Assessing the Evolutionary Origins of the Inorganic Phosphate-Metal Cotransporters PitA and PitH

Hugues Sayer
hs252@st-andrews.ac.uk

Supervised by Dr Joanne Boden
j.boden@st-andrews.ac.uk

School of Earth and Environmental Sciences
University of St Andrews

Contents

1. Introduction

There are 6 essential elements for life, commonly summarised by the acronym 'CHNOPS' which stands for Carbon, Hydrogen, Nitrogen, Oxygen, Phosphorus, and Sulfur. These make up the molecules found in all living organisms on Earth, each playing a distinct role in many biological processes. Without these elements, life would not exist. This research project focuses specifically on phosphorus, an important component of genetic material, cell membranes, and energy metabolism. As a result, living organisms have developed various mechanisms to extract Phosphorus from their environment. During the early evolution of life on Earth, Phosphorus was likely primarily present in the form of phosphate-metal complexes[1] (Brady et al., 2022). This form would have made phosphorus difficult for early life forms to access, suggesting that a particular group of proteins known as phosphate-metal cotransporters played a significant role in the evolution of early life. This research project aims to test the hypothesis that these proteins provided a solution to the potential phosphorus scarcity during the Archean eon. The proteins looked at in this project are PitA and PitH (Harris et al., 2001).

2. Methods

2.1. Literature Review

This research project started with a literature review with three main objectives. The first was to conduct background research about the conditions of early Earth during the Hadean and Archean eons, then find out what sort of life might have existed at the time. The second was learn more about the mechanism and importance of phosphate-metal cotransporters – PitA and PitH proteins being such cotransporters. For this I used the paper '*Characterization of Pita and Pitb from Escherichia Coli*' by Harris et al. 2001. The third objective of this literature review was to investigate the emergence of *Pseudomonadota* and *Actinobacteria*, groups of bacteria, to substantiate our hypothesis. This was accomplished by correlating the literature-derived data on the timing and evolution of these organisms with our own findings regarding the evolutionary history of PitA and PitH.

Hadean and Archaean Conditions

The Hadean (4.6-4.0 Ga) and Archean (4.0-2.5 Ga) are the two earliest eons in Earth's history. During the Hadean, the Earth was extremely hot, with a partially molten surface, and was frequently bombarded by meteorites (Sleep, 2010), with an atmosphere mainly consisting of volcanic gases (CO_2, N_2) and lacking oxygen (Kasting, 2014) . The Earth's crust was beginning to solidify, but was unstable and subject to intense volcanic activity. As for life, there is generally no conclusive evidence for it from the hadean Eon. At the dawn of the Archean Eon, the Earth's surface cooled enough for continents and stable oceans to form but the atmosphere remained anoxic[2] (Kasting, 2014). The Archean environment was still very volatile with frequent volcanic eruptions and tectonic activity (Catling & Zahnle, 2020). The Archean marks the emergence of life on Earth, with the first evidence of microbial life (prokaryotes, such as *Bacteria* and *Archaea*) appearing in the fossil record.

[1] Phosphate-metal complexes form when negatively charged phosphate binds with divalent metal cations like Fe^{2+}, which were abundant in early oceans (Smethurst et al., 2021; Moore et al., 2017).
[2] Lacking Oxygen

Function and Importance of Phosphate-Metal Cotransporters

The paper covers PitA and PitB, two types of phosphate-metal cotransporters. It explains that these cotransporters are membrane proteins that transport inorganic phosphate (P_i) into the cell. PitA is constitutively expressed, meaning it is always active regardless of external Pi levels. On the other hand, PitB is regulated by the pho regulon, which represses its activity under phosphate-limiting conditions. This allows the bacteria to adapt to varying environmental P_i concentrations. In summary, these transporters help ensure that the cell can maintain sufficient phosphate levels, even when external concentrations fluctuate. The presence of such transporters in their host organisms provide these with an adaptive advantage in environments with fluctuating phosphate levels. Since this paper was published we've discovered that PitB is a sub-type and very close relative of PitA so there are probably several PitA sequences which are regulated in response to external phosphorus concentrations even though the specific one investigated in this paper was not.

