

Annotation of the *Nanoarchaeum equitans* Kin4-M Genome from Locus Tags NEQ184 to NEQ186

Roman Novy-Marx, Margot Hilyard, Chuck Rutberg and Peter Hentschke
The Harley School Brighton, NY and the Western New York Genetics in Research and Health Care Partnership

Abstract

A group of three genes (NEQ184, NEQ185 and NEQ186) from the marine Archaea, *Nanoarchaeum equitans*, were annotated using the genome annotation website GENI-ACT. *Nanoarchaeum equitans* live in hydrothermal vents off the coast of Iceland. Each of the three genes was assessed for their general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. Searches from various databases (such as nr and SwisProt) for similar sequences gave possibilities of the gene's functions. Later, evidence for the location of the proteins were found using TMHMM, SignalP, PSORT-B, and other applications. Ultimately, we provide predictions for the gene products.

Introduction

The habitats of *Nanoarchaeum equitans* are submarine hydrothermal vents globally, alongside *Ignicoccus hospitalis*. *N. equitans* has no metabolic system of its own, and know way of producing small-molecule cellular precursors (lipids, amino acids, sugars, and nucleotides). Instead, it relies on physical contact with *I. hospitalis* for these necessities. The process by which these small-molecules are transferred between the species is still unknown. Their relationship has been labeled as both symbiotic and parasitic depending on the study, after finding that in lab conditions, *N. equitans* could not be propagated without the presence of *I. hospitalis*, and the effects of this relationship on *I. hospitalis* were negative, or neutral at best.

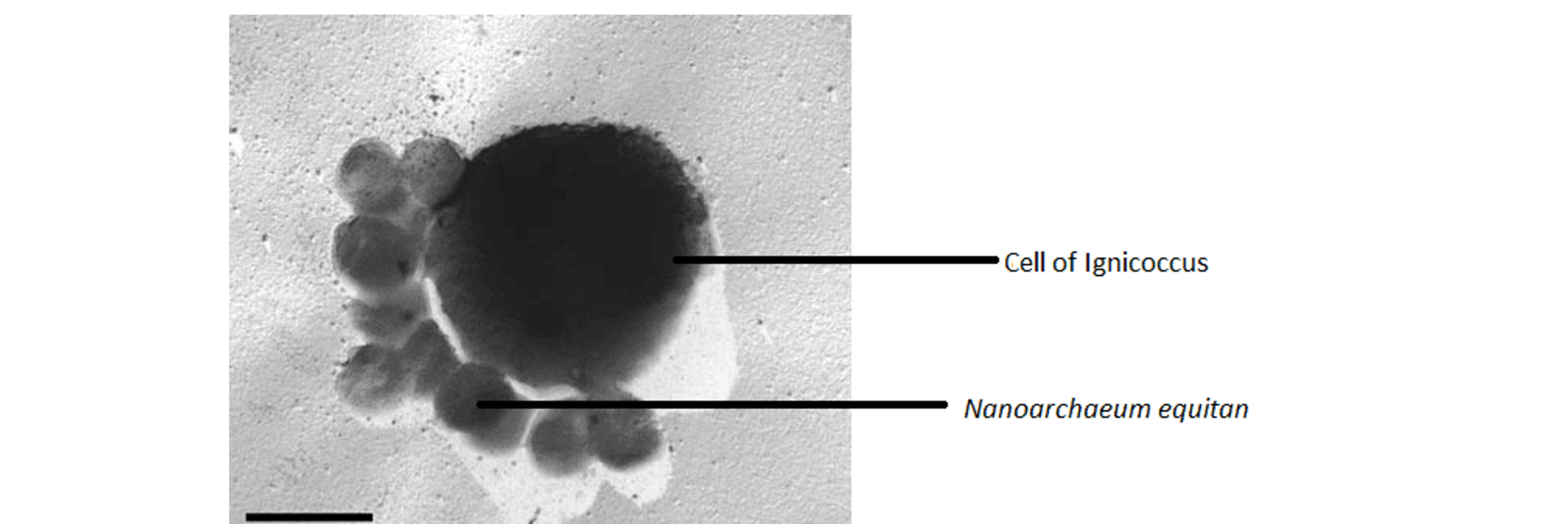


Figure 1 - Image of *N. equitans* attached to Ignicoccus cell.

