

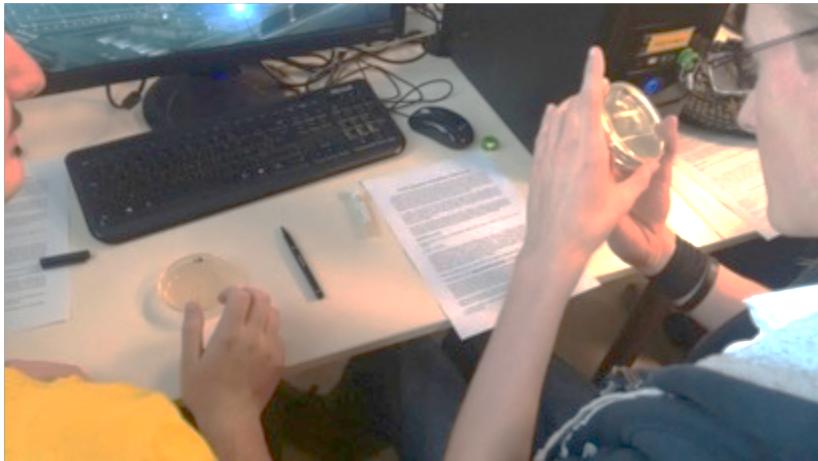
# Paris Update

H. Richard, I. Lafontaine

Lab. Computational and Quantitative Biology

Pierre and Marie Curie University

Bachelor Course: Introduction to Bioinformatics



## Le Métro:

- 371 stations, 14 lines
- Contacts with the MTA, waiting for accreditations (need 1 per line!)
  - CSD – Sampling at the entrance of the stations
  - 80 samples from 20 stations, now adding time points (2 / month)
  - Bachelor / Master Students in Biology and Comp. Sci.



# Functional domain annotation of metagenomics and metatranscriptomics samples

work by A. Ugarte, J. Bernardes and A. Carbone



**A. Carbone**  
Director, LCQB



Juliana Bernardes  
Assistant professor



Ari Ugarte  
PhD student



# Microbial communities



A  
B  
C  
D  
E  
F



Taxonomy

Biochemistry

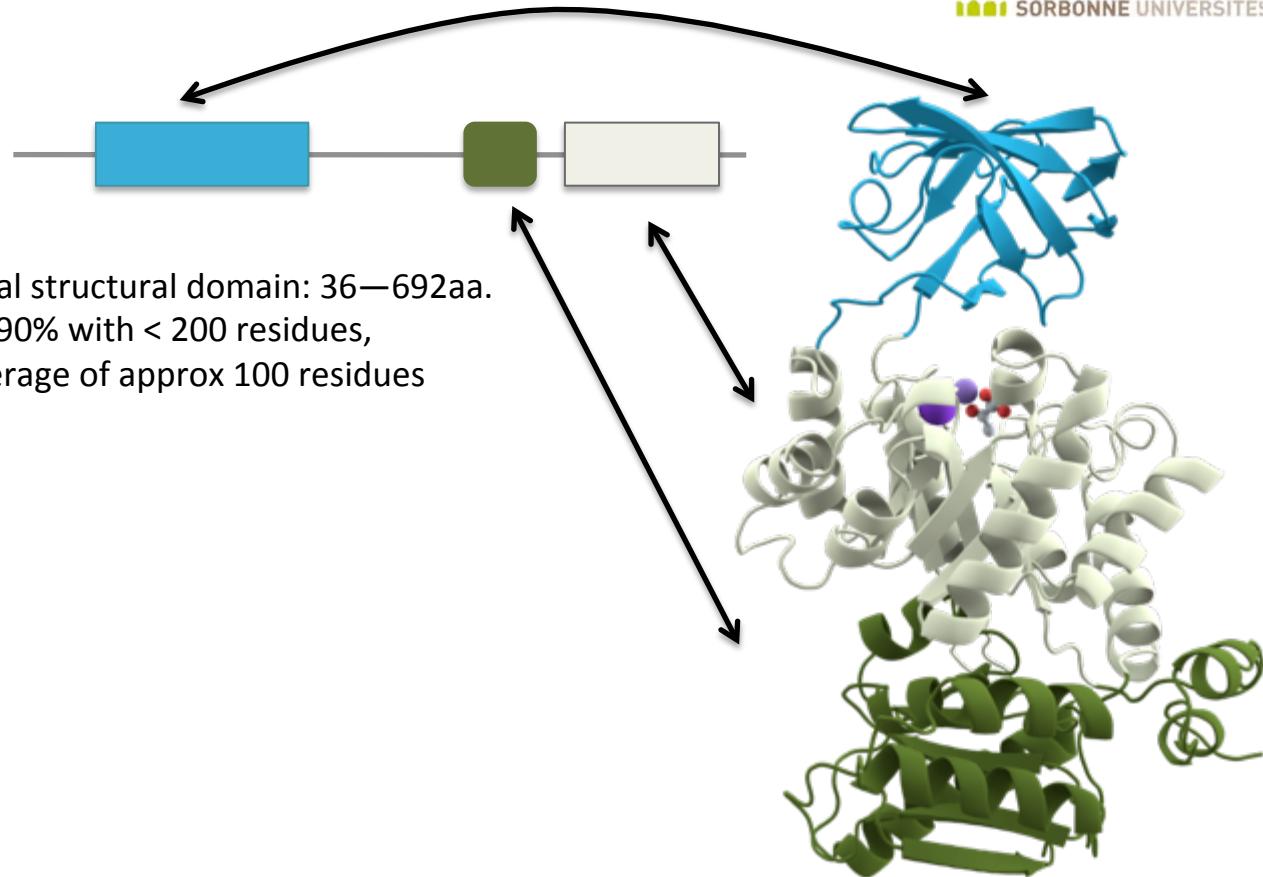
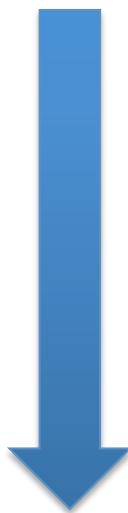
(on CDS)



