Investigating Nitrogen Fixation and Metabolic Pathways in the Plant Endophyte *Gluconacetobacter diazotrophicus*

A THESIS

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA BY

Erin Marie Schwister

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Dr. Brett M. Barney

July 2021

Acknowledgments

As I wrap up this thesis and my time in this program, it has never been more apparent that I would not have made it this far without the support of so many people in my life. I would first like to express my deepest gratitude to my advisor, Dr. Brett Barney, for his support, encouragement, and mentorship throughout the past two years. I also thank the members of my committee, Dr. Bo Hu and Dr. Neil Olszewski, for their guidance. I would like to further thank all the members of the laboratory for their advice and support during my time in lab. I thank especially Carol for her partnership and advice regarding coding and help with the Tn-seq study, Ben for his work on the genetic constructs and transposon library, and Natalia for her encouragement. To Mr. Leventhal, thank you for allowing me to see myself as a scientist. I would also like to thank my family and friends for being a wonderful and loving support system, for making me laugh, and allowing me to be my best self. To my parents and Katie, you are the best support system and have been so encouraging of every aspect of my life and I will never be able to thank you enough. And to everyone else who has impacted my life and my academic career: thank you.

Abstract

Reduction of the reliance upon synthetic fertilizers is critical for a more sustainable agricultural future. Plant endophytes, especially those that engage in biological nitrogen fixation, provide a potential route towards this end. In this work, a deeper understanding of the plant endophyte and nitrogen fixing bacterium *Gluconacetobacter diazotrophicus* is presented through the following four chapters. The first chapter provides contextual background on this important plant endophyte and its potential for use in agriculture. Following this introduction, a large-scale transposon library of *G. diazotrophicus* is explored in the second and third chapters through high-throughput sequencing in a transposon insertion sequencing (Tn-seq) study. This Tn-seq study explores gene essentiality relating to nitrogen fixation as well as nutrient metabolism in *G. diazotrophicus* through rapid characterization across the entire genome. Lastly, the fourth chapter describes genetic manipulations of several key genes in *G. diazotrophicus*, performed to improve extracellular ammonium production to construct a strain for potential use as a biofertilizer. This work provides a deeper understanding of the genetic mechanisms essential to nitrogen fixation in *G. diazotrophicus* as well as those essential for growth under varied environmental conditions. As a nitrogen-fixing plant endophyte, further exploration of *G. diazotrophicus* as a potential alternative to synthetic fertilization brings us closer to a more sustainable agricultural future.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Plant Endophyte *Gluconacetobacter diazotrophicus*

Plant endophytes

Plants have microbiomes consisting of those microbes that engage through association or interaction. An important aspect to the plant microbiome is endophytic bacteria. The definition of an endophyte has changed numerous times (1). Previously defined as non-pathogenic organisms that live within a plant for any part of their lifecycle (2), a more recent definition of plant endophyte specified any microbe that colonizes the interior of a plant, regardless of their relationship with the host (3). In general, endophytes are microbes that colonize plant tissues. Microbial host colonization can be localized to the point of entry, such as specific to the roots or leaves, or can spread systemically throughout the plant (4).

Regardless of their positioning within their host plant, plant endophytes can provide growth benefits to their hosts through features such plant hormone production or increased pathogen resistance. These beneficial or plant growth-promoting (PGP) characteristics are wide-ranging, dependent upon both the host and endophyte. Plant endophytes have PGP characteristics such as the ability to prime their host plant immune response, allowing for the plant to mount an accelerated defense response upon attack of a predatory microbe (5), protect the host from abiotic stresses (5), synthesize plant hormones (1), or even allow for increased host plant biomass, growth or yield.

Nitrogen fixation as a PGP

One key PGP characteristic is the ability of microbes to perform biological nitrogen fixation (BNF). BNF is the process of the conversion of nitrogen gas into a fixed form of nitrogen that host plants can use and is carried out through the activity of the enzyme nitrogenase (Equation 1).

$$N_2 + 8H^+ + 8e^- + 16\,ATP \rightarrow 2NH_3 + H_2 + 16\,ADP + 16\,P_i \quad (1)$$

Along with other essential nutrients for plant growth, such as carbon, oxygen and phosphorus, plants also require nitrogen. Nitrogen is a major component of chlorophyll as well as amino acids and nucleic acids, and as such is essential for sustaining life (6). However, while nitrogen is abundant in the atmosphere, it often remains extremely limiting in agricultural growth as plants are unable to use atmospheric nitrogen.

Therefore, reduced forms of nitrogen for growth, such as ammonium (NH_4^+) or ammonia (NH_3), must be provided to the plant.

A large-scale method of producing reduced plant-accessible nitrogen that does not use BNF is the Haber-Bosch process. Developed at the beginning of the 20th century by Fritz Haber and Carl Bosch (6), this process allows for the conversion of nitrogen gas into ammonium (Equation 2).

$$2N_2 + 3H_2 \rightarrow 2NH_3 \quad (2)$$

While this process was originally intended to be used to develop ammunition and fuel conflict, it also has allowed for drastic increase in crop yields through use of synthetic ammonium-containing fertilizers (7). The production of these agricultural fertilizers assisted the expansive population growth seen in the 20th century (7), as greater nitrogen availability resulted in increased agricultural productivity. Unfortunately, the addition of Haber-Bosch derived nitrogen to the soil has had unintended environmental consequences, including the contamination of groundwater and the eutrophication of waterways (6). Further, this process is energy-intensive (8). Given these issues, it is desirable to seek a more environmentally friendly method of providing plants with nitrogen, and BNF presents a well-understood solution.

Nitrogen-fixing bacteria, also known as diazotrophs, are diverse. These microbes can be free-living in the soil or engage in associative interactions with plants, and cover a wide range of phyla (9). While agricultural symbiotic nitrogen fixation is perhaps best known to occur in the root nodules of legumes, some diazotrophs do engage in associative or endophytic relationships with other nonlegumes (9). When found in association with plants, these diazotrophs are PGP, as aforementioned, since providing fixed nitrogen to the plant is a clear growth benefit.

Introduction to *Gluconacetobacter diazotrophicus* PA1 5

The bacterium *Gluconacetobacter diazotrophicus* PA1 5 is a plant endophytic diazotroph. Previously classified as *Acetobacter diazotrophicus* (10), *G. diazotrophicus* was such reclassified based on 16S rRNA analysis (11). A member of the Alphaproteobacteria class, it was originally isolated in Brazil from sugarcane roots and stems (12). Since this initial discovery, this microaerobic bacterium has been found as a

native host of numerous (13) other plant species, including agriculturally important crops such as sweet potato (14), pineapple (15), and coffee (16). In addition, *G. diazotrophicus* has been successfully inoculated into other plant species of which it is not known to be a native host (13), including corn (17, 18) and *Arabidopsis thaliana* (17, 19). Introduction of this bacterium into new plant species has the potential to further the understanding of endophytes, as well as explore the effects of its numerous PGP characteristics upon its host.

Colonization of *G. diazotrophicus* in numerous host plants has indicated growth and productivity benefits. In its native host sugarcane, this bacterium provides growth stimulation (20) and may be responsible for providing over half of the necessary nitrogen for growth without additional fertilizer application (21, 22). Beyond sugarcane, *G. diazotrophicus* colonization has been shown to stimulate fruit production in tomato plants (23), alleviate drought stress in red rice (24), benefit growth of strawberry plants (25), and even carry out an antagonistic role against plant pathogens (19, 26). Its ability to serve as a beneficial plant endophyte is due to its host of PGP abilities. Most notably, of course, is its ability to perform BNF. Differing from many other diazotrophs that associate with plants, *G. diazotrophicus* performs BNF without forming nodules (12, 27, 28). However, *G. diazotrophicus* has further PGP characteristics beyond BNF. It produces phytohormones critical for plant growth, including the auxin indole-3-acetic acid and several gibberellins (29, 30). The bacterium additionally produces hydroxamate type siderophores (31) and has other macro- and micronutrient solubilization capabilities (32). All together, these characteristics and more exemplify the numerous plant growth benefits of this endophytic bacterium.

