Reconstructing the History of the Proliferation of Genes involved in the Biological Nitrogen Cycle

Chris Parsons
Senior Integrative Exercise
December 29, 2017

Submitted in partial fulfillment of the requirements for a Bachelor of Arts degree from Carleton College, Northfield, MN
Table of Contents

Abstract
Introduction..........................................................................................................................1
Methods.................................................................................................................................8
Results ..................................................................................................................................10
Discussion.............................................................................................................................15
Conclusions.........................................................................................................................22
Acknowledgements ...........................................................................................................23
References Cited....................................................................................................................24
Appendix A ..........................................................................................................................29
Reconstructing the History of the Proliferation of Genes involved in the Biological Nitrogen Cycle

Chris Parsons
Carleton College
Senior Integrative Exercise
December 29, 2017

Advisor:
Dr. Rika Anderson, Carleton College

Abstract
Nitrogen is an essential element to life on Earth and a variety of bacteria and archaea structure their metabolisms around it. These various metabolisms and their associated chemical reactions come together into the biological nitrogen cycle, which has been shown to have been operating for at least 3.2 billion years. Each step in the nitrogen cycle is catalyzed by one or more specific enzymes. While it is possible to roughly date the origins of these enzymes, estimating their rate of proliferation through horizontal gene transfer over the course of Earth history can provide additional insights into the points in time at which these genes became particularly ecologically relevant. By comparing the phylogenies of the genes coding for these enzymes to a time-calibrated tree of life, we identified and timed loss, duplication, and horizontal gene transfer events for nitrogen-cycling genes. Our results show a major spike around 2.5 billion years ago in the rate of horizontal gene transfer for denitrification genes, which coincides neatly with the oxygenation of the atmosphere. These results suggest that the increased availability of oxidized forms of nitrogen stimulated the nitrogen cycle, activating the denitrification pathway. Additionally, we showed that biological nitrogen fixation by the molybdenum nitrogenase Nif has existed for almost the entirety of the history of life on Earth, and that a nitrous oxide-rich atmosphere was unlikely to have been the cause of elevated global temperatures in the Mid-Proterozoic.

Keywords: nitrogen cycle; horizontal gene transfer; geobiology; microbial evolution; nitrogenase
Introduction

*The Modern Nitrogen Cycle*

Nitrogen is a critical element to life on Earth, necessary for the construction of nucleotides and proteins, ubiquitous biological molecules. A variety of organisms structure their metabolisms around nitrogen, following various enzymatic pathways within the biological nitrogen cycle, which consists of a series of redox reactions (Figure 1). Nitrogen enters the nitrogen cycle through nitrogen fixation, the process of reducing atmospheric nitrogen (N\(_2\)) to ammonium (NH\(_4^+\)). Although this reaction is technically exergonic, 16 molecules of ATP are required to overcome the high activation energy (Zerkle and Mikhail, 2016). This step is catalyzed by nitrogenase, of which there are three varieties, based on the metals in their associated active site cofactors: Nif (Fe-Mo), Vnf (Fe-V), and Anf (Fe-Fe). Most organisms do not fix nitrogen for themselves, instead assimilating ammonium produced by limited populations of nitrogen fixers, or through ammonification, the process of acquiring ammonium from nitrogen-containing organic compounds (Canfield et al., 2010).

