



Plankton Genomics

Virtual lab

Alexandre SCHICKELE, Pavla DEBELJAK, Sakina-Dorothee AYATA, Lucie BITTNER,
Éric PELLETIER, Stéphane PESANT, Lionel GUIDI, Jean-Olivier IRISSON

Sorbonne Université, EMBL-EBI

Blue-Cloud Open Conference

Brussels, 8 December 2022

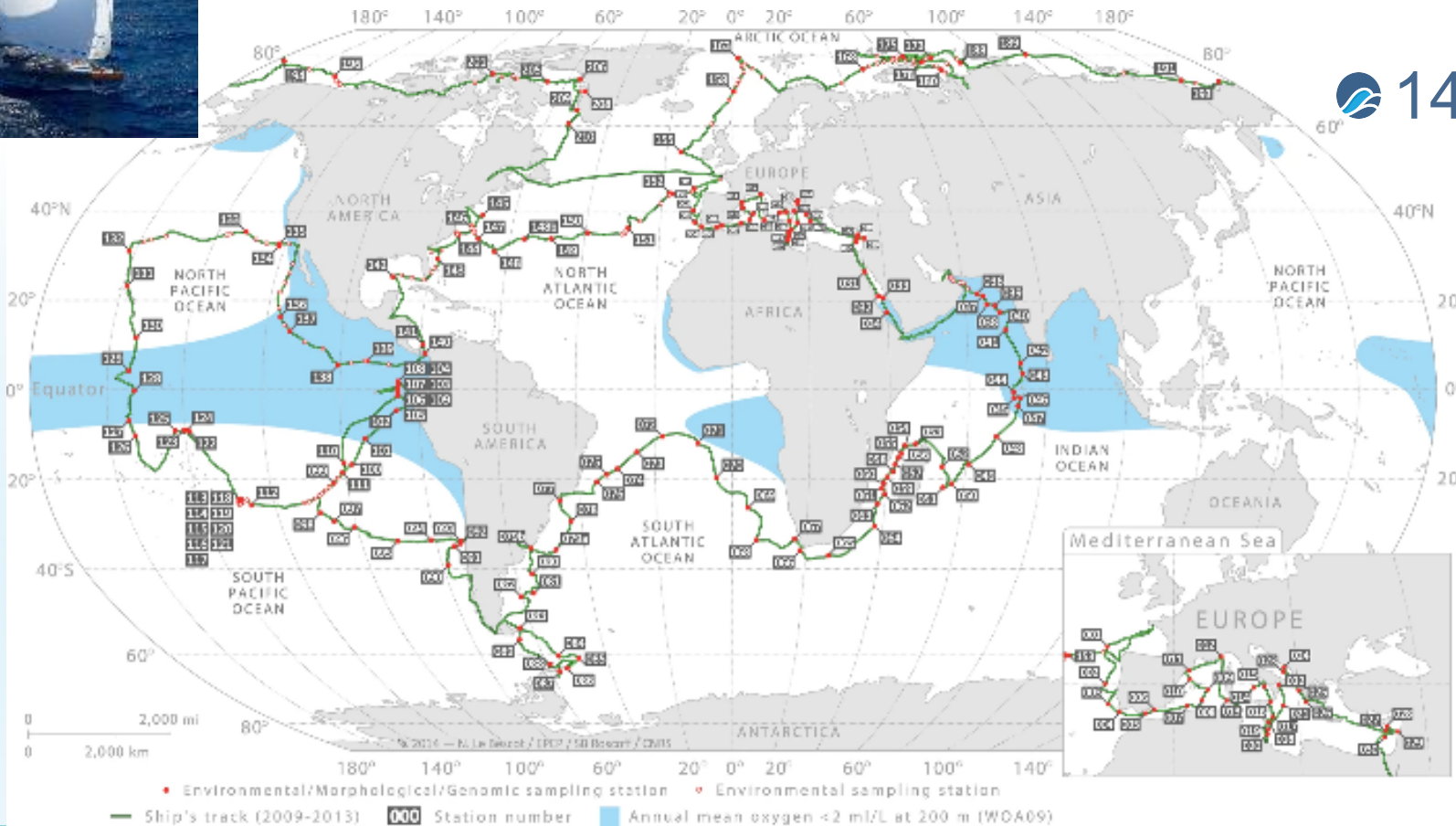


Blue-Cloud has received funding from the European Union's Horizon Programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, grant Agreement number 862409.