G. diazotrophicus: Biological nitrogen fixation

The perhaps best-known feature of *G. diazotrophicus*—its ability to carry out BNF—has been studied in much greater detail. As in all diazotrophs, the nitrogenase enzyme complex carries out the biological nitrogen fixation process (Equation 2). The molybdenum nitrogenase that is found in all diazotrophs (33, 34) is comprised of two proteins: the molybdenum-iron (MoFe) protein, and the iron (Fe) protein, its reductase (35–37). The genes *nifDK* code for the $\alpha_2\beta_2$ heterodimer MoFe protein, while *nifH* codes

for the homodimer Fe protein (36). These gene products are assembled by other *nif* genes in a complex system (36) (Figure 1). In brief, a conglomerate of the nitrogen fixation (*nif*) genes are found to be involved in the biosynthesis and assembly of nitrogenase. Critical to nitrogenase assembly are the genes *nifENB*. The *nifB* product catalyzes the first step in the synthesis of FeMo-co, and *nifEN* codes for the scaffold protein for the final synthesis of FeMo-co. Other important components in nitrogenase assembly include iron-sulfur clusters and electron delivery. While six genes (*nifHDK* and *nifENB*) were proposed to be required to predict the occurrence of nitrogen fixation in microbial genomes (38), in practice far more genes are required. Prior introduction of the nitrogen fixation genes into bacterial species that do not natively fix nitrogen has required more than a dozen genes (39, 40). Within *G. diazotrophicus*, these gene products and several other *nif* genes are centered in one genetic region termed the major *nif* cluster. The major *nif* cluster ranges from the regulatory *rpoN* to the molybdate transport-associated *modD* gene (41) and contains the bulk of the genes associated with BNF in *G. diazotrophicus*.

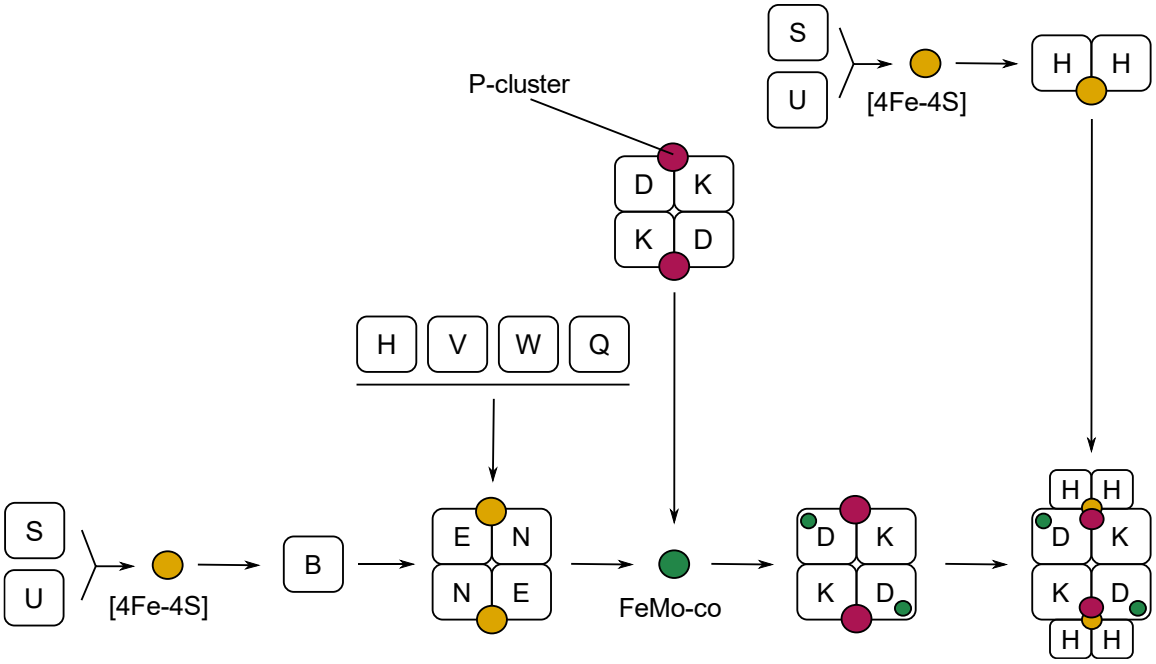

Figure 1. General nitrogenase assembly. Nitrogenase (nifHDK complex, bottom right) is assembled by the cohesion of the Fe protein (nifH) and the MoFe protein (nifDK). *nif* genes are represented by boxes with letters. *nif* gene complexes are those boxes shown together, and assembly cofactors 4Fe-4S, the P-cluster and FeMo are indicated by colored circles.

The expression of *nif* genes for Proteobacteria diazotrophs are controlled by the σ^{54}-dependent master transcriptional activator NifA, encoded by *nifA* (42). In many diazotrophs, including the gamma subgroup and beyond (43), the gene products of the *nifLA* operon regulate the transcription of *nif* genes encoding nitrogenase, *nifHDK* (44). In this system, NifA acts as the σ^{54}-dependent activator of nitrogenase expression and its partner protein, NifL, is the anti-activator and prevents NifA activity in response to nitrogen- or oxygen-rich environments (43). However, most diazotrophs within the alpha and beta subgroups lack a *nifL* homolog, and their NifA proteins are thought to be inherently sensitive to oxygen and level of fixed nitrogen (42, 45). *G. diazotrophicus* lacks a *nifL* homolog and its *nifA* is thought to be inherently sensitive to oxygen (46). While the expression of *nifA* here is not influenced by oxygen concentration (47), the activity of the NifA protein encoded by *nifA* is indeed oxygen-sensitive. This gene is also located within the major *nif* cluster in *G. diazotrophicus*.

The NifA protein contains three primary domains: a GAF domain at the N-terminus, an AAA+ domain central to the protein, and a helix-turn-helix (HTH) domain at the C-terminus (42, 48). The AAA+, or ATPase domain, contains the binding interface and itself activates transcription, and is highly conserved (48). The GAF domain has regulatory function in different diazotrophs, but in many species its function is poorly understood (42). Connecting the GAF and AAA+ domains is an interdomain region termed the Q-linker. So called due to the high content of glutamine amino acid residues (49), the Q-linker region has been associated with oxygen-sensitive NifA proteins (47).

Due to the highly energetically expensive process of BNF, requiring 16 ATP to produce $2NH_3$ (Equation 1), diazotrophic bacteria have a careful regulation process of both nitrogen fixation and nitrogen assimilation. This is done to ensure that ammonium is readily taken up by biomass, rather than wastefully excreted (44). Several prior reviews (42, 44) have tackled this topic. In general, this regulatory process involves components such as P_{II} signal transduction proteins and those interacting with the transcriptional regulator *nifA* (44). Prior gene expression analysis indicated that *nifA* expression and by extension the nitrogen-sensing system of *G. diazotrophicus* is in fact regulated in

response to nitrogen availability within the cell (47). While further work indicated that the nitrogenase in *G. diazotrophicus* is still active in a reduced capacity in the presence of ammonium (13), this regulatory system can be thought to be in place to ensure that BNF is only carried out when there is a demand for fixed nitrogen.

Therefore, under nitrogen-abundant and nitrogen-limiting conditions, the activity of these regulatory proteins are tightly controlled (Figure 2). Generally, the cell uptakes nitrogen as ammonium via diffusion across the membrane or through ammonium transporters, encoded by *amtB1* and *amtB2* in *G. diazotrophicus*. Nitrogen within the cell is largely sensed by the trimeric P_{II} signal transduction proteins. In *G. diazotrophicus*, these are encoded by *glnK1*, *glnK2* and *glnB* and are distributed throughout the genome. Under nitrogen-abundant conditions, or when there is sufficient fixed nitrogen present for cellular processes to be carried out, glutamine binds to the uridylyltransferase/uridylyl-removing enzyme encoded by *glnD* (42, 50), and the P_{II} proteins are subjected to modification. Under these conditions, the P_{II} proteins are deurydylated and attach to the ammonium transporters, blocking ammonium uptake through the transporters. Under nitrogen-limiting conditions, the opposite occurs to meet the demand for fixed nitrogen.