Ammonium produced from nitrogen fixation or ammonification, can then either be converted to organic forms of fixed nitrogen, such as amino acids, or, in the presence of oxygen, oxidized to nitrite (NO\(_2^-\)) or nitrate (NO\(_3^-\)) through the nitrification pathway (Figure 1). Amo and Hao are the enzymes responsible for conversion of ammonium to nitrite, and Nxr subsequently converts nitrite to nitrate (Canfield, 2010). Nitrification is a form of chemoautotrophy, in which nitrifiers gain energy from the electrons produced by the oxidation of nitrogen (Canfield et al., 2005).
Figure 1. The steps of the biological nitrogen cycle and their associated enzymes. Arrows are labelled with the pathway to which they belong. Adapted from Canfield et al. (2010).
In anoxic conditions, nitrate can be reduced to either nitrogen gas or ammonium through the denitrification and dissimilatory nitrate reduction to ammonium (DNRA) pathways, respectively. In these pathways, nitrogen is used as an electron acceptor in conjunction with the oxidation of carbon compounds, similarly to how oxygen is used as a terminal electron receptor in aerobic respiration (Canfield et al., 2005). Denitrification is a four-step process in which nitrate is converted to nitrite, by either Nar, Nap, or Nas. Nitrite is then converted into nitric oxide (NO) by NirK or NirS. Varieties of Nir can be distinguished by the metals in their active site cofactors, with NirK using copper and NirS using iron. Nitric oxide is then reduced to nitrous oxide (N\textsubscript{2}O) by Nor. At this point, nitrous oxide can either leave the system, as it is a gas, or be converted to nitrogen gas by Nos, completing the cycle. DNRA is similar to denitrification except that, following the reduction of nitrate to nitrite, the resulting nitrite can be converted directly to ammonium by Nrf. Nrf can also reduce downstream products of denitrification to ammonium. DNRA, while less understood, has been shown to be a major nitrate sink in a variety of aquatic systems, especially warm intertidal zones (Giblin et al., 2013).

*Atmospheric Nitrogen*

Nitrogen makes up approximately 78% of the Earth’s modern atmosphere by volume predominantly as nitrogen gas (Asimov, 1955). The early Earth’s atmosphere was probably also nitrogen-rich, with most nitrogen originating in the mantle near convergent plate margins as nitrogen gas and subsequently degassing into the atmosphere (Mikhail and Sverjensky, 2014). The abundance of atmospheric nitrogen in the early Earth is often proposed as a means to solve the “Faint Young Sun” paradox, in which it is
unclear how a 20% dimmer sun warmed the early Earth enough to allow for the existence of substantial amounts of liquid water in the Archean (Som et al., 2012; Sagan and Chyba, 1997). One solution is that high amounts of atmospheric N\textsubscript{2}, a non-greenhouse gas, may have amplified the greenhouse effect through a process known as pressure broadening (Goldblatt et al., 2009).

Similarly, in the mid-Proterozoic (~1000-1600 Ma), a buildup of biotically generated nitrous oxide, a true greenhouse gas, has been proposed as a method for warming the Earth, due to a limited supply of copper, an essential metal in Nos, the enzyme responsible for metabolism of nitrous oxide (Buick, 2007). In the absence of Nos, all nitrous oxide would then enter the atmosphere. Examining the evolutionary history of the nitrogen cycling enzymes responsible for the production or consumption of gaseous varieties of nitrogen can provide clues to the solution of this paradox.

**The History of Nitrogen Fixation**

An ongoing debate regarding the evolution of nitrogen fixation centers around the timing of its development, with some inferring it to have evolved early in life’s history, and others inferring an origin of nitrogen fixation following widespread oxygenation of the oceans and atmosphere. The crux of many such arguments is the extent to which abiotic sources of biologically available nitrogen, such as meteorite impacts and lighting, could sustain the early biosphere. It has been proposed that steady reduction in atmospheric CO\textsubscript{2} levels reduced the amount of lightning-generated fixed nitrogen, eventually leading to a nitrogen crisis around 2.2 Ga, which subsequently would have encouraged the development of biological nitrogen fixation (Navarro-González et al.,
However, several studies have concluded that abiotic nitrogen fixation in the early Earth would produce 50- to 5000-fold lower rates of nitrogen fixation than the modern ocean, making nitrogen highly limiting, even at higher CO₂ levels (Canfield et al., 2010). In agreement with studies which have determined abiotic nitrogen flux to have been to be too low to sustain the early biosphere, isotopic studies indicate that biological nitrogen fixation had evolved by at least 3.2 Ga (Stüeken et al., 2015). However, isotope ratios from this period are often affected by metamorphism, making very ancient nitrogen isotope ratios somewhat unreliable (Stüeken et al., 2016). Previous genomics-based studies have told conflicting stories regarding the evolution of nitrogenase, with some inferring the last universal common ancestor to have been capable of nitrogen fixation (Weiss et al., 2016) and others claiming a far more recent origin between 1.5 and 2.2 Ga (Boyd and Peters, 2013).