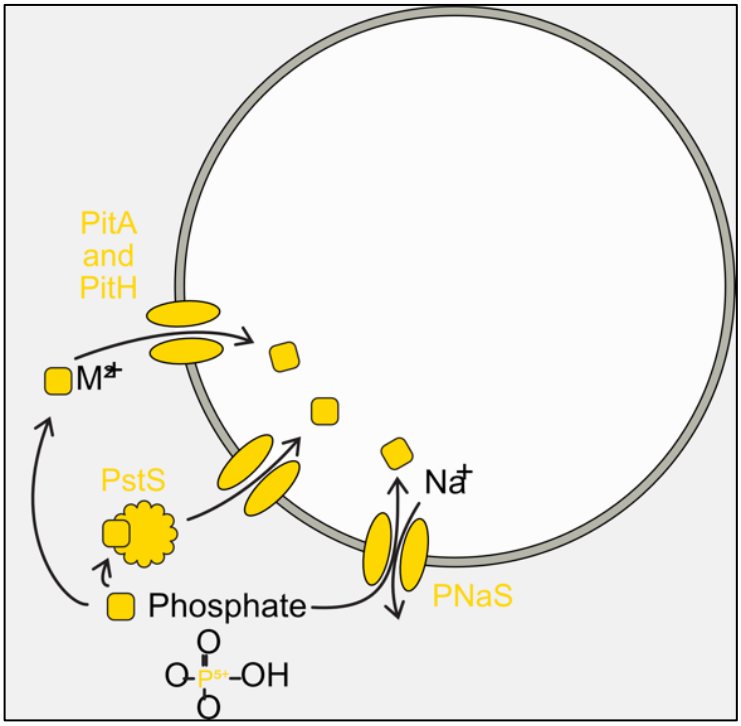

Figure 1 : Mechanism in which PitA and PitH facilitate the transport of phosphate into the cell.

This is an illustrative diagram of the mechanism through which PitA and PitH facilitate the transport of phosphorus (in the form of phosphate) into the cell (Boden et al, 2024). PitA and PitH are membrane proteins. These proteins typically form channels or pores through which phosphate can be transported into the organism's cell. Phosphate ions as represented by a yellow square carry a significant negative charge, meaning it is hard to transport them as they are repelled from the cell wall. When a divalent metal cation such as Fe^{2+} and Mg^{2+} binds to the phosphate ion, it partially neturalises its negative charge. It then becomes easier for the phosphate-metal complex to be transported by the PitA and PitH proteins through the cell wall.

2.2. Conducting BLAST searches for PitA and PitH Homologs

Mapping PitA & PitH Homologs

To identify homologs of PitA and PitH across various environments, BLAST[3] searches were conducted using the NCBI[4] database. The primary objective was to compile a comprehensive report detailing the geographic locations and host organisms of these homologs. The underlying principle is that if PitA and PitH homologs are found across a diverse range of environments and life forms, it would support the idea that these proteins have ancient evolutionary origins. Information regarding the species' collection sites, including host organisms, isolation sources, and geographic locations, was gathered from each relevant database. All collected data were compiled into two Excel documents, one for PitA homologs and another for PitH homologs.