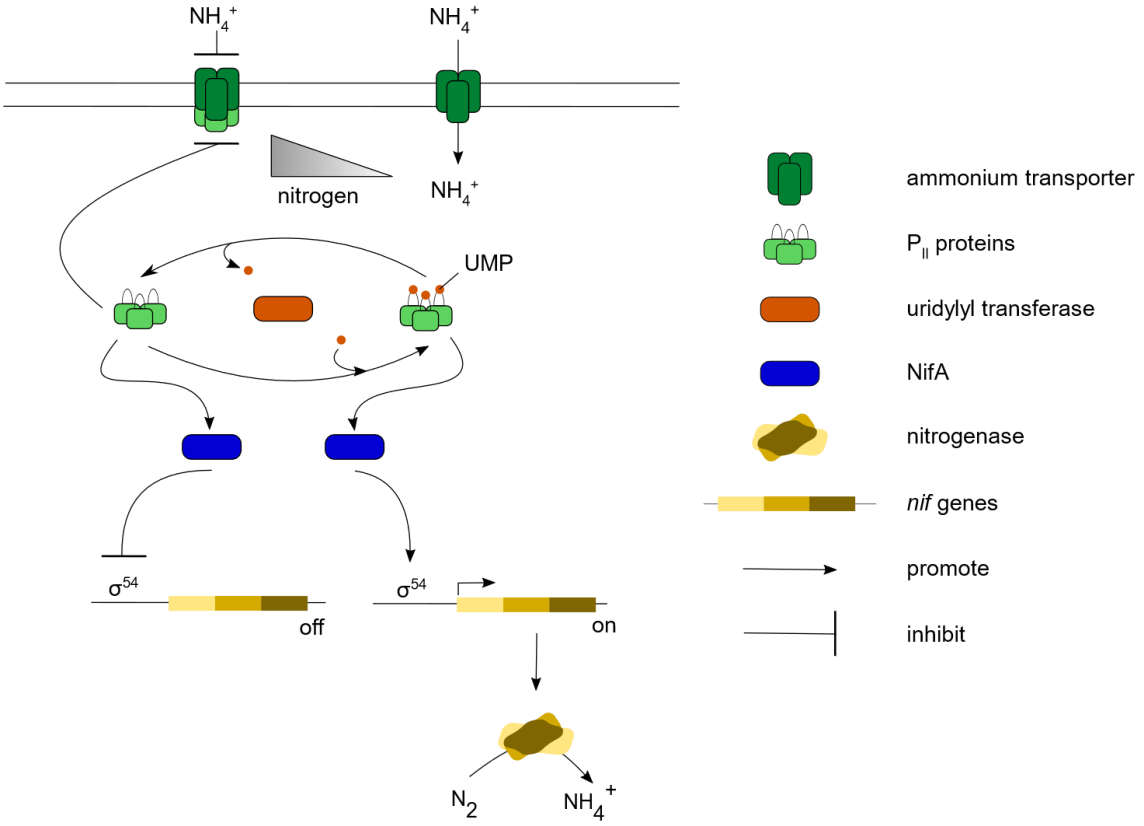

Figure 2. Regulation of nitrogen fixation in *G. diazotrophicus*. Under nitrogen-rich conditions, P_{II} proteins interact with the ammonium transporters to inhibit ammonium uptake into the cell. Under nitrogen-poor conditions, modified P_{II} proteins (uridylylated with UMP by action of the uridylyl transferase) do not interact with ammonium transporters and ammonium enters the cell. P_{II} proteins also directly interact with NifA, the master regulator of *nif* genes, to inhibit NifA activity under nitrogen-rich conditions to prevent BNF. NifA transcribes *nif* genes under nitrogen-poor conditions. Ammonium also diffuses across the membrane freely under sufficient conditions (not shown).

In other diazotrophs, the NtrBC system is activated by deurydylated P_{II} proteins and is essential for inactivating transcription of genes important to nitrogen metabolism (44). While the *ntrBCXY* operon is present in *G. diazotrophicus* and could be in operation here, prior work has indicated that the *ntrBC* proteins are not required for BNF in *G. diazotrophicus* (51). It has instead been suggested that *nifA* is targeted for inhibition by one or more non-modified P_{II} proteins when nitrogen is abundant (52). In a prior study, the *glnK2* gene product was required for complete *nif* repression, while *glnK1* and *glnB* relieved repression by *glnK2* both directly and indirectly (52).

To ensure that fixed nitrogen is quickly transported to biomass for use, any ammonium present in the cell is also heavily regulated. The ammonium that is fixed through BNF or otherwise taken up by the cell is assimilated by the glutamine synthetase/glutamate synthetase pathway (GS-GOGAT) (28, 41, 53). This has likewise been comprehensively reviewed (54). The activity of glutamine synthetase (*glnA*), which has previously been studied in *G. diazotrophicus* (51), is regulated by an adenyltransferase (*glnE*). A genomic analysis of *G. diazotrophicus* suggested that in addition to the GS-GOGAT pathway, alternative routes of ammonia incorporation could exist (46).

Importance of *G. diazotrophicus* with crops

As previously discussed, plants require bioavailable nitrogen for successful growth, and some of this essential nutrient can be provided by microbes through biofertilizers. Though about half of the total nitrogen used in agriculture is produced through the Haber-Bosch process (8), biofertilizers have the potential to reduce our global reliance upon synthetic fertilizer. Biofertilizers are microbial inoculants typically consisting of diazotrophs and other nutrient-providing strains. Potential biofertilizers have been explored previously (55), and crop seeds have long been inoculated with bacterial cultures to improve crop growth. Since the 1950s, many bacterial strains with biofertilization capabilities have been described and characterized through both research on individual strains and more recently metagenomics (56). *G. diazotrophicus* is a strain of particular interest for future biofertilizer use due to its unique nitrogen-providing characteristics. A prior study demonstrated its ability to secrete fixed nitrogen into growth medium and support greater than 40% of the nitrogen needs of an amylolytic yeast (57). As *G. diazotrophicus* is found in great numbers inside of sugarcane plant tissue (12), this bacterium may be capable of providing significant amounts of nitrogen as an endophyte (58). Use of this strain may reduce reliance upon synthetic fertilization as a potential biofertilizer for many crops, especially given the wide range of native and non-native host plants of *G. diazotrophicus*.

Significance of this work

In this work, a further understanding of the plant endophyte *G. diazotrophicus* is presented. Through transposon insertion sequencing (Tn-seq), a large-scale transposon library is first investigated through high-throughput sequencing to establish a baseline of genes found to be conditionally essential for growth under a variety of conditions, including under nitrogen fixation. A Tn-seq study allows for rapid gene characterization across an entire genome, and studies of this nature exploring *G. diazotrophicus* have not yet been reported. This work will inform future studies in *G. diazotrophicus* as well as in other plant endophytes through providing a snapshot of gene essentiality under interesting environmental conditions. Beyond Tn-seq, improved genetic techniques in *G. diazotrophicus* are additionally presented in this work. Genetic manipulations of several key genes related to nitrogen fixation and assimilation were performed with the intention to build a strain of *G. diazotrophicus* with improved fixed nitrogen excretion. This was done to construct a strain for potential use as a biofertilizer or otherwise further enhance plant growth. These genetic manipulations additionally allow for a further understanding of nitrogen fixation and regulation in *G. diazotrophicus*.

Reducing our reliance upon synthetic fertilizers is critical, given their origin in fossil fuels and potential for environmental contamination. One potential route to a more sustainable agricultural future is through BNF. As a nitrogen-fixing plant endophyte with many intriguing characteristics, further exploration of *G. diazotrophicus* is important. There is much still to learn from this microbe, from details surrounding its metabolism to the genes important to its successful endophytic colonization. Furthering our understanding of this agriculturally relevant microbe, through such methods as Tn-seq or other genetic manipulations as presented in this thesis, puts us one step closer to a more sustainable future.

Chapter 2: Exploring Fitness of Nitrogen Fixation Genes in Plant Endophyte *Gluconacetobacter diazotrophicus*

Introduction

The bacterium *G. diazotrophicus* is a model organism for the study of plant endophytes. As discussed above (Chapter 1), *G. diazotrophicus* was originally isolated from sugarcane (12). The bacterium serves as a beneficial plant endophyte due to its host of plant growth-promoting (PGP) characteristics. Chiefly, *G. diazotrophicus* carries out the process of biological nitrogen fixation (BNF), differing from many other diazotrophs that associate with root nodules. Given its status as a PGP bacterium, it should come as no surprise that the bacterium has also been isolated as a native host in plant species beyond sugarcane, including sweet potato (14) and coffee (16). *G. diazotrophicus* has additionally been found to colonize other plant species of which it is not known to be a native host (13).

Due to its status as a plant endophyte, some prior global analysis studies have been carried out on *G. diazotrophicus*. Following the complete genomic sequencing of *G. diazotrophicus* (46, 59), subsequent large-scale studies have addressed the regulation of genes and proteins. Quantitative proteomic analysis studies of *G. diazotrophicus* have examined growth phases and response of the bacterium to nitrogen levels (60) as well as while in interaction with a host plant (61). Further large-scale proteomic and genomic studies have explored proteins and genes involved in the association of the bacterium and its host plant (24, 62–64). Few other global analysis studies have been performed. To our knowledge, studies exploring *G. diazotrophicus* gene fitness through the tools provided by transposon insertion sequencing (Tn-seq) have not yet been reported.

Tn-seq is a variation of the transposon insertion sequencing methods introduced in the last decade (Figure 3). These studies aim to combine a large-scale library of transposon mutants with next-generation sequencing in order to examine the essentiality or fitness level of genes within a bacterial genome (65, 66). The construction of a large mutant library occurs first, through the exclusive use of the *Mariner* transposon (65, 67). Using this transposon, the insertion occurs only once at a random thymine-adenine dinucleotide (TA) site within the genome. This library size is constructed to ensure each

mutant contains a different transposon insertion within the genome. After subjecting the mutant library to test environmental conditions, the library is sequenced, and researchers then can calculate the fitness of each insertion site within the genome. Together, this method probes genes essential to growth under a range of conditions. However, Tn-seq studies are limited by potential insertional sites within the genome and coverage provided by the mutant library.

