



Dilantha Gunawardana ^{1,*} and Venura Herath ²

- ¹ Research Council, University of Sri Jayewardenepura, Gangodawila, Colombo 10250, Sri Lanka
- ² Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya,
- Kandy 20400, Sri Lanka; venura@agri.pdn.ac.lk

* Correspondence: dilantha@sci.sjp.ac.lk

Abstract: Nitrous oxide is a long-lived greenhouse gas that exists for 114 years in the atmosphere and is 298-fold more potent than carbon dioxide in its global warming potential. Two recent studies showcased the utility of Azolla plants for a lesser footprint in nitrous oxide production from urea and other supplements to the irrigated ecosystem, which mandates exploration since there is still no clear solution to nitrous oxide in paddy fields or in other ecosystems. Here, we propose a solution based on the evolution of a single cytochrome oxidase subunit II protein (WP_013192178.1) from the cyanobiont Trichormus azollae that we hypothesize to be able to quench nitrous oxide. First, we draw attention to a domain in the candidate protein that is emerging as a sensory periplasmic Y_Y_Y domain that is inferred to bind nitrous oxide. Secondly, we draw the phylogeny of the candidate protein showcasing the poor bootstrap support of its position in the wider clade showcasing its deviation from the core function. Thirdly, we show that the NtcA protein, the apical N-effecting transcription factor, can putatively bind to a promoter sequence of the gene coding for the candidate protein (WP_013192178.1), suggesting a function associated with heterocysts and N-metabolism. Our fourth point involves a string of histidines at the C-terminal extremity of the WP_013192178.1 protein that is missing on all other T. azollae cytochrome oxidase subunit II counterparts, suggesting that such histidines are perhaps involved in forming a Cu center. As the fifth point, we showcase a unique glycine-183 in a lengthy linker region containing multiple glycines that is absent in all proximal Nostocales cyanobacteria, which we predict to be a DNA binding residue. We propose a mechanism of action for the WP_013192178.1 protein based on our in silico analyses. In total, we hypothesize the incomplete and rapid conversion of a likely heterocystous cytochrome oxidase subunit II protein to an emerging nitrous oxide sensing/quenching subunit based on bioinformatics analyses and past literature, which can have repercussions to climate change and consequently, future human life.

Keywords: nitrous oxide; nitric oxide; cytochrome oxidase subunit II; urea; nitrous oxide reductase; Y_Y_Y domain; heterocysts

1. Introduction—Nitrous Oxide, Climate Change and Azolla

Before the oxygenation of the ancient world billions of years ago, prokaryotic lifeforms had to contend with nitrous oxide that saw its birth from corona discharge and not as lightning strikes—although there is literature attributing nitrous oxide emanation to streaks of lightning—which anucleated microorganisms reduced using an enzyme designated as nitrous oxide reductase [1]. Nitrous oxide is a greenhouse gas that subsists in the atmosphere for 114 years and is 298-fold more potent than the benchmark gas of carbon dioxide in global warming potential [2,3]. Nitrous oxide reductases are known to be of the same protein fold as cytochrome oxidases (subunit II), the former being the more pre-historical one and the latter evolving from the ancient fossil of a fold, although this is now disputed [4]. There are two biochemical pathways for the formation of nitrous oxide: the conversion of nitric oxide to nitrous oxide by nitric oxide reductases or by ammonia



Citation: Gunawardana, D.; Herath, V. A Hypothesis on How the Azolla Symbiosis Mitigates Nitrous Oxide Based on In Silico Analyses. *J* 2022, *5*, 166–185. https://doi.org/10.3390/ j5010013

Academic Editor: Elena Cristina Rada

Received: 28 December 2021 Accepted: 8 February 2022 Published: 4 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). oxidation [5]. Although nitrous oxide reductases are mostly found in heterotrophic bacteria that are capable of reducing nitrous oxide to nitrogen gas, they too are found in Archaea and in selective photosynthetic cyanobacteria. There is little information on the cyanobacterial *nosZ* gene containing microorganisms, where the *nosZ* gene product is responsible for the catalytic biogenesis of nitrogen gas from nitrous oxide that is known to be present in cyanobacteria [6]. Therefore, it would be interesting to search for candidates of nitrous oxide-reducing enzymes in cyanobacteria, even ones of which the catalytic activity on nitrous oxide is secondary.

In cyanobacteria, there are two kinds of cytochrome oxidase types: one that is expressed in all types of cells and ones which are specific for heterocysts, the cellular sites for nitrogen fixation [7]. They carry the designations, heterocyst-specific *coxBACII* operon and the globally expressed *coxBAC1* operon, with the former found in heterocysts as well as pro-heterocysts [7]. Nitrogenases, which are localized inside heterocysts of cyanobacteria, are known to be cross-reactive to a myriad of gases, namely nitrogen, nitrous oxide, acetylene and carbon monoxide that can become the substrate for the molybdenum-type nitrogenases [8]. Nitrous oxide, just the same as oxygen, is a highly oxidative gas, and the ambient presence of nitrous oxide by biological and anthropogenic means can lead to the evolution of systems that are capable of using nitrous gas as effector or substrate. One such source of nitrous oxide is urea, which was first synthesized in 1828 by Friedrich Wohler and has carried on its shoulder rice cultivation—especially the urea-responsive elite cultivars that were developed at the International Rice Research Institute (IRRI) during the Green Revolution of 1960–70s—for 150 years since the advent of the Industrial Revolution.

A conspicuous site in irrigated rice fields is water ferns of the genus *Azolla*, which are found omnipresent in most parts of Asia, particularly in the rice belt countries such as Vietnam and Thailand. Azolla has been used for 1000 years or more in rice cultivation, and farmers depend on Azolla for growing rice cultivars from the presence of a cyanobiont residing inside Azolla plants that is known for highly efficient nitrogen fixation [9]. The history of Azolla cultivation goes back even before urea was used, and the combined treatment of Azolla and urea is now a common sight due to attempts at Azolla-mediated mitigation of greenhouse gases due to rising global warming and the resultant effects of climate change [10,11]. The "how" of nitrous oxide mitigation by Azolla has remained elusive thus far. This communication is to remediate that notion by dissecting biology using bioinformatics.

The holistic nature of the Azolla system and recent studies have prophesized and shown the advantages of using Azolla as a complete substitute or as a supplement to urea usage [10,11]. Azolla has a cyanobiont living inside its fronds that is known by a trilogy of genera, namely *Anabaena*, *Nostoc* and *Trichormus* [12]. The nature of the cyanobiont is obligate and it is commonly inferred that the cyanobiont is in transit from being a free-living microorganism to an endosymbiont to an ultimate plastid status [13]. The cyanobiont inside Azolla possesses the nitrogenase enzyme that is molybdenum-reliant, and this furnishes copious amounts of ammonia at the expense of nitrogen gas in an energetic and efficient transformation [12].

It is mandatory that a solution to nitrous oxide is found from the biological arsenals that evolution has created. We show here in this bioinformatics manuscript that the Azolla symbiotic system can theoretically be a putative remedy for nitrous oxide emanation. We are especially interested in domain architecture of a single sensor protein, cellular localization, promoters, tertiary structures, linker regions, phylogeny, heterocyst microenvironment and physiology and plasticity of ancient folds that may have repercussions on accompanying an emerging function that we hypothesize is gaining in strength. The purpose of this manuscript is to connect the dots of theory as to how Azolla may be able to quench and possibly transform nitric/nitrous oxide as a solution to elevated nitrous oxide, as well as aiding scientists to better fashion such systems using newer methods such as CRISPR-Cas9, which may be able to enhance the function of an emerging protein.

2. Results and Discussion

2.1. Rationale for an In Silico Study of a Candidate Protein from Trichormus azollae

WP_013192178.1 was earlier identified by the first author of this study, using sequence comparison to a nosZ-containing cyanobacterial protein [3]. The comparison between WP_013192178.1 and the nosZ-containing protein yielded ~31.43% identity and ~55% sequence coverage [3]. Note that the assignment of nomenclature was based on computational methods and not wet experiments. According to the SCOP (structural classification of proteins) classification, a protein that has >30% sequence identity belongs to the same protein/enzyme family, with most arbitrational cutoffs determined for coverage taken as >50% [14]. According to SCOP, members of a protein superfamily have low sequence conservation (<30%) and share a common evolutionary origin, which puts nitrous oxide reductases and cytochrome oxidases (subunit II) in the same superfamily, although the 31.43% sequence identity and 55% sequence coverage identifies the WP_013192178.1 protein as belonging to the same family as the nitrous oxide reductases (nosZ gene products) [3]. It seems this cytochrome oxidase subunit 2 has partial features of a nitrous oxide reductase. The complete set of *T. azollae* cytochrome oxidases subunit 2 proteins are shown in Table 1. The c-terminal copper center of nitrous oxide reductases are conserved in the candidate protein under study in this in silico exploration (Figure 1).

