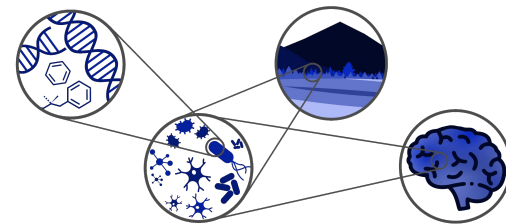


UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Functional Adaptations of Diatoms to Different Environments in TARA Oceans Data

Silvia Lazzardi, Kobe Siemons, Emanuele Pigani, Daniele Iudicone, Sandro Azaele, Samir Suweis

CSS 2025, Bari

15-17/09/2025

Samir Suweis - samir.suweis@unipd.it

Tara Oceans expeditions

Tara Ocean Foundation → scientific expeditions with the aim to:

- study **marine biodiversity**
- observe and anticipate the **impacts of climate change and pollution**



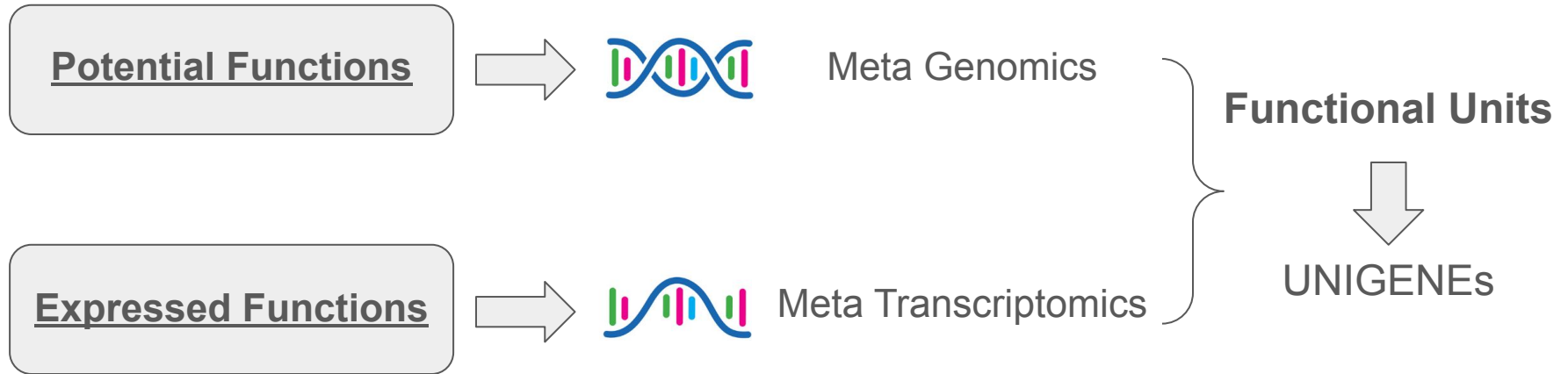
Diatoms - a class of Plankton

- Unicellular organisms
- Form massive annual springs
- Responsible for about half of photosynthesis in the global oceans
- Generate about 20 to 50 percent of the oxygen produced on the planet each year