It has also been proposed that limited molybdenum availability in the early oceans would have prevented the evolution of Nif, which uses molybdenum in its active site cofactor. Therefore, alternative nitrogenases, such as Anf and Vnf, which use iron and vanadium in their cofactors, respectively, have been suggested to have developed prior to Nif, due to the lack of molybdenum (Raymond et al., 2004). However, these genes are only present in a small number of modern organisms and the organisms that do possess them also possess a gene for Nif, indicating some dependence on Nif (Boyd and Peters, 2013). This study attempts to provide another potential estimate for the timing of the development of biological nitrogen fixation, as well as the identities of the prevailing nitrogen-fixing enzymes at any given point in early Earth history.
The History of Nitrification, Denitrification, and DNRA

The high Fe\(^{2+}\) concentration of the Archean ocean, combined with the lack of free oxygen, created a highly reducing environment, causing the predominant form of dissolved nitrogen in the early ocean to exist as ammonium, the most reduced variety (Stüeken et al., 2016). Following the development of oxygenic photosynthesis, oxidized forms of nitrogen would have steadily become more common, allowing for the development of the rest of the modern nitrogen cycle. Isotopic studies have demonstrated appreciable levels of atmospheric oxygen by 3.0 Ga, but still only approximately \(3 \times 10^{-4}\) times modern levels (Crowe et al., 2013). Immediately following the development of oxygenic photosynthesis, the majority of dissolved nitrogen likely would still have existed as ammonium, but localized regions of high cyanobacterial abundance, termed oxygen oases, may have allowed for active nitrifying and denitrifying pathways (Olson et al., 2013). These pathways would likely have become more widespread as oxygen concentrations increased and as oxygen oases grew.

Horizontal Gene Transfer

Horizontal gene transfer (HGT) is the process by which an organism can gain genetic information from another, possibly highly unrelated, organism, either by assimilation of foreign DNA in solution, direct cell-to-cell transfer of DNA, or viral introduction of new DNA (Thomas and Nielsen, 2005). Especially in microbial lineages, HGT is a crucial evolutionary mechanism by which a clade of organisms can develop new useful phenotypes, without the evolutionary cost associated with independently evolving genes that are already in existence (Gogarten and Townsend, 2005). HGT can
allow clades of bacteria to invade new ecological niches, as, if a gene becomes fixed in a population, it likely provides some kind of selective advantage. As such, studying the history of HGT of specific genes can provide insights into the points at which possession of such genes provided substantial selective advantages, which can then be used as a metric for the availability of the substrates of the enzymes which particular genes code for.

Genes associated with both nitrogen fixation and denitrification have been shown to have undergone substantial HGT, making them good candidates for this study (Jones et al., 2008; Raymond et al., 2004). This methodology works by comparing the phylogenies for specific genes, gene trees, to one based on concatenations of highly-conserved ribosomal proteins, the species tree. As the species tree should be reflective only of vertical transfer of genes, due to the highly-conserved natures of these proteins, any discrepancies between the species tree and a given gene tree can be attributed to HGT, gene loss, and/or gene duplication.