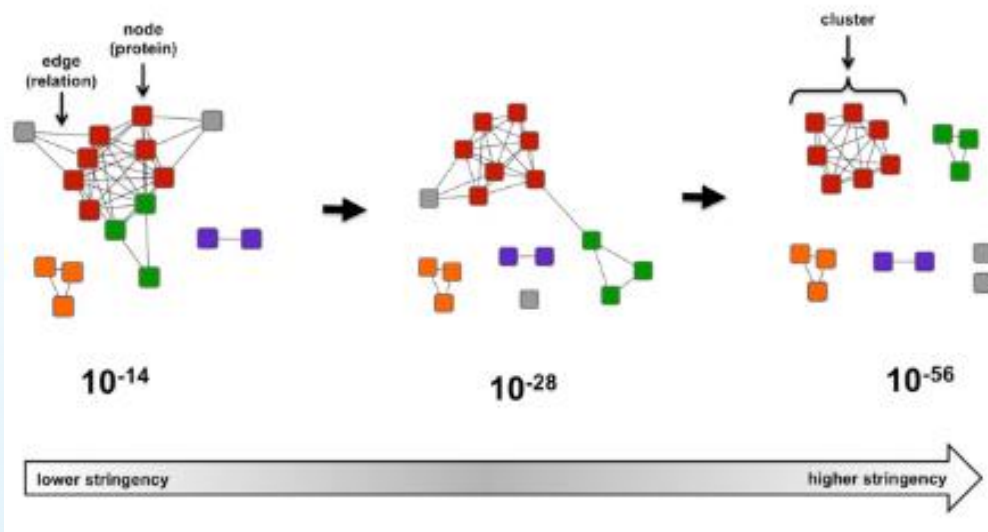
The data: Tara Oceans

 143 stations



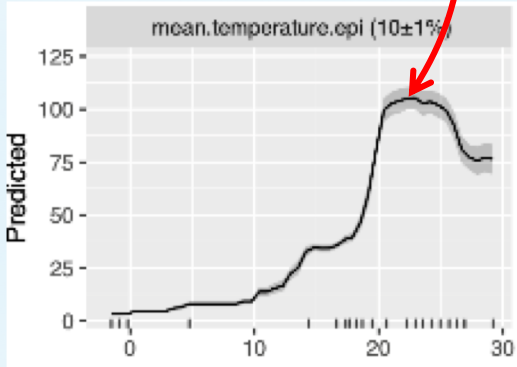
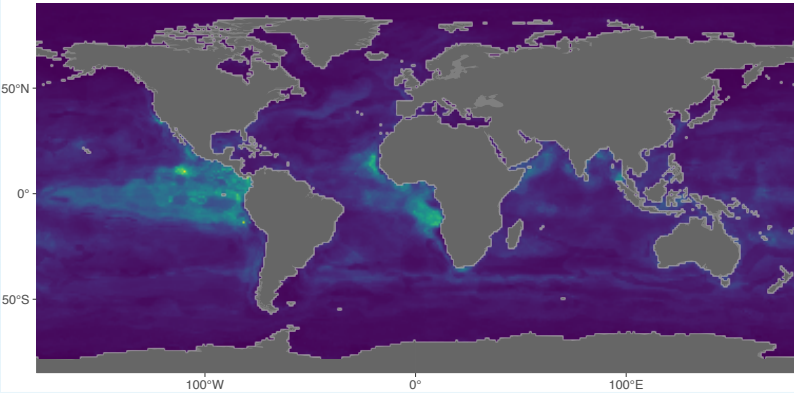
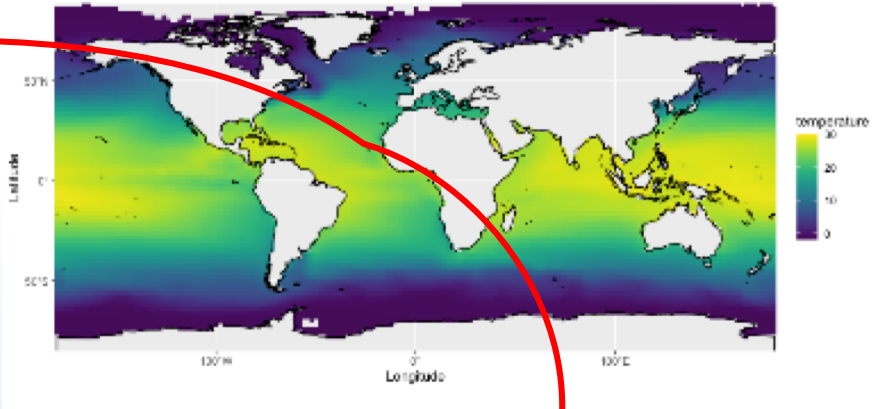
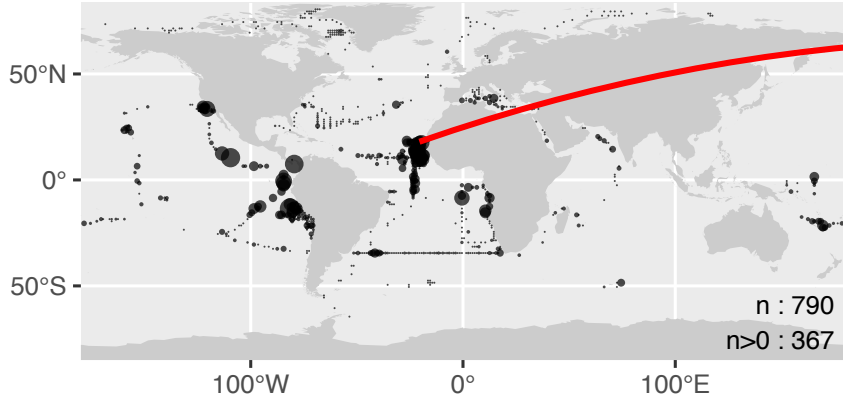
Pesant et al, 2015