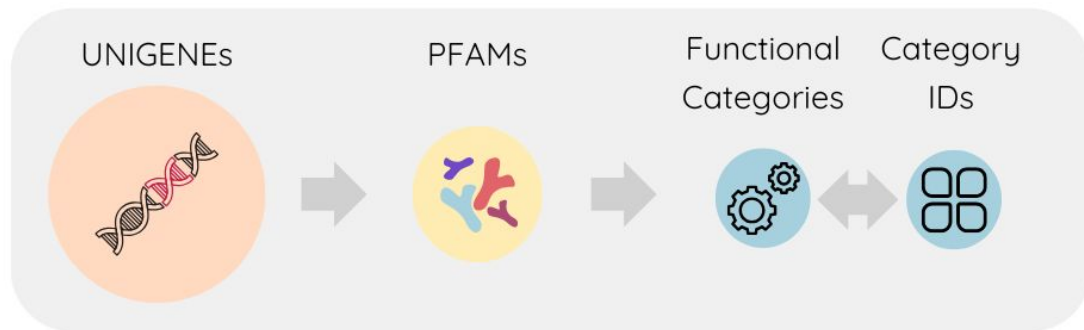


To understand their functional adaptations to environmental changes is crucial

Biological Functions

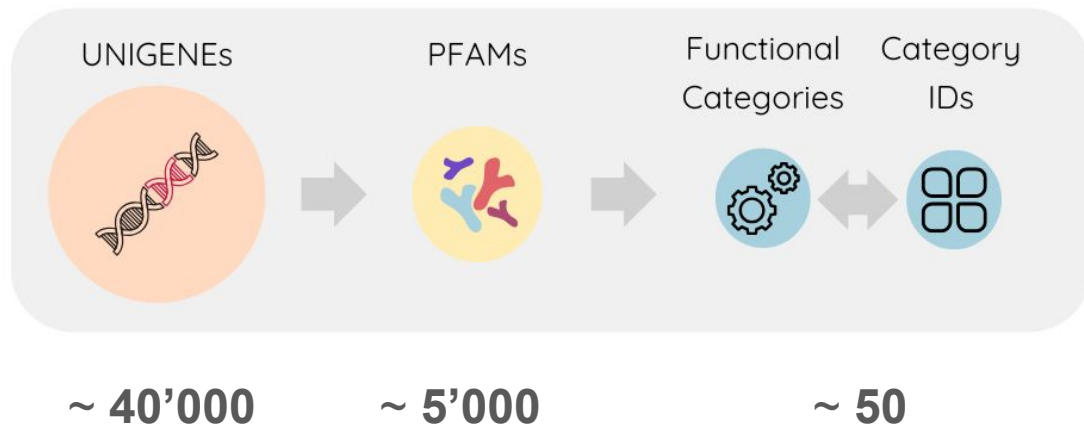


From UNIGENEs to Functions



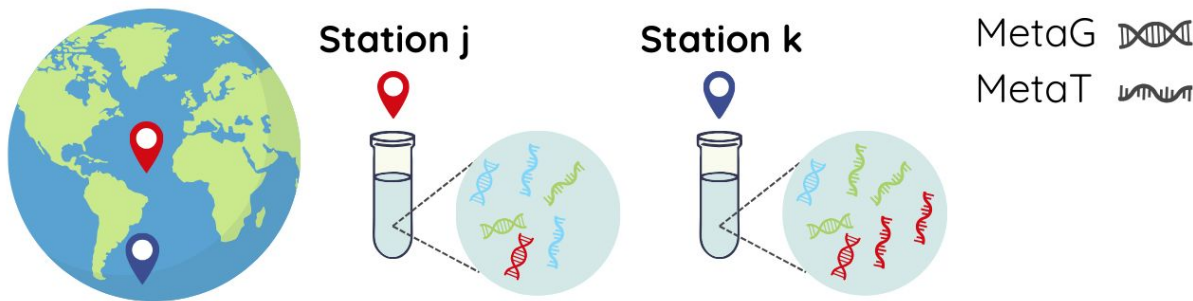
- Genes clustered into **UNIGENEs**
- Mapped to **Pfam domains**
- Pfams grouped into **Functional Categories**

From UNIGENEs to Functions

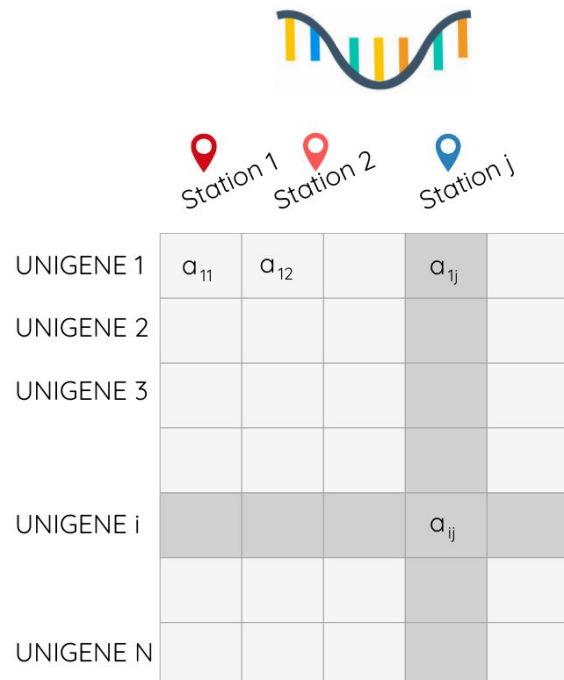
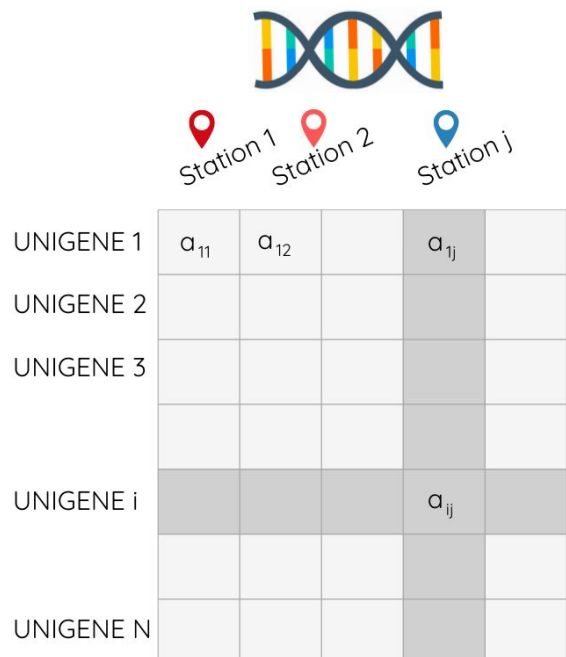


Tara Oceans Project – Datasets Overview

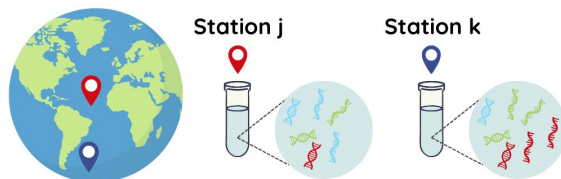
- **77 sampling stations** worldwide → uniform sampling
- **DNA** and **RNA** from ocean water samples
- **Genomes and transcriptomes** reconstructed from DNA/RNA sequences
- ~ 40,000 **UNIGENES** (clustered gene sequences) identified



Datasets format



Datasets format



	Station 1	Station 2	Station j	Station R
UNIGENE 1	a_{11}	a_{12}	a_{1j}	
UNIGENE 2				
UNIGENE 3				
UNIGENE i			a_{ij}	
UNIGENE N				a_{NR}

$$M_j = \sum_{i=1}^N a_{ij}$$

$$\left\{ \begin{array}{l} \langle f_i \rangle = \frac{1}{R} \sum_{j=1}^R \frac{a_{ij}}{M_j} \\ O_i = \frac{1}{R} \sum_{j=1}^R \begin{cases} 1 & \text{if } a_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \end{array} \right.$$