Organism	Location	Host	Isolation Source	Domain	Phylum
Serratia (102)					
Serratia	multiple	multiple	multiple	Bacteria	Pseudomonadota
Serratia entomophila	New Zealand	Costelytra Zealandica	Diseased grass grub larva	Bacteria	Pseudomonadota
Serratia fonticola	France + Portugal	-	freshwater	Bacteria	Pseudomonadota
Serratia grimesii	-	-	-	Bacteria	Pseudomonadota
Serratia inhibens	Denmark: Zealand	-	potato	Bacteria	Pseudomonadota
Serratia liquefaciens	-	-	-	Bacteria	Pseudomonadota
Serratia marcescens	multiple	multiple	multiple	Bacteria	Pseudomonadota
Serratia microhaemolytica	China: Guangzhou	-	Artifical Lake	Bacteria	Pseudomonadota
Serratia odorifera	France: Bordeaux	-	Sputum	Bacteria	Pseudomonadota
Serratia oryzae	China: Hunan	Rice	-	Bacteria	Pseudomonadota
Serratia plymuthica	France	-	Water	Bacteria	Pseudomonadota
Serratia proteamaculans	-	-	-	Bacteria	Pseudomonadota
Serratia quinivorans	France: Paris	Mushroom	-	Bacteria	Pseudomonadota
Serratia rhizosphaerae	South Korea	sia japonica subsp. littoricola	Rhizospheric soil	Bacteria	Pseudomonadota
Serratia rubidaea	France	omo Sapiens (Bacterial Culture)	-	Bacteria	Pseudomonadota
Serratia silvae	France: Champenoux	-	Forest	Bacteria	Pseudomonadota
Serratia symbiotica	-	-	-	Bacteria	Pseudomonadota
Serratia ureilytica	India: West Bengal: Torsa River	-	River Water	Bacteria	Pseudomonadota
sified Serratia (in: enterobact	-	-	-	Bacteria	Pseudomonadota
Serratia sp. JSRIV002	Japan: Shiga (35 N 136 E)	-	River Water	Bacteria	Pseudomonadota
Serratia sp. UGAL515B_01	USA: Georgia, Athens	Aedes aegypti	Larvae Gut	Bacteria	Pseudomonadota
Serratia sp. DD3	Germany: Constance	Daphnia magna	Gut	Bacteria	Pseudomonadota
Serratia sp. M24T3	Portugal	Bursaphelenchus xylophilus	-	Bacteria	Pseudomonadota

Figure 2 : Segment of compiled data for PitA.

Organism	Location	Host	Isolation Source	Domain	Phylum
WP_310872585.1	unclassified Streptomyces	n/a	n/a	Bacteria	Actynomycetota
WP_345592753.1	Streptomyces marokkonensis	Morocco	f the argan tree (Argania spin	Bacteria	Actynomycetota
WP_100047000.1	Streptomyces sp. Alain-F2R5	United Arab Emirates: Alain	mato rhizosphere soil - sample	Bacteria	Actynomycetota
WP_121747788.1	Streptomyces sp. E2N166	Algeria: Fetzara Lake	algeria wetland - sample type:	Bacteria	Actynomycetota
WP_216723651.1	Streptomyces sp. A108	Philippines: Albay Province	whole organism - environmental	Bacteria	Actynomycetota
WP_279563791.1	Streptomyces sp. DH41	Egypt	ntext: egyptian sinai desert, s	Bacteria	Actynomycetota
WP_031044146.1	Streptomyces	n/a	n/a	Bacteria	Actynomycetota
WP_107478700.1	Streptomyces	n/a	n/a	Bacteria	Actynomycetota
WP_210963640.1	Streptomyces sp. b94	SA: Wisconsin OR Turkey: Samsun	ion: 28C° - USA (43.06 N 88.13	Bacteria	Actynomycetota
WP_345635953.1	Streptomyces thinghirensis	Algeria: Mila	e: raw wastewater - sample type	Bacteria	Actynomycetota
WP_019326776.1	Streptomyces	n/a	n/a	Bacteria	Actynomycetota
WP_043382067.1	Streptomyces mutabilis	osamples: only 1 : China: Xinji	isolation source: soil	Bacteria	Actynomycetota
WP_055417636.1	Streptomyces pactum	n/a	sample type: cell culture	Bacteria	Actynomycetota
WP_064728641.1	Streptomyces	n/a	n/a	Bacteria	Actynomycetota
WP_125495584.1	Streptomyces sp. WAC 04229	A: Clearwater, FL 27.96 N 82.8	environmental medium: soil	Bacteria	Actynomycetota
WP_217143394.1	Streptomyces sp. AC627_RSS907	SA: Carlsbad Caverns, New Mexic	host: Tadarida Brasiliensis	Bacteria	Actynomycetota
WP_217166648.1	Streptomyces sp. AC512_CC834	SA: Carlsbad Caverns, New Mexic	host: myotis thysanodes	Bacteria	Actynomycetota
WP_076973740.1	Streptomyces sp. M1013	Turkey: Izmir	ion source: legume rhizospheres	Bacteria	Actynomycetota
WP_121701811.1	unclassified Streptomyces	n/a	n/a	Bacteria	Actynomycetota
WP_030191652.1	Streptomyces violaceorubidus	n/a	sample type: cell culture	Bacteria	Actynomycetota

Figure 3 : Segment of compiled data for PitH.