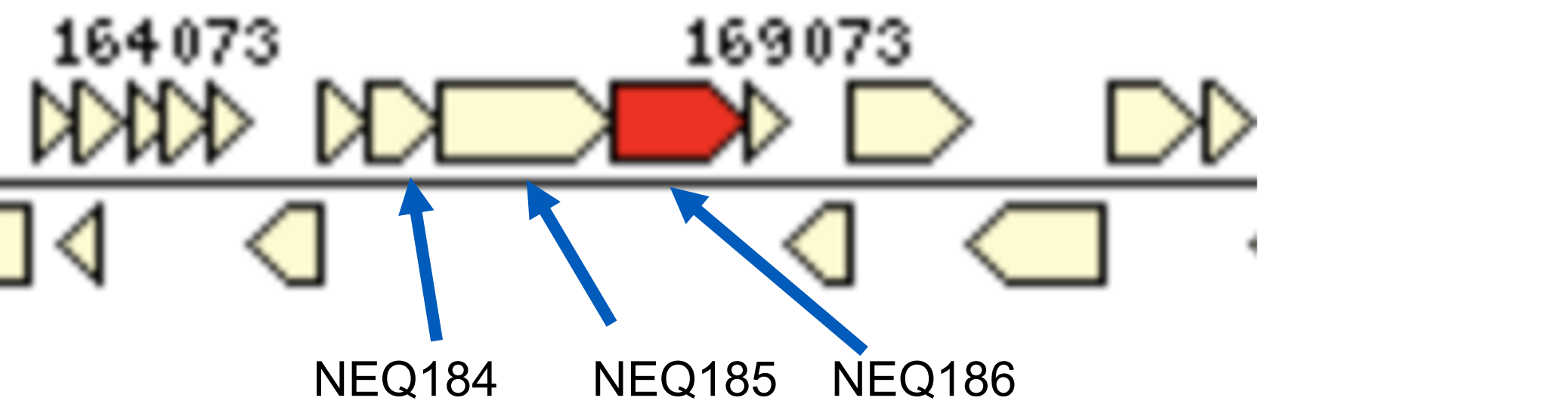


Figure 2 - The locus tags and relative position of the genes under investigation in this research.

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Nanoarchaeum equitans* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

NEQ184:
The gene has coordinates of 166567..i67163, making its length 597 nucleotides or 198 amino acids. Its first BLAST hit, which was found in the nr database was exosome complex component Rrp4, from the organism *Archaeoglobus fulgidus* DSM 4304. Next, through the programs T-COFFEE and WebLogo, it was gathered that the gene has not been well conserved. After that, Pfam, TIGRFAM, and PDB were used to determine the potential structure of the protein. Finally, TMHMM and SignalP were used to test for transmembrane helices and signal peptides, respectively, neither of which were found before LipoP, PSORT-B, and Phobius helped form the hypothesis that the protein is cytoplasmic.