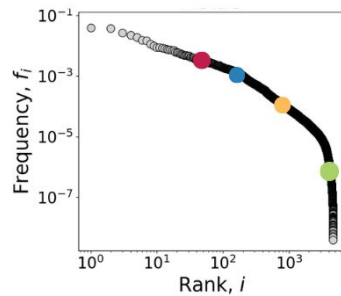
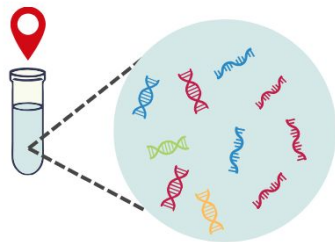
Different approaches to answer the same question

- **Null Model**
- Community Detection & Clustering
- PCA analysis

Null Model and Simulations

Null Hypothesis: $a_{ij} \sim \text{Poisson}(M_j \langle f_i \rangle)$

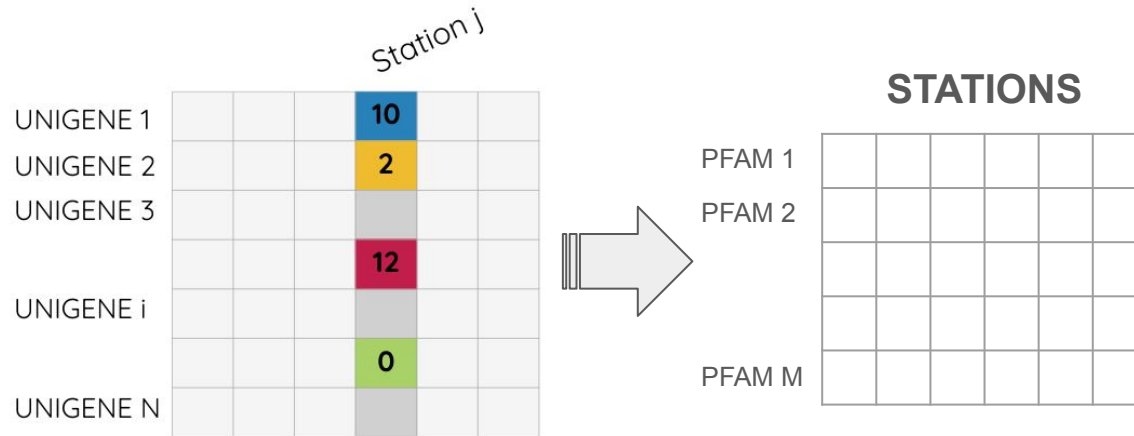
Station j



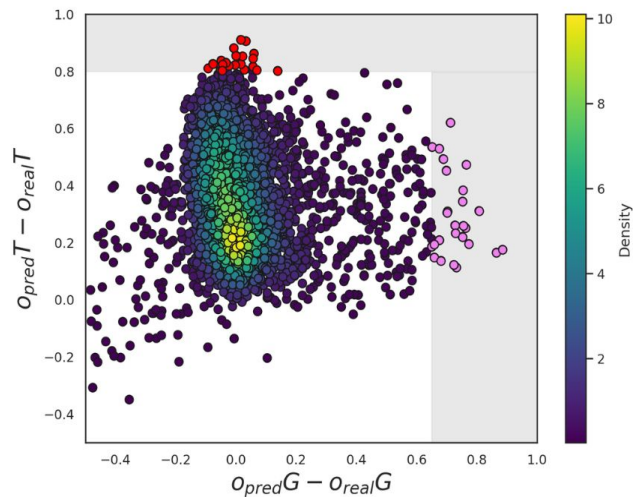
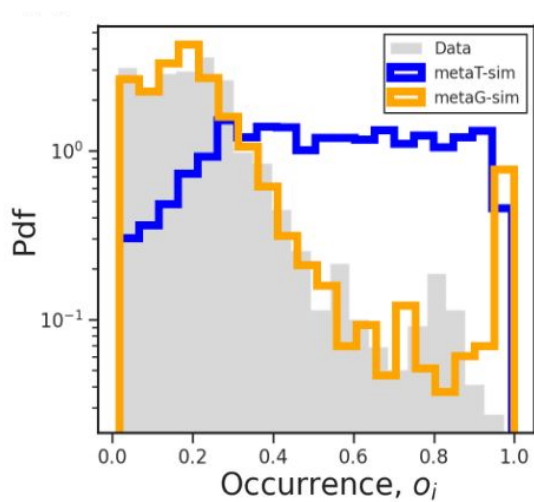
Simulated matrix

	Station j				
UNIGENE 1			10		
UNIGENE 2			2		
UNIGENE 3					
UNIGENE i			12		
			0		
UNIGENE N					

Null Model and Simulations



Occurrence analysis

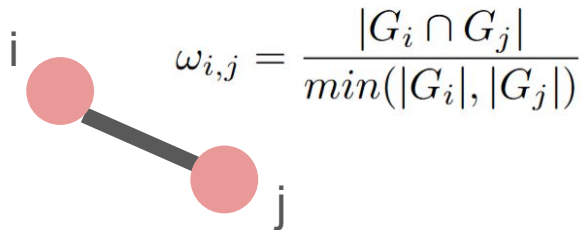


Different approaches to answer the same question

- Null Model
- **Community Detection & Clustering**
- PCA analysis

Different approaches to answer the same question

- Null Model
- **Community Detection & Clustering**
- PCA analysis

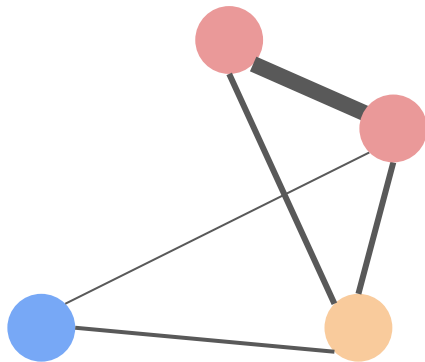


A diagram showing two red circular nodes, labeled 'i' and 'j', connected by a thick black line representing an edge. Node 'i' is positioned slightly higher and to the left of node 'j'.

