A pan-genome view of the Asgard archaea, the closest relatives of eukaryotes

Saioa Manzano-Morales[1,2], Toni Gabaldón[1,2,3],
[1]Barcelona Supercomputing Centre (BSC-CNS), Plaça Eusebi Güell, 1-3, 08034, Barcelona, Spain
[2]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028, Barcelona, Spain
[3]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain.

INTRODUCTION

Asgard archaea were only discovered in the last decade and have since been broadly accepted as the closest relatives of eukaryotes, yielding important insights into the evolutionary origin of eukaryotic cells. The availability of over 200 Asgard group genomes provides an opportunity to analyze the diversity of this clade of prokaryotes from a pangenome perspective. A pangenome, which can be defined as the non-redundant set of all genes (clusters of orthologs) found in all genomes of a taxon, and is potentially altered by both habitat and phylogeny. The main processes shaping pangenomes are gene duplication and loss during vertical inheritance and gene acquisition via horizontal gene transfer (HGT). Therefore, a pangenome analysis, coupled and complemented with phylogenomics, can yield unprecedented insight into the evolutionary forces shaping Asgard genomes.

We report a **pangenomic analysis of the Asgard group,** with focus on the *Candidatus Prometheoarchaeum syntrophicum* proteome. We further analyze the **pangenome architecture** of the Asgard group, assessing its **openness**, as well as characterizing the patchiness of the presence of **Eukaryotic Signature Proteins** (ESPs) across the organisms of this clade. Additionally, we analyze the **Gene Ontology (GO) enrichment** of protein functions for the core and accessory gene categories.

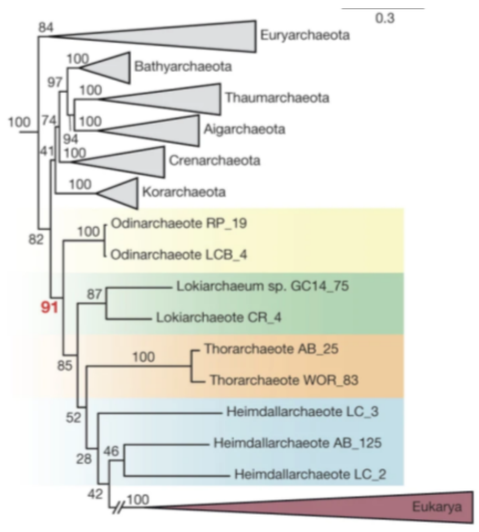

Figure 1: Asgard archaea as the closest clade to eukaryotes (Zaremba-Niedzwiedzka et al., 2017)