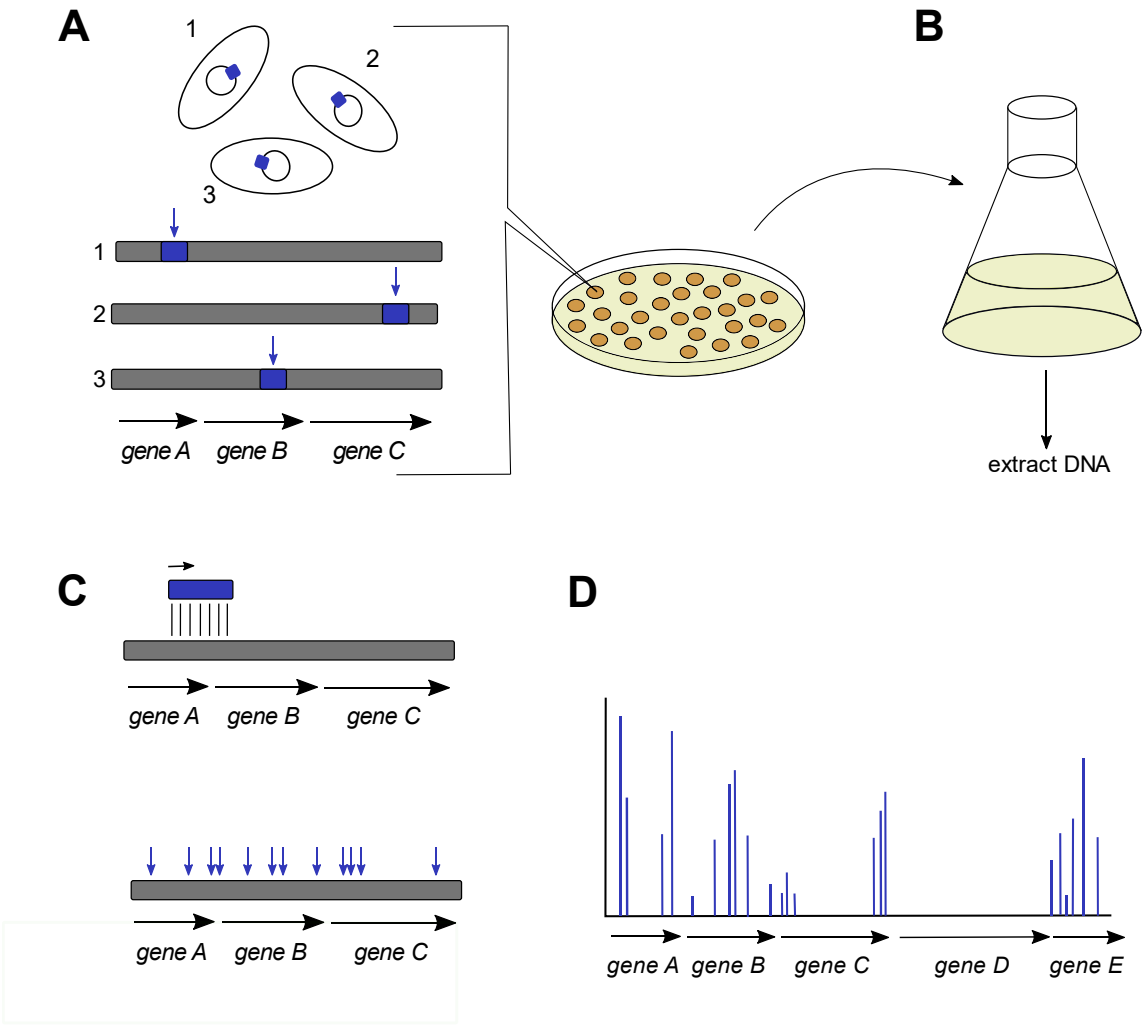

Figure 3. Tn-seq process schematic. **A**) Construction of transposon mutant library. **B**) Growth of transposon mutant library under select media conditions. Extraction of genomic DNA of the library in each condition for sequencing. **C**) Sequencing of library and mapping of TA insert locations within the genome. **D**) Calculation of frequency of transposon insertion in different TA insert locations.

Tn-seq allows for rapid characterization of gene function across an entire genome. Traditional methods of characterizing genes can be time-intensive and laborious, leading to a restricted understanding of the function of genes within an organism (68). Tn-seq and other high-throughput analyses allow for widescale understanding of complex microbial mechanisms in a short period of time. In agriculture, Tn-seq has the potential to determine important components of plant-microbial interactions, especially those that can be used to increase crop yields (69). Several diazotrophs, both free-living and endophytes, have been previously characterized using Tn-seq (70, 71). Exploring the genome of *G. diazotrophicus* through Tn-seq provides rapid insight into genetic interactions and pathways of this agriculturally relevant bacterium.

In this study, we conducted a Tn-seq study in *G. diazotrophicus* to examine gene fitness in those genes within this species that were relevant to BNF. Like all global analysis studies of microbes, Tn-seq has its limitations. However, this method provides insight to the broader research community into the untapped knowledge pertaining to the genes within this plant endophyte. As few large genomic data studies have been performed on this organism, this study can be used to inform future work with *G. diazotrophicus* and nitrogen-fixing endophytes. Here, we present an analysis of 3,000+ genes and their related fitness values when grown under nitrogen-sufficient and nitrogen fixation conditions.

Methods

Bacterial strains and growth conditions

Gluconacetobacter diazotrophicus PA1 5 (ATCC 49037) was obtained from Cedric Owens and was grown aerobically at 30ºC on GAD basal medium and modified B medium unless otherwise specified. GAD medium was adapted from DYGS basal medium (35) (g L^{-1}: glucose, 2; yeast extract, 2; tryptone, 1.5; $MgSO_4$, 0.5; glutamic acid, 1.5; adjusted to pH 6.2 with NaOH before sterilization). The modified B medium was adapted from Burk's (B) medium (72) (50 mM citrate, g L^{-1}: sucrose, 20; $MgSO_4$, 0.2; $CaCl_2$, 0.09; Na_2MoO_4, 0.025; $FeSO_4$, 0.05g; 2 mL 100X phosphate buffer composed of 20 g KH_2PO_4 and 80 g K_2HPO_4 in 1 L dH2O; adjusted to pH 6.2 with NaOH before sterilization). The modified B medium was supplemented with 10 mM $NH_4(SO_4)_2$ for

growth with provided nitrogen. For growth under nitrogen fixation conditions, the medium was not supplemented with a nitrogen source. All media was additionally supplemented with 100 mg/L tetracycline unless otherwise noted.

Escherichia coli WM3064 was used for conjugation with the plasmid pBB298, which contained the Mariner transposon, and was grown on lysogeny broth (LB) medium (67, 73).

Genetic constructs

The plasmid pBB298 containing the *Mariner* transposon was constructed prior to Tn-seq library construction. All relevant plasmids (Table 1) and primers (Table 2) are described regarding its construction.

Library construction

A transposon library was created through the transformation of *G. diazotrophicus* with *E. coli* WM3064 containing plasmid pBB298. Briefly, ~50 uL *G. diazotrophicus* cells were removed from a GAD plate and suspended in 0.5 mL LB medium. Similarly, ~50 uL *E. coli* WM3064 cells containing pBB298 were removed from a LB plate supplemented with 0.338 M tetracycline (100 uL of a 15 mg/mL stock) and 100 uM 2,6-di-aminopimelic acid (DAP; 50 uL of a 10 mg/mL stock) and suspended separately in 0.5 mL LB medium. From these stocks, 5 uL *E. coli* WM3064 and 100 uL *G. diazotrophicus* were combined and mixed with a pipettor before being spotted onto GAD plates supplemented with 100 uM DAP. These plates were incubated overnight at 30C. Each dot on the plate was then transferred into 50 mL GAD medium in a 125 mL flask and incubated overnight at 30ºC on a shaker table at 180 rpm. Following growth in GAD medium, 1 mL of each culture was removed and pelleted before removal of 900 uL supernatant. The cells were resuspended in the remaining media and plated onto GAD medium supplemented with 2.25 M tetracycline (100 uL of a 100 mg/mL stock). Plates were then incubated at 30ºC for several days until colonies formed. After sufficient colony growth, the cells were removed and flash frozen in liquid nitrogen before storage at -80ºC.

This process was repeated until a library of $\geq$ 100,000 mutant colonies was created. The final iteration of $\geq$ 18,000 mutants was not frozen, but instead inoculated

directly into 250 mL GAD medium as indicated below to account for variation among fitness of genes associated with freezing the cells.

Library growth conditions

The library was first inoculated into one flask containing 250 mL GAD medium. The library was inoculated both directly from GAD agar plates and from previously frozen libraries. Frozen libraries were thawed at room temperature before inoculation in GAD medium. The entire library was combined and mixed before growth in GAD medium at 28ºC with 160 rpm shaking under air, due to multiple aliquots of the library from various frozen stocks.