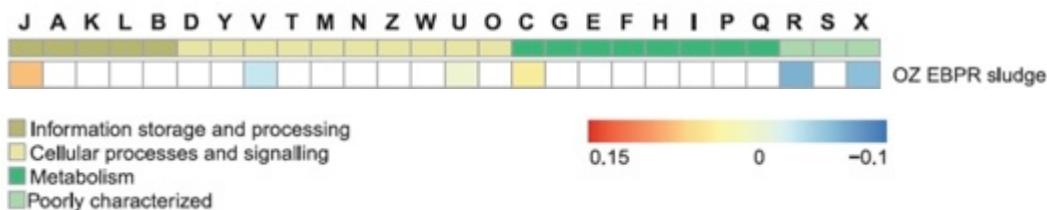
- What is the structure of the community?  
What is functionally done by the **community**?  
What is functionally done by the **subcommunities**?



## Domain annotation of CDS fragments



## Identification of the community metabolic preferences

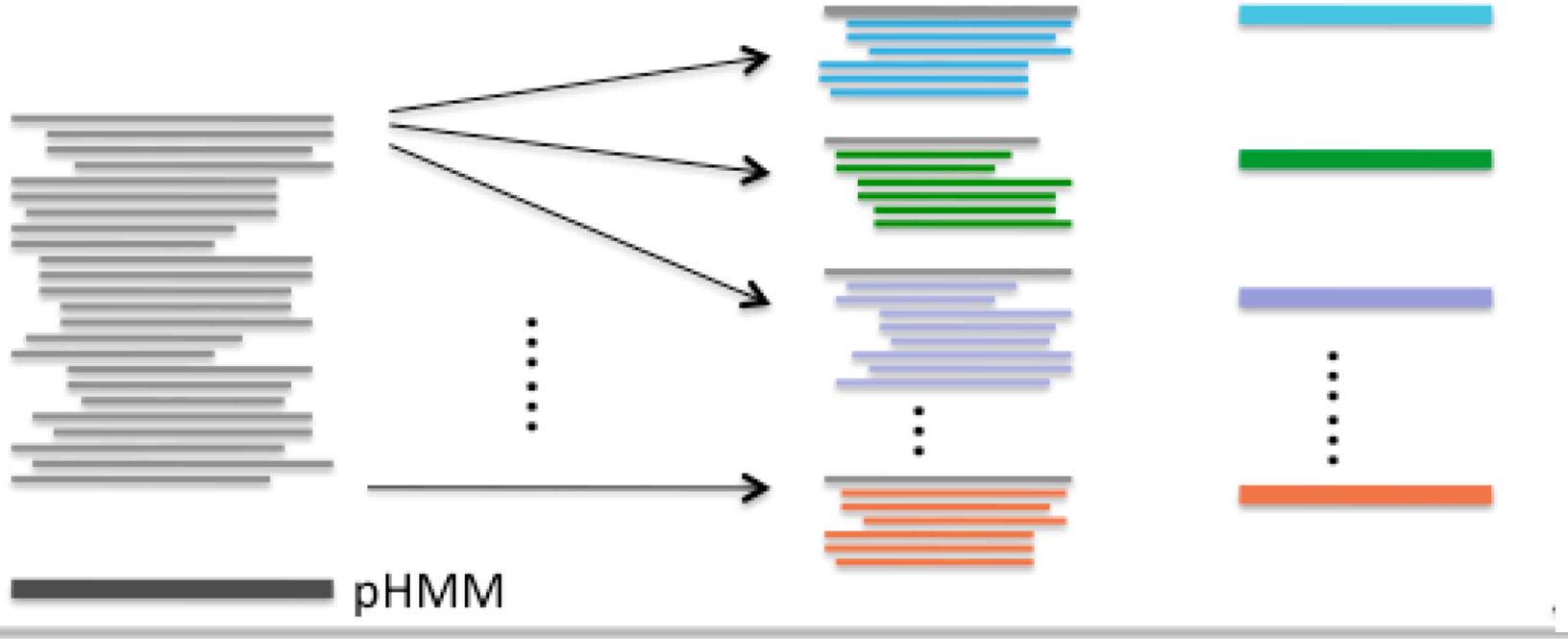


Genetic information is fragmented, lost, isolated, new.



# Domain models: profile HMM

clade-centered  
models (CCM)



PFAM domains  
