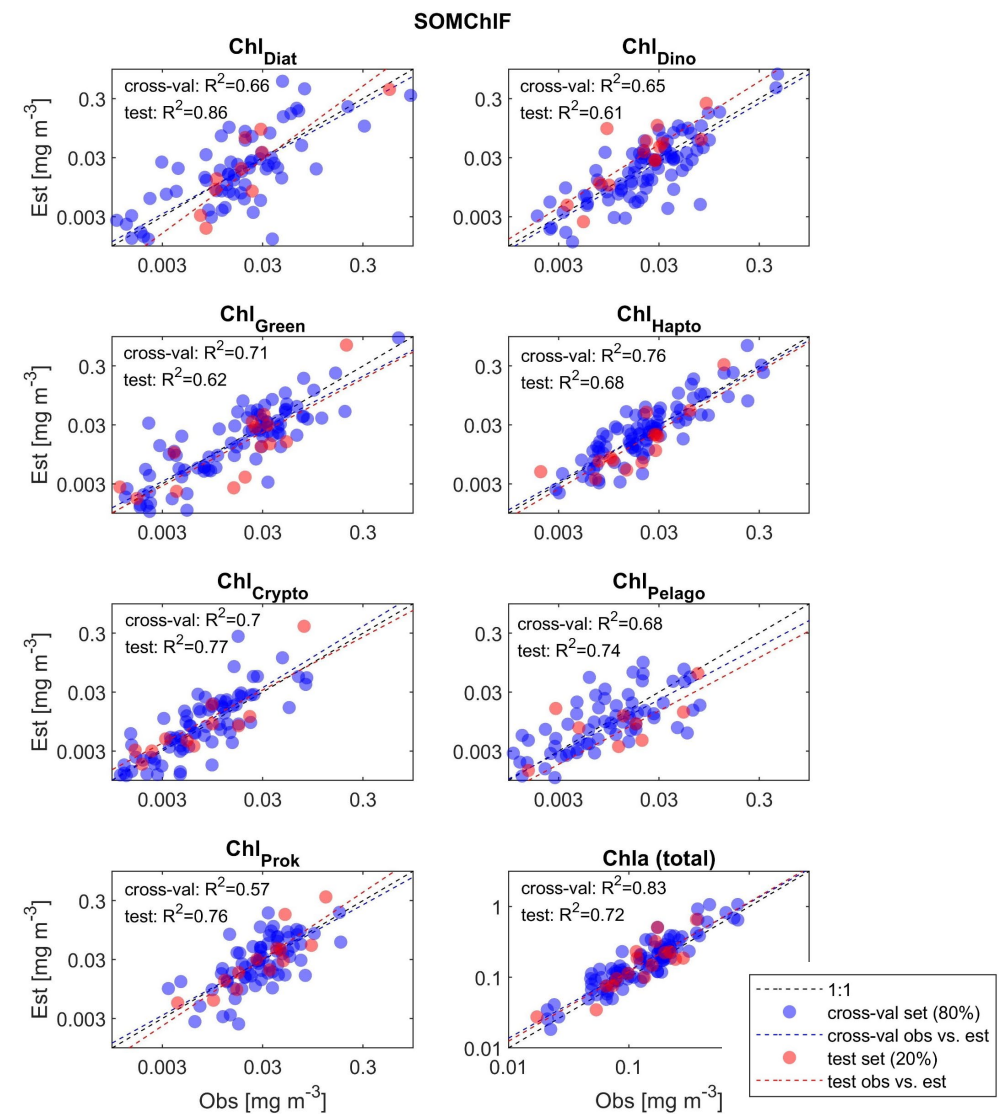
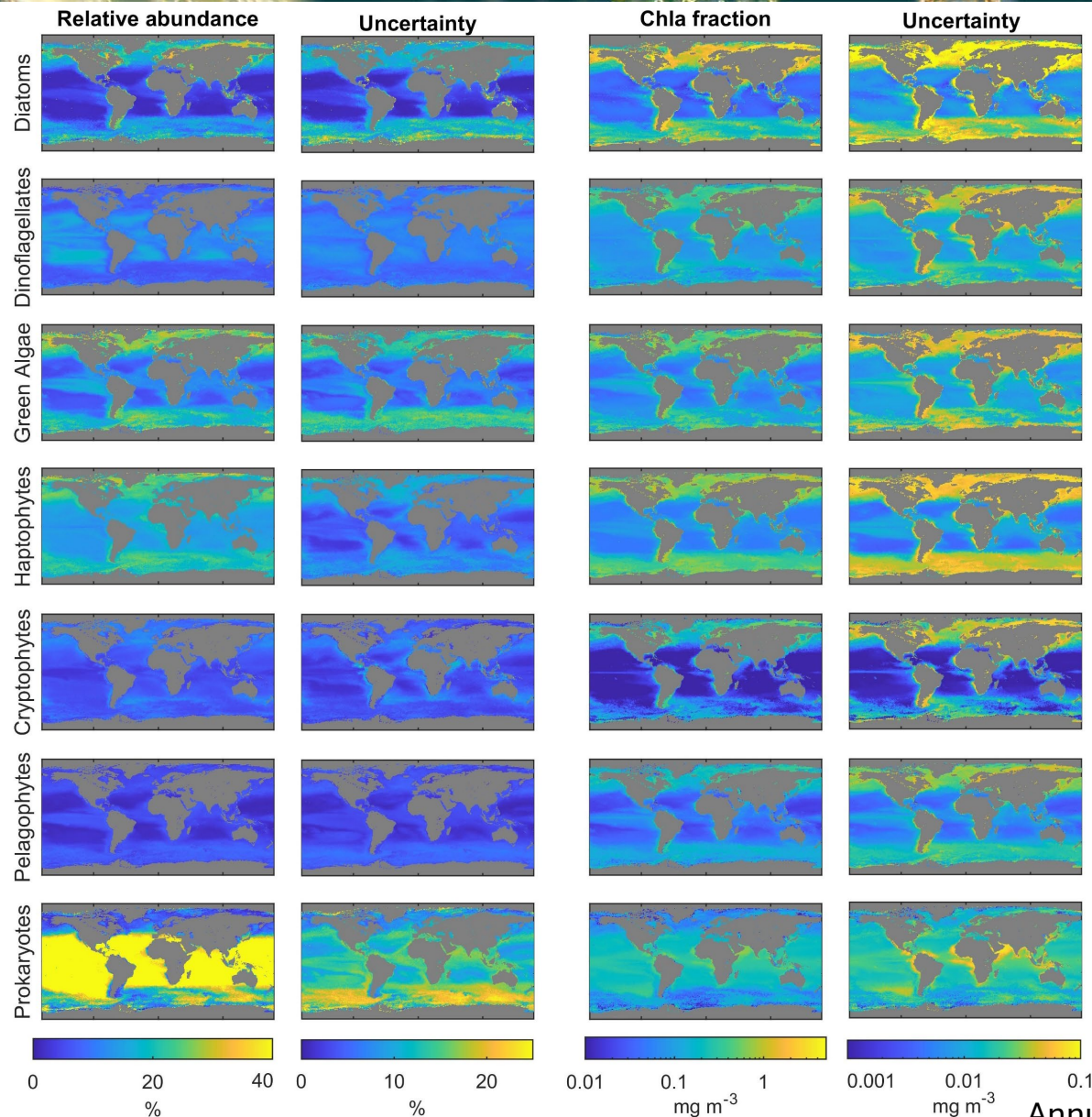


