

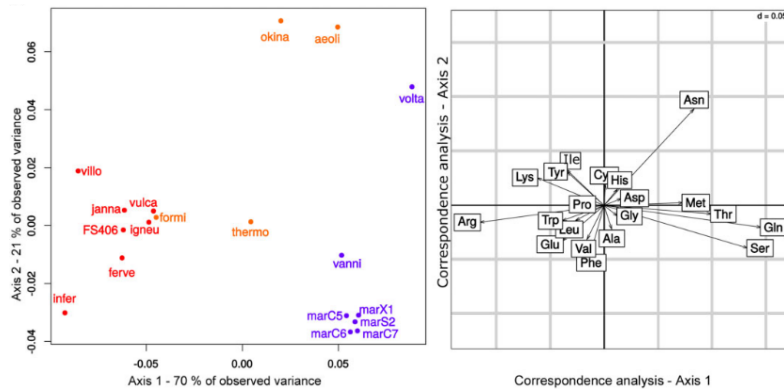
## Titre du stage :

# Étude des processus évolutifs impliqués dans la thermoadaptation des protéines

## Contexte

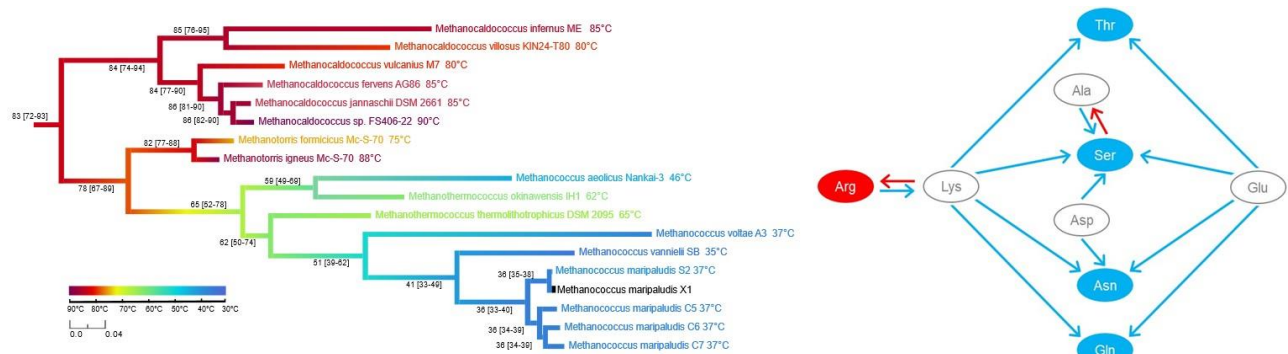
La vie sur terre s'est développée dans tous les environnements, y compris les plus extrêmes. La colonisation de ces environnements a nécessité des adaptations très spécifiques. Notre équipe cherche à comprendre les mécanismes impliqués dans l'adaptation des protéines à la température environnementale (thermoadaptation). Plusieurs études montrent, que chez les procaryotes, la température environnementale affecte fortement la composition en acides aminés des protéines : certains acides aminés sont surreprésentés ou sous-représentés dans les protéines en fonction de la température optimale de croissance (OGT) des organismes considérés [1, 2]. La difficulté majeure lorsqu'on s'intéresse à cette question est que l'OGT n'est pas le seul facteur ayant un impact sur la composition en acides aminés des protéines. Parmi les autres facteurs, la proportion de bases G et C dans les génomes (GC% génomique) est le facteur dominant, expliquant à lui seul plus de 50% de la variation de la composition en acides aminés des protéines, alors que l'OGT n'explique que 10 à 15%.

Une étude pilote menée par notre équipe a montré que **les archées du groupe des *Methanococcales* est un modèle très prometteur** [3]. En effet, nous avons montré que l'OGT est le facteur dominant, expliquant à lui seul plus de 70% de la variation de la composition en acides aminés des protéines de ces organismes (Fig. 1), alors que le GC% génomique n'en explique que 21%. Ceci s'explique par le fait que le GC% génomique est très stable au sein de ce groupe, ce qui limite l'impact de ce facteur sur l'abondance des acides aminés dans les protéines.



**Fig. 1:** Panel de gauche : Analyse des correspondances de l'abondance des acides aminés dans les protéines des *Methanococcales* (rouge = organismes hyperthermophiles, orange = thermophiles, bleu = mésophiles). La position des organismes le long de l'axe 1 est corrélée aux OGTs. À noter que « aeoli » a un mode vie à la limite entre mésophilie et thermophilie. L'axe 1 explique à lui seul 70% de la variance observée de la composition en acides aminés des protéines des *Methanococcales*, alors que l'axe 2, qui explique 21% de la variance, est corrélée au GC% génomique. Panel de droite : Distribution des acides aminés le long de ces deux axes mettant en évidence une forte corrélation entre l'abondance des acides aminés arginine, sérine, thréonine, glutamine et asparagine et les OGTs.