The data: sequence similarity network

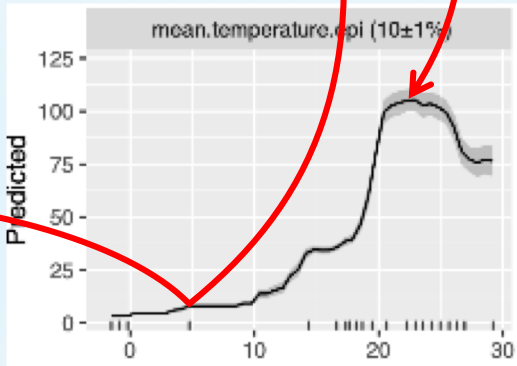
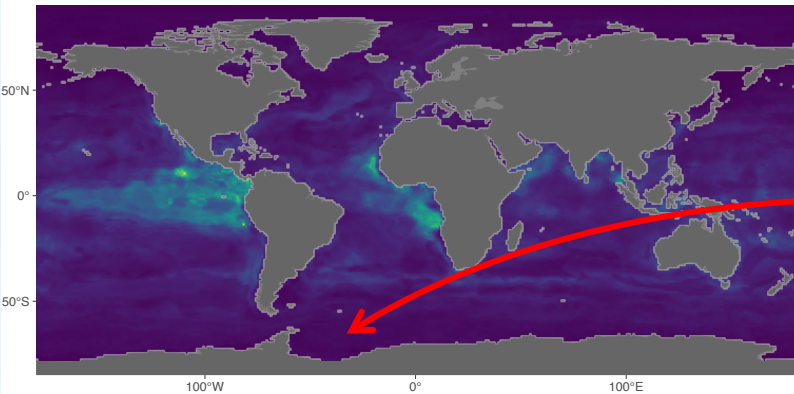
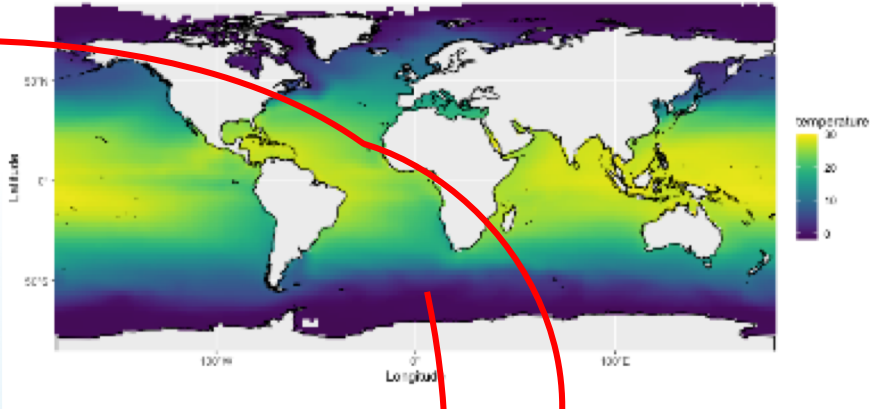
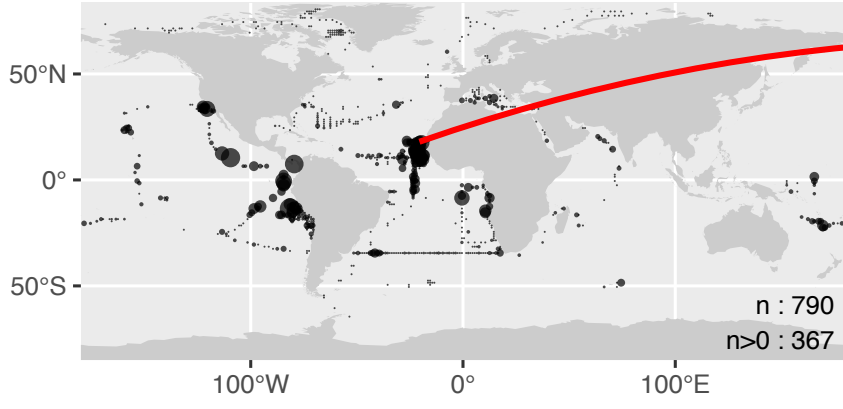