**SOM classification procedure**

### Self-Organizing Maps (SOM)

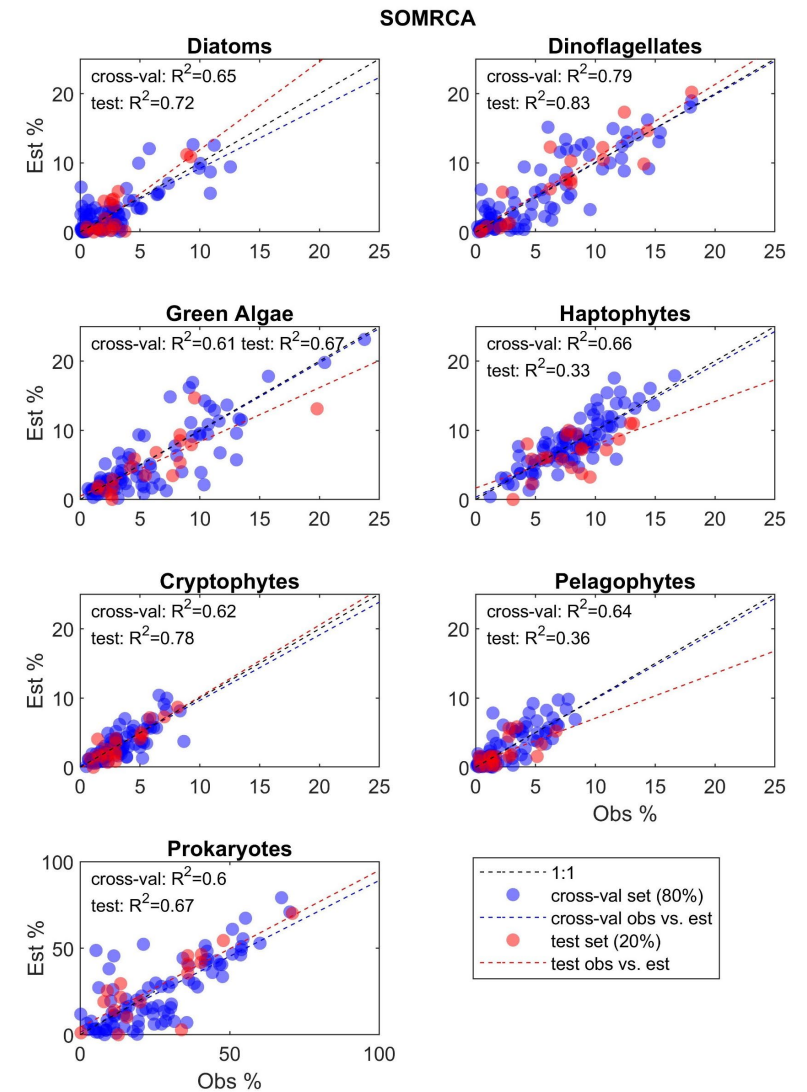
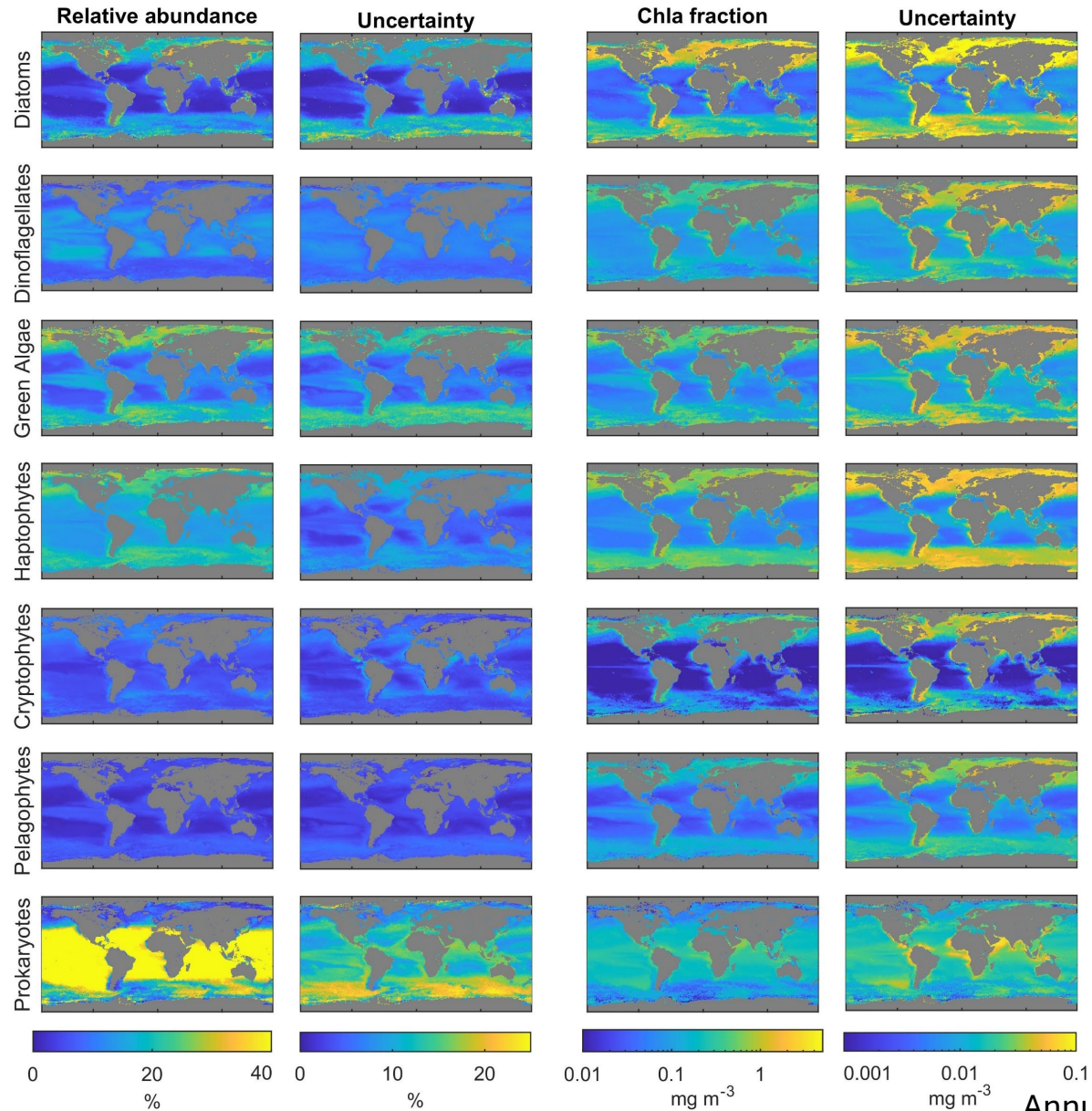
- Unsupervised neural network clustering method
- Allow the projection of a high-dimension database on a 2-D discrete space
- **Efficient interpolation/imputation method due to the multivariate topology conservation**