$$\omega_{i,j} = \frac{|G_i \cap G_j|}{\min(|G_i|, |G_j|)}$$

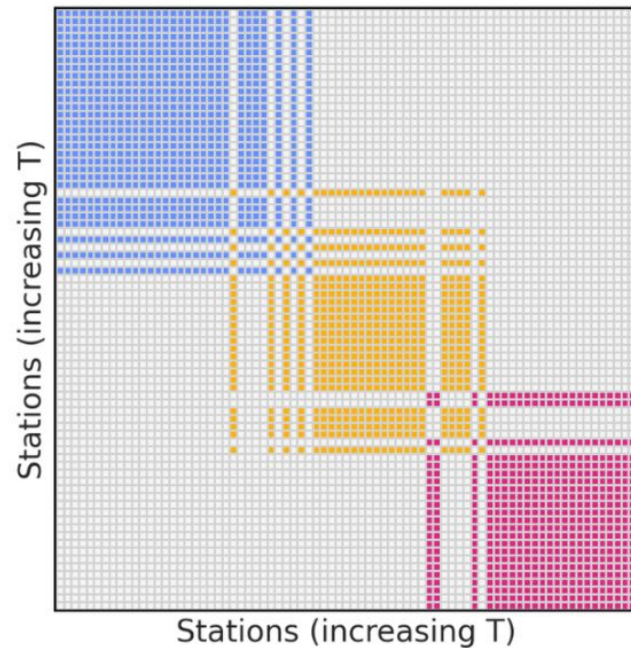
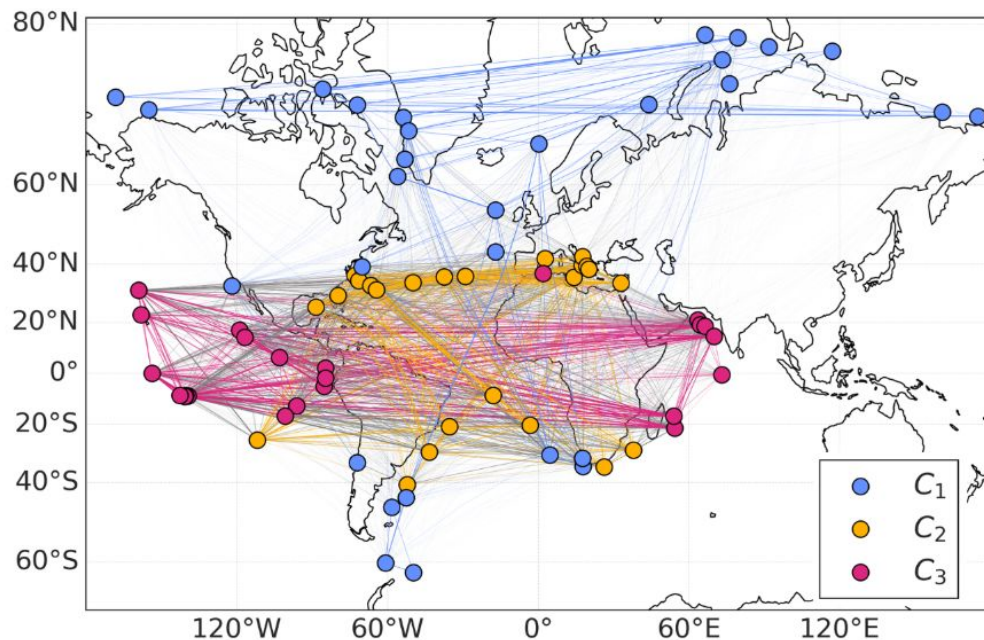
Different approaches to answer the same question

- Null Model
- **Community Detection & Clustering**
- PCA analysis



Clusters and Temperature

Clusters → 3 bio-geographical regions



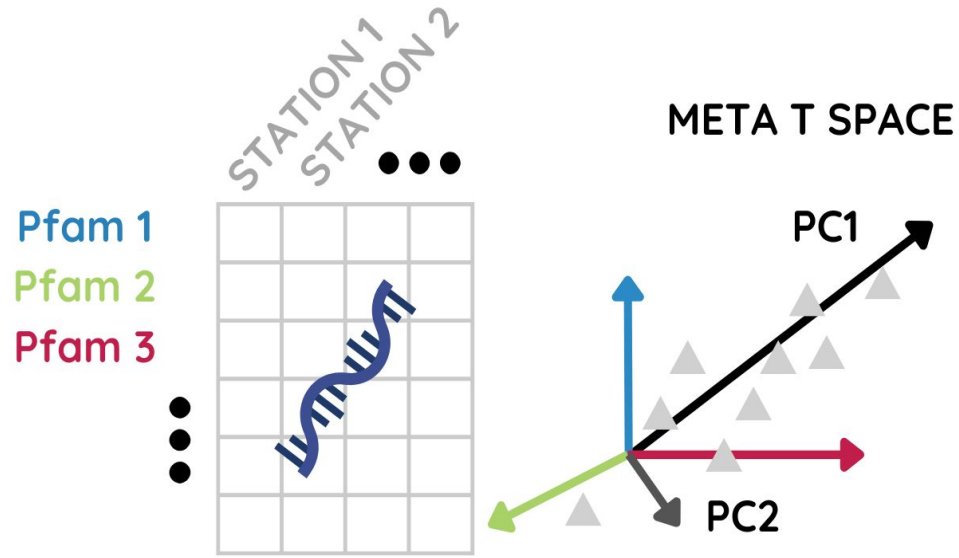
Deviating Pfams are concentrated in metaT but not in metaG