- 80% coverage, 80% similarity
- 1.2M clusters
- Functionally and taxonomically annotated

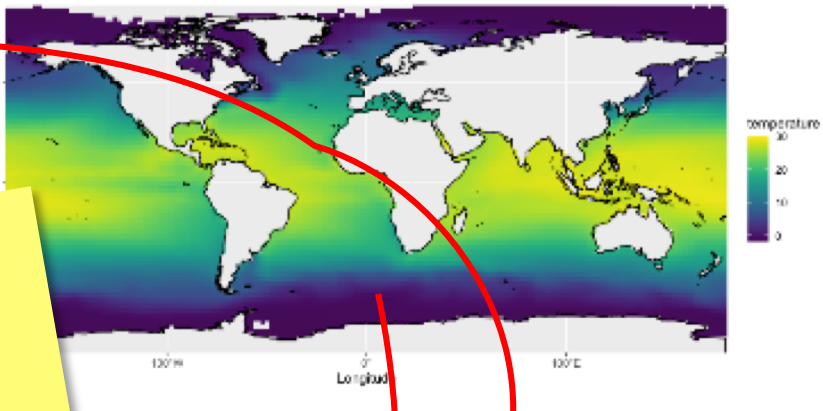
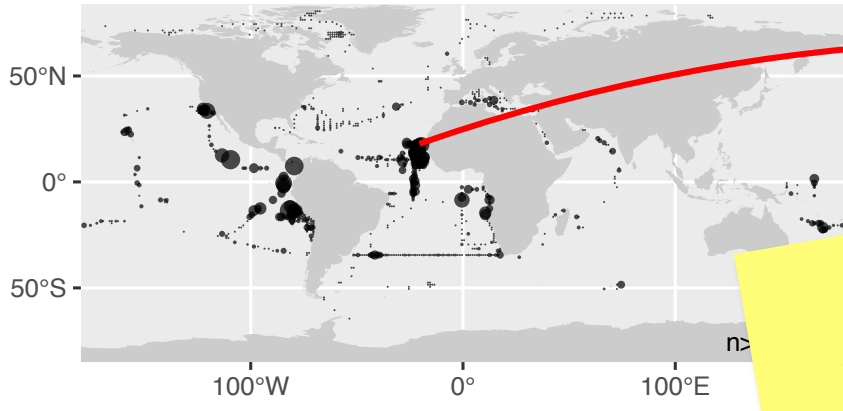
The process