# From genes to global-scale patterns



Annual composites (1997-2022)

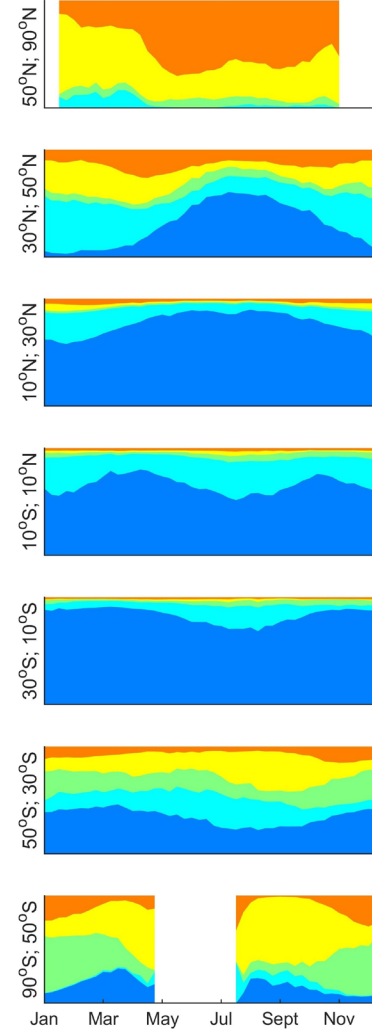
# From genes to global-scale patterns



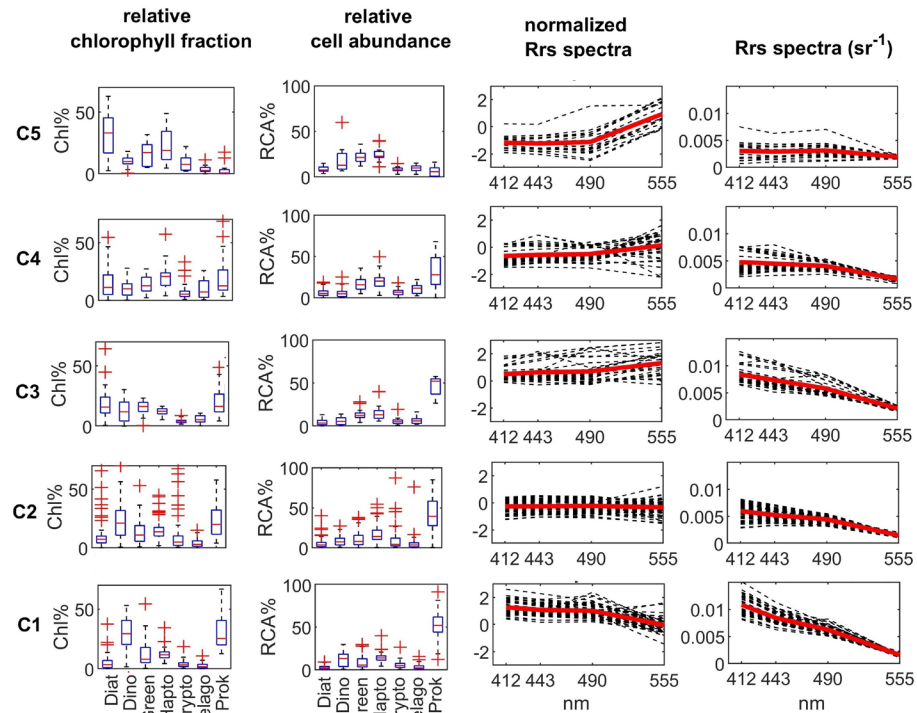
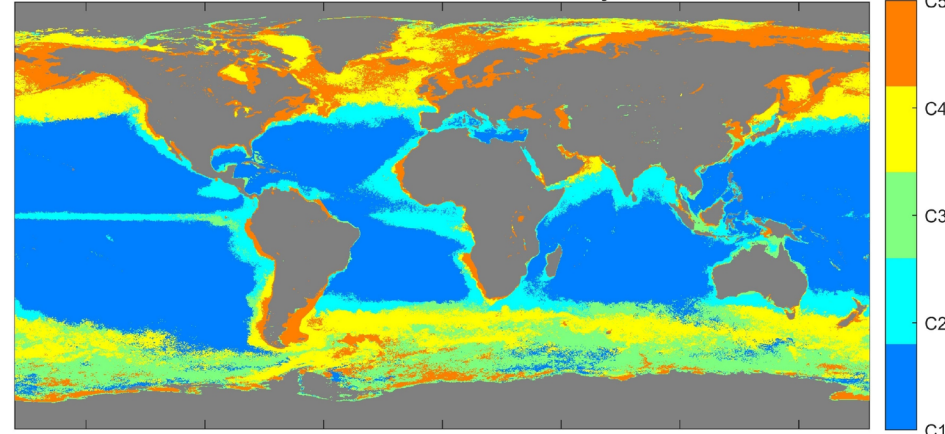