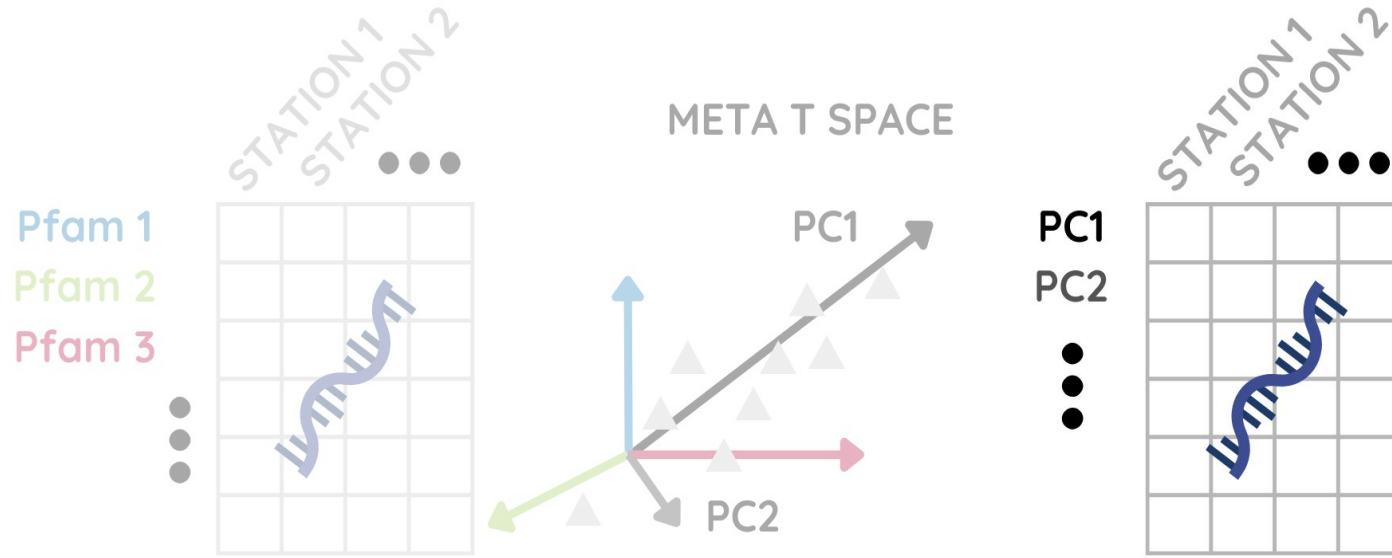
Different approaches to answer the same question