Interpretation of the results of this study is based on the underlying assumption that rates of HGT can be correlated directly with ecological relevance of the genes being transferred, as well as substrate availability for the enzymes coded for by those genes. While this is likely to be generally true, it is occasionally the case that one taxonomically confined, yet abundant clade of organisms is responsible for the entirety of a certain metabolic pathway. In such circumstances, this study would observe low rates of transfer for genes involved in that metabolism, and conclude that pathway to be inactive, due to the inability to infer population sizes. Therefore, results from this study should be interpreted with such possible scenarios in mind.
**Materials and Methods**

*Genome Selection and Compilation*

The initial collection of genomes was constructed by locating the genomes associated with a database of *nifH* genes (Gaby and Buckley, 2014), and downloading them from the NCBI genome database (NCBI Resource Coordinators, 2016). This dataset was expanded by adding representative genomes from the NCBI database and ggKBase’s collection of 2500 curated genomes (Anantharaman et al., 2016) for any bacterial or archaeal phyla not already present in the sample, as defined by Hug et al. (2016), in order to create a tree fully representative of modern organismal diversity. Eukaryotic genomes were not included in this dataset due to computational limitations, preventing the identification of eukaryotic nitrogen-cycling genes. Due to the small number, and late origins, of eukaryotic nitrogen fixers, this should not substantially affect the results of this study, especially prior to 2 Ga.

*Gene Isolation*

To isolate the highly conserved amino sequences for the species tree, bacterial and archaeal genomes were mined for single copy ribosomal protein sequences L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, and S19 using Phylosift (Darling et al., 2014), with the isolate and best hit command line flags. Genomes were removed from the dataset if the aligned output contained more than 50% gaps. Ribosomal protein sequences from Hug et al. (2016) were used in the case of eukaryotes. Nitrogen-cycling amino acid sequences for the gene trees were identified by conducting BLASTP (Altschul et al., 1990) searches of the open reading frames of every genome from the
previously generated collection, which were identified using Prodigal (Hyatt et al., 2010). Curated gene database queries for individual genes were generated based on KEGG orthologies (Kanehisa and Goto, 2000) and downloaded from the UniProt database (The UniProt Consortium, 2017). The maximum e-value cutoff for BLAST hits was $10^{-12}$ and matches were excluded if the length of the local alignment was less than 50% of the length of the query sequence.

Tree Construction

Genes were aligned using MUSCLE (Edgar, 2004), and phylogenies were constructed from the resulting alignments using RAxML with 100 rapid bootstraps (Stamatakis, 2014). Trees for nitrogen-cycling genes (gene trees) were constructed using the PROTGAMMALG substitution matrix, whereas the PROTCATLG matrix was used for the tree based on reference marker genes (the species tree). The root for the species tree was defined as the Bacteria.

Chronogram Construction

The species tree was converted into four species chronograms using PhyloBayes (Lartillot et al., 2009). Two sets of calibration points were used, one liberal (earlier divergence) and one conservative (later divergence), to test the sensitivity of methodology (Table 1). Similarly, two clock models were tested, a log normal model and a CIR process. Two clocks were run in parallel for each set of parameters, so that the two concurrent runs could be compared to one another as a test of convergence. Simultaneous
runs generally converged after around 10,000 cycles and chronograms were generated with burn-ins of around 1,500 cycles.

**Gene Tree and Species Chronogram Reconciliation**

Gene trees were reconciled with the different chronograms using AnGST (David and Alm, 2011). Event penalties were set to hgt: 3, dup: 2, los: 1, and spc: 0. Ultrametric was set to True in order to constrain events temporally. 100 gene tree bootstraps were fed into each run, in order to increase accuracy. Individual event timings were defined as the midpoint of the temporal region during which a given event could occur.