Cells from this initial growth of the library were collected through centrifugation at 15,000 rpm for 90 seconds and the subsequent removal of supernatant. The cell pellets were then inoculated in duplicate into 250 mL modified B media. The original optical densities (OD_{600}) of the media were 0.025. For growth with supplemented nitrogen, the medium was incubated at 28ºC with 160 rpm shaking under air. For growth in media without nitrogen, bacteria were grown in a turbidostat reactor (without medium addition) at 28.9ºC with constant stirring in a controlled 2.5% oxygen atmosphere (0.020 L/min argon, 0.370 L/min nitrogen and 0.010 L/min oxygen). Both cultures were incubated until harvest. The OD_{600} values of each culture and time of harvest are provided in Table A1.

Preparation of samples for sequencing

Upon reaching the desired optical density, cells from each growth condition were collected through centrifugation and the removal of supernatant. Cells pellets were flash frozen in liquid nitrogen and stored at -80ºC until analysis. Genomic DNA from each library was isolated with ZR Fungal/Bacterial DNA MiniPrep kits (Zymo Research), quantified with a nanodrop, and subsequently stored at -20ºC.

Sequencing

For each sample, 100 ng of genomic DNA was fragmented to a size of 350 bp by an acoustic DNA shearing device (Covaris). NEBNext Tn-seq Illumina libraries were prepared from the sheared DNA though use of a custom method (74). All libraries were pooled and sequenced on one lane of a 2x150 bp run on the Illumina NovSeq 6000

system with a SP flow cell type. A transposon specific (Sleeping Beauty) primer was used for library sequencing.

Analysis of Tn-seq data

All TA sites within the genome of *G. diazotrophicus* were located using custom and adapted (75) Python scripts. TA sites that were non-unique and non-permissive were located and removed. Through use of Cutadapt (version 1.18) (76) and Bioawk, raw sequencing reads of the library grown under the differential growth conditions were filtered and trimmed. Reads without the transposon were discarded. The remaining 16 bp sequences were mapped to the full genome (CP001189.1, genome; CP001190.1, plasmid within *G. diazotrophicus*; JGI assembly (59) used as reference library and annotation in analysis), allowing for identification of TA insertion locations. Frequency of insertion was determined through counting and summing reads mapping to each gene. Fitness (W) was calculated through the below equation (Equation 3) (65), where N_{t_1} and N_{t_2} were the proportion of the gene before (t_1) and after (t_2) growth under selective conditions, and d was the growth expansion factor calculated via optical density readings as $OD_{600,t2}/OD_{600,t1}$ (Equation 3).

$$W = \frac{\ln\left(N_{t_2} * \frac{d}{N_{t_1}}\right)}{\ln\left(\left(1 - N_{t_2}\right) * \left(\frac{d}{1 - N_{t_1}}\right)\right)} \quad (3)$$

All subsequent data analysis and visualization were performed in R through the tidyverse, Gviz, forcats, aplot, gggenes and patchwork packages.

Table 1. Key plasmids used in Tn-seq.

Plasmid	Relevant manipulations[a]	Parent vector	Source or reference
pBB295	pTnMm1 with addition of BamHI sites	pTnMme1	This study
pBB296	pBB296 with addition of BamHI sites	pBB295	This study
pBB298	Cloned Tet^r cassette from pBBTET6 into pBB296; used for construction of Tn-seq libraries	pBB296	This study
pBBTET2	pUC19 with Tet^r instead of Amp^r	pUC19	(73)
pBBTET6	pUC19 with Tet^r with BamHI sites	pBBTET2	This study
pTnMme1	Plasmid containing *Mariner* transposon and transposase pEB001		(67)

[a] Tet^r, tetracycline resistance; Amp^r, ampicillin resistance.

Table 2. Key primers used in Tn-seq.

Primer	Sequence (5'-3')	Purpose[a]
BBP1282	GCGTATCACGAGGCCCTTTCGTCTTCAAG	Confirm insertion of Tet^r into pBB298
BBP3084	CGCCAAGCTTGCAT GGATCC AGGTCGACTCTAGATATC	Add BamHI site to pBBTET2
BBP3085	GATATCTAGAGTCGACCTGGATCCATGCAAG CTTGGCG	Add BamHI site to pBBTET2
BBP3091	CTTGACGAGTTCTTCTGAGCGGGATCCTGGGG TTCGCGGAATTAATTC	Add BamHI site to pTnMme1
BBP3092	GAATTAATTCCGCGAACCCCAGGATCCCGCTC AGAAGAACTCGTCAAG	Add BamHI site to pTnMme1
BBP3093	GGTTAATTAAGGGCTGCAGGGATCCGATATC AAGCTTATCG	Add BamHI site to pTnMme1
BBP3094	CGATAAGCTTGATATCGGATCCCTGCAGCCCT TAATTAACC	Add BamHI site to pTnMme1
Illumina primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGNNNGTGTATGTAAACTTCCGACTTCAACT* G	Sequence Tn-seq libraries

[a] Tet^r, tetracycline resistance.

Results and Discussion

Library growth

We did not achieve the expected culture growth density of the library under growth in the turbidostat reactor without nitrogen. In conditions of growth under air with

provided nitrogen, we achieved higher levels of optical density (Table A1), approximately 7 generations of growth. Under diazotrophic growth, we achieved only over 4 generations of growth. This lack of density may be due to growth limitation under oxygen-limited diazotrophic conditions. Growth and resulting fitness values may further change upon omission of tetracycline, which was not performed here.

Library statistics

The complete genome of *G. diazotrophicus* has 58,303 TA sites. Of these sites, 57,574 were within the chromosome (CP001189.1) and 729 were within the plasmid (CP001190.1) (JGI assembly (59) used as reference library and annotation). A total of 8,979 of all TA sites were found to be homologous while 3,773 were non-permissive. In some cases, TA sites were both homologous and non-permissive. Due to the restrictive features of these regions, we removed a total of 8,668 TA sites (14.87%), 8,641 from the chromosome and 27 from the plasmid. After this filtering step, a total of 49,635 TA sites remained, which covered 3,337 protein genes out of 3,501 (95.32%). There was a total of 164 genes that were covered by unusable TA sites, which contained either homologous or non-permissive TA sites, or both. These genes were thus unable to be analyzed through this method. Genes without TA sites were additionally unable to analyzed by this method. Due to the restrictive size of our library, not every gene—especially those with few TA sites—received a transposon insertion by chance. This is an inherent limitation in any Tn-seq experiment which we attempted to minimize here through construction of a large mutant library.

Characterization of t_1 mutant library

Data calculations on the t_1 mutant library allowed us to probe genes essential for growth of *G. diazotrophicus* on GAD medium. The three sets of t_1 data were made up of approximately 22.4M, 19.0M and 20.6M reads (Table A2). Following initial filtering to yield 49,635 available TA sites remaining in the dataset following initial filtering, it was found that 33,112 TA sites were hit with a transposon insertion (66.71% insertion rate). Within the 3,337 protein genes that were covered by the available TA sites, 3,260 genes had TA insertions at least one site, leaving 77 genes that did not have any TA insertions. Of these genes with insertions, we flagged those that had less than 50% of their TA

insertions within the center 80% of the gene. Various prior Tn-seq studies discarded TA sites that fell within the tip or tail of a protein gene (70, 77) and only examined the central ORF of genes. Though transposon insertions at the beginning or end of a gene may not disrupt gene functioning (78), we chose not to filter out insertions into TA sites in these regions as to not make assumptions about gene function across the genome. We found that 2,604 genes within *G. diazotrophicus* had at least 50% of their TA insertion counts located in the center 80% of the gene.

We simulated an "expected" library to compare the results of our library. As we built a library of an estimated ≥ 100,000 mutant colonies, we simulated the available TA sites with replacement 100,000 times. Within our simulated library, we observed 33,941 unique TA hit sites. This is comparable in size to what we observed in our actual library. While this simulated TA hit value is a bit higher (2.44%), we can attribute this difference to insertions into essential genes, as the simulation did not differentiate between essential and nonessential genes. This simulated library size confirms our approximate t_1 library size of ≥ 100,000 mutant colonies and allows for a deeper understanding of the library saturation.

To further explore, we compared our simulated and actual data for instances of TA insertion in genes with five or fewer possible TA sites (Figure 4). Here, our actual data has a comparable number of sites that were hit with the simulated library. As the simulated or expected library did not differentiate between gene essentiality, it follows that there would be a slight difference between the real and expected datasets. The TA sites hit within these genes do seem to follow roughly the same distributions for total insertions within genes with few TA sites present, an encouraging observation for the robustness of our actual library. Further, to identify genes with no or few insertions, we compared the values of the actual and expected data for actual counts and $\log_2$-transformed values (Figure 5) (70, 79). We observed roughly similar distributions between the two datasets, albeit more widespread for the actual data. A second slight curve was seen in the $\log_2$-transformed data around a value of 0, which can be expected to be the essential genes (Figure 5).