Table 1. Cytochrome oxidase subunit 2 enzymes in *T. azollae*. All sequences were extracted from the NCBI protein database using the WP_013192178.1 sequence as the query.

Cyanobiont	Host Plant	Cytochrome Oxidase Subunit II Enzymes
Trichormus azollae	Genus Azolla (Monilophyte/Pteridophyte)	WP_013192178.1 WP_013191541.1 WP_013192238.1 WP_013193310.1

In two recent studies, the researchers showed a significant reduction in nitrous oxide in urea/biochar-treated rice paddies in the presence of grown Azolla, which suggests that Azolla cyanobiont may have means to negate the nitrous oxide that is produced due to urea application and denitrification [10,11]. We suggest here that WP_013192178.1 is capable of nitrous oxide sensing and/or quenching and we use bioinformatics and literature to further our claims.

On this note, key studies have demonstrated that nitric oxide/nitrous oxide is susceptible to interactions with terminal cytochrome oxidases occupying sites found within the structure of the enzyme, although it is not the primary substrate [15]. Nitric oxide has been shown to be a substrate, inhibitor and effector of candidate cytochrome oxidases (Figure 2) [16]. At lower concentrations of oxygen, as is the case for heterocysts, the cytochrome oxidases are reduced, allowing for the accumulation of NO in the near vicinity (microenvironment) of the enzyme and for not consuming NO [16]. The above sequestering reaction has been termed physiologically relevant.

It is also proposed that nitrous oxide may be converted into dinitrogen gas in the presence of enzymes [15]. In fact, both mammalian (bovine) and plant (bean) mitochondrial cytochrome oxidases have been shown to be inhibited by nitrous oxide [15]. The chemistry of the interaction between cytochrome oxidase II and nitrous oxide is shown to be participatory to multiple sites of the bovine heart cytochrome oxidase, although the method of participation is not suggested to involve ligand donation or oxidase function [15]. The cross-reaction of cytochrome oxidases with nitrous oxide has been reported in other studies, as a measure of concentration of cytochromes.

Therefore, the rationale of this study is that the SCOP-based assessment of likely family and putative (surrogate) sensor activity—i.e., nitric/nitrous oxide sensing—requires exploration in silico first, prior to the study of biological and biochemical parameters from lab-based experimentation. This is such an attempt.

WP_013192178.1 ALE15278.1 CAA37714.2	MSNEETKMRLNRRQMLGTTAFMAAAGAVG-AGGALTLSGGTATPARAQETSG MSDKDSKNTPQVPEKLGLSRRGFLGASAVTGAAVAATALGGAVMTRESWAQAVKESK	0 51 57
WP_013192178.1 ALE15278.1 CAA37714.2	SSYEVKPGELDEYYVFFSSGQSGEIRIVGAPSMREMMRIPVFNRCSATGWGQTNESRKVM QKIHVGPGELDDYYGFWSGGHQGEVRVLGVPSMRELMRIPVFNVDSATGWGLTNESRHIM	0 111 117
WP_013192178.1 ALE15278.1 CAA37714.2	TEGLLPETVEFLKDQGGLYLNGDLHHPHPSFTDGTYDGRYLYANDKSNSRVCRIRLDVMK GDSAKFLNGDCHHPHISMTDGKYDGKYLFINDKANSRVARIRLDIMK	0 171 164
WP_013192178.1 ALE15278.1 CAA37714.2	CDKIIQLPNQHTVHGLRVQKYPKTGYVFCNGEDAVPVPNDGKTMGDKN-SYQAIFTAVDG CDKMITVPNVQAIHGLRLQKVPHTKYVFANAEFIIPHPNDGKVFDLQDENSYTMYNAIDA	0 230 224
WP_013192178.1 ALE15278.1 CAA37714.2	MKIPSSIWTLLIGIVLTL ETMEVAWQVMVDGNLDNVDADYQGKYCFATCYNSEEGFTLADMMASEQDWVVIFNLKRIE ETMEMAFQVIVDGNLDNTDADYTGRFAAATCYNSEKAFDLGGMMRNERDWVVVFDIHAVE * . *.::.:	18 290 284
WP_013192178.1 ALE15278.1 CAA37714.2	ASLWYGQNHGLL EAVAKGDYKEIGGVPVLDGRKGSPYTRYVPVPNSPHGINPAPDGIHVVANGKLS AAVKAGDFITLGDSKTPVLDGRKKDGKDSKFTRYVPVPKNPHGCNTSSDGKYFIAAGKLS *	30 344 344
WP_013192178.1 ALE15278.1 CAA37714.2	PTAASDEAVLIDGLFNAMMTVSTSIFLIVEGVLV PTVTVFDVRKFDDLFDDKIQ-ARDTVVAEPELGLGPLHTAYDGKGNAYTTLF-IDSQVCK PTCSMIAIDKLPDLFAGKLADPRDVIVGEPELGLGPLHTTFDGRGNAYTTLF-IDSQVVK ** : : : .** : . *	64 402 403
WP_013192178.1 ALE15278.1 CAA37714.2	YSIFKYRRRAGDNEDGPPIEGNVPLEILWTAIPAIIVIGISVYSFDVYNE WNIEDA-KRAYAGEKVDPIRHKLDVHYQPGHNHTSMGQTKEADGKWLISLNKFSKDRYLN WNMEEA-VRAYKGEKVNYIKQKLDVHYQPGHLHASLCETNEADGKWLVALSKFSKDRFLP :.: . ** .*. *. :: : ::::: :* * :	114 461 462
WP_013192178.1 ALE15278.1 CAA37714.2	IGGFNPHTAHEAAIMQESISMPGTAIAVTLSDTPGSTEPNLNQEKSDEAMQDPSTAAVPN VGPLKPENDQLIDISGDEMVLVHDNPTFAEPHDATIVHASKINPVHVWN VGPLHPENDQLIDISGDEMKLVHDGPTFAEPHDCIMARRDQIKTKKIWD :* ::* : * :.: : * * :*: . : : : : : :	174 510 511
WP_013192178.1 ALE15278.1 CAA37714.2	PDTPQMHNGSGVGSVAPTLGASREKEGQASQLVVNVSALQYAWIFTYPDTGVTT RDDPFFVDAVAQAKADNIDLMVDSEVIRDGNKVRVYMTSAAPAFGL RNDPFFAPTVEMAKKDGINLDTDNKVIRDGNKVRVYMTSMAPAFGV : * : : : : : : : : : : : : : : : : : :	228 556 557
WP_013192178.1 ALE15278.1 CAA37714.2	GELHVPIGREVQMNMTANDVIHAFWIPEFRLKQDVIPGRQTEIRFTPKTPGDYSLI DDFTVKQGDEVTVYVTNIDEVEDLTHGFCIVNYGINMEVAPQATASVTFKASRPGVYWYY QEFTVKQGDEVTVTITNIDQIEDVSHGFVVVNHGVSMEISPQQTSSITFVADKPGLHWYY :: * * ** : :* :* :* : : : : : : : * : : * : * :	284 616 617
WP_013192178.1 ALE15278.1 CAA37714.2	CAELCGPY <mark>H</mark> GAMRTQVVVESQPAFDSWMQEQLVASKDTLNQAVSINPANNMSVDKFLAPY CTWFCHAMHMEMKGRMLVEAQGA CSWFCHAL <mark>H</mark> ME <mark>M</mark> VGRMMVEPA *: :* * * :::**	344 639 638
WP_013192178.1 ALE15278.1 CAA37714.2	SKDMGIHPEMLHQVHH 360 639 638	

Figure 1. The *T. azollae* WP_013192178.1 aligned against two nitrous oxide reductases (*nosZ* gene products), showcasing the key residues (in color) that are conserved in the motif $H-X_{34}$ -CXEXCX₃HX₂M. The putative Y_Y_Y domain is shown in a box. Stars indicate identical fully-conserved residues between the three sequences.





Figure 2. The likely functional aspects of cytochrome oxidases as a consequence of oxygen, nitric oxide and oxidative potential. The above illustration was adapted partially from a past drawing [16].