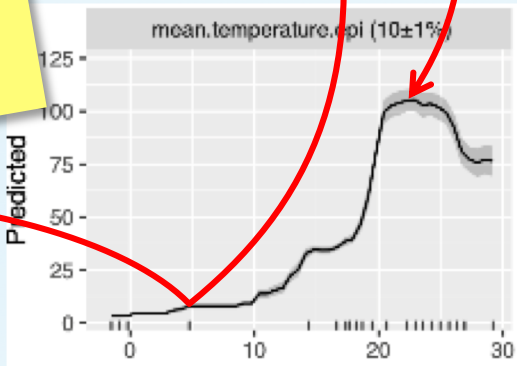
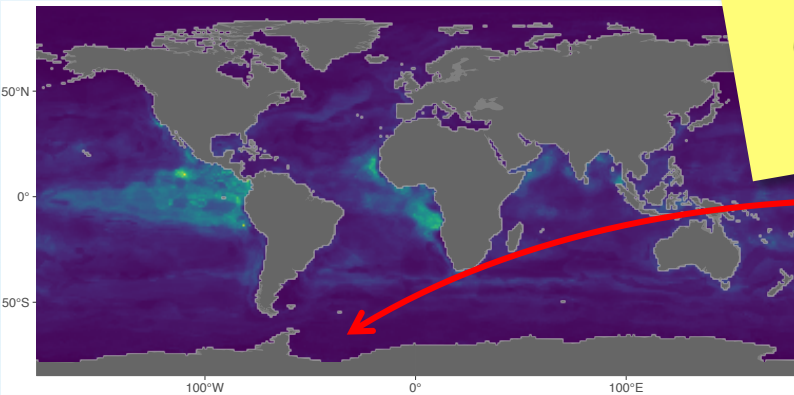
The process



The process

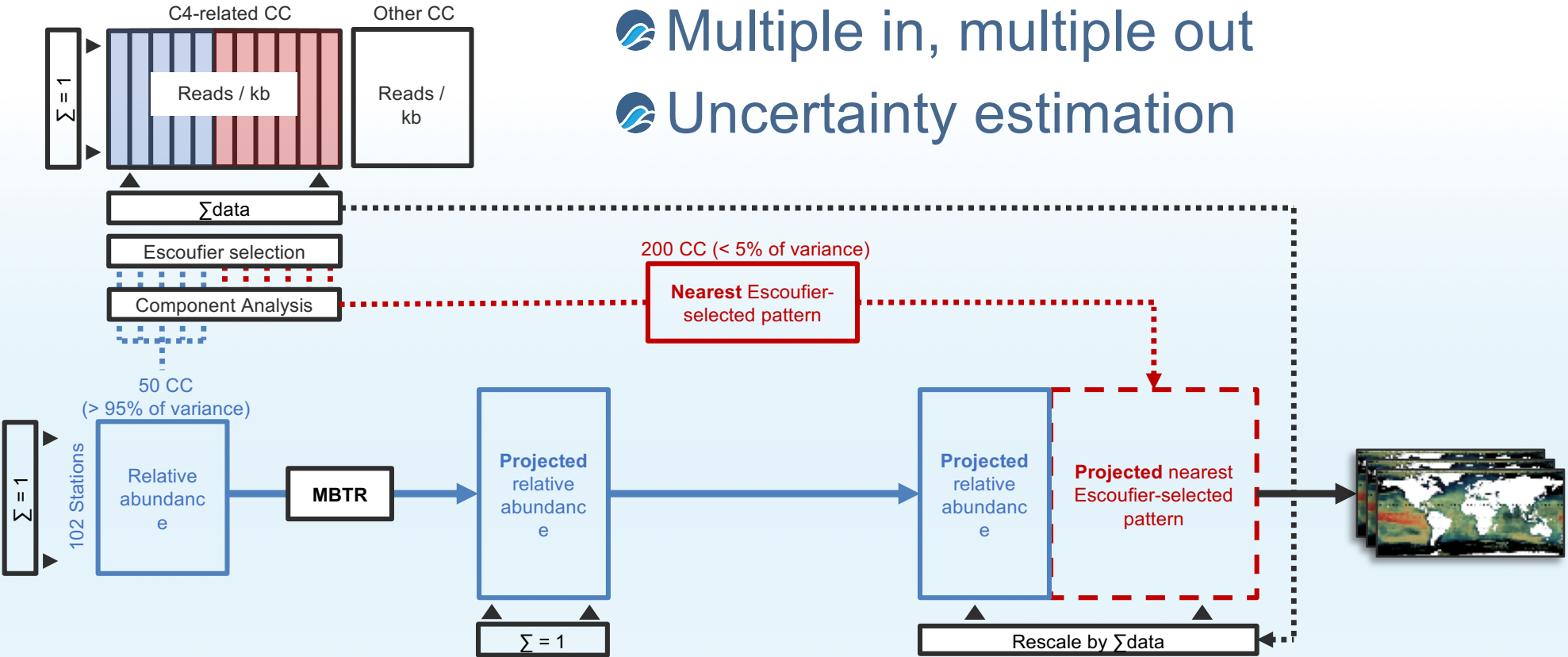


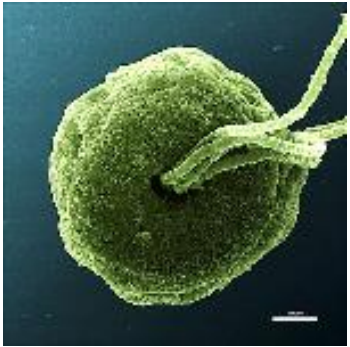
But multi-
multi-variate!