Annual composites (1997-2022)

# Summarizing common pattern between both estimates

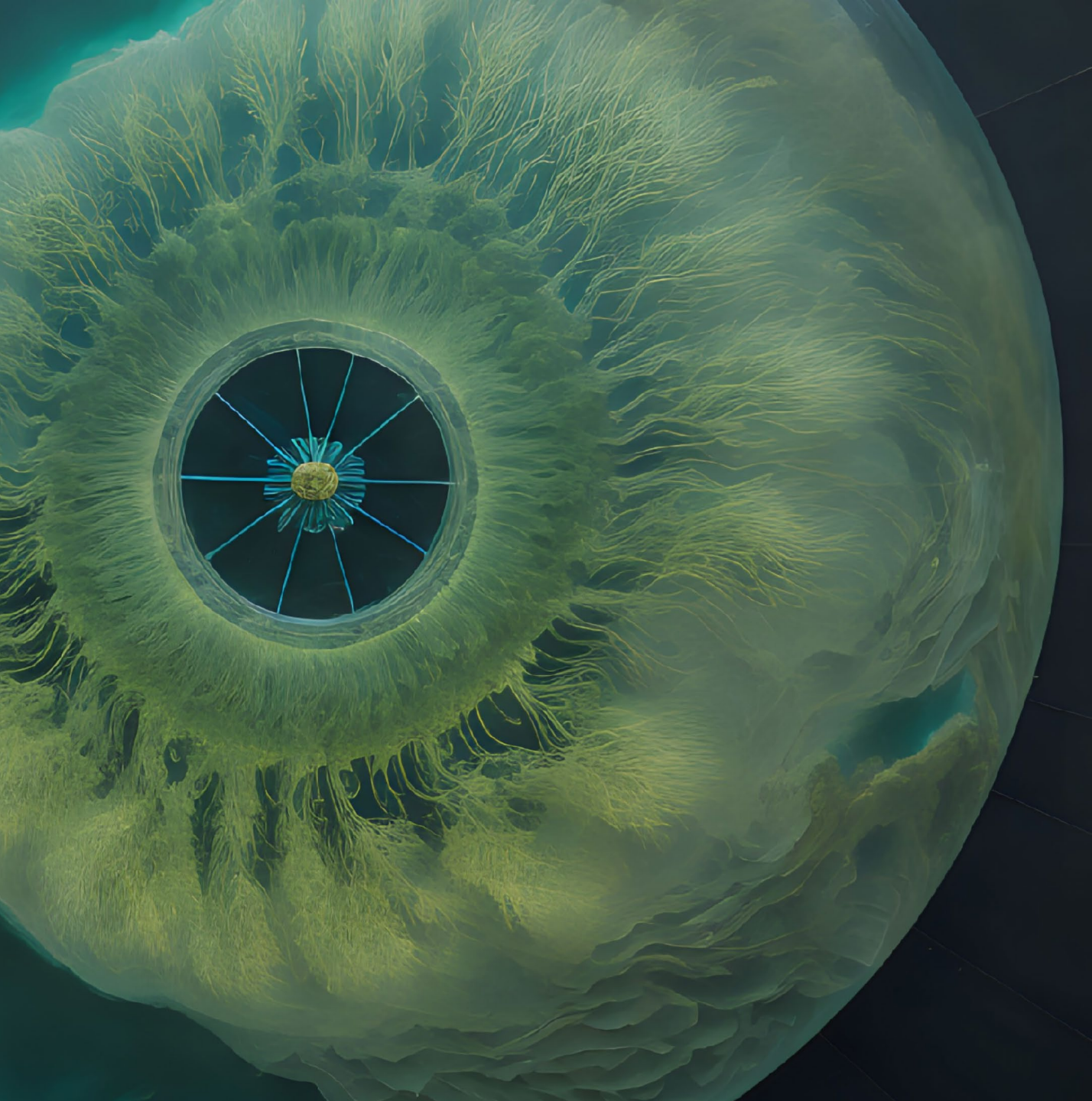
Distribution of the phytoplankton community structures per latitudinal band