2.2. The Sensory Periplasmic Y_Y_Y Domain

A unique sequence motif yielded a stretch of three amino acids (KTP) in a short loop and a single helix (GDYSLI). The above nine residue sequence motif is distinct from the same stretch in closely associating proteins, and the highlighted mutations (TP and S) are not all simple wobble (third) position mutations, which emphasizes their integral and likely functional nature. The above stretch of 9 amino acids is found in a region containing a periplasmic sensor Y_Y_Y domain (residue from 260 to 284) that is shown in Figure 3, of which the function has remained elusive to the scientific community, although it is said to be involved in bacterial signal transduction. The periplasmic localization of this Y_Y_Y domain tallies with the partial periplasmic localization of the WP_013192178.1 protein, which consolidates the insights drawn from motif analyses. The Y_Y_Y protein is synonymous with nitric oxide sensing and the modification of this domain in the WP_013192178.1 protein hints at a sensory function for nitric oxide [17]. The Y_Y_Y domain is only found in the WP_013192178.1 protein and is missing in the closest allied protein sequences in T. azollae that it aligns with, suggesting a recent modification to transform into a nitric/nitrous oxide sensor as suggested in this study. Out of 506 Y_Y_Y containing proteins in the PFAM database, 237 contain the histidine kinase domain outside of the Y_Y_Y domain [18]. Other domains found with Y_Y_Y motifs are GGDEF adenylyl cyclase signaling domain, as well as the SpoIIE sporulation domain, while 26/506 Y_Y_Y proteins were not two-component regulators in sequence [18].

When the 24 residue Y_Y_Y domain of the WP_013192178.1 protein is searched against the family "cyanobacteria" using the BLASTp search tool, the results reveal many proteins, with the top two being cytochrome oxidase subunit II proteins from Nostoc sp. FACHB-110 that when searched for putative conserved domains contained the Y_Y_Y domain. Full length sequences in the list "after" Nostoc sp. FACHB-110 (Figure 4) did not contain the Y_Y_Y domain. *Nostoc* sp. FACHB-110 is a cyanobacterium that is found as "organism specific" in China. Therefore, we present a hypothesis here that all three cytochrome oxidases that have a putative Y_Y_Y domain are likely to have undergone some degree of evolution to handle nitric/nitrous oxide from the immediate ecosystem. Whether the nitric oxide/nitrous oxide binding is used for the concomitant transformation of the gases cannot be inferred from this study alone. None of the other full-length cytochrome oxidase subunit II proteins in *Trichormus azollae* contain the Y_Y_Y domain (Figure 5).



Pfam ID:

Y_Y_Y

Description:

PF07495, Y_Y_Y domain

Appearance:

Position	260284
Alignment Query Database	KQDVIPGRQT-EIRFTPKTPGDYSLI keWveasdysqeasytnlppGdYtlk
i-Evalue	0.24

Sequence:

MKIPSSIWTLLIGIVLTLASLWYGQNHGLLPTAASDEAVLIDGLFNAMMTVSTSIFLIVE GVLVYSIFKYRRAGDNEDGPFIEGNVPLEILWTAIPAIIVIGISVYSFDVYRBIGGFNP HTAHBAAIMQESISMPGTAIAVTLSDTPGSTEPNLNQEKSDEAMQDPSTAAVPNPDTPQM HNGSGVGSVAPTLGASREKEGQASQLVVNVSALQYAWIFTYPDTGUTTGELHVPIGREVQ MMMTANDVIHAFWIPEFRLKQDVIPGRQTEIRTTPKTPGDYSLICAELCGPYHGAMRTQV VVESQPAFDSWMQEQLVASKDTLNQAVSINPANNMSVDKFLAPYSKDMGIHPEMLHQVHH

Figure 3. Domain/motif architecture of the *T. azollae* cytochrome oxidase subunit II protein (WP_013192178.1 protein) as indicated by the PFAM database. The Y_Y_Y domain is mostly found at the end of the beta propellers (Pfam:PF07494) in a family of two component regulators [19,20]. The region in red font showcases the putative Y_Y_Y domain.

Sequences producing significant alignments Download V Res Select columns V Show 10							100 🌱 🔞			
Select all 100 sequences selected			GenPept Graphics Dist	ance t	ree of	results	Multi	<u>ple align</u>	ment	New MSA Viewer
		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
		cytochrome c oxidase subunit II [Trichormus azollae]	Trichormus azollae	83.8	83.8	100%	1e-18	100.00%	360	WP_013192178.1
		cytochrome c oxidase subunit II.[Nostoc sp. FACHB-110]	Nostoc sp. FACHB-110	73.2	73.2	100%	6e-15	88.00%	345	MBD2440653.1
		cytochrome c oxidase subunit II [Nostoc sp. FACHB-110]	Nostoc sp. FACHB-110	73.2	73.2	100%	6e-15	88.00%	335	WP_199336312.1
		cytochrome c oxidase subunit II [Fischerella sp. NIES-4106]	Fischerella sp. NIES-4106	71.9	71.9	100%	2e-14	88.00%	361	WP_096673871.1
		cytochrome c oxidase subunit II [Fischerella sp. PCC 9339]	Fischerella sp. PCC 9339	71.9	71.9	100%	2e-14	88.00%	361	WP_017312851.1
		cytochrome c oxidase subunit II [Pelatocladus maniniholoensis HA4357-MV3]	Pelatocladus maniniholoensis HA4357	71.9	71.9	100%	2e-14	88.00%	361	MBW4435131.1
		cytochrome c oxidase subunit II [Westiellopsis prolifica]	Westiellopsis prolifica	71.9	71.9	100%	2e-14	88.00%	360	WP_131118407.1
		cytochrome c oxidase subunit II [Fischerella sp. PCC 9431]	Fischerella sp. PCC 9431	71.9	71.9	100%	2e-14	88.00%	360	WP_026723307.1
		cytochrome c oxidase subunit II [Hapalosiphon sp. MRB220]	Hapalosiphon sp. MRB220	71.9	71.9	100%	2e-14	88.00%	360	WP_053458487.1
	•	cytochrome C oxidase subunit II [Hapalosiphonaceae cyanobacterium JJU2]	Hapalosiphonaceae cyanobacterium JJU2	71.9	71.9	100%	2e-14	88.00%	360	RAM49556.1

Figure 4. The top ten sequences when the putative Y_Y_Y domain (KQDVIPGRQTEIRFTPK-TPGDYSLI) is searched against the family cyanobacteria using the BLASTp search tool. Beyond *Nostoc* sp. FACHB-110 (sequence 2 and 3 above)—a cyanobacterium found in China—the Y_Y_Y domain is absent, i.e., only three cyanobacterial sequences are likely to contain the putative Y_Y_Y domain in cytochrome oxidase subunit II proteins, and all appear to be found associated with the rice growing belt, namely *T. azollae* and *Nostoc* sp. FACHB-110 (Note: this is when the full-length sequence of the corresponding protein is searched for likely functional domains and not domains by themselves).



Figure 5. The likely domains of the three cytochrome oxidases from *T. azollae* that do not contain the Y_Y_Y domain.

2.3. The Phylogeny of the Cytochrome Oxidase Subunit II Enzymes

Phylogeny was inferred using the maximum likelihood method with bootstrap support from 500 replicates. Phylogenetic kinships revealed WP_013192178.1, the cytochrome oxidase subunit 2 sequence of *T. azollae*, as an evolving protein and *nosZ* gene products of bacterial candidates as supplementary enzymes that evolved from a single origin (Figure 6). A single archaeal contender was used as the outgroup to anchor the phylogenetic tree (Figure 6). Cluster 3 of the phylogenetic tree contained sequences from all three cyanobionts, with an internal duplication event (of *N. punctiforme*) (Figure 6). There is high bootstrap support for Cluster 1 and Cluster 3, intraspecifically, and this shows that the phylogenetic inferences are likely to be accurate for the above eight sequences, organized as two clusters. Cluster 1 appears to have committed to terminal outcomes in all three cyanobiont species using two consecutive symmetrical bifurcation events (both prior to Bryophyte evolution), but not cluster 2 and 3, which evolved in three stages. As for cluster 2, which too has conserved members of the three cyanobionts, there appears to be two evolutionary outcomes to *N. punctiforme* before the temporally distinct division of the common clade into its nodal tips carrying Cycas and Azolla cyanobiont sequences.