Figure 4. Comparison of TA insertions in simulated transposon library and actual transposon library. Abbreviations: number of TA sites per gene, NTA.

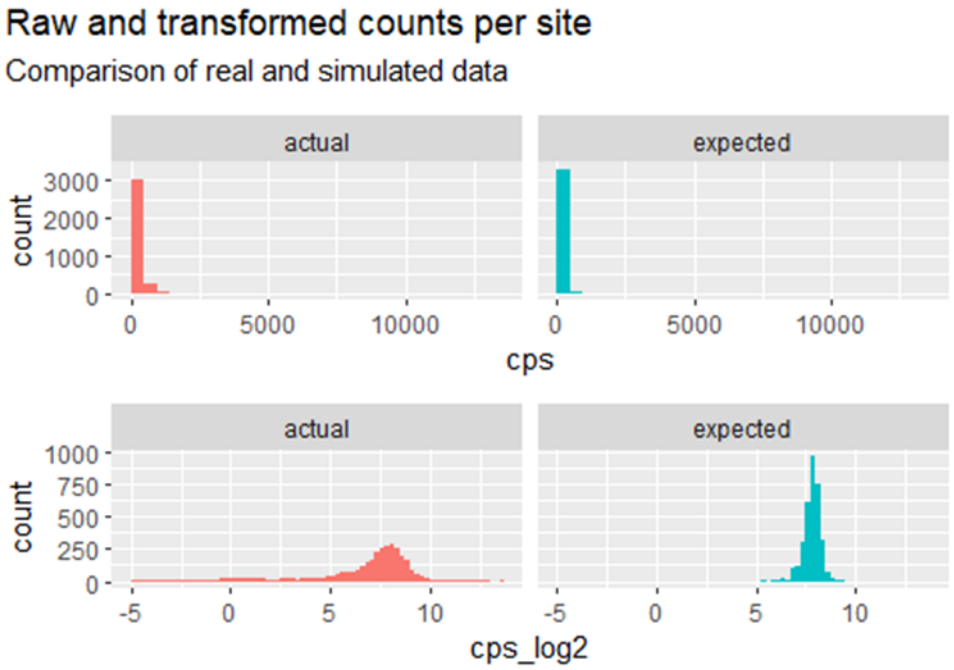

Figure 5. Comparison of actual counts and $\log_2$-transformed values for simulated transposon library and actual transposon library. Abbreviations: counts per site, cps.

In this dataset, essential genes were defined as those within the t_1 library that did not tolerate any insertions, with no transposons inserted into any TA site within the gene, or those genes that contained transposon insertions that were rare when compared to the population. To determine this latter set of genes, a similar method was followed as in prior studies (70, 79). A normal distribution was fit to the top 50% of log_2-transformed insertion counts per site (75), and the probability of each gene fitting this distribution was calculated. The resulting probability values were adjusted via the Holm procedure to account for multiple testing (Figure 6), and genes with an adjusted *P* value of less than 0.05 across replicates were subsequently identified as essential for growth on GAD (75).

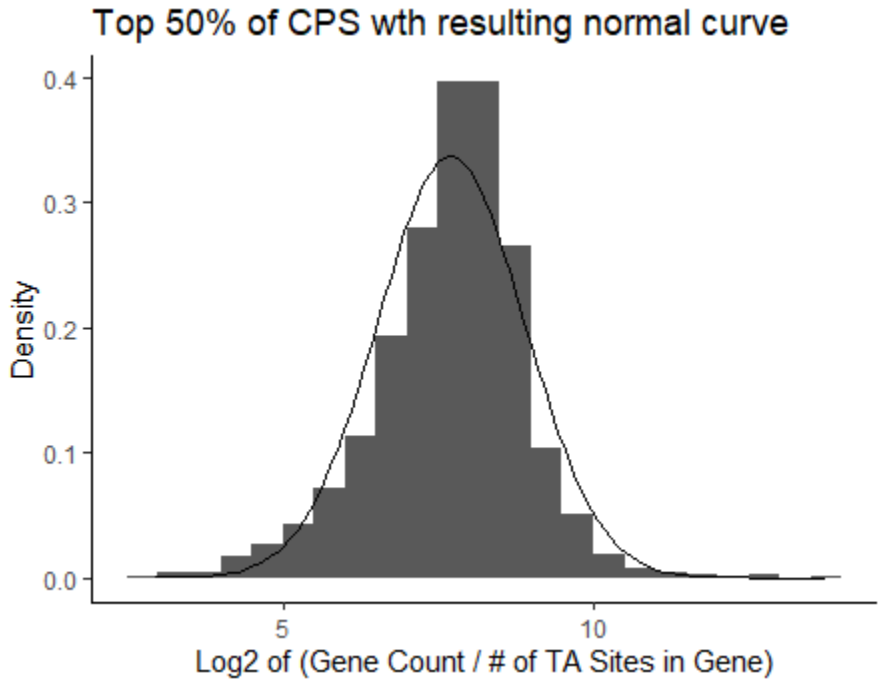

Figure 6. Normal distribution of the top 50% of the log_2-transformed insertion counts per site of the transposon insertion library. Probability values adjusted via the Holm procedure to account for multiple testing.

This analysis of the t_1 mutant library identified 417 genes in *G. diazotrophicus* as essential for growth in our GAD medium. Prior Tn-seq studies regarding essential genes identified core essential genes (70, 71, 75, 79, 80) as ranging between 300 and 400 genes. We have identified a greater number of genes as essential in *G. diazotrophicus* for growth. As this library of ≥ 100,000 mutant colonies were isolated on the same rich

media under identical conditions, we are unable to explicitly differentiate core essential genes and genes that are media specific. Therefore, we present both genes critical for growth and survival as well as genes required for growth on GAD media as essential genes. The remaining 2,920 genes in *G. diazotrophicus* are nonessential for growth.

Many of the essential genes we identified in *G. diazotrophicus* occur sequentially in the genome within genetic clusters. Occurring in clusters of upwards of 15 genes together, these genes appear to be made up of genes that share similar features. One identified cluster of 35 genes (*Gdia_2956-Gdia_2991*) (Figure 7) can largely be categorized as carrying out genetic information processing features. This cluster contains genes coding for an RNA polymerase (*Gdia_2960; Gdia_2961*) and other critical ribosomal proteins. It follows that these large clusters of genes within *G. diazotrophicus* (Figure 7) would be essential, given that many protein products are related to the flow of genetic information. Further characterization of the essential genes in *G. diazotrophicus* through comparison with other known endophytes may uncover genetic similarities. However, to determine genes essential to the endophytic lifestyle, growth of the strain in an endophytic lifestyle would be necessary.

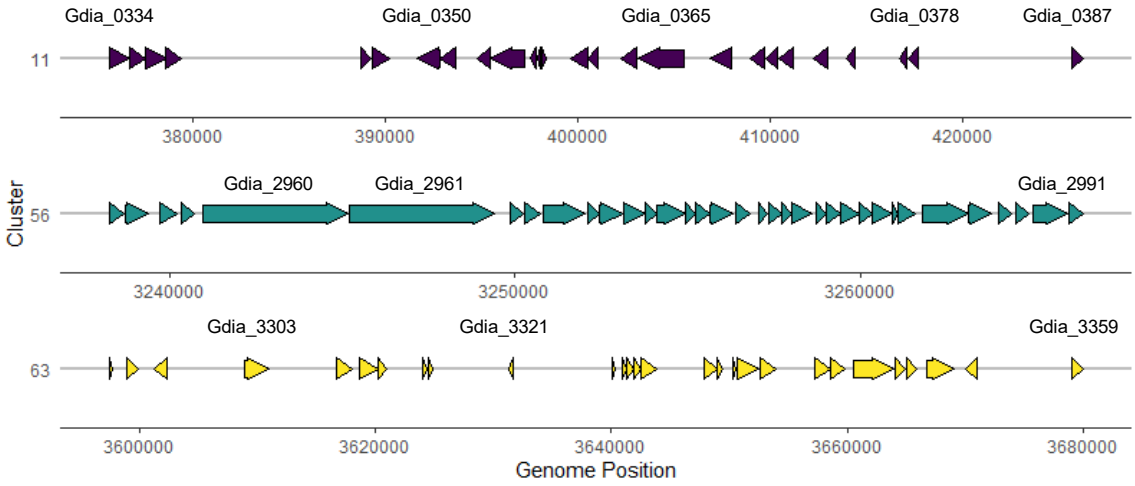

Figure 7. Three selected essential gene clusters in *G. diazotrophicus*. Gaps between genes indicated as some genes were not examined further due to either a lack of TA insertion or contained less than 50% of their TA insertions in the center 80% of the gene. Genome position and gene locus tags are noted. Gene functions of each noted gene from KEGG database: Top (purple): *Gdia_0334*, arabinose-5-phosphate isomerase; *Gdia_0350*, succinyl-diaminopimelate desuccinylase; *Gdia_0365*, outer membrane

protein YaeT; *Gdia_0378*, 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase; *Gdia_0387*, superoxide dismutase. Center (green): *Gdia_2960*, DNA-directed RNA polymerase, beta subunit; *Gdia_2961*, DNA-directed RNA polymerase, beta subunit; *Gdia_2991*, ribosomal protein L17. Bottom (yellow): *Gdia_3303*, translation elongation factor G, *Gdia_3312*, putative integral membrane protein; *Gdia_3359*, protein of unknown function DUF1022.