The process

- Multiple in, multiple out
- Uncertainty estimation

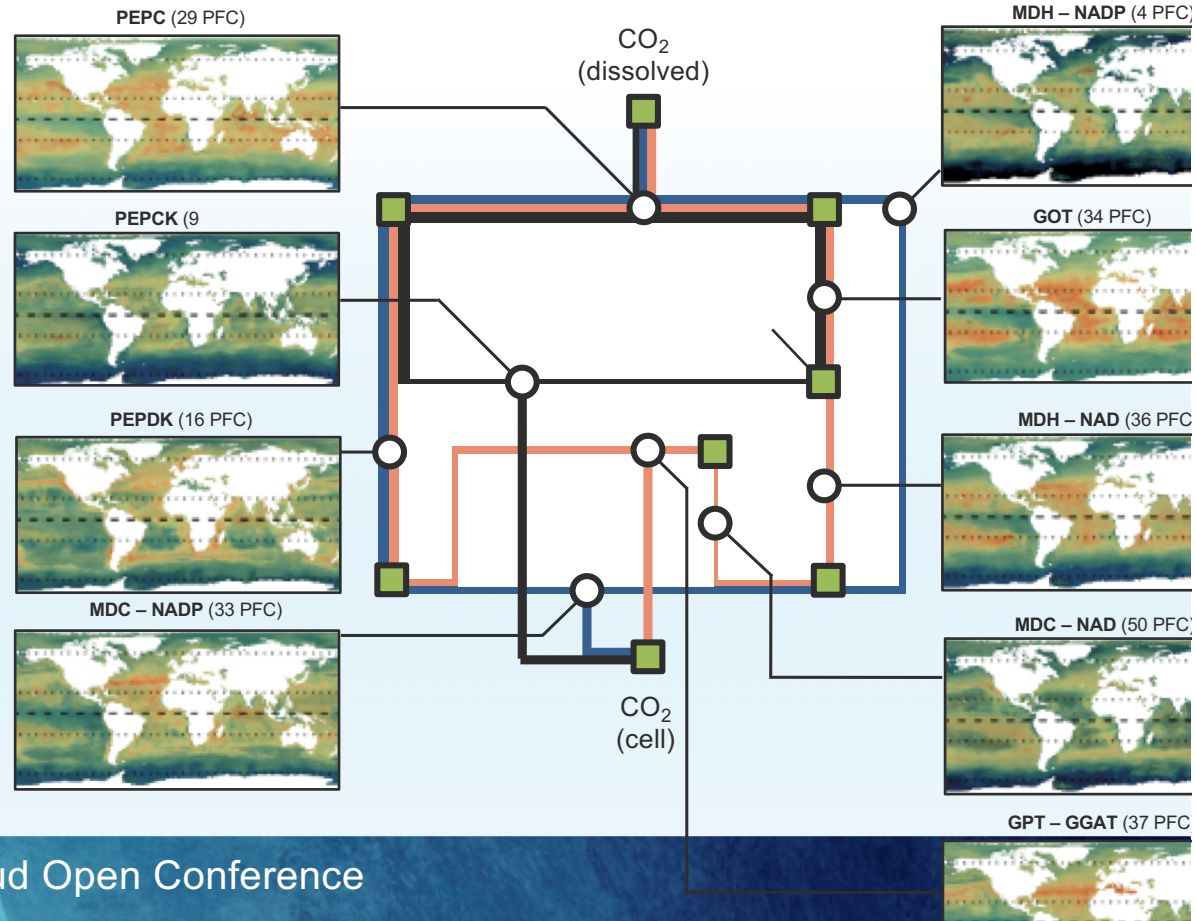
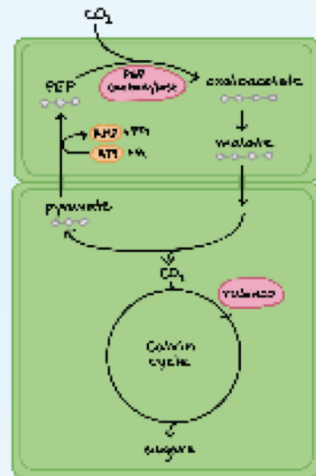
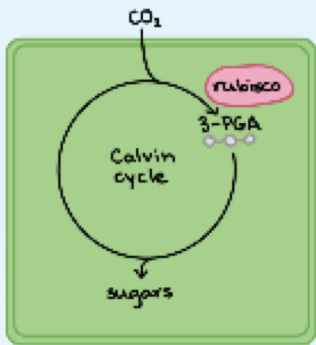