However, the key protein of emphasis (WP_013192178.1) in this study is poorly organized into a cluster of low bootstrap support (<41%) (Figure 6) so that we are unable to draw strong inferences from that clade. However, the sequence deviation supports a case of



divergent evolution which we hypothesize to involve contact/exposure with highnitrous oxide levels.

Figure 6. The phylogenetic tree of cyanobiont (*T. azollae, Nostoc punctiforme* and *Nostoc cycadea*) cytochrome oxidase subunit II sequences with bacterial nitrous oxide reductases (nosZ gene products) and one archaeal protein sequence (outgroup) for the inference of phylogeny from the above sequences to resolve the evolutionary relationships of proteins that possess the same fold. Clusters are named 1, 2 and 3 and as bacterial *nosZ* gene products. *Haloarcula rubripromontori* belongs to the kingdom Archaea.

2.4. Gene Synteny Analyses

When a gene synteny analysis was performed on *coxBAC* operons of *T. azollae* with 3 other canonical model organisms (a single strain of each *N. punctiforme, Anabaena cylindrica,* and *Trichodesmium erythraem*), it could be seen that *T. azollae* had the most copies of the *coxBAC* operons (four complete operons) (Figure 7). Visualization of the synteny map was performed using SimpleSynteny v1.4 using default options [21].



Figure 7. Gene synteny analyses of *coxBAC* operons in cyanobacterial genomes, including *T.* (*Nostoc*) *azollae. A. cylindrica PCC_7122* is a free-living species, *T. erythraem_*IMS101 is a subsection III cyanobacterium that lives in oligotrophic marine waters and *N. punctiforme* PCC 73102 is a cyanobiont.

2.5. Promoter for the NtcA Transcription Factor

One key regulator of nitrogen fixation, heterocyst development and function, is the NtcA transcription factor, which is thought to regulate the downstream function of >2000 mostly nitrogen-metabolism-related genes [22]. We searched for the NtcA binding promoter region in the WP_013192178.1 encoding gene to see whether it was under top-down regulation. We found a 100% match for the *NtcA* promoter at -66 to -52 region upstream of the WP_013192178.1 coding gene (Figure 8). The sequence reads 5' GTA gcctgcac TAC 3' (Figure 8), which was accurate for the signature binding sequence (5' GTA n(8) TAC 3') of the NtcA transcription factors [23]. Therefore, the WP_013192178.1 encoding gene is likely to be part of the nitrogen response transcriptome and interactome and functioning inside heterocysts. Heterocysts are oxygen depleted, photosystem-2 inactivated and ATP-producing (from photosystem 1), which synthesize ammonium ions and transport such nitrogen currencies to the neighboring vegetative cells as glutamine. None of the upstream regions of the closest gene homologs of the focal gene under scrutiny here (WP_013192178.1 coding gene), contain the NtcA binding sequence, showcasing that perhaps this modification is recent in origin.

Under nitrogen-rich conditions, 2-oxoglutarate counts are low and NtcA is found in an inactive form [22]. In a recent publication, it was shown that the NtcA transcription factor of *T. azollae* putatively uses a putative disulfide bond to perform its robust function in modifying gene expression using *NtcA* responsive promoters [22]. We propose here that nitrous oxide released from urea can gradually replace or function secondarily to the canonical function of reducing oxygen to water by the WP_013192178.1 enzyme due to extremely low levels of oxygen found inside heterocysts. However, cellular respiration inside heterocysts has to be taken into consideration, where the WP_013192178.1 enzyme can contribute as a terminal oxidase by keeping the oxygen concentrations extremely low so as not to inhibit the nitrogenase function, although the light-dependent photosystem-1based Mehler activity in heterocysts is known to supersede respiration [24], again giving the cytochrome oxidase subunit II the freedom to acquire a newer substrate through rapid evolution. In fact, Cyt *c*6 is the key donor for both cytochrome oxidase enzymes as well as the engine for cyclic photosynthesis of photosystem 1 in heterocysts, while in vegetative cells, there is the division of labor by Cyt *c*6 and plastocyanin as electron donors [25].

- 1 tgcccgcttg aggtagatat tettaaaatt taetatttga teagteeaga eetattteea
- 61 ccctaaccta tgagctaagg tttaaaaatt gcctttcccc tatactttga atcactctag
- 121 ccacttttag ctgtaagatt taggtatttt ttaagatgtg aataattaaa tgttggcaat
- 181 tatttagatt tteetaaata attgecaaca tttteettaa gttttggtgt taattateat
- 241 ttattgttga caacatcaat atgaatacct ttactcaaga ttagtaacct agtcaaaaaa
- 301 attaataaaa atttcagata tttataatct atattccGTA gcctgcacTA Ccttagtaag
- 361 tgagttagtt tgaaataacg taatcaattc cattaagtaa gccgtgaaaa ttccaagttc





Figure 8. The NtcA transcription factor binding promoter of the WP_013192178.1 encoding gene shown inside an orange box magnified with the translation start site shown highlighted in yellow above the local gene arrangement surrounding the candidate gene.

2.6. Terminal Histidines

There are four histidines that are found at the C-terminal extremity of the WP_013192178.1 protein (Figure 9). The copper center in bacterial nitrous oxide reductases is formed by six-seven histidine residues. Therefore, the presence of a string of histidines, which are completely unavailable in all other *T. azollae* cytochrome oxidase subunit II enzymes (Figure 9), suggests that this may be an adaptive function to aid in the handling of nitrous oxide in periplasmic spaces, which serendipitously happens to be the localization of all classical nitrous oxide reductases. The MBD2440653.1 and WP_199336312.1 sequences from *Nostoc* sp. FACHB-110, which possess the putative Y_Y_Y domain, contain the same histidines as the WP_013192178.1 protein. The C-terminal half of the three proteins are highly conserved, suggesting a likely significant function (Figure 10). Clustered histidines are known to donate ligands for copper binding [26] We state that our bioinformatics-based inferences require further functional validation.

116 117 118 120	LIVEGVLVYSIF <mark>KYRR</mark> AGDNEDGPPIEGNVPLEILWTAIPAIIVIGISVYSFDVYNEIG LVVEGTILIFLFMYRRRPGDNSDGTPIKGNLALEIFWTAVPTVIVVVLGIYSVNIYNQMG LGLLGMMVYSILFFRAPKNDYSEGHPSRGDIKLEILWTASPTLLVLWIAWQGFNIYQQLN LAVTSTVLYSITFHRAKNYDMSDGPHIEGNVTLEVVWTAIPVMLVLWIAAYSYQVYEQMG	WP_013192178.1 WP_013191541.1 WP_013193310.1 WP_013192238.1
176	* : . :: : .* * .:* .* .* .*:: **:.*** *.::*: : ::*:::. GFNPHTAHEAAIMQESISMPGTAIAVTLSDTPGSTEPNLNQEKSDEAMQDPSTAAVPNPD GLEPGSHDHGHVAHVSGTAIAASMSDGS0	WP_013192178.1
140 14 143	ILGLGQI'HHLPTSLEKPA-YGLVN IQGPSHLVHLHNPMQMQTGYAQE	WP_013193310.1 WP_013192238.1
234	TPQMHNGSGVGSVAP TLGASREKEGQASQLVVNVSALQYAWI FTYPDTGVTTGELHVP	WP_013192178.1
177	NDKPKPASETIEVEVKONDWSFRYPNNV-ISYELHLP	WP 013193310.1
184	EETPAETVSEPLEKIDVLAKQWAWVFHYPEKGITSTELHLP . ::* *: * ***:. : ***:*	WP_013192238.1
294	IGREVQMNMTANDVIHAFWIPEFRLKQDVIPGRQTEIRFTPKTPGDYSLICAELCGPYHG	WP 013192178.1
256	VGADVQLNLSADDVIHSFWVPQFRLKQDAIPGVPTELRFVATKPGTYPVVCAELCGGYHG	WP_013191541.1
237	eq:vnlrtrlnlhakdvlhsfyvsefrlqqyivpgrdidlvvtpnrigqyklkdalfsgtyfa	WP_013193310.1
244	SDRRVRLALQSEDVLHGFYIPAFRLKQDIVPGRTINFEFTPIRPGKYSLTDSQYSGTYFA:: :**:*.*:: ***:* :** :: * * : * *	WP_013192238.1
354	AMRTQVVVESQPAFDSWMQEQLVASKDTLNQAVSINPANNMSVDKFLAPYSKDMGI <mark>H</mark> PEM	WP_013192178.1
315	${\tt SMRSQVIV} H {\tt TQGDFDSWLTENQLAQQQGKPQVVAVNPA-DLSTSEFLTPYIND {\tt MGVGAAT}$	WP_013191541.1
287	LMDANINVESLEQYNQWVTQAK-QKPIAMKNHAVDEYTQPPKTLFKSNWYT	WP_013193310.1
295	TMQANVIVETPEEYHEWLTKAASHIPTPAKNQAASEYAQTSNQRIKTGWGT * ::: *.: : :*: : : :::: : :	WP_013192238.1
	L <mark>H</mark> QV <mark>HH</mark> 360	WP_013192178.1
	LSQIEK 321	WP_013191541.1
	VAPKQSSIANSK- 299	WP_013193310.1
	VKPAAPPVVNFSG 308	WP 013192238.1

Figure 9. Sequence alignment of the putative cytochrome oxidase subunit II enzymes in *T. azollae*. The exclusive presence of four histidines (colored in yellow) on the C-terminus of the WP_013192178.1 protein in *T. azollae*, perhaps priming the presence of a copper center. The basic locality (positively charged residues) containing the twin arginines is also shown in pink and cyan. (NOTE: MBD2440653.1 and WP_199336312.1 sequences from Nostoc sp. FACHB-110 contain the same histidines). Stars indicate identical fully-conserved residues between the four sequences.