Annual distribution of dominant community structures

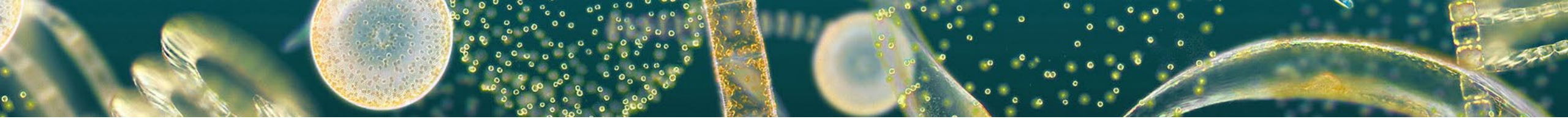


El Hourany, R. et al., Linking satellites to genes with machine learning to estimate phytoplankton community structure from space, *Ocean Sci.*, 20, 217–239, <https://doi.org/10.5194/os-20-217-2024>, 2024. 14

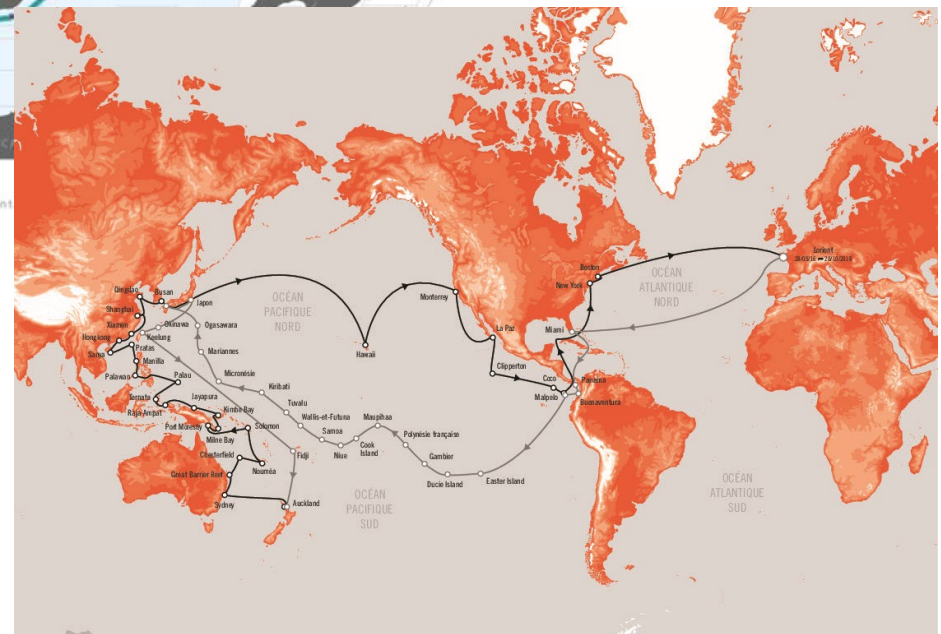
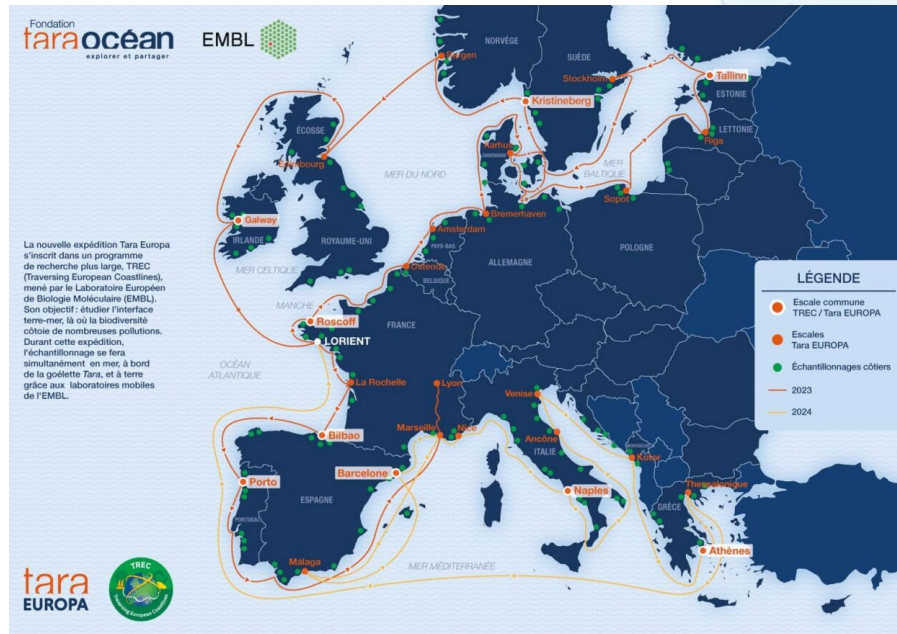
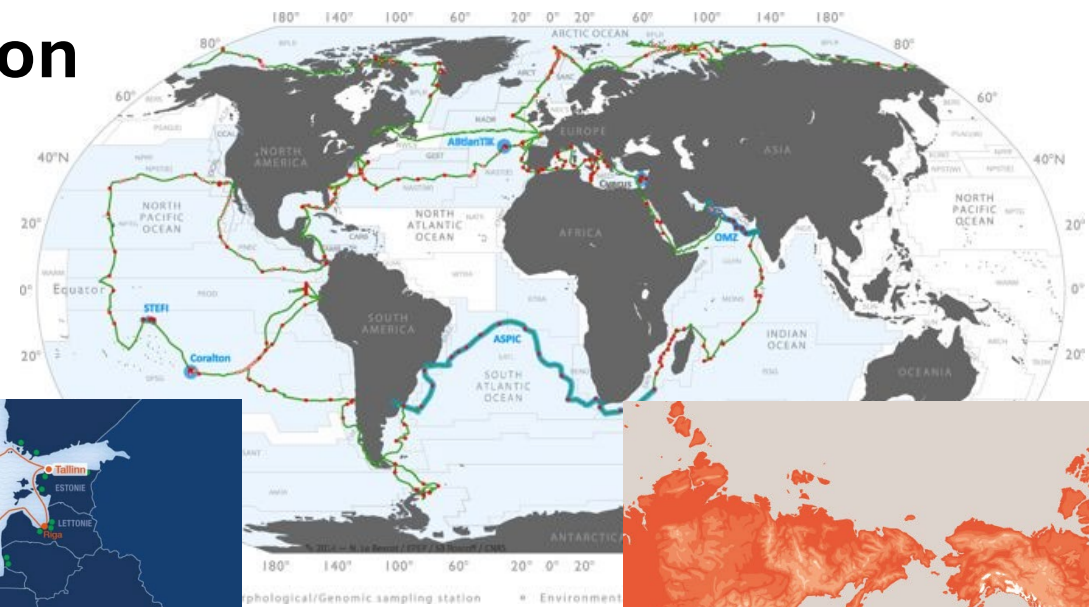