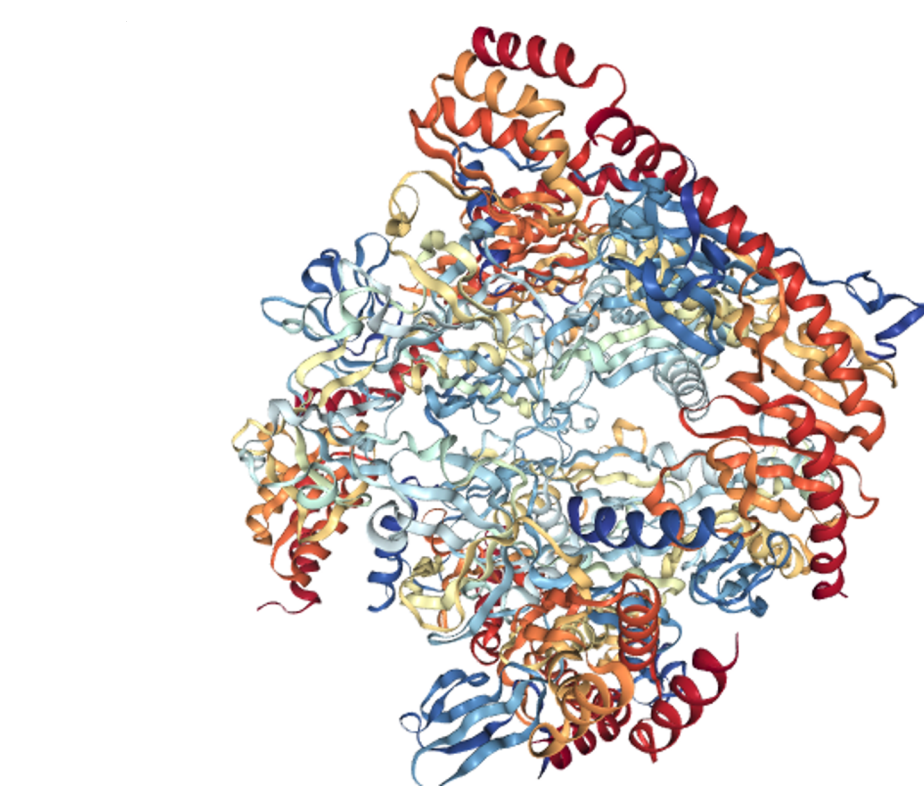


Figure 3 - This is a 3D rendering of the top PDB hit for NEQ184. It is called archaeal exosome core and also comes from *Archaeoglobus fulgidus*, just like the top Blast hit for the protein.

NEQ185:
This gene's coordinates are 167163..168512. The nucleotide sequence was predicted to be 1350 bases long, and the amino acid sequence 499 amino acids long. BLAST results indicated the most similar proteins to be from the organism *Nanoarchaeota archaeon*, and relating to charging tRNA(Asn). Then, TIGRFAM, Pfam, and PDB were used to determine the structure of the protein and structurally related proteins. Results from TMHMM and SignalP indicated that there were no transmembrane helices or signal peptides present in the protein, meaning that the protein is most likely not found in the membrane, or excreted from the cell. Results from LipoP, PSORT-B, and Phobius confirmed that the protein most likely resides in the cytoplasm.