WP_013192178.1 MBD2440653.1 WP_199336312.1	MKIPSSIWTLLIGIVLTLASLWYGQNHGLLPTAASDEAVLIDGLFNAMMTVSTSIFLIVE MKIPSSIWTLLIGIVLTLVSIWYGQNHGLLPTAASDEAVLVDGLFNSMMTVSTGIFLIVE MIGIVLTLVSIWYGQNHGLLPTAASDEAVLVDGLFNSMMTVSTGIFLIVE :*******.*:	60 60 50
WP_013192178.1 MBD2440653.1 WP_199336312.1	GVLVYSIFKYRRRAGDNEDGPPIEGNVPLEILWTAIPAIIVIGISVYSFDVYNEIGGFNP GVLIYCVIKYRRPGDRQDGPPVEGNVPLEILWTAIPAIIVIGISVYSFEVYNDIGGFDP GVLIYCVIKYRRPGDRQDGPPVEGNVPLEILWTAIPAIIVIGISVYSFEVYNDIGGFDP ***:*.::***** **.:****	120 120 110
WP_013192178.1 MBD2440653.1 WP_199336312.1	HTAHEAAIMQESISMPGTAIAVTLSDTPGSTEPNLNQEKSDEAMQDPSTAAVPNP-DTPQ HAGHNAPMAQSSMTMPGSAIAATLSDAQAQDPATAEVRNTDQIPQ HAGHNAPMAQSSMTMPGSAIAATLSDAQAQDPATAEVRNTDQIPQ *:.*:* : *.*::***:***:.	179 165 155
WP_013192178.1 MBD2440653.1 WP_199336312.1	MHN <mark>G</mark> SGVGSVAPTLGASREKEGQASQLVVNVSALQYAWIFTYPDTGVTTGELHVPIGREV HQD <mark>A</mark> PGVGSVAPSIGSSPDKAGQPPELLVNVTGLQYAWLFTYPETGVTTGELHVPIGREV HQD <mark>A</mark> PGVGSVAPSIGSSPDKAGQPPELLVNVTGLQYAWLFTYPETGVTTGELHVPIGREV ::. *******::*:* :* ** :*:***:	239 225 215
WP_013192178.1 MBD2440653.1 WP_199336312.1	QMNMTANDVIHAFWIPEFRLKQDVIPGRQTEIRFTPKTPGDYSLICAELCGPYHGAMRTQ QINMTANDVIHAFWVPEFRLKQDAIPGRQTEIRFTPKTAGDYTLICAELCGPYHGAMRSQ QINMTANDVIHAFWVPEFRLKQDAIPGRQTEIRFTPKTAGDYTLICAELCGPYHGAMRSQ *:***********	299 285 275
WP_013192178.1 MBD2440653.1 WP_199336312.1	VVVESQPAFDSWMQEQLVASKDTLNQAVSINPANNMSVDKFLAPYSKDMGIHPEMLHQVH VVVETPAAFEKWMQEQLVASHETLNKAVAVNPAD-LSPDEFLAPYTKDMGIQPEILHQVH VVVETPAAFEKWMQEQLVASHETLNKAVAVNPAD-LSPDEFLAPYTKDMGIQPEILHQVH ****: **:.*****************************	359 344 334
WP_013192178.1 MBD2440653.1 WP_199336312.1	H 360 H 345 H 335	

Figure 10. Sequence alignment of the WP_013192178.1 protein, with MBD2440653.1 and WP_199336312.1 from Nostoc sp. FACHB-110. Glycine-183 is shown in yellow. Stars indicate identical fully-conserved residues between the three sequences.

2.7. Translocation of the Periplasmic Domains

Periplasmic proteins such as nitrous oxide reductases conventionally require the presence of a transporter (TAT–Twin Arginine Translocation) to ferry the protein across from the cytoplasm to the periplasmic spaces and are identified by a leader sequence that includes a region "SRRXFLK" in bacteria. Although such a region is not found residue to residue, there are twin arginines in the N-terminal half surrounded by a strongly basic region (in the immediate vicinity) in our focal protein of this study (Figure 9).

Leader sequences more than 60 residues are found in bacteria. For example, the *Paracoccus denitrificans* nitrous oxide reductase, and at least three other homologs from bacteria, have unusually lengthy leader sequences [27]. The leader sequence of nitrous oxide reductase of the organism, *Paracoccus denitrificans*, is cleaved between amino acid 57 and 58 [27][•] However, we could not find a translocation signal peptide using the SignalP 5.0 web server, indicating that our inferences may be coincidental

2.8. Proposed Functional Mechanism of the Candidate Protein WP_013192178.1

A co-functional dependence has been attributed to nitrite reductases and NO reductases in heterotrophic bacteria [28]. They can be elaborated as the following [28]: (i) localization of NO reductase on the cytoplasmic membrane to preclude the diffusion of toxic NO to the cytoplasm; (ii) highly-efficient NO binding to NO reductase for the concomitant catalytic conversion of NO to N₂O, a stable product.

Therefore, the candidate protein under study here is likely to be present partially in the periplasmic spaces and to be effectuated by the NO present in that peri-cytoplasmic locality due to myriad processes. In the Azolla symbiosis, nitric oxide is made by conversion of arginine, while nitrous oxide is synthesized by ammonia oxidation and mitochondrial cytochrome oxidases activity in plants, while the microbiome in Azolla can also contribute nitrous oxide (nitrification and denitrification) while also taking into consideration urea breakdown in the ambience [29]. In fact, it is reported that aseptically grown plants emit nitrous oxide from mitochondrial nitric oxide conversion by terminal cytochrome oxidases [29] which strengthens our hypothesis.

We infer that these are the likely steps of functions based exclusively on bioinformatic inferences and past literature

- 1. The NtcA transcription factor binds to the promoter of the gene to produce mRNA that will be translated into the corresponding WP_013192178.1 protein
- The periplasmic domain of the WP_013192178.1 protein is exported into the periplasmic space.
- 3. The nitric oxide sensing function will negotiate periplasmic NO as the trigger using the Y_Y_Y domain to detect toxic NO in the periphery of the heterocyst membrane. Note, we also envisage direct binding of N₂O to the Y_Y_Y domain, since they are both small molecules and known to bind to the same proteins
- 4. The binding of a Y_Y_Y domain to nitric oxide has been reported at least once before to disentangle the sensor protein from the membrane localization, and this dissociation from a fixed membrane position, has been linked to subsequent binding to DNA to effectuate downstream functions [17]. We propose a similar mechanism here.
- 5. The mid-section of the candidate protein (WP_013192178.1), comprising a loosely random-coiled section as shown in the homology model (Figure 11), is likely to be functional as a DNA binding surface. DNA binding could be involved in transcription of nitric/nitrous oxide quenching/transforming ORFs or other regulatory functions. The evolutionarily "elderly" global fold common to nitrous oxide reductases and cytochrome oxidases (subunit II) may be more adaptable due to its long vigil in time and may even accommodate novel adaptive catalytic properties [30]. Long coiled segments are characteristic of DNA and RNA binding proteins, including DNA and RNA chaperones [31] and even as oligomerization interfaces.