An obstacle in performing this was that many of the biosamples of these homologs were spread into several different locations, hosts, and isolation sources – or even had none. This hinders our ability to assign the homolog to a specific location/environment.

[3] Basic Local Alignment Search Tool
[4] National center for Biotechnology Information

After compiling the terrestrial data from the NCBI database, the presence of marine samples was also investigated. This involved downloading the HMM[5] profiles of all PitA homologs and uploading them to OceanGeneAtlas, a specialized marine database. This helped compile the following map with samples from both databases. It was not possible to download an HMM profile for PitH as it was not available on the NCBI database.

N.B. As it was not possible to generate an HMM profile for PitH Homologs, the rest of the methods chapter will only involve research using PitA Homologs.

Taxonomic Distribution of PitA Homologs

This process involved uploading the HMM profile of the PitA homologs into Ocean Gene Atlas using Krona which is a software that helps display hierarchical data such as multiple taxonomic layers. This is done to see in which organisms PitA homologs are mainly found and can help us determine how far back these homologs evolved. It was not possible to create a pie chart for the taxonomic distribution of PitH homologs as it also needs an HMM profile to be generated.

2.3. Constructing a Phylogenetic Tree

BASH, MAFFT, AliView

Once having merged the 2 homolog files (from NCBI and Ocean Gene Atlas) into a .faa file, the next step was to align the amino acid sequences of the PitA homologs using MAFFT[6]. MAFFT was used using BASH[7]. The alignment process was done using the following command:

```
mafft PitAHomologs.faa > PitAHomologs.aligned.faa
```

The output file was then uploaded into Aliview to visualise the changes made by MAFFT.

IQ-TREE

Subsequently, using a remote computing cluster (supercomputer) and BASH, an evolutionary tree for PitA homologs was generated by executing the following IQ-TREE command (Minh et al., 2020):

```
./../../iqtree-2.3.4-Linux-intel/bin/iqtree2 -s PitAHomologs.aligned.trim85.faa.fas
-bb 1000 -alrt 1000 -ntmax 10 -nt AUTO -nm 5000 -m MFP -mset LG,WAG -mrate G,R -
cmin 4 -cmax 8 -madd Q.pfam+C20+G4,Q.pfam+C20+G4+F -pre PitATree
```

Important in this command is the '`-bb 1000`' option (1000 ultrafast bootstraps) which tests the robustness of the tree. This function statistically validates the evolutionary relationship between two PitA Homologs.

[5] Hidden Markov Model profile
[6] Multiple Alignment using Fast Fourier Transform.
[7] Bourne Again SHell

Another significant option in this command is the substituion model '`Q.pfam+C20+G4+F`' chosen according to BIC[8]. This is used to describe the process by which one amino acid in PitA is replaced (or substituted) by another over time.

TreeViewer

The output file generated by the IQ-TREE function was then uploaded into TreeViewer. Due to the numerous strains and varying branch lengths, the tree was difficult to interpret, so it was converted into a cladogram. To distinguish the marine sequences (from Ocean Gene Atlas) from the NCBI sequences, the marine sequences were colored blue. This process involved opening Aliview, sorting the sequences by name, copying them into a comma-delimited file, and then re-uploading this file into TreeViewer to apply the colour differentiation to the tree.

3. Results

Mapping PitA & PitH Homologs

The data gathered for PitA Homologs through BLAST searches in the NCBI and Ocean Gene Atlas databases helped produce the following figure. PitA was found in many places including the United States, France, India, China, Japan, and New Zealand. PitH, not shown in the following figure was even more widespread, having been found in every continent save for Antarctica. As PitA and PitH can be found across a diverse range of environments and life forms, this supports the idea that these proteins have ancient evolutionary origins.

Figure 3 : Geographic Distribution of PitA Homologs.