Further quality analysis of t_2 mutant library

Following calculation of fitness for the t_2 library, we further analyzed the data for overall quality. We had previously flagged the genes in the t_1 dataset that had less than 50% of their TA insertions within the central 80% of the gene as potentially misleading. We chose not to examine those same genes within the t_2 data, as the insertions of the t_1 data were indicative of the library depth of the entire dataset. To further examine all t_2 data more accurately, we additionally identified those genes in the t_2 dataset with a coefficient of variance, also known as relative standard deviation, of greater than 70. Due high variability in the gene insertion rate between t_2 replicates in a fraction of the data, we performed this additional step as we could not accurately draw conclusions regarding the fitness values of these highly variable genes.

Major* nif *cluster genes

The focus of this project was to explore the fitness values of genes related to molybdenum-dependent nitrogen fixation in *G. diazotrophicus* when grown in the presence of biologically available ammonium and under nitrogen fixation conditions. Fitness values of single genes allow for rapid synthesis of the large volume of information typical with high-throughput datasets, and additionally allow for understanding that is beyond traditional laborious methods of gene characterization. Here, we present non-exhaustive results of genes associated with BNF in *G. diazotrophicus*. We focus here on the main genes associated with BNF, as prior work has been done in this organism relating to its BNF abilities (Chapter 1).

Unlike some other diazotrophs such as *Azotobacter vinelandii*, *Pseudomonas stutzeri* and *Rhodobacter sphaeroides* (81), the major genes associated with molybdenum-dependent nitrogen fixation in *G. diazotrophicus* are arranged in one cluster, ranging from the sigma factor *rpoN* to the molybdate transport-associated gene

modD (41). The fitness values of this major gene cluster of *nif* and *nif*-related genes are presented in Figure 8. These genes were assigned to one of several categories based upon their fitness value that was averaged across all TA sites within the gene. The categories were as follows: a large growth defect (purple), moderate defect (blue), no or minimal defect (green), and growth promotion (yellow). Genes that were flagged due to high variability within the t_2 data and are not further discussed are also indicated (white). Genes that were not sampled by the transposon library (white) are additionally indicated.

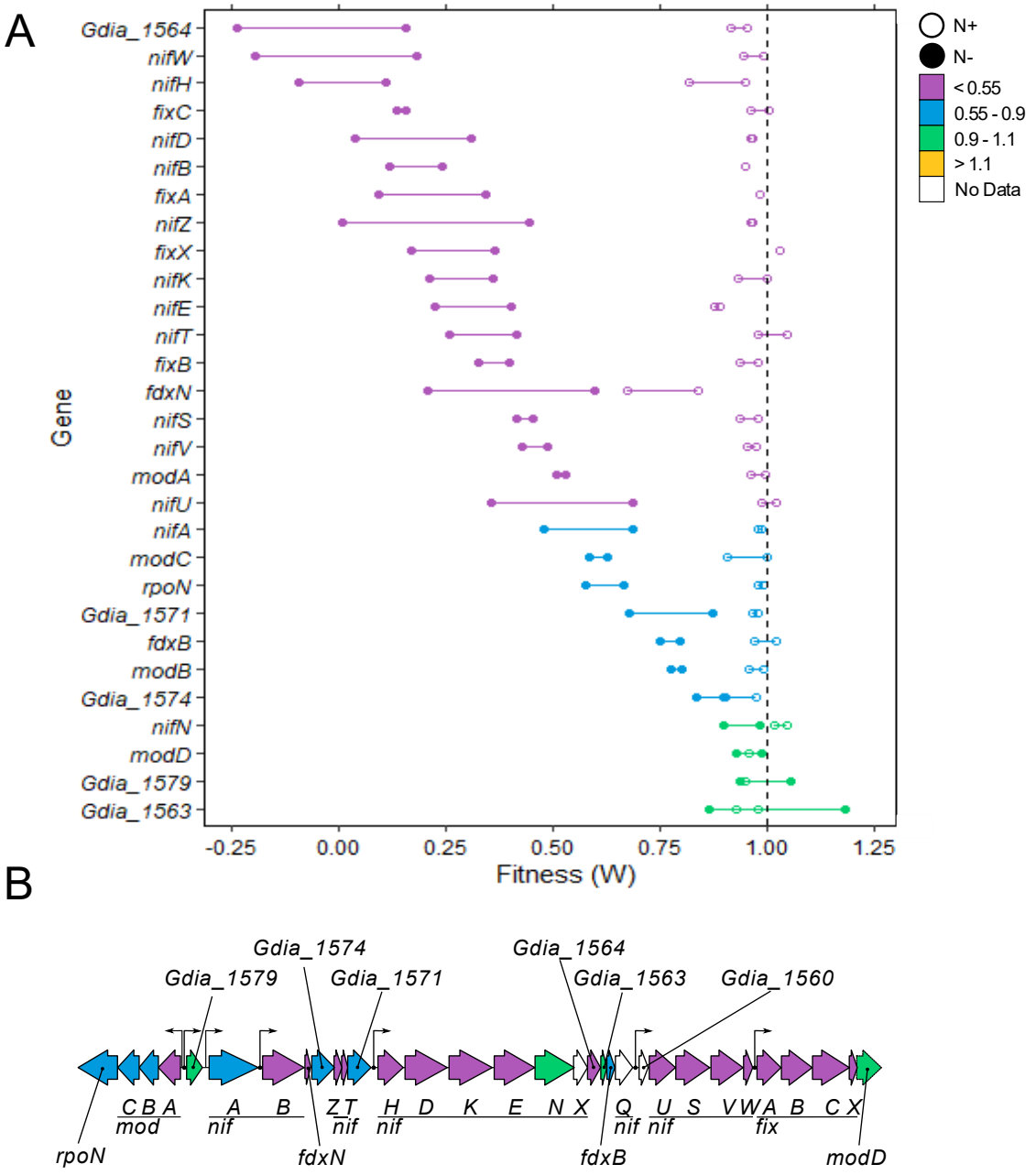

Figure 8. Select genes from the major *nif* cluster. **A**) Gene fitness. Replicates shown per condition. Closed circles shown are without nitrogen (N-), while open circles are with nitrogen provided as ammonium sulfate (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order. Fitness of genes *nifX*, *nifQ* and *Gdia_1560* omitted due to high variability between replicates. **B**) Gene schematic of major nif cluster shown in A. Transcription units denoted by arrows; locations approximated as in (41).

G. diazotrophicus has little functional redundancy in genes related to nitrogen fixation. While functional redundancy is often considered to be a hallmark of organisms carrying out niche functions such as nitrogen fixation, this microbe solely performs BNF through molybdenum-dependent nitrogenase (46, 59). It does not have alternative nitrogenases, such as the vanadium-dependent *vnf* or iron-only *anf* systems (82), so all nitrogen fixation is dependent upon Mo-nitrogenase. This is unlike other well-studied diazotrophs such as *A. vinelandii* and *Rhodopseudomonas palustris*, which both have all three nitrogenases active in their genetic system (82). Some redundancy is in fact found in *G. diazotrophicus* regarding nitrogen fixation genes, but this redundancy is largely limited to the dual ammonium transporter *amtB* homologs and the two *fix* gene clusters. However, *G. diazotrophicus* could be considered a simpler system in which to study nitrogen fixation due to this lack of genetic redundancy that is often found in other systems. Additionally, the major genes associated with nitrogen fixation in *G. diazotrophicus* are arranged in one cluster (81), unlike other diazotrophs like *A. vinelandii* (41). This microbe has a relatively small cluster of *nif* genes related to nitrogen fixation, with the absence of genes critical in other systems such as *nifJ* and *nifL*. To further reiterate this point, *G. diazotrophicus* also lacks genes important in other systems like *naf* and *dra*. This concise arrangement of nitrogen fixation genes allows for the study of a simpler and more streamlined genetic system, and makes *G. diazotrophicus* well-suited for study through Tn-seq. The lack of redundancy evident for diazotrophic growth permits a clear understanding of gene essentiality.