**Results**

**Maximum Likelihood Species Tree**

The final species tree was based on concatenated protein sequences from 308 organisms (Figure 2). Monophyly was preserved for most major phyla, with three notable exceptions: (i) Tenericutes is contained within Firmicutes, (ii) the PVC superphylum contains Omnitrophica, and (iii) Lentisphaerae is nested within Verrucomicrobia. The species tree is a three-domain tree, in contrast to recent studies which have shown the addition of the Asgard Archaea to the tree of life to cause Eukarya to group within Archaea, creating a two-domain tree (Zaremba-Niedzwiedzka et al., 2017). The Asgard Archaea did form a monophyletic group, however. The relationship of the three domains to one another should not substantially affect the final results of this study, due to the infrequency of inter-domain HGT, as well as the fact that eukaryotic nitrogen-cycling genes are not included in this study.
Table 1. Fossil calibration points used in Phylobayes runs. Liberal points are earlier, and therefore less agreed upon.

<table>
<thead>
<tr>
<th>CALIBRATION EVENT</th>
<th>CONSERVATIVE (GA)</th>
<th>LIBERAL (GA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUCA</td>
<td>&gt;3.5 (Noffke et al., 2013)</td>
<td>&gt;3.8 (Mojzsis et al., 1996)</td>
</tr>
<tr>
<td>ORIGIN OF OXYGENIC PHOTOSYNTHESIS</td>
<td>&gt;2.45 (Bekker et al., 2004)</td>
<td>&gt;3 (Lyons et al., 2014)</td>
</tr>
<tr>
<td>ORIGIN OF METHANOGENESIS</td>
<td>&gt;2.7 (Eigenbrode and Freeman, 2006)</td>
<td>&gt;3.46 (Ueno et al., 2006)</td>
</tr>
<tr>
<td>ORIGIN OF EUKARYOTES</td>
<td>&gt;1.7 (Pang et al., 2013)</td>
<td>&gt;3.2 (Javaux et al., 2010)</td>
</tr>
</tbody>
</table>

Table 2. Number of BLAST hits for nitrogen-cycling genes within 254 bacterial and archaeal genomes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>BLAST Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>amo</td>
<td>5</td>
</tr>
<tr>
<td>anfG</td>
<td>20</td>
</tr>
<tr>
<td>hao</td>
<td>174</td>
</tr>
<tr>
<td>napA</td>
<td>120</td>
</tr>
<tr>
<td>narG</td>
<td>129</td>
</tr>
<tr>
<td>nasA</td>
<td>138</td>
</tr>
<tr>
<td>nifH</td>
<td>210</td>
</tr>
<tr>
<td>nirK</td>
<td>88</td>
</tr>
<tr>
<td>nirS</td>
<td>60</td>
</tr>
<tr>
<td>norB</td>
<td>116</td>
</tr>
<tr>
<td>nosZ</td>
<td>35</td>
</tr>
<tr>
<td>nrfA</td>
<td>54</td>
</tr>
<tr>
<td>nxrA</td>
<td>97</td>
</tr>
<tr>
<td>vnfG</td>
<td>17</td>
</tr>
</tbody>
</table>
Figure 2. Maximum likelihood phylogeny generated using RAxML with a PROTCATLG substitution matrix, based on an alignment of 308 concatenations of highly-conserved proteins. Well-represented bacterial and archaeal phyla are labelled in black text. Bootstrap values (from 100 rapid bootstraps) greater than 50 are shown as transparent gray circles, with larger circles representing higher bootstrap values.
Additionally, there are some discrepancies regarding the higher order arrangement of bacterial clades between this tree and the tree of life presented in Hug et al. (2016). Especially notable is that Hug et al. (2016) shows the recently-discovered Candidate Phyla Radiation (CPR) bacteria to be the outgroup to the bacterial phyla, whereas this tree shows the CPR bacteria to be more deeply nested within the bacteria, as a sister group to the Cyanobacteria. Regardless, any minor discrepancies in the topology of the species tree should not greatly affect the results of this study, due to the high number of events inferred.