Figure 11. (**Top**) Graphical representation of the common motif architecture of the four cytochrome oxidases subunit II proteins using Multiple Em for Motif Elicitation (MEME). (**Middle**) The secondary structural elements of the middle region (42 residues that do not align with other homologs) of the candidate WP_013192178.1 protein with the glycine-rich regions showcased in a green box; (**Below**) SWISS-MODEL Homology Model of the WP_013192178.1 protein using the PDB template (2yev.1.B)—a cytochrome oxidase subunit II—as reference. The extended middle section (Residue 118 to Residue 205) is shown in a red box.

2.9. DNA Binding and/or Oligomerization

It is of note that when the protein sequence WP_013192178.1 was queried for RNA and DNA binding residues using DRNApred server [32], there was a single glycine (Glycine-183) in a glycine-rich region that was identified as a DNA binding residue. Surprisingly, the MBD2440653.1 and WP_199336312.1 from *Nostoc* sp. FACHB-110 (Figure 10) lack this glycine (Glycine-183), which is mutated to an alanine in both proteins. When a simple PSI-BLAST search was performed on the linker region, it was clear that Glycine-183 was unique and unfounded in any of the closest matches, suggesting a residue that is functionally significant to the WP_013192178.1 protein. A single glycine found in a hinge region has been shown to be crucial for DNA binding for the HapR protein from *Vibrio cholerae*, where a single G39D mutation abolished the DNA binding property [33].

The glycine-rich region is found in a connecting "lengthy" region amounting to 43 amino acids between the N-terminus and the C-terminal cupredoxin-like fold (Figure 11). Such glycine-rich linker regions within DNA binding domains are found in proteins such as HapR, a quorum-sensing master regulatory protein in *Vibrio cholerae*, where they connect two helices and have been demonstrated to be significant for function. The 43-residue linker region when searched using PSI-BLAST only manifests ~76% sequence identity with its nearest match, showcasing a divergent nexus, and is significantly longer to fellow counterparts from *T. azollae* (Figure 11)

Glycine-rich regions are also used to form protein dimers or oligomers. In fact, linkers that are glycine rich are flexible and provide a linker to connect functional domains without interfering with the functional aspects of the individual domains [34]. The lengths of glycine-rich linkers vary from 2 to 31 amino acids and can play a role that has minimal constraints to separated functionalities of evolutionary-fused proteins [34].

The glycine-rich region in WP_013192178.1 (Figure 11) is likely to be part of a DNA binding mechanism and whether it is the regulation of specific genes, i.e., as a transcription factor, needs to be established. It could even be that the *coxBAC* operon in heterocysts is regulated by this DNA binding activity stemming from glycine-183.

2.10. Phosphorylation of Serine Residues

There were 6 serine phosphorylation sites in total in the WP_013192178.1 protein with two possessing compelling likelihoods: Ser 105 (0.719 probability) and Ser 304 (0.757 probability) (Figure 12). Out of the above, Ser-105 is unfound in all the other cytochrome oxidases found in *T.azollae*, while the position 304 is held by a serine found in the proximal 3-D space of the Y_Y_Y domain. Even the serine residue at position 5 is also unfounded in all other protein sequences. The presence of such unique serines could be used for signal transduction via protein phosphorylation by relevant kinases.



Figure 12. Putative serine phosphorylation sites in the WP_013192178.1 protein.

2.11. Interpreting a Hydropathy Profile of the WP_013192178.1 Protein

The hydropathy profile from the Kyte and Doolittle method indicates that there are two highly hydrophobic regions that are likely to form helices crossing the inner membrane of the *T. azollae* heterocysts (Figure 13). The two highly hydrophobic (positive > 1.6) transmembrane domains are likely to appear in the first 110 residues, with a strongly hydrophilic "surface exposed" region separating the two (a surface exposed region between residues 65 and 85) that can be seen in the homology model in Figure 13 where the two helices are depicted in the color "blue". However, the region 110-200 amino acids are weakly hydrophilic, indicating a region that is immersed into the cytoplasmic side of the cell. In addition, when the "outside negative" rule is implemented, we arrive at the periplasmic region (220–300) rich in aspartates (D) and glutamates (E), suggesting their translocated nature (i.e., present in the periplasmic space), including that of the Y_Y domain [35]. Still, the region 260-284 is stronger in its hydrophobicity compared to the regions in the near adjacent vicinities within the periplasmic domain, perhaps hinting at a binding function for an inert gas such as nitrous oxide or the bulky reduction intermediates that may show a preference for hydrophobic residues. In this, three valines (locations 300-302) in the near structural vicinity (3D space) of the Y_Y_Y domain, stand out.



Figure 13. Calculated hydropathy profile of the WP_013192178.1 protein as a function of sequence.

3. Methods

3.1. Phylogenetic Reconstructions

The non-redundant downloaded amino acid sequences (as FASTA files) from each query were first aligned with the ClustalW algorithm using MEGA version X (default parameters) and phylogenetic reconstruction was performed using the neighborhood joining/maximum likelihood methods with support from 500 bootstrap replications. There was also the assignment of an archaeal outgroup [36].

3.2. Multiple Sequence Alignments

The non-redundant downloaded amino acid sequences (as FASTA files) were employed for sequence alignment using the ClustalW algorithm using the MEGA X and ClustalW servers available in the web such as Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). (Accessed on 22 January 2022).

3.3. Determination of Protein Families/Superfamilies

The PSI-BLAST search was also used to find suitable homologs. As stated in the structural classification of proteins (SCOP) database [37], 30% sequence identity was considered as the lower cutoff for protein families and 15% for protein superfamilies. Coverage was taken as 50%, which was an arbitrary value. Together, they formed the basis for the identification of true protein homologs belonging to a single family.

3.4. Protein Motif Analyses

Protein motif analyses were performed using the default parameters of the web servers, MOTIF Search (https://www.genome.jp/tools/motif/) (accessed 22 December 2021) and MEME (https://meme-suite.org/meme/tools/meme) (accessed 22 December 2021) [38].

3.5. Gene Synteny Analyses

Genomes of *Nostoc azollae* 0708 (NC_014248.1), *T. erythraeum IMS101*, *N. punctiforme PCC* 73102 (NC_010628.1) and *A. cylindrica* PCC 712 (NC_019771.1) were retrieved from the NCBI Nucleotide Database (https://www.ncbi.nlm.nih.gov/nuccore) (accessed 15 December 2021). Synteny analysis was carried out using BLAST (Ver. 2.2.26) E-value: 0.001 cutoff with minimum query coverage cutoff of 50% with gapped alignment settings using Simple Synteny Server ver. 1.3 (https://www.dveltri.com/simplesynteny/index. html) (accessed 15 December 2021).

3.6. Building of Homology Models

Homology models were built using the SWISS-MODEL (https://swissmodel.expasy. org/) (accessed 22 December 2021] server utilizing the default parameters of the search box for searching the PDB-stored structures of homologous proteins to be used as templates [39]. The highest homology structure was used as template.

3.7. Phosphorylation Site Prediction

The web address (http://www.cbs.dtu.dk/services/NetPhosBac/) (accessed 22 December 2021) hosting the Netphos 1.0 [40] was used for the identification of the putative phosphorylation sites.

3.8. Prediction of DNA Binding Residues

DNA binding residues were predicted using the DRNApred server [32].

3.9. Hydropathy Profile

The hydropathy profile was determined using the Kyte and Doolittle method via the ProtScale server (https://web.expasy.org/protscale/) (accessed 28 January 2022) [41].

4. Conclusions

While viruses such as coronaviruses mutate in days, weeks and months, antibioticresistant bacterial strains are known to have occurred in the last 60–80 years ever since the golden age (1940–1960) of antibiotics discovery, and there are evolutionary mal-adaptations that are stemming from polluters such as Tributylin (discovered in the 1950s) that are changing the reproductive biology/organs of the marine biota (such as mollusks). There is no assured way of deciding on a sensory reception module (Y_Y_Y motif) that may be gaining strength in the *T. azollae* cytochrome oxidase subunit II protein or to state whether the evolution is ongoing or terminal. Due to the proximity of the sequence motif to the conserved Cys-XXX-Cys sequence and taking into consideration the nature of the mutations as non-wobble-position mutations and the periplasmic presence of the Y_Y_Y motif, it is fair to propose that the emergence of this prototypical protein is likely to have perhaps occurred in the last 150 years from exposure to enhanced urea and other synthetic N-fertilizer usage.