**many species**

**one profile**

CLADE domains  
**many species**

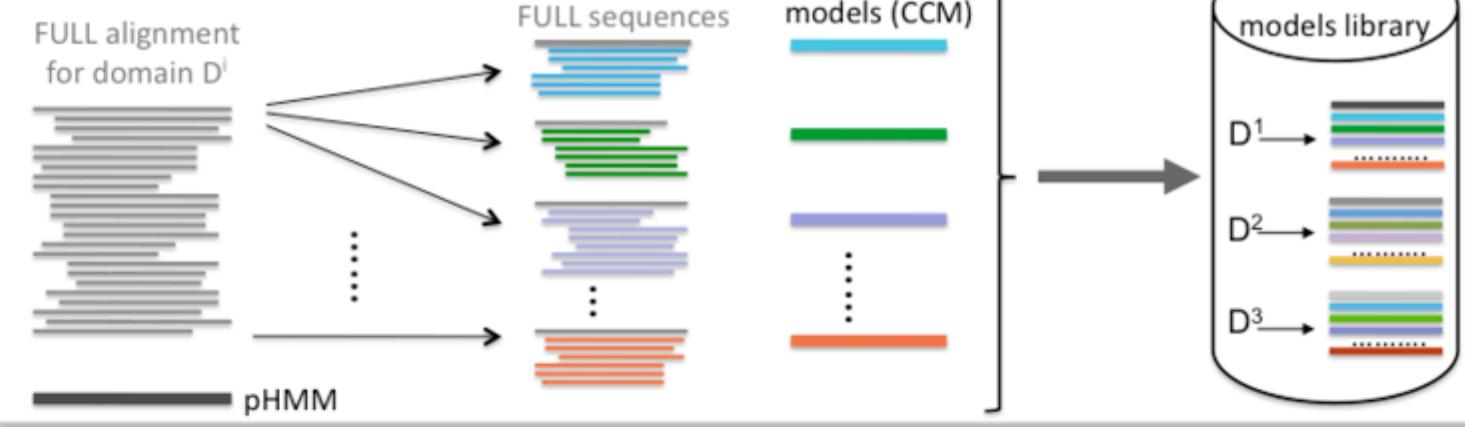
**population of profiles  
along the phylogeny**

Bernardes et al. Improvement in protein domain identification is reached by breaking consensus, with the agreement of many profiles and domain co-occurrence - **PLoS Computational Biology** (In Press).  
Bernardes et al. A multi-objective optimisation approach accurately resolves protein domain architectures. **Bioinformatics** (2015).