Our results

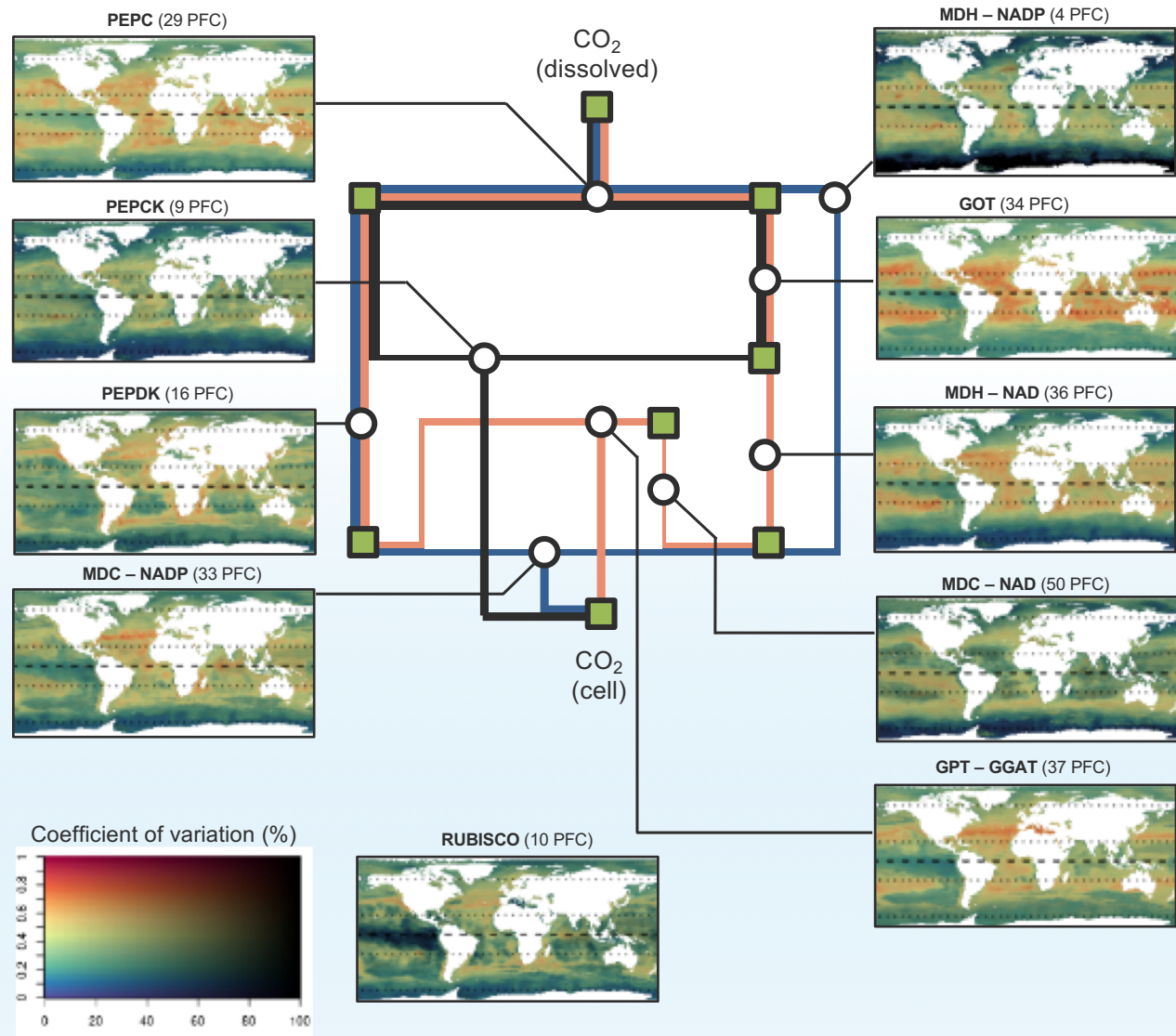


Carbon fixation in small green algae (0.2-5 μ m)