Nitrogen containing compounds such as thalidomide (C13H10N2O4) have also had immediate footprints on the human health landscape, while even biogenic cyanotoxins,

being nature's arsenal of nitrogenous molecules, have resulted in animal and human mortalities. Nitric oxide is also a potent toxic molecule and a reactive nitrogen species, as well as being a signaling molecule that needs to be handled carefully when it can putatively harm organisms such as cyanobacteria when produced by nitric oxide synthases [42]. Inert nitrous oxide biogenesis is a solution to nitric oxide toxicity.

While urea may not have the same chemistry as toxins, we do not know the actual effect urea has on organismal physiology. Some plants contain ureases that are able to hydrolyze urea, similar to a spectrum of soil bacteria without which urea will tend to accumulate in the ambience and in plant parts/organs if internalized [43]. The genus *Oryza* contains urease enzymes but such catalysts appear to be lacking in the genus *Azolla*. Azolla is also poor in its internalization of nickel, which is a cofactor for ureases.

Receptive proteins, due to their presence in the periplasmic spaces, form interactions with environmental polluters and may be susceptible through exposure to developing a system to remediate the anthropogenically magnified molecule. This is using the same fundamental reasoning as the acquisition of antibiotic resistance that is acquired through lateral transfer events as plasmid vectors, as well as due to mutations (due to increased evolvability) as a means to resist an antibiotic in the immediate medium, primarily through SOS mutagenesis where the microbes produce error-prone DNA polymerases [44]. Furthermore, endosymbionts change faster than free-living organisms due to genome erosion, pseudogenization, and stronger substitutions that are common in prototypical symbionts that have not yet become a full-fledged plastid [13]. To conclude, it is stated by the authors here that the WP_013192178.1 protein from *T. azollae* is possibly emerging as a possible candidate for nitric/nitrous oxide sensing (via an emerging Y_Y_Y domain).

The evolution of enzymes has been shown to involve the conservation of the key residues of the catalytic site while allowing for the modification of residues in the substrate binding site [45]. In an enzyme superfamily such as Nudix hydrolases, the amino acid arrangement of the Nudix box active site and the overall protein fold is retained, while the substrate binding site is tinkered to allow for diverse substrate recognition [46].

Heterocysts are known for ecological specialization such as the thermophilic *Fischerella* spp., modifying the impermeability of the heterocyst envelopes to ensure anoxic conditions, agreeing with the plasticity-first hypothesis, which states that natural selection can customize ancestrally "plastic" features to fine-tune, to fit or replace with a locally optimized trait [47]. The order of events took place—as we hypothesize—mediated by nitrous oxide synthesized abiotically, paving way to biotic oxygen, and now, nature is evolving a specific cytochrome oxidase subunit II enzyme to possibly quell the urea-based nitrogen pollution by sensing nitric/nitrous oxide. Dollo's principle states that evolution discourages reinventing (exactly) an ancient trait from an eclipsed unidirectional fate, and reforming an enzymatic function is no easy feat. Many mutations are likely to be necessary to reimagine the same fold and the same superfamily with a reinvented substrate specificity—albeit one found abundantly in nature—to make sense as a member of a receptive function.

Rapid evolution is proposed to occur even inside the span of a century, just about the time we have used urea this far and is thought to have an imprint on ecological interactions that are already underway [48,49]. Rapid evolution can take place due to a plethora of anthropogenic measures, such as Peppered Moths and the Industrial Revolution, or due to natural means, the latter being synonymous with Darwin's Finches in the face of fluctuating rainfall [49]. We gently state that the likely heterocyst-functioning WP_013192178.1 is expedited in "rapid" evolution.

We suggest that the (1) emerging sensory Y_Y_Y domain, (2) the accumulation of four histidines at the C-terminal extremity, (3) a putative lengthy linker region rich in glycines including a likely unique DNA-binding glycine-183, (4) the NtcA-responsive promoter/binding site, (5) the secondary nature of cellular respiration in heterocysts to the Mehler activity, (6) the weak phylogeny to familial counterparts and (7) The promiscuity of ancient folds and structures are all indicative of the putative role of nitrous oxide sensing and effectuation of DNA-reliant downstream functions by an evolving cytochrome oxidase subunit II protein that may not be an optimal design as of the contemporary, but may be one of utility.

There is no solution to nitrous oxide in the atmosphere and we hope that the clues and theories presented in the paper will be a trigger for the characterization of the candidate protein's many activities, the resolution of the three-dimensional structure, gene knockout studies and precision gene-editing (CRISPR-Cas9) that may respectively validate function and augment the utility of this enzyme to the Azolla symbiosis and the climate change-impacted crucible of mankind.

To end, due to *T. azollae* being a strong nitrogen-fixing obligate cyanobiont, the addition of nitrous oxide conversion to dinitrogen gas could make this biological system one of invaluable utility to both climate change mitigation as well as crop nutrition. Determination of novel catalytic chemistries require "wet lab" biochemical experiments supported by structural biology studies.

Author Contributions: D.G. was responsible for the conceptualization, designed the *in silico* study, performed the bioinformatics-based explorations, analyzed the data and wrote the manuscript. V.H. performed the gene synteny analyses and produced the accompanying literature. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data pertaining to the publication will be made available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mvondo, D.N.; Navarro-González, R.; McKay, C.; Coll, P.; Raulin, F. Production of nitrogen oxides by lightning and coronae discharges in simulated early earth, venus and mars environments. *Adv. Space Res.* 2001, 27, 217–223. [CrossRef]
- BBC. 2021. Available online: https://www.bbc.com/future/article/20210603-nitrous-oxide-the-worlds-forgotten-greenhousegas (accessed on 28 January 2022).
- 3. Gunawardana, D. An Exploration of Common Greenhouse Gas Emissions by the Cyanobiont of the Azolla–Nostoc Symbiosis and Clues as to Nod Factors in Cyanobacteria. *Plants* **2019**, *8*, 587. [CrossRef]
- 4. Saraste, M.; Castresana, J. Cytochrome oxidase evolved by tinkering with denitrification enzymes. *FEBS Lett.* **1994**, *341*, 1–4. [CrossRef]
- Prosser, J.; Hink, L.; Gubry-Rangin, C.; Nicol, G.W. Nitrous oxide production by ammonia oxidisers: Physiological diversity, niche differentiation and potential mitigation strategies. *Glob. Change biol.* 2020, 26, 103–118. [CrossRef] [PubMed]
- 6. Weathers, P.J.; Niedzielski, J.J. Nitrous oxide production by cyanobacteria. Arch. Microbiol. 1986, 146, 204–206. [CrossRef]
- Jones, K.; Haselkorn, R. Newly Identified Cytochrome c Oxidase Operon in the Nitrogen-Fixing Cyanobacterium *Anabaena* sp. Strain PCC 7120 Specifically Induced in Heterocysts. *J. Bacteriol.* 2002, 184, 2491–2499. [CrossRef]
- Christiansen, J.; Seefeldt, L.; Dean, D.R. Competitive Substrate and Inhibitor Interactions at the Physiologically Relevant Active Site of Nitrogenase. J. Biol. Chem. 2000, 275, 36104–36107. [CrossRef]
- Pushpakumara, B.; Gunawardana, D. Preliminary data on the presence of an alternate vanadium nitrogenase in a culturable cyanobiont of Azolla pinnata R. Brown: Implications on Chronic Kidney Disease of an unknown etiology (CKDu). *Data Brief* 2018, 21, 2590–2597. [CrossRef]
- 10. Kimani, S.M.; Bimantara, P.O.; Hattori, S.; Tawaraya, K.; Sudo, S.; Xu, X.; Cheng, W. Co-application of poultry-litter biochar with Azolla has synergistic effects on CH₄ and N₂O emissions from rice paddy soils. *Heliyon* **2020**, *6*, e05042. [CrossRef]
- Malyan, S.K.; Bhatia, A.; Tomer, R.; Harit, R.C.; Jain, N.; Bhowmik, A.; Kaushik, R. Mitigation of yield-scaled greenhouse gas emissions from irrigated rice through Azolla, Blue-green algae, and plant growth–promoting bacteria. *Environ. Sci. Pollut. Res.* 2021, 28, 51425–51439. [CrossRef]
- 12. Pereira, A.L.; Vasconcelos, V. Classification and phylogeny of the cyanobiont Anabaena azollae Strasburger: An answered question? *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 1830–1840. [CrossRef] [PubMed]
- Ran, L.; Larsson, J.; Vigil-Stenman, T.; Nylander, J.A.A.; Ininbergs, K.; Zheng, W.-W.; Lapidus, A.; Lowry, S.; Haselkorn, R.; Bergman, B. Genome Erosion in a Nitrogen-Fixing Vertically Transmitted Endosymbiotic Multicellular Cyanobacterium. *PLoS ONE* 2010, *5*, e11486. [CrossRef]
- 14. Conte, L.L.; Ailey, B.; Hubbard, T.J.; Brenner, S.E.; Murzin, A.G.; Chothia, C. SCOP: A Structural Classification of Proteins database. *Nucleic Acids Res.* 2000, *28*, 257–259. [CrossRef]

