Annotation of the *Nanoarchaeum equitans Kin4-M* Genome from Locus Tags NEQ316 to NEQ318

Caroline McGraw, Rafael Ferguson, Lucy Ray and Peter Hentschke
The Harley School, Rochester New York and the Western New York Genetics in Research and Health Care Partnership

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

Abstract

A group of consecutive 3 genes from the microorganism *Nanoarchaeum equitans* (NEQ316 to NEQ318) were annotated using the collaborative genome annotation website GENI-ACT. GENI-ACT did not propose a gene product name for each gene. Each sequence was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. Gene products are proposed based on this research.

Introduction

Nanoarchaeum equitans is known as the smallest known of archaeon. It was discovered by Karl Stetter in 2002 in a hydrothermal vent near the coast of Iceland. It tends to grow best in almost boiling environments of 80℃ and environments with a pH of around 6. It is very similar to an ectoparasite, meaning it strictly relies on physical contact with its host to obtain important molecules for metabolism. *N. equitans* likes to occupy itself in *Ignicoccus hospitalis*, which is the only known example of a specific pairing between two Archaea species. How they recognize and pair up with each other is unknown. But it is known that they facilitate gene transfer within each membrane. *N. equitans* obtains its nucleotides, amino acids, and lipids from its host cell, whereas *I. hospitalis* does not get any benefits from their pairing (Heimerl, 2017). Some experiments have been done to uncover more about this relationship, however not much has explained how or why these two organisms pair up with each other. As a result, *N.equitans* still remains a changing and mysterious organism.

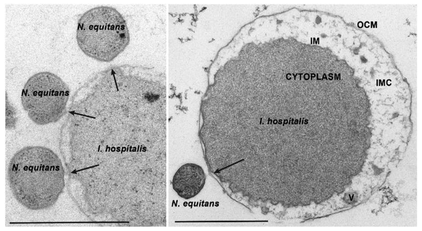

Figure 1 – A TEM of the interaction between *N.equitans* and *I.hospitals* (Dombrowski, 2019)

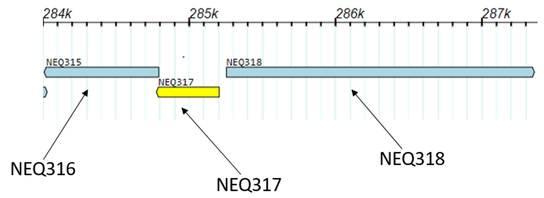

Figure 2. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Nanoarchaeum equitans Kin4-M* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What are the coordinates, DNA sequence, and Amino Acid sequence of your protein?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank? What types of proteins is yours most similar to?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	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? Most likely in the cytoplasm, transmembrane, non cytoplasm, or signal peptide?

Results

NEQ316:

This gene is located at coordinates 283658...284254., and it codes for a sequence of 198 amino acids. The BLAST search identified the gene product name to be dCTP Deaminase. In Module Three, TIGERFAM and Pfam both had similar results to what Geni-act had proposed for the characteristics of their families. The top PDB hit (PDB Code 2QLP) is for a Bifunctional dCTP deaminase. (Figure 3). The applications in Module 4 however, such as TMHMM, SignalP, PSORT-B and Phobius showed no detected transmembrane helices, signal peptides, or a specific location of the sequence. Therefore it is hypothesized that this protein lies in the cytoplasm of the cell.

Figure 3. The proposed crystal structure of NEQ316 from the PDB.

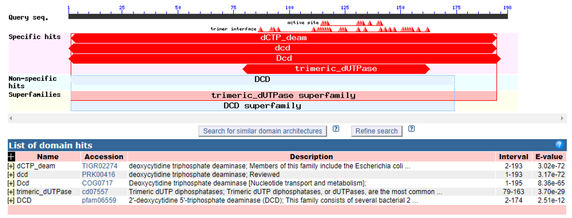

Figure 4. The CDD results for NEQ316.

NEQ317:

This gene has the coordinates 285014..285430. The sequence is 138 amino acids long. The top BLAST hit identified the gene product as a 50S ribosomal protein L15P (uncultured archaeon). CDD, T-Coffee, and WebLogo all supported this. This sequence did not have any TIGRFAM hits, but the one Pfam hit it had also said it is most likely a ribosomal protein. TMHMM predicted 0 transmembrane helices. SignalP found no likely cleavage site, and LipoP found less than 4 putative cleavage sites. SignalP also predicted that its likelihood of being one of the three main protein types (Signal peptide, TAT signal peptide, and Lipoprotein signal peptide), it said that is was most likely to be none of those and a different, unstated one. LipoP, PSORTb, and Phobius all suggested that this sequence was found in the cytoplasm.

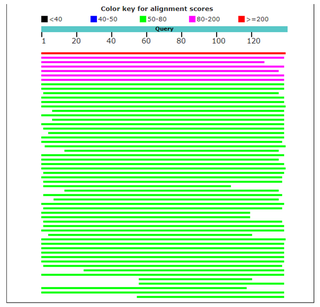

Figure 5. The nr BLAST results for NEQ317, which show scores that are moderately significant.

NEQ318:

This gene has the coordinates 285479..287589. The sequence is 701 amino acids long. The first two BLAST hits identified the gene product as a reverse gyrase. The CDD and TIGRFAM agreed with this result, however my Pfam suggested *DNA topoisomerase*. My TMHMM predicted 0 transmembrane helices. SignalP found no likely cleavage site, and LipoP found less than four putative cleavage sites. SignalP also predicted that it is highly unlikely to be any of the main protein types, leaving a solid line predicting another unknown protein. PSORTb suggests that the gene is located in the cytoplasm. Phobius confirmed the absence of transmembrane helices and signal peptides, supporting the assumption that it is cytoplasmic.

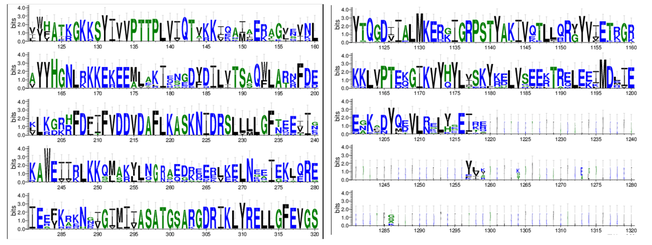

Figure 6- WebLogo Multiple Sequence Alignment for NEQ318 with blue as basic amino acids, green as polar, and black as hydrophobic (note absence of acidic amino acids).

Conclusion

Based on the work done analyzing these genes, the proposed annotation for the gene products are shown in the table below.

Locus Tag	Proposed Annotation
NEQ316	dCTP deaminase
NEQ317	LSU ribosomal protein L15P
NEQ318	Reverse gyrase

References

Heimerl et al., 2017. A complex endomembrane system in the archaeon *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. Front Microbiol 8:1072.

Nina Dombrowski et al., 2019. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiology Letters 366(2) January 2019, fnz008, https://doi.org/10.1093/femsle/fnz008

University at Buffalo The State University of New York
Jacobs School of Medicine and Biomedical Sciences
buffalo.edu

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209