- Null Model
- Community Detection & Clustering
- **PCA analysis**

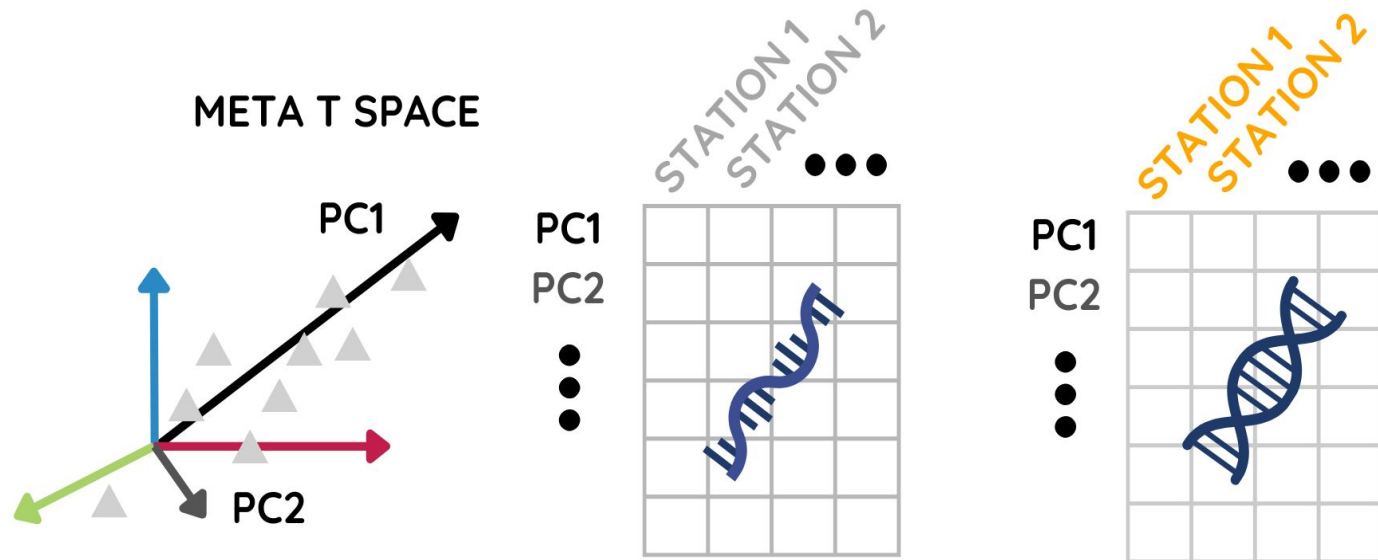
PCA analysis



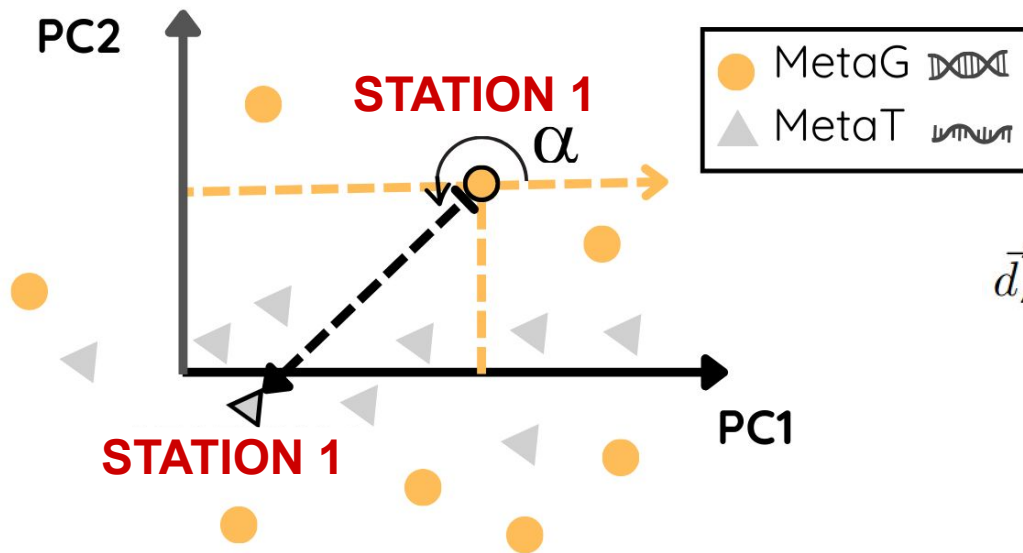
PCA analysis



PCA analysis



PCA analysis

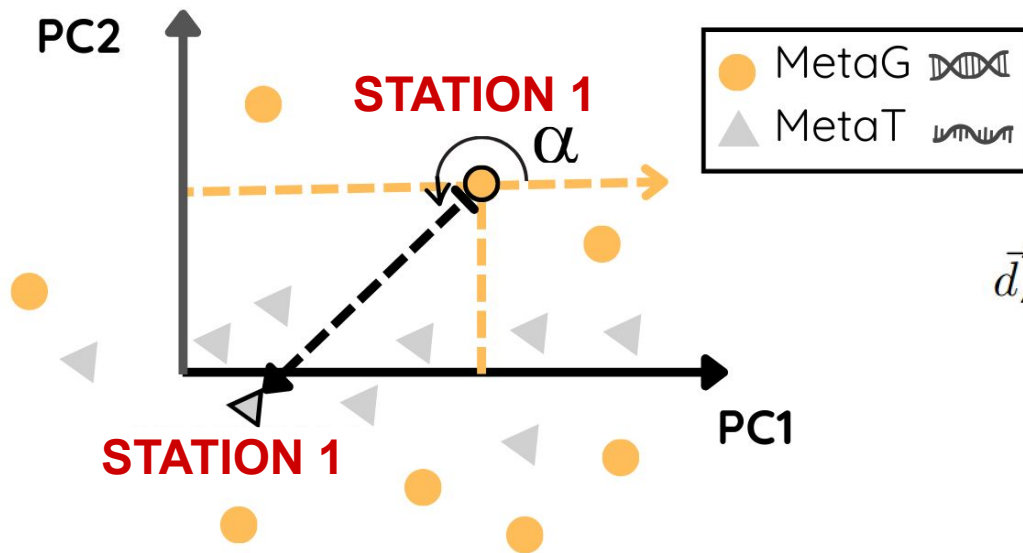


What is α ?

$$\vec{\Delta}_j = \vec{f}_G^j - \vec{f}_T^j$$

$$\vec{d}_{\Delta_j} := \frac{\vec{\Delta}_j}{|\vec{\Delta}_j|} = \cos(\alpha_j) \vec{p}c_1 + \sin(\alpha_j) \vec{p}c_2$$

PCA analysis



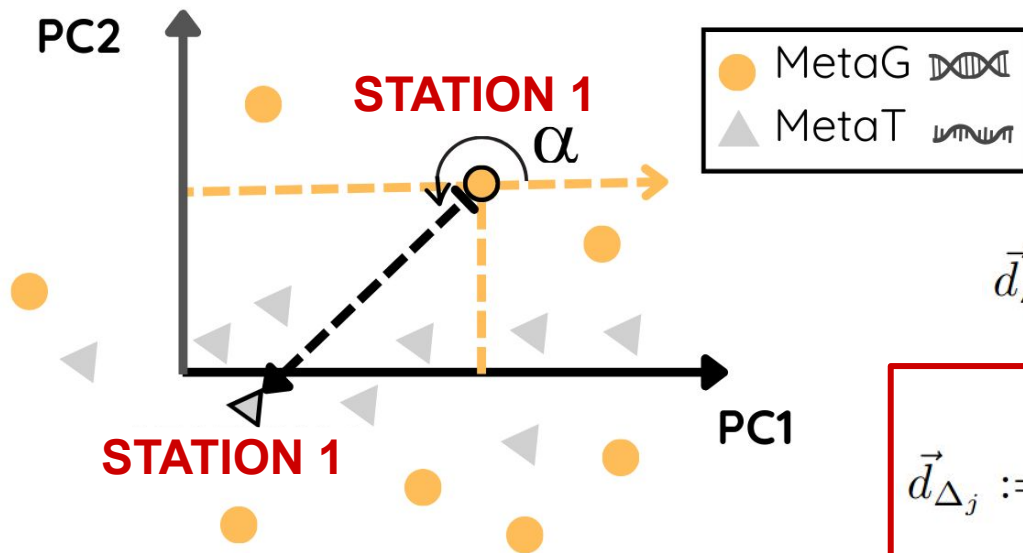
What is α ?