## Takeaway points and perspectives

- Using genomic-based data, such as *psbO*, is a step towards a better definition of phytoplankton diversity from space
- More omics data are in view, noting Tara Pacific, and Tara Europa



# New *psbO* datasets soon



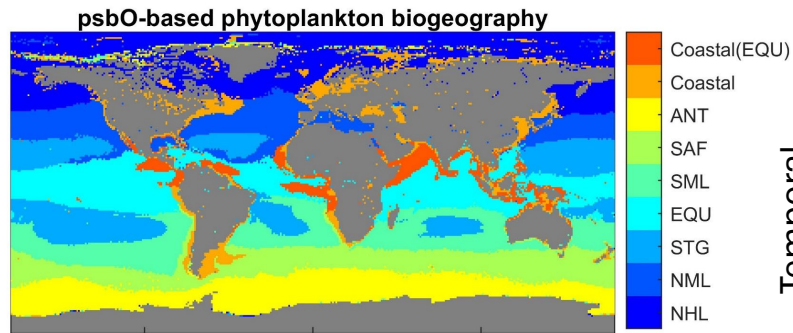
Tara Europa 2024

Tara Pacific 2016-2018

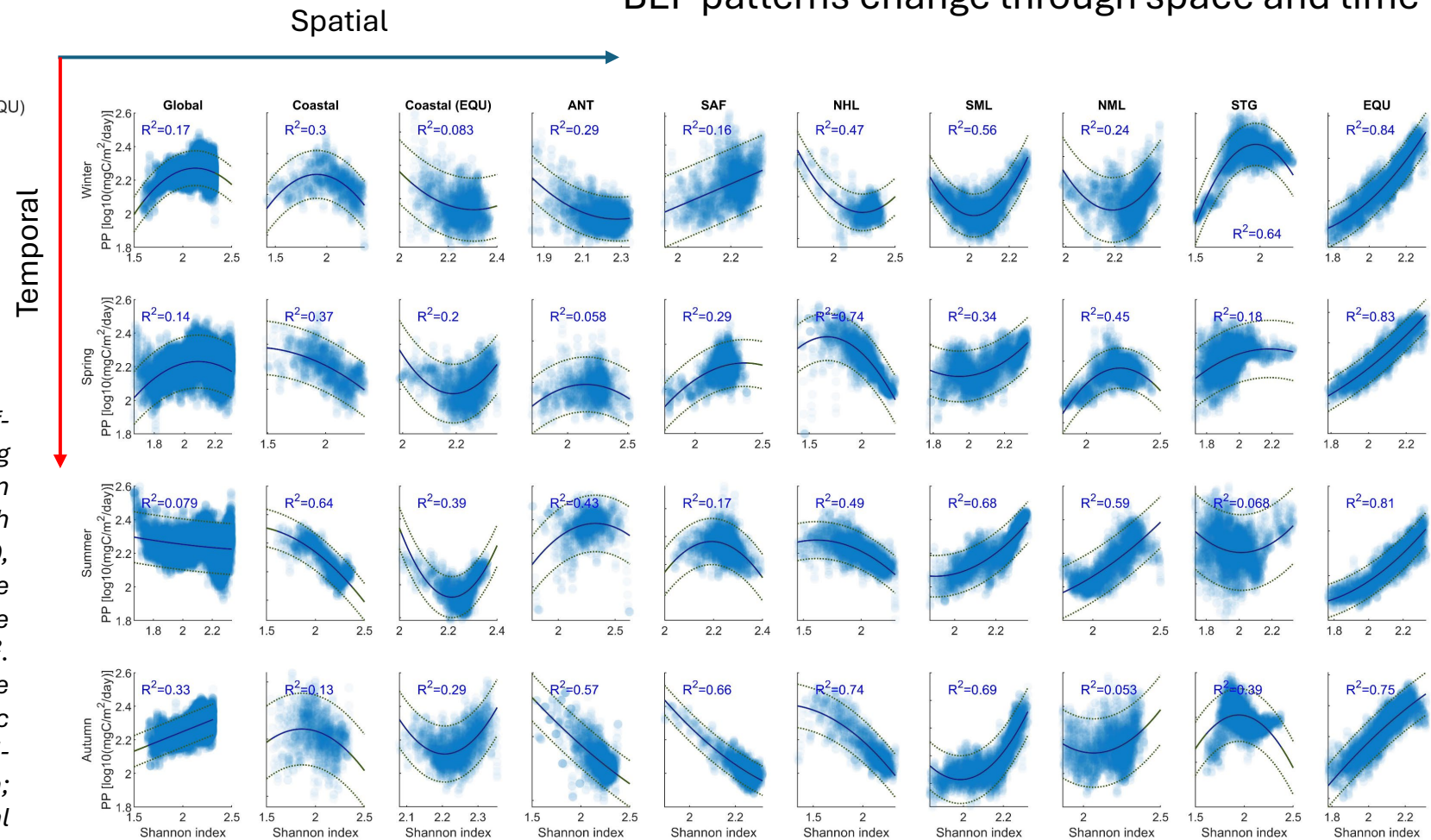


# Testing Biodiversity-Ecosystem Functioning hypothesis

BEF patterns change through space and time

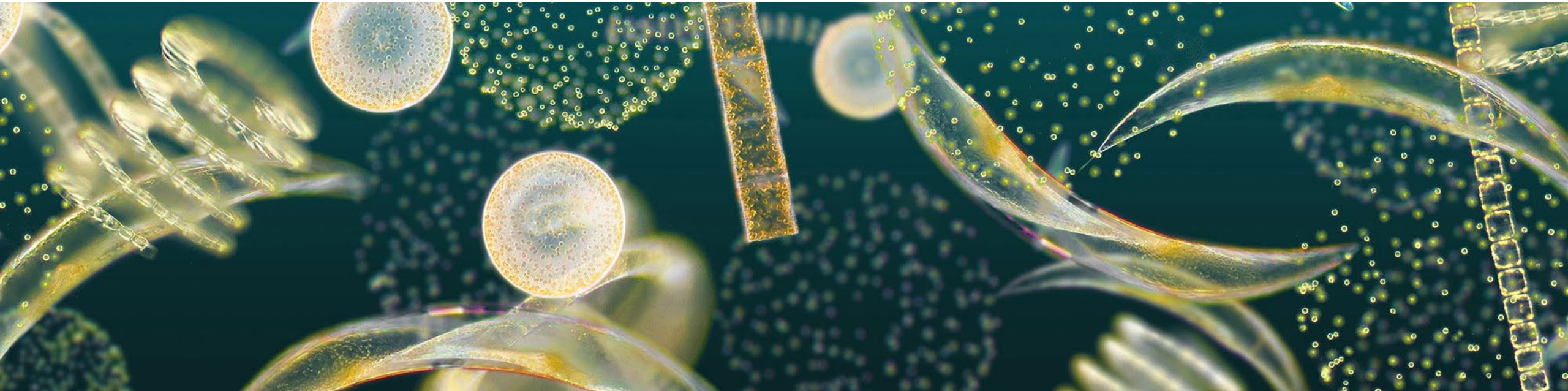


Phytoplankton BEF provinces delimited using Self-Organizing Maps (SOM) and Hierarchical Ascending Clustering (HAC) algorithms (a). The relation between PP and diversity was analyzed globally and for each province based on PG products (El Hourany et al., 2019, 2024). The curve represents a polynomial fit to illustrate the BEF relationship. The goodness of fit of each curve was evaluated based on the regression coefficient  $R^2$ . The dashed line represents the 95% confidence bounds. SAF: Sub-Antarctic Front, ANT: Antarctic Ocean, STG; Sub-Tropical Gyres; SML; Southern Mid-Latitude Region; NHL; Northern High-Latitude Region; NML; Northern Mid-Latitude Region; EQU; Equatorial region; Coastal (EQU): Equatorial coastal region; Coastal: Coastal region.



# Recommendations

- Integrate, harness cross-disciplinary expertise, push forward collaborations to better observe diversity from space
- AI and machine learning technics can solve non-linear, intricate problems such as the homogenization of multi-source data, to extract common patterns between remote sensing and ground truth data
- Incorporate and elaborate EBVs linked to the oceanic ecosystem