[8] Bayesian Information Criterion

Taxonomic Distribution of PitA Homologs

The following pie chart generated with Ocean Gene Atlas and Krona shows us in which organisms PitA homologs were found:

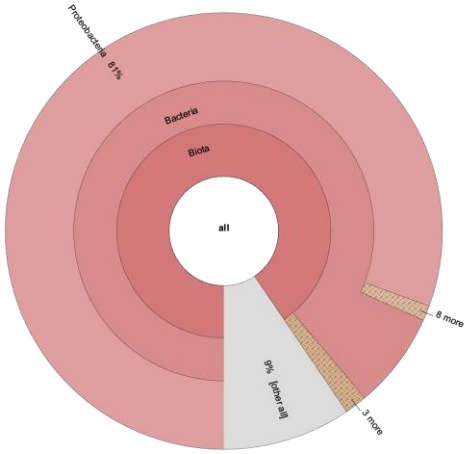

Figure 4 : Taxonomic Distribution of PitA Homologs.

As we can see *Proteobacteria* (*Pseudomonadota*) are the majority group containing 81% of existing PitA homologs. This result can be associated with the third literature review segment alluded to earlier in the methods section.

Given that *Pseudomonadota* is among the earliest-diverging bacterial lineages, it is plausible that PitA cotransporters first emerged during the early evolutionary history of this group. The later estimates for *Pseudomonadota* suggested an origin of up to 2 Ga (Mahendrarajah et al. 2023), (Betts et al. 2018). These estimates place the origins of these lineages after the Archean eon, which doesn't support the hypothesis that these cotransporters may have developed under the ancient environmental conditions of that era. However, two more recent papers looked at (Boden et al, 2024), (Davin et al, 2023) both gave older estimates of 3.1 Ga and 3.0-2.6 Ga which in in turn does help to prove our hypothesis.

During this literature review it was also found that PitH is found in a similar early-diverging bacterial lineage *Actinomycetota* (*Actinobacteria*). This lineage is estimated to be 2.3 Ga according to Gutierrez et al. (2023) and between 3.5-2.8 Ga according to Davin et al (2023).

BASH, MAFFT, Aliview.

Here is a visualiation of the alignment process done with MAFFT shown thanks to AliView.