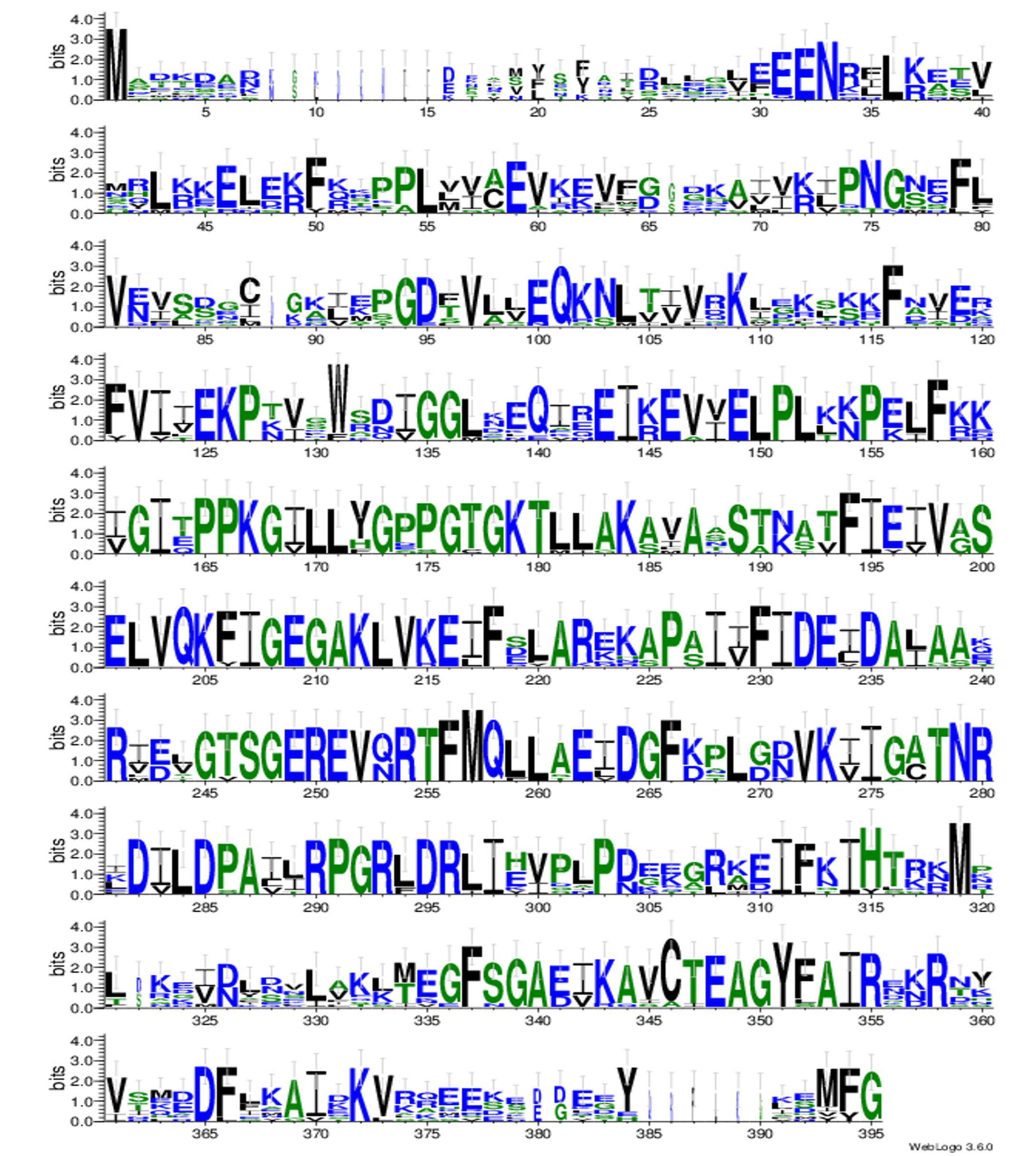


Figure 4 – WebLogo of NEQ186 and the top 10 BLAST hits of similar sequences. At the start and the end of the sequences, particularly on the N-terminal half of the first row and the C-terminal half of the last row, the sequences are not well conserved. However, on average, there is much more conservation of the genes over the next 8 rows.