*Time-Calibrated Maximum Likelihood Species Tree*

As would be expected, the use of liberal calibration points pushed the timing of most divergence events earlier. The only major difference between the chronograms constructed by a CIR process and those constructed by a log-normal model is that the log-normal chronograms, especially the one based on conservative calibration points, showed bacterial diversification to have occurred much closer to LUCA than the CIR chronograms. The timing of LUCA varied slightly with clock model, and substantially (~500 Ma difference) with calibration points selection. The main results presented in this study utilize the log-normal conservative chronogram (Figure 3).

*Gene Trees*

14 nitrogen-cycling genes were searched for within the bacterial and archaeal genomes represented in the species tree (Table 2). One nitrifying gene, *amo*, was scarce within the genome dataset, and so was excluded from the analysis. Ultimately, 13
Figure 3. Species chronogram. Generated by running a log-normal relaxed molecular clock in PhyloBayes, using a conservative set of calibration points (see Table 1)
maximum likelihood gene trees were generated. Monophyly was rarely preserved for most major phyla, when compared with the species tree.

Gene and Species Tree Reconciliation

AnGST inferred high numbers of horizontal gene transfer events for each gene in this study, when compared to speciation, duplication, and loss events. The shapes of the resulting histograms were generally unchanged by the use of liberal time points instead of conservative time points in construction of the species tree, as well as by the use of a CIR process instead of a log-normal clock (See Appendix A).

11 studied genes were inferred to be fairly ancient, with origins prior to 3.5 Ga, but 6 showed a dramatic increase in their rate of HGT between 2 and 3 Ga: nirK, nirS, nrfA, nxrA, norB, and nosZ (Figure 4a). 5 genes showed a fairly steady increase in the rate of horizontal gene transfer since 3.5 – 4 Ga: napA, narG, nasA, nifH, and hao (Figure 4b). Substantially later birth events (~2.5 Ga) were inferred for anfG and vnfG, so the HGT events for these genes were inferred to be much more recent (Figure 5).

Discussion

Nitrogen Fixation

These data show nitrogen fixation through the use of molybdenum nitrogenase (Nif) to be an ancient process, with an origin prior to the LCA of all bacteria. This finding is consistent with recent nitrogen isotope studies, which have concluded that biological nitrogen fixation had arisen by 3.2 Ga (Stüeken et al., 2015), as well as studies which have concluded that abiotic sources of fixed nitrogen were likely too weak to
Figure 4. Histogram plots of the timings of horizontal gene transfer events for (a) genes which show generally steady increases in the rate of HGT since LUCA and (b) genes which show a substantial spike in the rate of HGT between 2 and 3 Ga. AnGST runs shown are based on the log-normal clock tree of life with the conservative set of calibration points (see Table 1 for calibration points; results under other models and time points in Appendix A).
Figure 5. Histogram plots of the timings of horizontal gene transfer events for the alternative nitrogenases, which were inferred to have been born at ~2.5 Ga. AnGST runs shown are based on the log-normal clock tree of life with the conservative set of calibration points (see Table 1 for calibration points; results under other models and time points in Appendix A).
sustain a thriving biosphere (Canfield et al., 2010). With these data, it can be concluded that biological nitrogen fixation is a primordial process, possibly necessary to sustain the Archean biosphere.

Additionally, the inferred late origin of iron and vanadium nitrogenases (Anf and Vnf, respectively) indicates that these alternative nitrogenases evolved much later than Nif, not prior to it, in agreement with other genomics-based studies (Boyd and Peters, 2015). Perhaps localized regions of higher molybdenum concentrations existed, allowing for the development of Nif, and proliferation of nitrogen fixers, even while the ocean at large was molybdenum-depleted.

**Nitrogen Cycling Prior to the Oxygenation of the Oceans and Atmosphere**

In the anoxic ocean that existed before the development of oxygenic photosynthesis, the most widespread form of dissolved inorganic nitrogen was ammonium, the most reduced form of nitrogen (Stüeken et al., 2016). Denitrification and nitrification would be expected to be minor, or highly taxonomically confined, processes in such an environment, due to the limited availability of oxidized forms of nitrogen. However, the results of this study show active transfer of genes for nitrite-producing enzymes, through both oxidative and reductive pathways (Figure 6). Further complicating the situation, rates of HGT for genes which take nitrite, an intermediate in both pathways, as a substrate are relatively low during this period.

