# M2 - Exploring the genomic diversity of representative species of Archaea

Genome-resolved metagenomics have profoundly reshaped our understanding of the distribution, functionalities and roles of Archaea. Within the domain, major supergroups are Euryarchaeota, which includes many methanogens, the TACK, which includes Thaumarchaeaota that impact ammonia oxidation in soils and the ocean, the Asgard, which includes lineages inferred to be ancestral to eukaryotes, and the DPANN, a group of mostly symbiotic small-celled archaea. These archaea are not restricted to extreme habitats, but are widely distributed in diverse ecosystems [1-4].
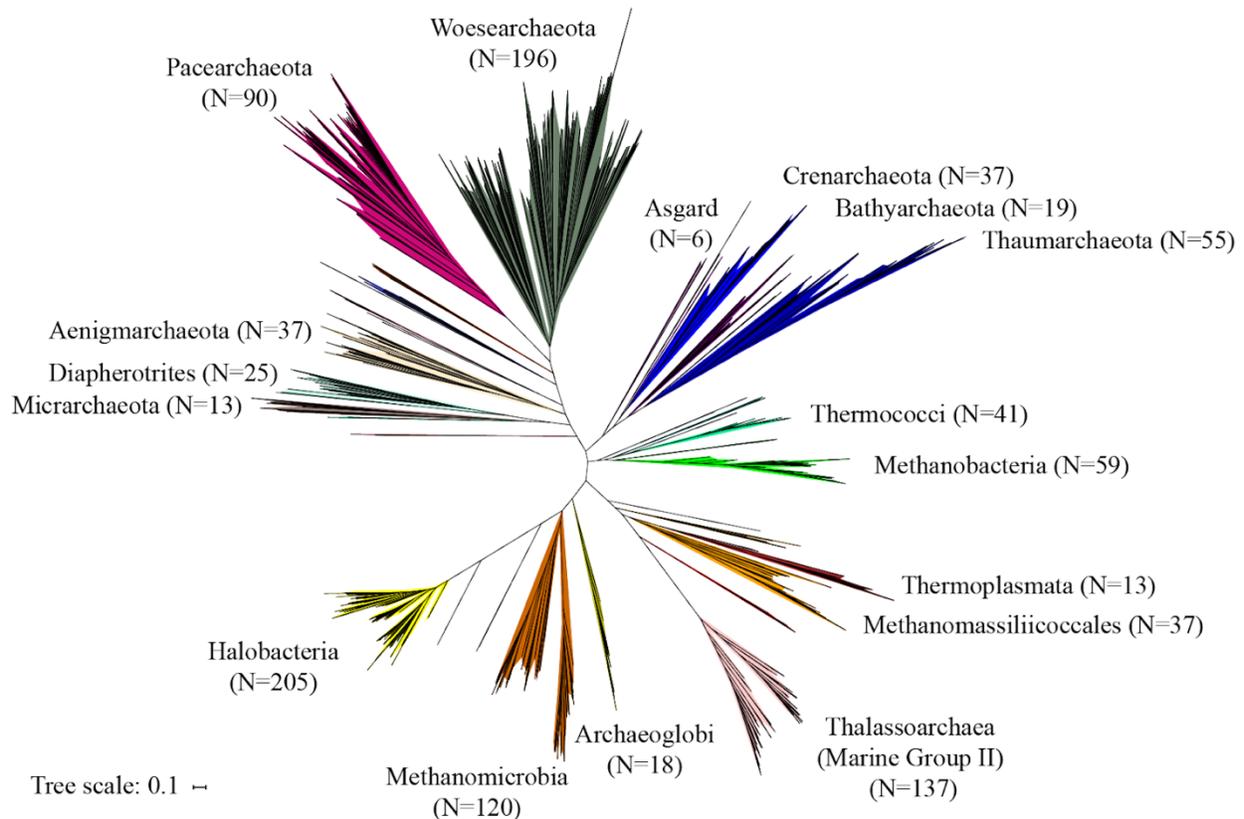
We now have enough data to explore the intra-species genetic diversity of several archaeal species. We propose to explore the pangenomes of several archaeal species using the tool PPanGGOLiN we recently developed in our lab [5]. The objective will be to define sub-species groups based on their gene content and then investigate associations between ecology and metabolic capacities using read recruitment from metagenomic data [6]. Another project would be to focus on the variable part found in the genomes of several archaeal species using the tool PanRGP to detect genomic islands [7]. Most of the work will consist in exploring the functions of the genes nested within the genomic islands. One objective will be to identify novel genes and pathways with features suggestive of utility for genome editing and biotechnology. We also are interested in the discovery of new proviruses [8].

We are looking for a highly motivated student in microbiology, ecology, genomics or bioinformatics. The successful candidate will be responsible for the pangenome analysis and the

functional annotation of the detected genomic islands. She/He will be helped by the tools developed in the lab as well as the expertise of the LABGeM team on microbial genomics (https://labgem.genoscope.cns.fr). As the internship is fully bioinformatics-focused, a minimal set of skills in scripting and data manipulation would be highly appreciated (e.g bash, R, python...). The start and end dates of the internship can be changed.

For more information, you may contact Raphaël Méheust (raphael.meheust@gmail.com) and David Vallenet (vallenet@genoscope.cns.fr). The position will be located at the Genoscope (http://jacob.cea.fr/drf/ifrancoisjacob/Pages/Departements/Genoscope.aspx) in Evry.

References:

1. Adam, P. S., Borrel, G., Brochier-Armanet, C., & Gribaldo, S. (2017, November 1). The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. ISME Journal, Vol. 11, pp. 2407–2425.
2. Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., & Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. Nature Microbiology, 1–14.
3. Spang, A., Caceres, E. F., & Ettema, T. J. G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science.
4. Méheust, R., Castelle, C. J., Jaffe, A. L., & Banfield, J. F. (2020). Early acquisition of conserved, lineage-specific proteins currently lacking functional predictions were central to the rise and diversification of archaea. BioRxiv, 2020.
5. Gautreau, G., Bazin, A., Gachet, M., Planel, R., Burlot, L., Dubois, M., … Vallenet, D. (2020). PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. PLOS Computational Biology, 16(3), e1007732.
6. Delmont, T. O., Kiefl, E., Kilinc, O., Esen, O. C., Uysal, I., Rappé, M. S., … Eren, A. M. (2019). Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. ELife, 8.
7. Bazin, A., Gautreau, G., Médigue, C., Vallenet, D., & Calteau, A. (2020). panRGP: a pangenome-based method to predict genomic islands and explore their diversity. BioRxiv, 2020.03.26.007484. Bioinformatics, in press, doi:10.1093/bioinformatics/btaa792
8. Prangishvili, D., Bamford, D. H., Forterre, P., Iranzo, J., Koonin, E. V., & Krupovic, M. (2017, December 1). The enigmatic archaeal virosphere. Nature Reviews Microbiology, Vol. 15, pp. 724–739.