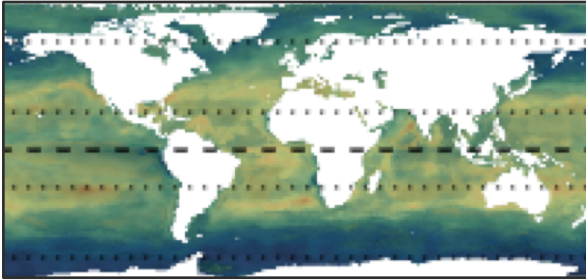
08/12/2022

Blue-Cloud Open Conference

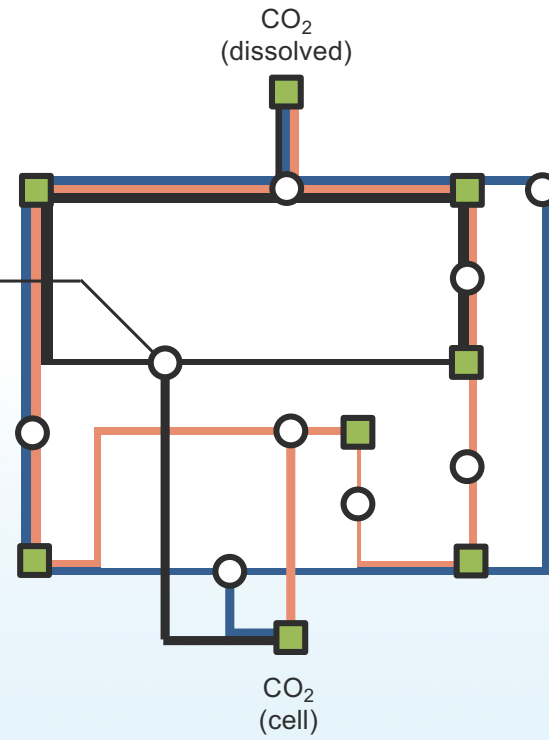
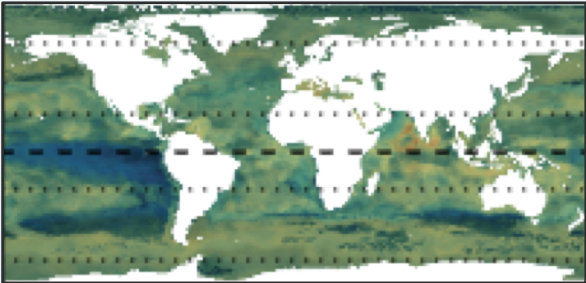


PEPCK (9 PFC)

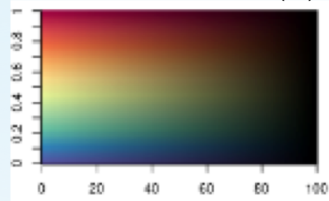
Potential distribution of function



Potential distribution of function *in organisms*



Coefficient of variation (%)



The future

- Used 310 clusters, within 1.2M (0.02% of data) ⇒ much left to do for other **genomic/plankton researchers**
 - Explore other pathways/functions
 - Focus on unknown clusters (~80%) rather than annotated ones
- First implementation of MBTR, first genomic niche modelling in the cloud ⇒ much to **improve**
 - More efficient parallel processing on the infra
 - Add systematic estimation of model quality
 - Generalise to other data types

The future

- Used 310 clusters, within 1.2M (0.02% of data) ⇒ much left to do for other **genomic/plankton researchers**
 - Explore other pathways/functions
 - Focus on unknown clusters (~80%) rather than annotated ones
- First implementation of MBTR, first genomic niche modelling in the cloud ⇒ much to **improve**
 - More efficient parallel processing on the infra
 - Add systematic estimation of model quality
 - Generalise to other data types





Blue-Cloud

Unlocking
Open Science
in support of the
EU Green Deal

Website: www.blue-cloud.org

E-mail: info@blue-cloud.org

Twitter: [@BlueCloudEU](https://twitter.com/BlueCloudEU)

LinkedIn: [Blue-Cloud Org](https://www.linkedin.com/company/blue-cloud-org)

FINAL CONFERENCE