If the rates of HGT for specific genes are to be interpreted as metrics for the availability of their enzymes’ respective substrates, these data indicate that, prior to ~2.8 Ga, ammonium and nitrate were far more abundant than nitrite. It has been proposed that,
Figure 6. Histogram plots of the timings of horizontal gene transfer events for genes coding for enzymes which take (a) nitrate, (b) nitrite, and (c) nitrite and downstream denitrification products (NO and N\textsubscript{2}O) as substrates. AnGST runs shown are based on the log-normal clock tree of life with the conservative set of calibration points (see Table 1 for calibration points).
while the surface ocean remained mostly anoxic, localized regions may have had thriving cyanobacterial populations (Olson et al., 2013). In such pockets, oxygen concentrations would be high, causing nitrate to be the prevailing form of nitrogen, rather than ammonium. Any nitrate leaving these pockets would be reduced to nitrite by a variety of microorganisms and any ammonium entering would be oxidized to nitrite. Because genes for nitrite-metabolizing enzymes were not being actively transferred during the early Archean, however, it appears that nitrite-metabolizing genes were not widespread across the tree of life, indicating that few, if any, varieties of microbes subsequently converted this nitrite into nitrate or ammonium, implying that the conversion likely happened abiotically. Fe$^{2+}$, abundant in the early anoxic ocean, would have readily reduced nitrite to ammonium or nitrogen gas outside of the proposed oxic bubbles (Canfield, 2010). Regardless of the mechanism, it can generally be concluded that nitrite levels during this time period were too low to cause substantial proliferation of genes for enzymes which take nitrite as a substrate, while nitrate and ammonium concentration were high enough to allow for biological metabolism of these molecules.

*Effects of Increased Oxygen Levels on the Nitrogen Cycle*

By 2.75 Ga, marine environments have been shown to have been substantially oxygenated (Stüeken et al., 2016). At around the same time, the rate of HGT for Nrf, which reduces a variety of oxidized forms of nitrogen to ammonium through the DNRA pathway, began to increase. It is possible that this generalized nitrogen reduction pathway prevailed through the early oxygenation of the ocean, before the more complex denitrification pathway had evolved. At around 2 Ga, however, the rates of transfer for
enzymes involved in the modern denitrification pathway also sharply increased. Additionally, Nxr, which oxidizes nitrite into nitrate, showed a very similar increase in transfer rates at 2 Ga.

The starting point of all of the substrates for the enzymes which demonstrated this sudden increase in the rate of HGT is nitrite, indicating that the oxygenation of the ocean caused increased nitrite availability. A possible mechanism for the increase in nitrite concentrations is that, following the widespread oxidation of Fe$^{2+}$ in the oceans, abiotic reduction of nitrite substantially decreased, to the point at which it became metabolically favorable to use nitrite as an electron donor, as less nitrite was being reduced abiotically by Fe$^{2+}$, increasing overall nitrite concentrations. Alternatively, with increasing cyanobacterial proliferation, more local oxic environments arose, causing increased interfacing between reducing and oxidizing environments. At such interfaces, nitrite concentrations are particularly high, as it is an intermediary between ammonium and nitrate. As more interfaces developed, global nitrite concentrations reached the threshold at which it became feasible for organisms to use nitrite in the DRNA and denitrification pathways. At the very least, these results indicate that the oxygenation of the oceans and atmosphere stimulated the global nitrogen cycle by increasing the availability of oxidized varieties of nitrogen.

Atmospheric Nitrogen

The primordial nature of Nif shown in the data indicates that there were appreciable levels of atmospheric nitrogen gas in the early Earth’s atmosphere, in
concordance with most knowledge of the history of the atmosphere (Stüeken et al., 2016; Mikhail and Sverjensky, 2014).