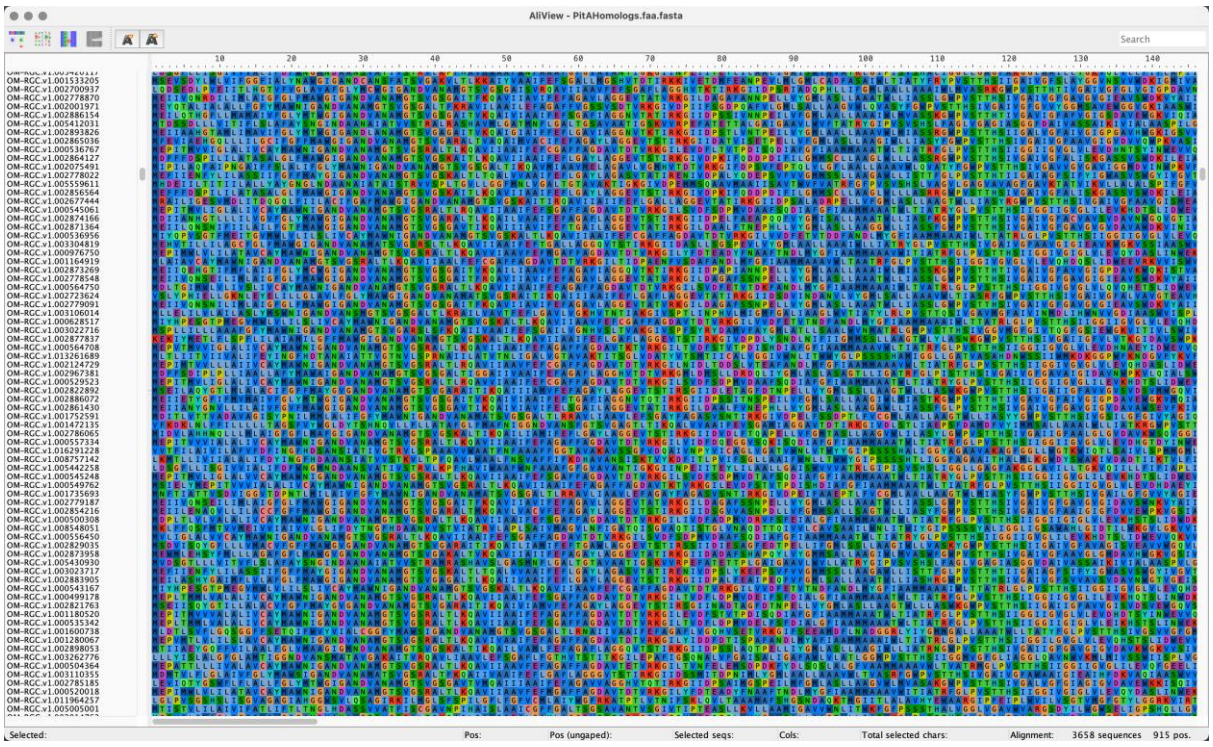

Figure 5: Before MAFFT Alignment Process.

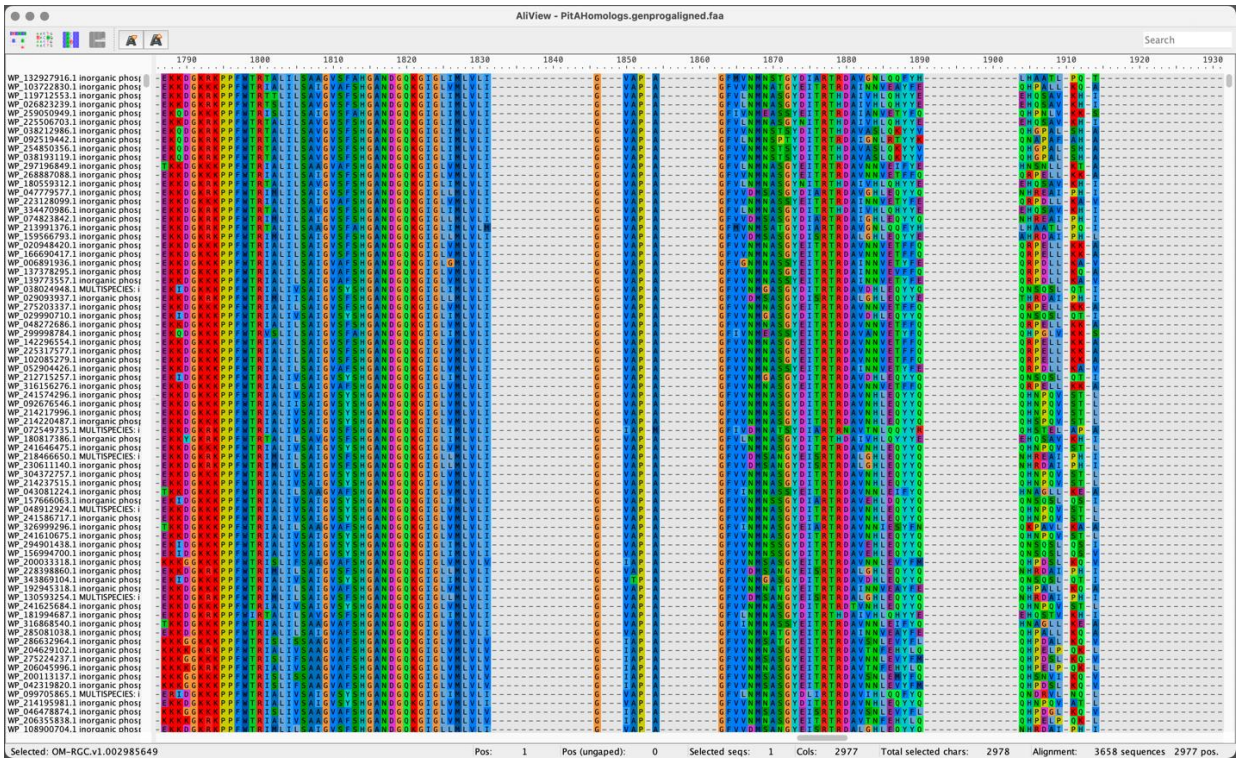

Figure 6: After MAFFT Alignemnt Process.