$$\vec{\Delta}_j = \vec{f}_G^j - \vec{f}_T^j$$

$$\vec{d}_{\Delta_j} := \frac{\vec{\Delta}_j}{|\vec{\Delta}_j|} = \cos(\alpha_j) \vec{p}c_1 + \sin(\alpha_j) \vec{p}c_2$$

$$\vec{p}c_k = \sum_{i=1}^N p_{ki} PF\vec{A}M_i$$

PCA analysis



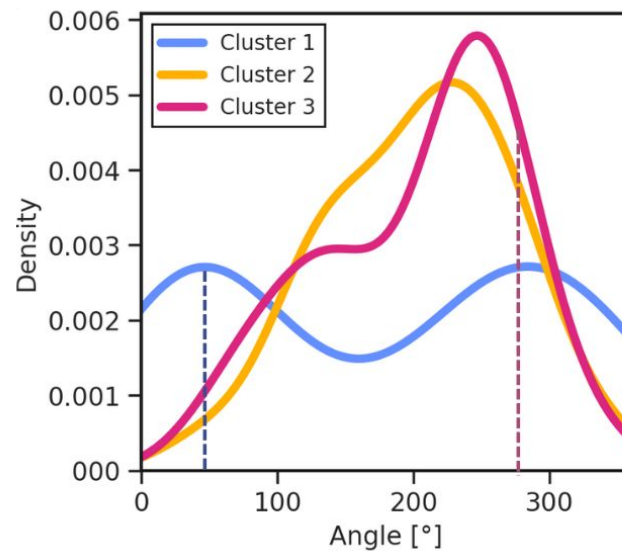
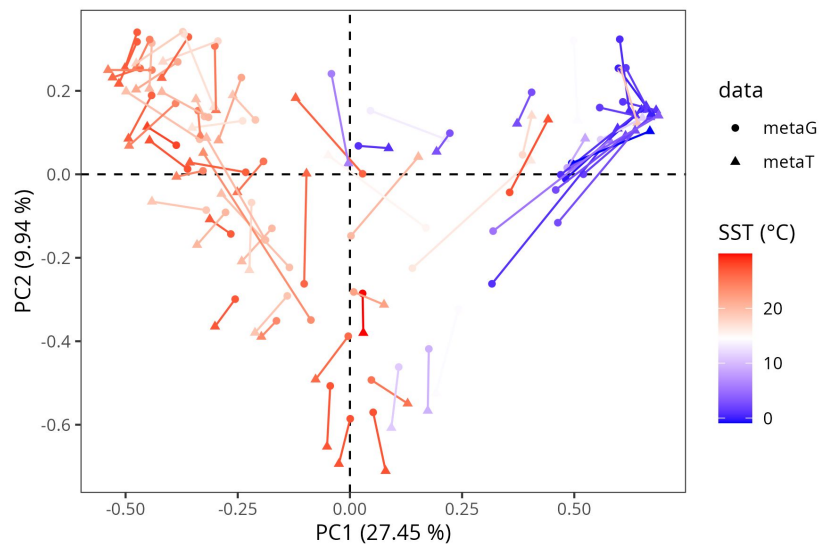
What is α ?

$$\vec{\Delta}_j = \vec{f}_G^j - \vec{f}_T^j$$

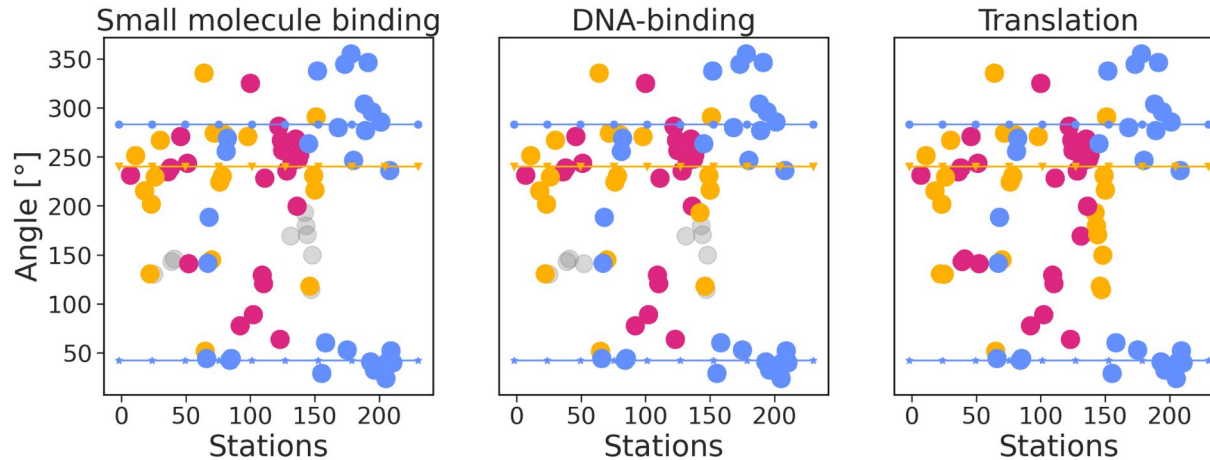
$$\vec{d}_{\Delta_j} := \frac{\vec{\Delta}_j}{|\vec{\Delta}_j|} = \cos(\alpha_j) \vec{pc}_1 + \sin(\alpha_j) \vec{pc}_2$$

$$\vec{d}_{\Delta_j} := \sum_{i=1}^N c_{ki} PF \vec{A} M_i \longrightarrow \text{weighted sum of biological functions}$$