Thank you

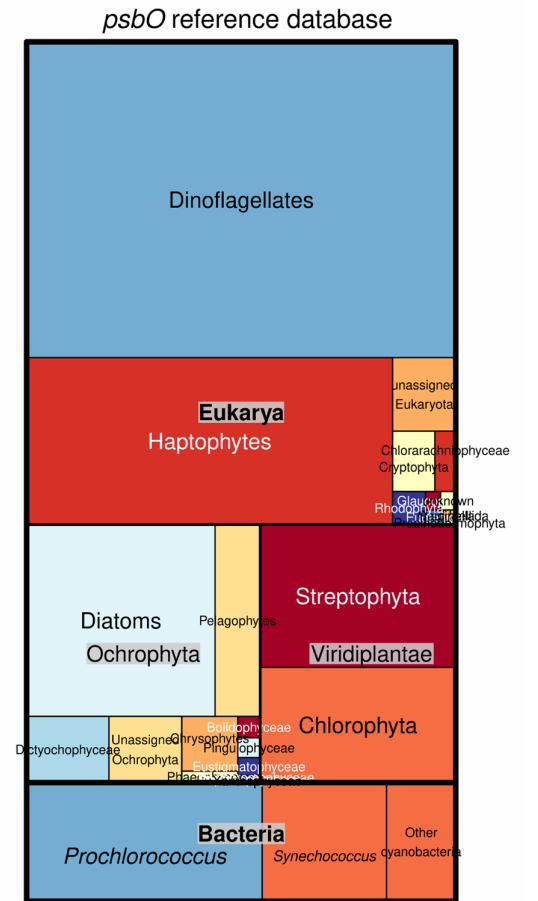
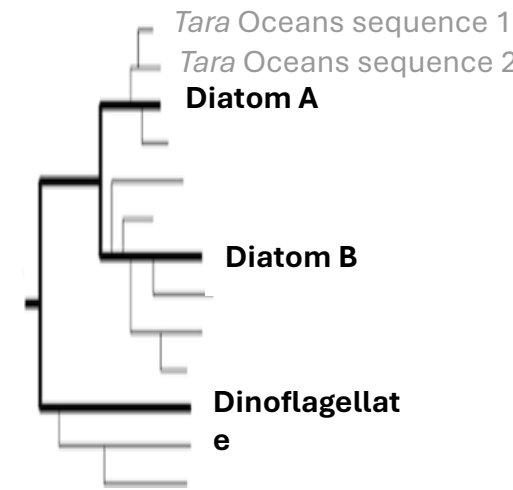
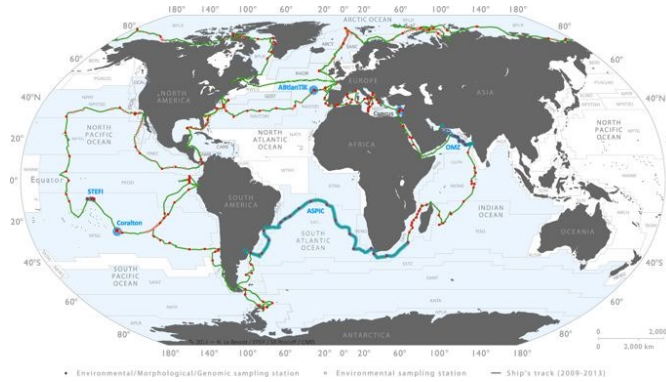
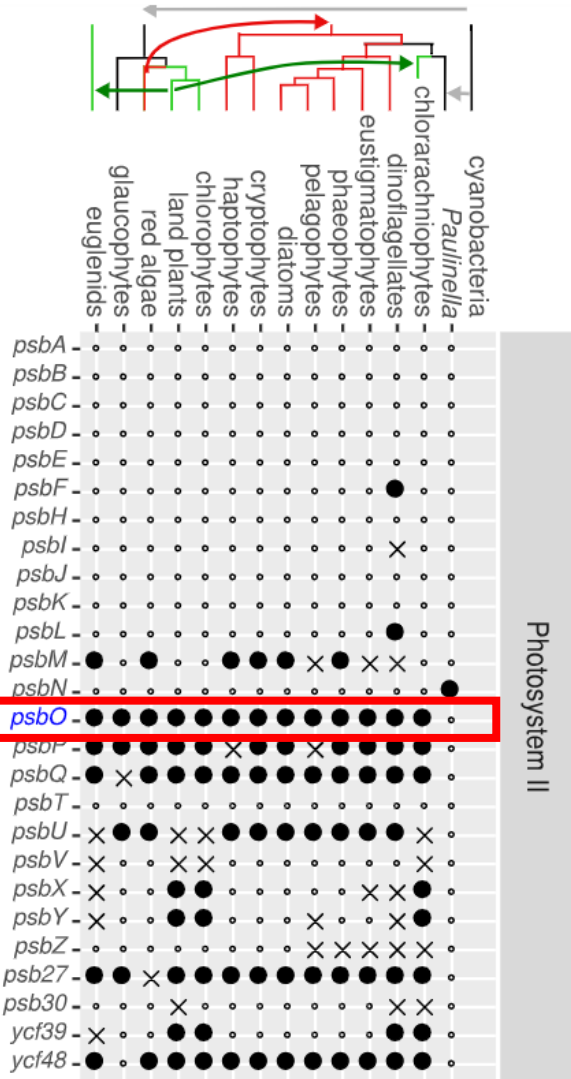


# Alternative methodology covering the entire community structure

The *psbO* protein; a core subunit of photosystem II (PSII), unique to organisms that carry out oxygenic photosynthesis.

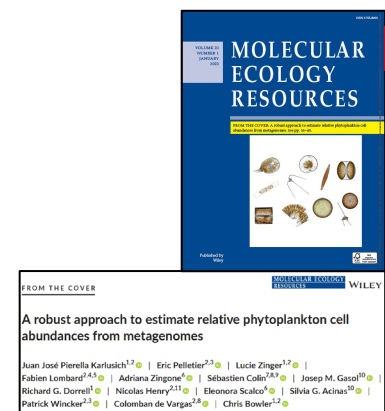
The *psbO* is a **single-copy gene**, present in both eukaryotes and prokaryotes groups

= Proxy of cell abundance



Curated database of >18k unique *psbO* sequences

Pierella et al., 2022, <https://doi.org/10.1111/1755-0998.13592>



# Uncertainty and applicability

Reliability index: Testing the belonging of each satellite multivariate pixel of an image, to the distribution of the *Tara* Oceans matchup data

Is a multivariate pixel an outlier to the data that was used to train the SOM?