TreeViewer

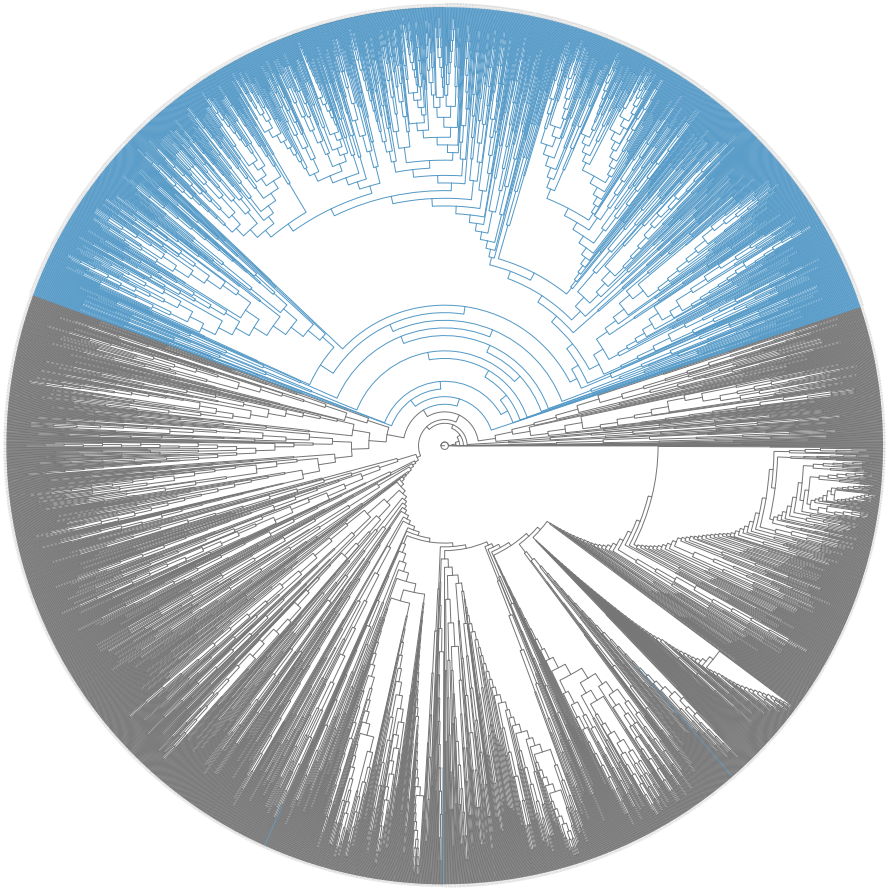

Figure 6 : Cladogram Produced.

As can be seen in the figure above the marine sequences of PitA homologs from Ocean Gene Atlas are clustered in one place on the cladogram. We can therefore conclude that the ability to use of PitA to collect phospahte-metal complexes from their environment was a single evolutionary event which happened in the Archean ocean.

4. Discussion

The results suggest that PitA could have played a crucial role in addressing potential phosphorus scarcity during the Archean. However, the research period was insufficient to accurately determine when PitA and PitH evolved. To achieve a more precise estimate, I recommend using a molecular clock. By comparing the rate of genetic mutations across a wide range of species, using a molecular clock could provide both an upper and lower limit for when these proteins likely evolved. This method is more accurate than our current approach, which only offers an upper limit based on the divergence of specific lineages like *Pseudomonadota*. To achieve a similar analysis for PitH, I recommend reconstructing an HMM profile for PitH. This can be done by gathering a diverse set of known PitH sequences, aligning them using a tool like MAFFT, and then using software like HMMER to build the HMM profile.