RESULTS

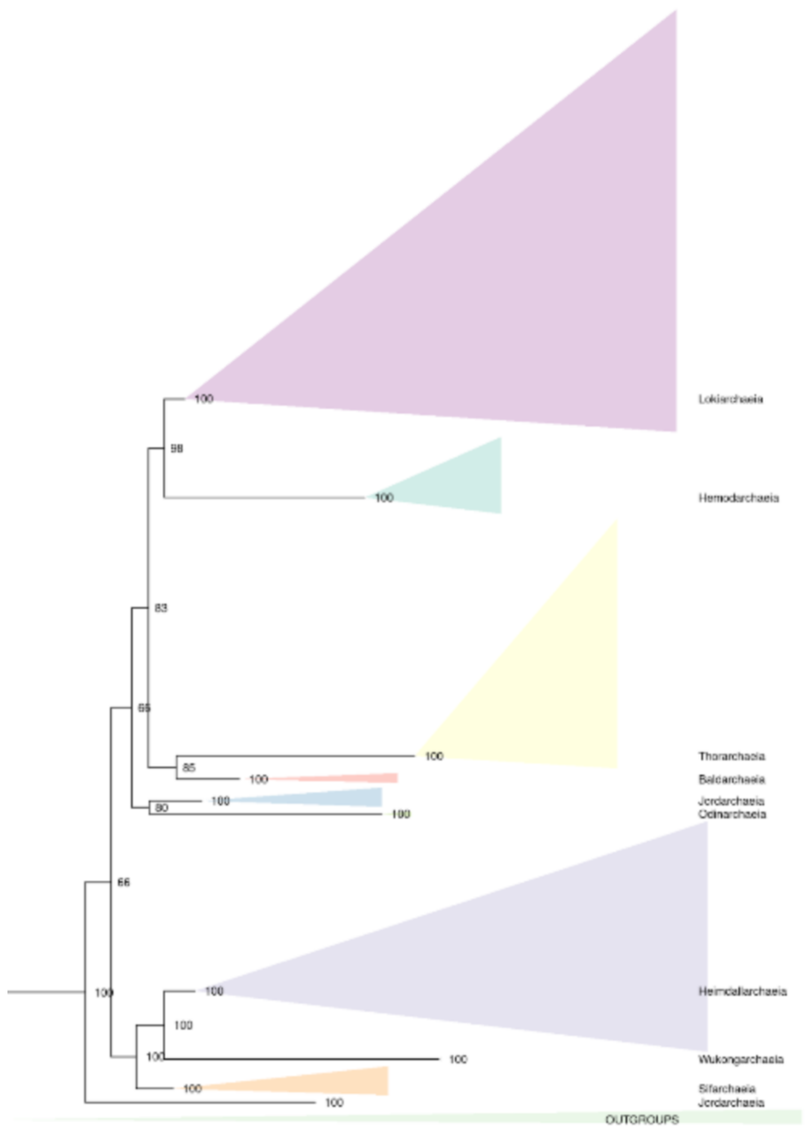

Figure 3: Concatenated marker tree of the Asgard clade. Phylogenomic tree of the Asgard archaea built by aligning with MUSCLE, trimming with trimAl and concatenating the 53 GTDB markers, reconstructed with iQTree.

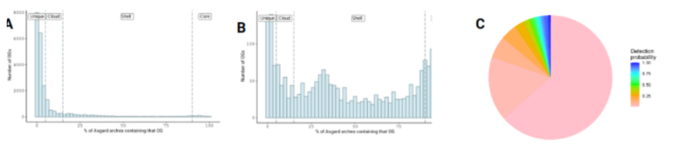

Figure 3: Asgard group pangenome composition. A) Orthogroup size distribution of the OGs in the Asgard pangenome. It is composed of 22850 families, most of them of small size. B) Orthogroup size distribution of the OGs from the Asgard pangenome present in Prometheoarchaeum syntrophicum, showing a bimodal distribution. C) Pan-genome family probability distribution. The optimal is 16 components, estimating a core of 13 and a predicted size of 23957.

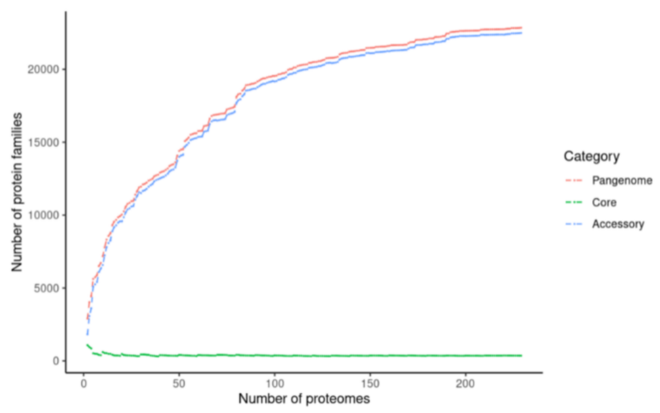

Figure 4: Pangenome openness. Number of protein families with each increased addition of a proteome, for the total pangenome (red), the core (green) and the accessory (blue). The pangenome appears open (Heaps law alpha=0.72) with the estimation from Chao's lower bound being 23132