- 15. Einarsdóttir, O.; Caughey, W.S. Interactions of the anesthetic nitrous oxide with bovine heart cytochrome c oxidase. Effects on protein structure, oxidase activity, and other properties. *J. Biol. Chem.* **1988**, *263*, 9199–9205. [CrossRef]
- Taylor, C.; Moncada, S. Nitric Oxide, Cytochrome C Oxidase, and the Cellular Response to Hypoxia. *Arter. Thromb. Vasc. Biol.* 2010, 30, 643–647. [CrossRef] [PubMed]
- Guttierez, E.; Long, P.F. Abstract. Available online: https://aaas.confex.com/aaas/2012/webprogram/Paper8031.html (accessed on 28 January 2022).
- 18. Menke, M.; Berger, B.; Cowen, L. Markov random fields reveal an N-terminal double beta-propeller motif as part of a bacterial hybrid two-component sensor system. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4069–4074. [CrossRef]
- 19. Murphy, M.E.P.; Lindley, P.F.; Adman, E.T. Structural comparison of cupredoxin domains: Domain recycling to construct proteins with novel functions. *Protein Sci.* **1997**, *6*, 761–770. [CrossRef]
- Chen, C.K.-M.; Chan, N.-L.; Wang, A.H.-J. The many blades of the β-propeller proteins: Conserved but versatile. *Trends Biochem. Sci.* 2011, *36*, 553–561. [CrossRef]
- Veltri, D.; Malapi-Wight, M.; Crouch, J.A. SimpleSynteny: A web-based tool for visualization of microsynteny across multiple species. *Nucleic Acids Res.* 2016, 44, W41–W45. [CrossRef]
- Gunawardana, D. An in silico Study of Two Transcription Factors Controlling Diazotrophic Fates of the Azolla Major Cyanobiont Trichormus azollae. *Bioinform. Biol. Insights* 2020, 14, 1177932220977490. [CrossRef]
- Vázquez-Bermúdez, M.F.; Flores, E.; Herrero, A. Analysis of binding sites for the nitrogen-control transcription factor NtcA in the promoters of Synechococcus nitrogen-regulated genes. *Biochim. Biophys. Acta-Gene Struct. Expr.* 2002, 1578, 95–98. [CrossRef]
- 24. Milligan, A.J.; Berman-Frank, I.; Gerchman, Y.; Dismukes, G.C.; Falkowski, P.G. Light-dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in nitrogenase protection1. *J. Phycol.* **2007**, *43*, 845–852. [CrossRef]
- Torrado, A.; Ramírez-Moncayo, C.; Navarro, J.A.; Mariscal, V.; Molina-Heredia, F.P. Cytochrome c6 is the main respiratory and photosynthetic soluble electron donor in heterocysts of the cyanobacterium Anabaena sp. PCC 7120. *Biochim. Biophys. Acta* 2018, 1860, 60–68. [CrossRef] [PubMed]
- Schilling, K.M.; Tao, L.; Wu, B.; Kiblen, J.T.; Ubilla-Rodriguez, N.C.; Pushie, M.J.; Britt, R.D.; Roseman, G.P.; Harris, D.A.; Millhauser, G.L. Both N-Terminal and C-Terminal Histidine Residues of the Prion Protein Are Essential for Copper Coordination and Neuroprotective Self-Regulation. J. Mol. Biol. 2020, 432, 4408–4425. [CrossRef]
- Hoeren, F.U.; Berks, B.C.; Ferguson, S.J.; McCarthy, J.E.G. Sequence and expression of the gene encoding the respiratory nitrousoxide reductase from Paracoccus denitrificans. New and conserved structural and regulatory motifs. *Eur. J. Biochem.* 1993, 218, 49–57. [CrossRef] [PubMed]
- 28. Cutruzzolà, F. Bacterial nitric oxide synthesis. Biochim. Biophys. Acta Bioenerg. 1999, 1411, 231-249. [CrossRef]
- 29. Timilsina, A.; Zhang, C.; Pandey, B.; Bizimana, F.; Dong, W.; Hu, C. Potential Pathway of Nitrous Oxide Formation in Plants. *Front. Plant Sci.* **2020**, *11*, 1177. [CrossRef]
- Rani, S.; Dasgupta, B.; Bhati, G.K.; Tomar, K.; Rakshit, S.; Maiti, S. Superior Proton-Transfer Catalytic Promiscuity of Cytochrome c in Self-Organized Media. *ChemBioChem* 2020, 22, 1285–1291. [CrossRef]
- 31. Tompa, P.; Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **2004**, *18*, 1169–1175. [CrossRef]
- 32. Yan, J.; Kurgan, L. DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. *Nucleic Acids Res.* 2017, 45, e84. [CrossRef]
- Dongre, M.; Singh, N.S.; Dureja, C.; Peddada, N.; Solanki, A.K.; Ashish, D.; Raychaudhuri, S. Evidence on How a Conserved Glycine in the Hinge Region of HapR Regulates Its DNA Binding Ability. *J. Biol. Chem.* 2011, 286, 15043–15049. [CrossRef] [PubMed]
- 34. Reddy Chichili, V.P.; Kumar, V.; Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci.* 2013, 22, 153–167. [CrossRef] [PubMed]
- Baker, J.A.; Wong, W.-C.; Eisenhaber, B.; Warwicker, J.; Eisenhaber, F. Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule". BMC Biol. 2017, 15, 66. [CrossRef]
- Sudhir, K.; Glen, S.; Michael, L.; Christina, K.; Koichiro, T. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 2018, 35, 1547–1549.
- 37. Andreeva, A.; Kulesha, E.; Gough, J.; Murzin, A.G. The SCOP database in 2020: Expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res.* 2019, *48*, D376–D382. [CrossRef] [PubMed]
- Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology; AAAI Press: Menlo Park, CA, USA, 1994; pp. 28–36.
- Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; De Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018, 46, W296–W303. [CrossRef]
- 40. Miller, M.L.; Soufi, B.; Jers, C.; Blom, N.; Macek, B.; Mijakovic, I. NetPhosBac-A predictor for Ser/Thr phosphorylation sites in bacterial proteins. *Proteomics* 2009, *9*, 116–125. [CrossRef]

- 41. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.
- 42. Xue, L.-G.; Li, S.W.; Xu, S.-J.; An, L.-Z.; Wang, X.-L. Alleviative effects of nitric oxide on the biological damage of spirulina platensis induced by enhanced ultraviolet-B. *Acta Microbiol. Sin.* **2006**, *46*, 561–564.
- 43. Polacco, J.C.; Mazzafera, P.; Tezotto, T. Opinion–Nickel and urease in plants: Still many knowledge gaps. *Plant Sci.* 2013, 199-200, 79–90. [CrossRef]
- 44. Revitt-Mills, S.A.; Robinson, A. Antibiotic-Induced Mutagenesis: Under the Microscope. *Front. Microbiol.* 2020, *11*, 2611. [CrossRef]
- 45. Petsko, G.A.; Kenyon, G.L.; Gerlt, J.A.; Ringe, D.; Kozarich, J.W. On the origin of enzymatic species. *Trends Biochem. Sci.* **1993**, *18*, 372–376. [CrossRef]
- 46. Gunawardana, D.; Cheng, H.-C.; Gayler, K.R. Identification of functional domains in Arabidopsis thaliana mRNA decapping enzyme (AtDcp2). *Nucleic Acids Res.* 2007, *36*, 203–216. [CrossRef] [PubMed]
- Miller, S.R.; Longley, R.; Hutchins, P.R.; Bauersachs, T. Cellular Innovation of the Cyanobacterial Heterocyst by the Adaptive Loss of Plasticity. *Curr. Biol.* 2020, 30, 344–350. [CrossRef] [PubMed]
- 48. Thompson, J.N. Rapid evolution as an ecological process. Trends Ecol. Evol. 1998, 13, 329–332. [CrossRef]
- 49. Hairston, N.G.; Ellner, S.P.; Geber, M.A.; Yoshida, T.; Fox, J.A. Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* **2005**, *8*, 1114–1127. [CrossRef]