However, these results do not support the possibility of a nitrous oxide-rich atmosphere in the mid-Proterozoic. The argument for a so-called “laughing gas greenhouse” hinges on the assumption that widespread euxinia following the Great Oxidation Event would have reduced marine copper levels through sulfide stripping (Buick, 2007). It then follows that Nos, the terminal enzyme in the denitrification pathway, which requires copper in its active site, would be too difficult to produce in such a copper-limiting environment, leading to a buildup of Nor-generated nitrous oxide. These data suggest a different story, showing the high rates of mid-Proterozoic transfer for Nos, as well as for NirK, another denitrifying enzyme with copper in the active site, indicating that copper levels were not sufficiently low to prevent fixation of horizontally transferred copper-containing enzymes.

Conclusions

Five general conclusions can be drawn from this study: (i) Biological nitrogen fixation is a primordial process, possibly necessary for the early evolution of life on Earth. (ii) The first nitrogenase was Nif, the variety with a molybdenum cofactor in the active site. (iii) Prior to the oxygenation of the atmosphere, nitrate and ammonium were the prevailing forms of dissolved nitrogen in the oceans, with little to no nitrite. (iv) Following the oxygenation of the atmosphere, the abundance of oxidized varieties of nitrogen, especially nitrite, increased. (v) Limited copper availability did not cause Mid-
Proterozoic atmospheric nitrous oxide levels to be substantially higher than present levels.

In future, this process can be applied to other nitrogen-related genes, either those involved in bioavailable nitrogen uptake or the poorly-understood anammox pathway, in order to paint a clearer picture of nitrogen availability throughout Earth’s history.

**Acknowledgements**

I would like to thank Dr. Rika Anderson, who served as both the principal investigator for this research project and my advisor for the resulting write-up. This paper would not have been nearly as successful without her expertise in bioinformatics and incredibly helpful feedback on my writing. Her enthusiasm for genomics made this research exciting and encouraged me to pursue a career in the field. I would also like to thank Dr. Eva Stüeken at the University of St. Andrews, who helped to bring geological context into an otherwise biologically-focused study. Additionally, I would like to thank my parents at home and my friends at Carleton for their continued support throughout my time in college.
References Cited

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., 1990, Basic local alignment search tool: Journal of Molecular Biology, v. 215, p. 403-410


Buick, R., 2007, Did the Proterozoic ‘Canfield Ocean’ cause a laughing gas greenhouse?: Geobiology, v. 5, p. 97-100


Eigenbrode, J.L., and Freeman, K.H., 2006, Late Archean rise of aerobic microbial ecosystems: PNAS, v. 103, no. 43, p. 15759-15764


NCBI Resource Coordinators, 2016, Database resources of the National Center for Biotechnology Information: Nucleic Acids Research, v. 44, p. D7-D19

Noftke, N., Christian, D., Wacey, D., and Hazen, R.M., 2013, Microbially induced sedimentary structures recording an ancient ecosystem in the ca. 3.48 billion-year-old Dresser Formation, Pilbara, Western Australia: Astrobiology, v. 13, no.12, p. 1103-1124

Olson, S.L, Kump, L.R., and Kasting, J.F., 2013, Quantifying the areal extent and dissolved oxygen concentrations of Archean ocean oases: Chemical Geology, v. 362, p. 35-43


Zerkle, A.L., and Mikhail, S., 2016: The geobiological nitrogen cycle: From microbes to the mantle: Geobiology, v. 15, p. 343-352
Appendix A

Results of the study under alternative molecular clock models and calibration points. See Table 1 for calibration points.
Log-Normal Clock with Liberal Time Points
CIR Process with Conservative Time Points
CIR Process with Liberal Time Points

![Bar charts showing the distribution of various bacterial genes over time.](chart.png)