5. Conclusion

The objective of this research project was to assess the distribution of the phosphate-metal cotranspoters PitA and PitH in microbial communities from different environments in order to provide a first-order perspective on their evolutionary origins. This was done to investigate the hypothesis that PitA and PitH provided a solution to the potential Phosphorus scarcity during the Archean. The methods, chiefly: conducting a literature review, BLAST searches, as well as constructing an evolutionary tree, revealed that PitA is found primarily in the early bacterial lineage; *Pseudomonadota* whereas PitH is found primarily in a different bacterial lineage; *Actinomycetota*. Previous research has estimated that these lineages emerged in the Archean eon: for *Actinomycetota* at the earliest around 3.5-2.8 Ga and for *Pseudomonadota* at the earliest around 3.1 Ga. It is possible that these proteins emerged during the Archean eon, although further research is needed to test this. The results also show that there was a widespread distribution of these proteins across various environments including deep oceans, forests, agricultural and volcanic soils, wetlands, and deserts which shows their evolutionary importance and adaptive advantage where phosphate availability was/is limited. The phylogenetic tree constructed further suggests that PitA originated from a single evolutionary event in a marine environment. In conclusion, the evidence presented in this research project supports that PitA and PitH were important in the survival and proliferation of early life forms in phosphorus-limited environments, showing a significant evolutionary innovation in early life on Earth.

Acknowledgements

I would like to express my gratitude to the Laidlaw foundation and Lord Laidlaw for their support; for the opportunity I was given to expand my intellectual horizons alongside an exceptional group of peers. My special thanks to Dr Boden for her englightening and patient guidance.

6. References

Journals

Betts, H.C. *et al.* (2018) 'Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin', *Nature Ecology & Evolution*, 2(10), pp. 1556–1562. doi:10.1038/s41559-018-0644-x.

Boden, J.S. *et al.* (2024) 'Timing the evolution of phosphorus-cycling enzymes through geological time using phylogenomics', *Nature Communications*, 15(1). doi:10.1038/s41467-024-47914-0.

Catling, D.C. and Zahnle, K.J. (2020) 'The Archean atmosphere', *Science Advances*, 6(9). doi:10.1126/sciadv.aax1420.

Davín, A.A. *et al.* (2023) *An evolutionary timescale for bacteria calibrated using the great oxidation event* [Preprint]. doi:10.1101/2023.08.08.552427.

Harris, R.M. *et al.* (2001) 'Characterization of pita and pitb from *escherichia coli*', *Journal of Bacteriology*, 183(17), pp. 5008–5014. doi:10.1128/jb.183.17.5008-5014.2001.

Kasting, J.F. (2014) 'Atmospheric composition of Hadean–early Archean earth: The Importance of Co', *Earth's Early Atmosphere and Surface Environment* [Preprint]. doi:10.1130/2014.2504(04).

Mahendrarajah, T.A. *et al.* (2023) 'ATP synthase evolution on a cross-braced dated tree of life', *Nature Communications*, 14(1). doi:10.1038/s41467-023-42924-w.

Martinez-Gutierrez, C.A., Uyeda, J.C. and Aylward, F.O. (2023) 'A timeline of bacterial and archaeal diversification in the Ocean', *eLife*, 12. doi:10.7554/elife.88268.3.

Minh, B.Q. *et al.* (2020) 'IQ-tree 2: New models and efficient methods for phylogenetic inference in the genomic era', *Molecular Biology and Evolution*, 37(5), pp. 1530–1534. doi:10.1093/molbev/msaa015.

Moore, E.K. *et al.* (2017) 'Metal availability and the expanding network of microbial metabolisms in the Archaean Eon', *Nature Geoscience*, 10(9), pp. 629–636. doi:10.1038/ngeo3006.

Sleep, N.H. (2010) 'The Hadean-archaean environment', *Cold Spring Harbor Perspectives in Biology*, 2(6). doi:10.1101/cshperspect.a002527.

Smethurst, D.G.J. and Shcherbik, N. (2021) 'Interchangeable utilization of metals: New perspectives on the impacts of metal ions employed in ancient and extant biomolecules', *Journal of Biological Chemistry*, 297(6), p. 101374. doi:10.1016/j.jbc.2021.101374.

Websites

PitA Homologs – NCBI Page
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/NF033774/

PitH Homologs – NCBI Page
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/NBR010556/

Ocean Gene Atlas
https://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/