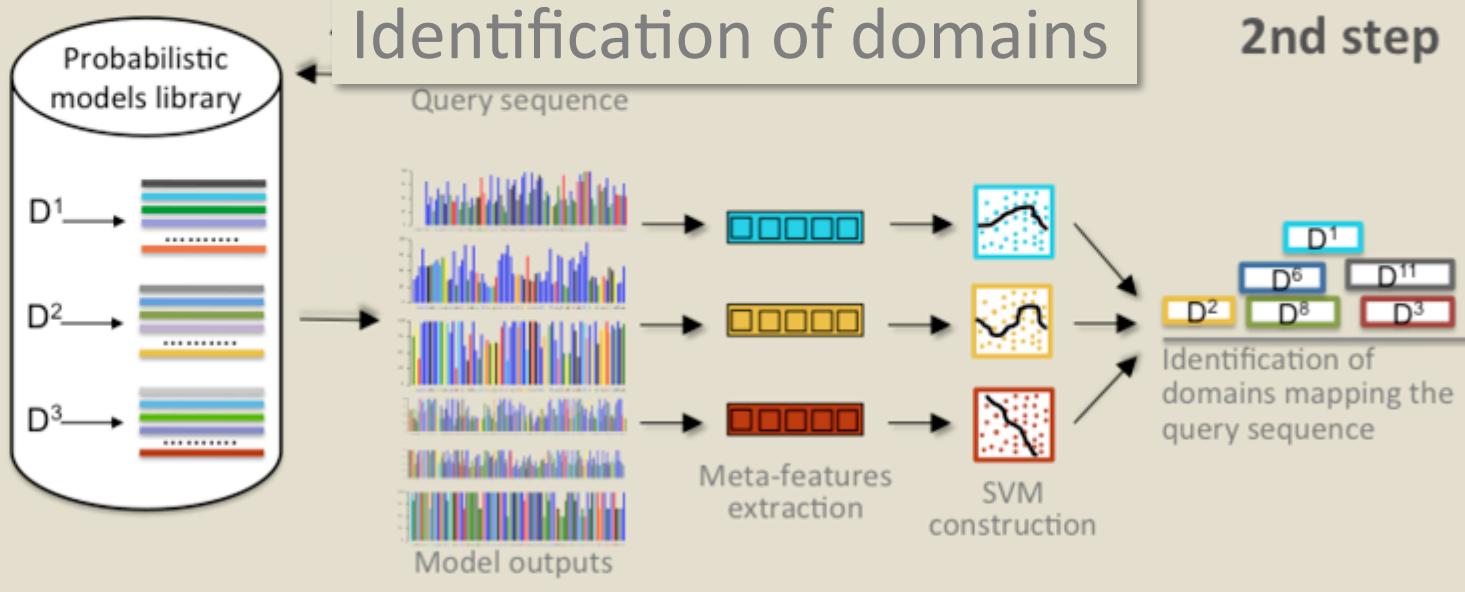
## 1st step

# Construction of the models



## 2nd step

### Identification of domains



## 3rd step

### Selection of domains



# Annotation of P. Falciparum whole genome

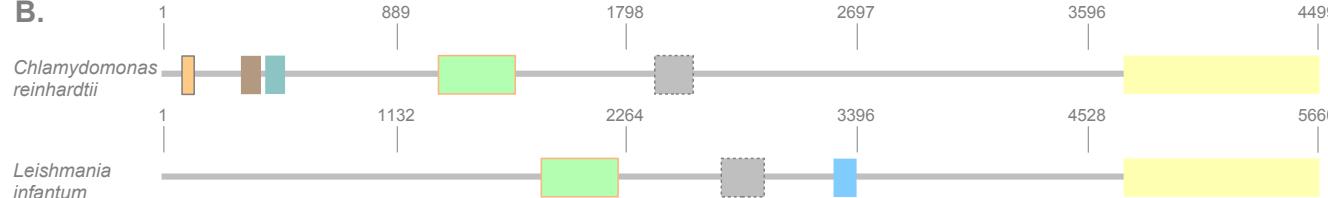
PF3D7 1122900 (PF11\_0240)