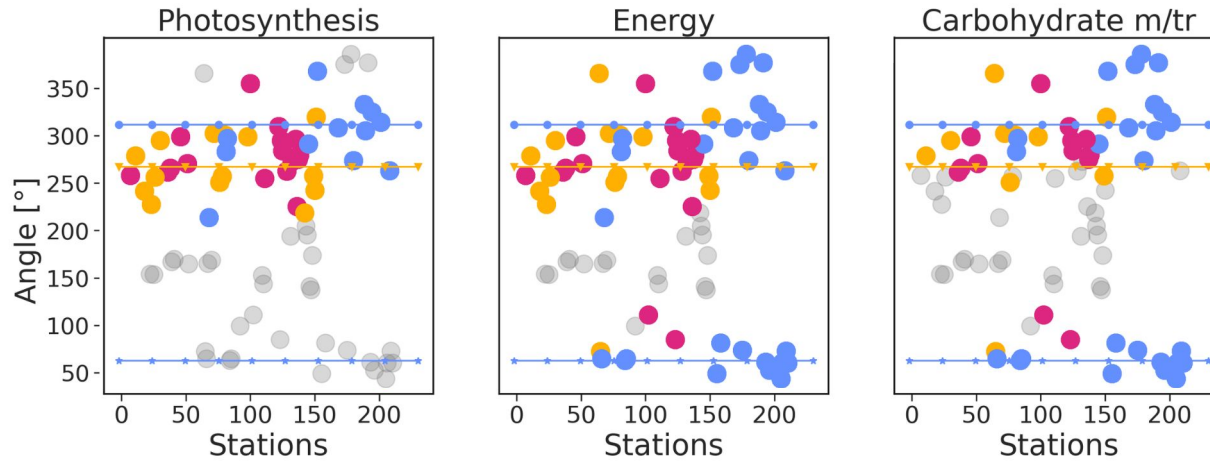
PCA analysis reveals preferential trajectories from metaG to metaT



General functional categories → expected to be expressed everywhere



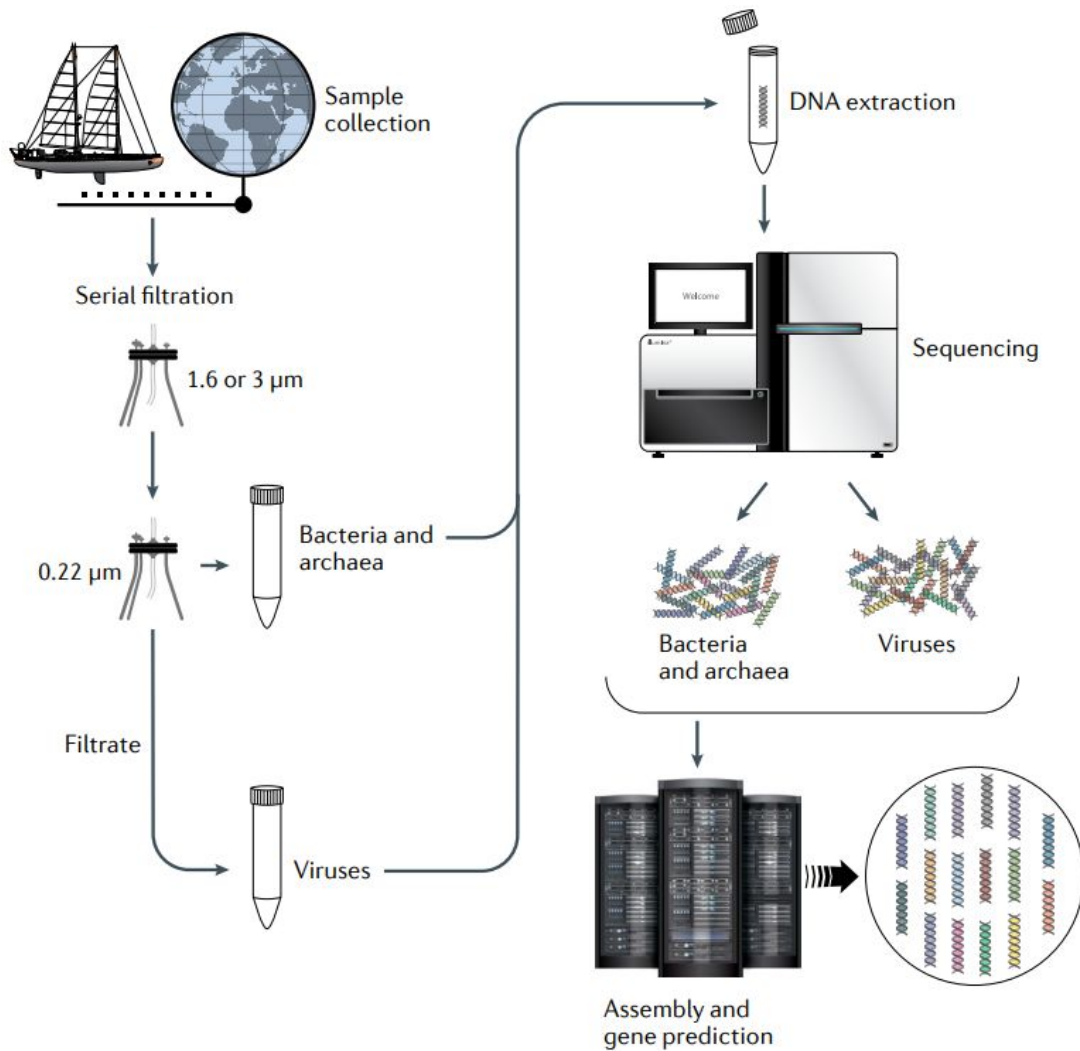
Energy and Photosynthesis related functional categories



Conclusions

- Distinct **biogeographic clusters** reveal functional structure
- Fine-scale **unigenes/PFAMs patterns** show **environmental selection**
- **Localized functional traits** suggest **ecological specialization**
- **MetaG and MetaT divergence** captures **acclimation** trajectories

Thank you for your attention!



Functional categories redundancy across clusters