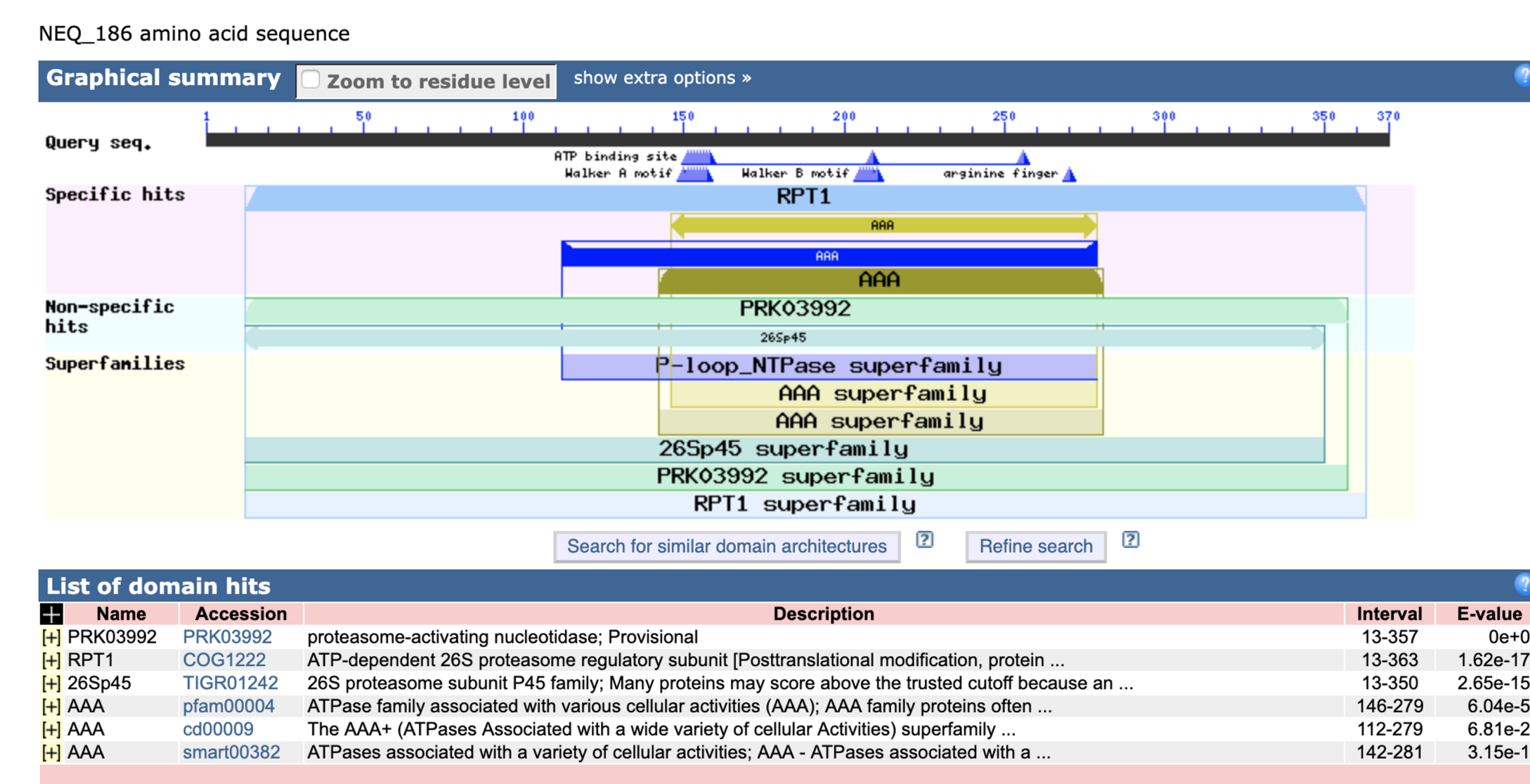


Figure 5 – Conserved domain results for NEQ186. Top hits include three ATPases, which is the predicted function of this gene.

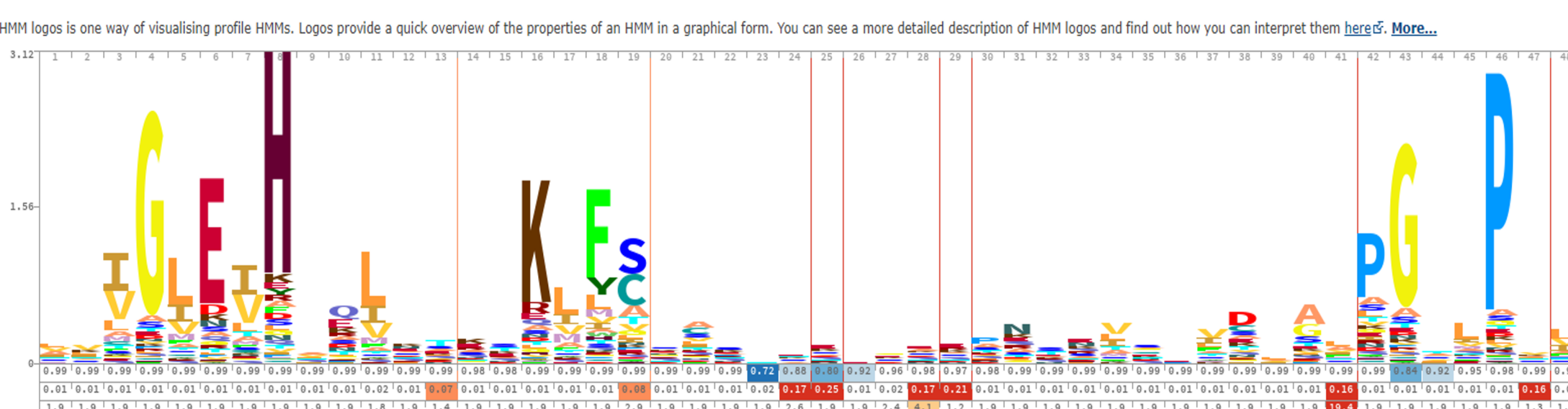


Figure 6 – HMM logo from the Pfam results of NEQ185. Shows minimal conservation.

NEQ186:
This gene has the coordinates 168505..169614 and is located on the top strand as predicted by GENI-ACT. The sequence is 1110 nucleotides or 369 amino acids long. BLAST results of similar sequences proposed the product of this gene to be an ATPase, which is an enzyme that catalyzes the decomposition of ATP into ADP and a free phosphate. Top CDD hits of this gene also supported it's function to be an ATPase. Using TMHMM and SignalP, NEQ186 was determined to have no transmembrane helices and no signal peptides. This ruled out the possibility of the protein being a membrane protein or secreted, meaning it would likely be located in the cytoplasm. PSORT-B results confirmed this possibility giving the protein created by the gene to have a very high chance of residing in the cytoplasm.

Conclusion

Proposed gene products were found from BLAST and CDD hits of similar sequences to the query sequences. These proposed functions were determined from definite functions of genes found to be similar when aligned to NEQ184, NEQ185, and NEQ186. The gene products proposed for these genes are shown below.

Locus Tag	Proposed Products
NEQ184	ribosomal RNA-processing protein RRP4
NEQ185	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
NEQ186	Proteasome-activating nucleotidase or ATPase

References

- Wikipedia., *Nanoarchaeum equitans*, BioLib:163391
- Giannone et al., 2014. Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. ISME J. Volume 9(1):101–114. doi:10.1038/ismej.2014.112