Pfam<sub>24</sub>

CLADE<sub>24</sub>

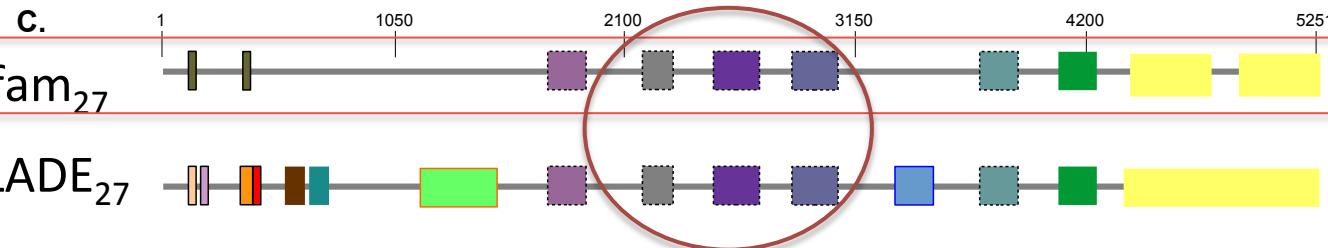
B.



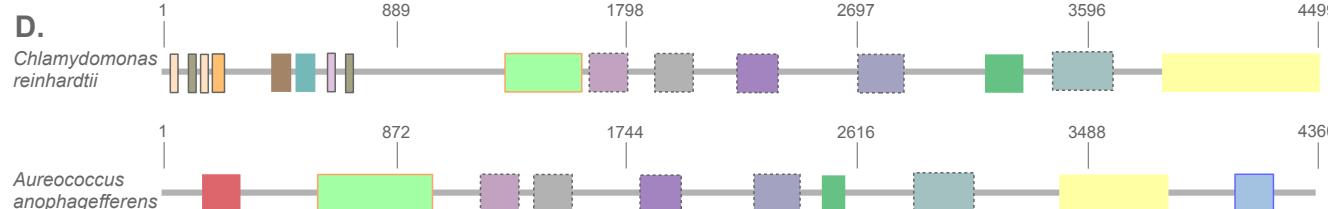
Pfam<sub>27</sub>

CLADE<sub>27</sub>

C.



D.



- Kelch\_1 → Kelch motif
- Kelch\_2 → Kelch motif
- Kelch\_3 → Kelch motif
- Kelch\_4 → Kelch motif
- Kelch\_5 → Kelch motif
- TIG → IPT/TIG domain
- Filamin → Filamin/ABP280 repeat

- DHC\_N2 → Dynein
- AAA → ATPases Associated proteins
- AAA\_5 → ATPases Associated proteins
- AAA\_6 → ATPases Associated proteins
- AAA\_7 → ATPases Associated proteins
- AAA\_8 → ATPases Associated proteins
- AAA\_9 → ATPases Associated proteins

- MT → Microtubule-binding stalk of dynein motor
- Dynein\_heavy → Dynein
- DUF947 → Domain of unknown function
- DUF572 → Domain of unknown function



# MetaCLADE for MetaSub

- More sensitive annotation of domains
- Annotation of domains co-occurrences (need longer fragments)
  - gene assembly from metagenomic data
  - correlation of coverage counts across samples
- Computing ressources:
  - Supercomputer IVANOE, ISCD@UPMC
  -  **world community grid**

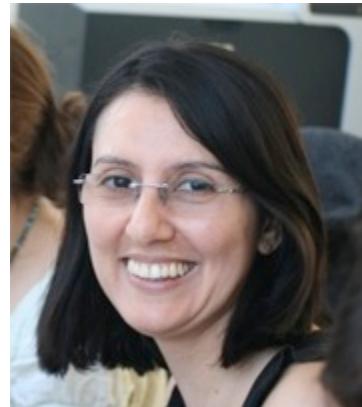


# Acknowledgments



**Alessandra  
Carbone**

Tina Alaeitabar (Master student UPMC, IR at INRA)  
Jinxin Wang and Jodie Gauvain (Master students, UPMC)



Juliana Bernardes

**CLADE**



Ari Ugarte

**metaCLADE**



## Ingrid Lafontaine and sampling team:

Bouly Jean-Pierre, Jaubert Marianne, Baudon François,  
Agier Nicolas, Delmas Stéphane , Bittner Lucie, Arnaud Meng,  
Anne-Sophie Benoiston, Quentin Letourneau, Carbajo Miguel,  
Sarah Shalaby, Florian, Manon Loubens, Desert Alexandre,  
Thiébaut Antonin, Fleiss Aubin, Collin Sylvie, Boeri Julia,  
Laine Elodie, Laurent David.