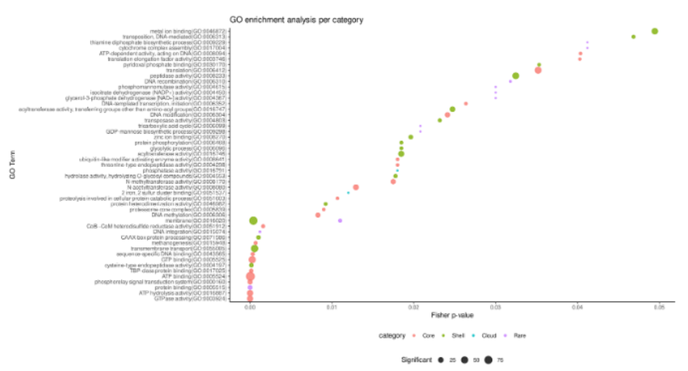

Figure 5: GO enrichment. Enriched GO terms for the core and accessory components, indicating the ontology category, the number of significant proteins and the Fisher p-value. Most of the terms enriched in the core are related to DNA and metabolism (interestingly, methanogenesis)

96 *Giardia* ESPs have hits, in **196 orthogroups**, with the vast majority of hits being **grouped in the same orthogroup** (63/96). Out of these orthogroups, **almost all (129) are core,** with only 15 belonging to the shell category and 39 being rare. In some cases, these hits belonging to rare orthogroups may be due to over-splitting of a core OG by orthofinder, but there seem to be cases where the ESP has **hits only to proteins** belonging to a OG of the **rare category** and therefore the ESP is *bona fide* infrequent in the Asgard group. 25 of the ESPs have hits in *P. syntrophicum,* 24 of which belong to a COG in the core category. This further highlights that the **ESPs seem to be prevalent in the Asgard group.**

This research was supported by Gordon and Betty Moore Foundation (Grant GBMF9742)

METHODS

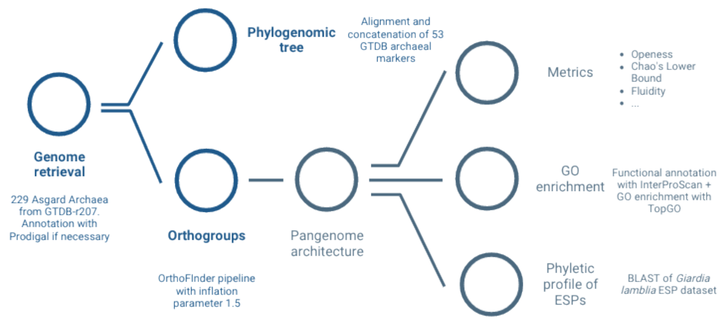

FIgure 2: Work pipeline. 229 Asgard proteomes were retrieved from GTDB. After that we built the phylogenomic tree, generated Orthogroups with OrthoFinder and analyzed the pangenome architecture

CONCLUSIONS

- The Asgard group pangenome **appears open**, indicating that sequencing of new members and of new taxonomic clades will result in an increase of the pangenome: the **full genomic (and metabolic) potential** of the Asgard archaea may be yet **unrevealed,** despite core genes being identified.
- Members of the Asgard group show a **remarkable degree of divergence,** with a small amount of proteins being present in all: this may be artefactual and as more genomes are sequences COGs will increase in size and the bimodal distribution will be recovered
- The Asgard **core** is enriched in functions related to DNA and basal metabolism
- Most of the **ESPs** are found in the **core genome,** revealing a pervasiveness of these proteins across the Asgard group

REFERENCES

1. Imachi, H., M. K. Nobu, N. Nakahara, Y. Morono, M. Ogawara, Y. Takaki, Y. Takano, et al. 2020. "Isolation of an Archaeon at the Prokaryote-Eukaryote Interface." *Nature* 577 (7791)
2. Parks, Donovan H., Maria Chuvochina, Christian Rinke, Aaron J. Mussig, Pierre-Alain Chaumeil, and Philip Hugenholtz. 2021. "GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy." *Nucleic Acids Research* 50 (D1): D785–94.
3. Zaremba-Niedzwiedzka, K., E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, et al. 2017. "Asgard Archaea Illuminate the Origin of Eukaryotic Cellular