En utilisant la phylogénie moléculaire, la génomique comparée et la reconstruction de séquences ancestrales, nous avons identifié 5 acides aminés majeurs dont l'abondance dans les protéines est fortement corrélée à l'OGT chez les *Methanococcales* : l'arginine, la sérine, la thréonine, la glutamine et l'asparagine, et nous avons dévoilé le schéma substitutionnel sous-jacent [3] (Figs. 1 et 2). Ce schéma substitutionnel est organisé autour de la lysine (et dans une moindre mesure du glutamate) qui joue le rôle d'un hub central permettant de contrôler la fréquence des 5 acides aminés dont l'abondance est très fortement impactée par l'OGT (Fig. 2).



**Fig. 2.** Panel de gauche : Évolution de l'OGT des *Methanococcales* au cours de leur histoire évolutive. Panel de droite : Schéma substitutionnel associé aux changements d'OGT chez les *Methanococcales* (extraite de [3])

## Objectifs du stage

L'étude pilote réalisée chez les *Methanococcales* nous a permis de mettre en place une méthodologie rigoureuse permettant d'étudier le processus de thermoadaptation des protéines à la température environnementale chez les procaryotes.

Notre objectif est maintenant de **déterminer si le schéma substitutionnel observé chez les *Methanococcales* est spécifique de ce groupe d'archées ou, au contraire, si il est universel**. L'existence ou non d'un processus universel sous-tendant l'adaptation des protéines à la température environnementale fait l'objet de nombreux débats. Apporter une réponse à cette question représente à la fois un challenge et un enjeu fort. Pour apporter des éléments de réponses à cette question, nous nous proposons d'étendre notre études à d'autres groupes taxonomiques.

La première étape consistera à explorer les bases de données publiques (e.g. Tempura [4], ThermoBase [5], RefSeq [6]) afin d'identifier les groupes taxonomiques de bactéries ou d'archées pertinents pour aborder cette question. Ces groupes devront présenter les mêmes caractéristiques que les *Methanococcales*, à savoir : être constitués d'organismes ayant des OGT différents, mais ayant des GC% génomiques proches. L'étude des groupes taxonomiques retenus sera réalisée en appliquant une approche similaire à celle développée pour l'étude des *Methanococcales*.

La mise en œuvre du projet va nécessiter la mise en place d'un pipeline informatique permettant d'automatiser les différentes étapes de l'analyse afin de pouvoir analyser un grand nombre de groupes taxonomiques.

## Profil recherche

Nous recherchons un candidat en cinquième année ayant un intérêt fort pour la biologie évolutive et possédant des compétences en programmation et en statistiques, avec une maîtrise de Python et de R. Le contenu du stage pourra cependant être adapté pour un étudiant motivé de quatrième année.

## Modalités pratiques

Le stage s'inscrit dans le cadre du projet ANR ThermAdapt qui associe le Laboratoire de Biométrie et Biologie Évolutive (LBBE, UCBL Lyon), l'Institut Camille Jordan (ICJ, UCBL Lyon), le Laboratoire Microbiologie, Adaptation, Pathogénie (MAP, INSA Lyon) et l'institut de Biologie Structurale (IBS, Grenoble). Le stagiaire sera donc amené à interagir et à participer aux réunions des membres du consortium.

Le stage se déroulera au LBBE, sur le campus Lyon-Tech La Doua, dans l'équipe Bioinformatique, Phylogénie et Génomique Évolutive (<https://lbbe.univ-lyon1.fr/>).

## Contact

Céline Brochier-Armanet

Adresse électronique : [celine.brochier-armanet@univ-lyon1.fr](mailto:celine.brochier-armanet@univ-lyon1.fr)

Site web : <https://lbbe.univ-lyon1.fr/fr/annuaire-des-membres/brochier-armanet-celine>

## Références bibliographiques

- [1] B. Boussau, S. Blanquart, A. Necșulea, N. Lartillot, M. Gouy, Parallel adaptations to high temperatures in the Archaean eon, *Nature*, 456 (2008) 942-945.
- [2] M. Groussin, M. Gouy, Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea, *Molecular biology and evolution*, 28 (2011) 2661-2674.
- [3] M. Lecocq, M. Groussin, M. Gouy, C. Brochier-Armanet, The Molecular Determinants of Thermoadaptation: Methanococcales as a Case Study, *Molecular biology and evolution*, 38 (2021) 1761-1776.
- [4] Y. Sato, K. Okano, H. Kimura, K. Honda, TEMPURA: Database of Growth TEMPeratures of Usual and RAre Prokaryotes, *Microbes Environ*, 35 (2020).
- [5] J. DiGiacomo, C. McKay, A. Davila, ThermoBase: A database of the phylogeny and physiology of thermophilic and hyperthermophilic organisms, *PloS one*, 17 (2022) e0268253.
- [6] W. Li, K.R. O'Neill, D.H. Haft, M. DiCuccio, V. Chetvernin, A. Badretdin, G. Coulouris, F. Chitsaz, M.K. Derbyshire, A.S. Durkin, N.R. Gonzales, M. Gwadz, C.J. Lanczycki, J.S. Song, N. Thanki, J. Wang, R.A. Yamashita, M. Yang, C. Zheng, A. Marchler-Bauer, F. Thibaud-Nissen, RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation, *Nucleic acids research*, 49 (2021) D1020-D1028.