Under diazotrophic growth, the structural genes encoding nitrogenase (*nifHDK*) (46) were found to have large fitness losses, which was to be expected. These genes associated are associated with the α- and β-subunits of the MoFe protein (*nifDK*) and the electron donor to the MoFe protein (*nifH*) (36, 47). Similarly to *nifHDK*, genes *nifB* and *nifE,* both essential for MoFe cofactor biosynthesis (41, 47), exhibited significant fitness defects upon insertion of the *Mariner* transposon under nitrogen fixation conditions. Given the importance of these genes in nitrogenase assembly and activity, their significant fitness defects under BNF conditions are to be expected.

Interestingly, *nifN*, one half of the $\alpha_2\beta_2$ tetramer typically formed with *nifE* and essential for the synthesis of the MoFe protein through the formation of the NifB-co maturation scaffold (36), did not exhibit a fitness defect under diazotrophic conditions. *nifN* is thought to be the sixth gene essential for BNF across all diazotrophs for identifying computationally (38), so this was a surprising result. Upon closer inspection, however, it can be seen that the bulk of the TA sites with high levels of insertion under nitrogen fixation conditions were located at the 3' end of the gene (Figure 9). Insertions at the end of a gene are less likely to disrupt the functioning of the gene (78). We can gather that the high level of insertion at the 3' end of the gene resulted in this lack of fitness differential, as the fitness values presented here are averaged across the entire gene. As such, the insertion sites at the end of *nifN* could be thought to be nonsignificant. The remaining TA sites present within the front of the gene at the 5' end had low levels of insertion under diazotrophic growth (Figure 9), implying that the functioning of the remainder of *nifN* is also conditionally essential here as in other genes such as *nifE*. Considering that *nifEN* are thought to function in concert to form the scaffold for FeMo-co synthesis and eventual nitrogenase maturation, it is reasonable to assume that the 3' end of *nifN* is nonessential for protein activity and that *nifN* is otherwise essential for BNF.

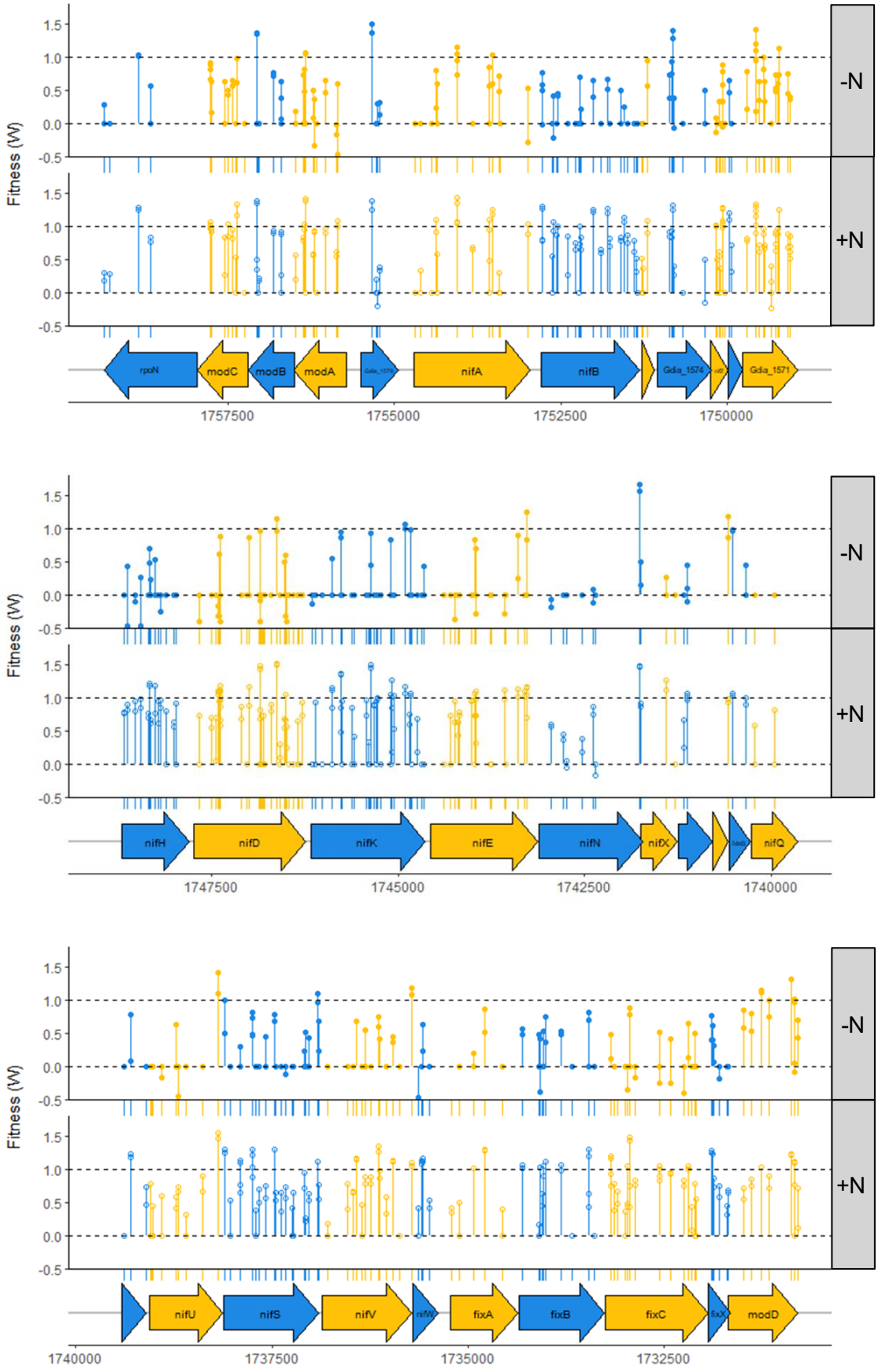
Fitness (W)
1.5
1.0
0.5
0.0
-0.5
-N
+N
rpoN
modC
modB
modA
nifA
nifB
Gdia_1574
Gdia_1571
1757500
1755000
1752500
1750000
nifH
nifD
nifK
nifE
nifN
nifX
nifQ
1747500
1745000
1742500
1740000
nifU
nifS
nifV
nifW
fixA
fixB
fixC
fixX
modD
1740000
1737500
1735000
1732500

Figure 9. Fitness of each TA site in the major *nif* cluster. Genes are colored by alternating colors (blue and yellow). Two replicates shown per condition.

Other genes within the main *nif* cluster that displayed large fitness losses included those that are largely involved in maturation and activation of nitrogenase. These genes included two associated with iron-sulfur cluster assembly (*nifU* and *nifS*), a homocitrate synthase (*nifV*) involved in FeMo-co synthesis, and a protein associated with the MoFe protein prior to P-cluster maturation (*nifW*). These genes are transcribed together within the major *nif* cluster in addition to an iron-sulfur cluster assembly accessory protein (*Gdia_1560*). These genes are certainly essential for nitrogenase assembly in other microbes (36), so their significant fitness defects exhibited here are of no surprise. Additionally, *nifA*, the major transcriptional regulator of all *nif* genes, displayed a moderate fitness loss, as did the transcriptional regulator *rpoN*. These genes appeared to still be important for BNF activity, but these proteins may have greater extraneous regions that can experience disruptions without majorly affecting gene function (Figure 9).

As described above, the nitrogen fixation of *G. diazotrophicus* relies upon the molybdenum-dependent nitrogenase system. To limit potential molybdenum limitation, this diazotroph as well as many others employ highly efficient molybdate transporters, encoded by *modABC* (83). These transporters import the bioavailable form of molybdenum, MoO_4^{2-} (84), allowing for FeMoCo assembly. This gene cluster is found in the major *nif* cluster in *G. diazotrophicus* (*Gdia_1579*-*Gdia_1581*) and displayed large fitness deficits under BNF. However, the *modD* gene (*Gdia_1551)* did not display differential fitness. It is important to note that the *mod* genes in *G. diazotrophicus*, while all found in the major *nif* cluster, are not arranged in one consecutive operon. Instead, *modD* is transcribed in a separate operon. In *E. coli*, the *modABCD* cluster is regulated by the transcription factor *modE*. While these genes are not in one operon in *G. diazotrophicus*, it is possible that all four *mod* genes are likewise regulated by the putative transcriptional regulator *modE* (Gdia_1131) that exhibited a fitness defect under BNF (Figure 10). *G. diazotrophicus* additionally possesses molybdenum cofactor biosynthesis proteins *moaABCDE*. In other organisms such as *E. coli*, these proteins are

associated with molybdenum accumulation for synthesis of molybdoenzymes other than Mo-nitrogenase (83). However, under BNF conditions, *moaAB* (*Gdia_0124*; *Gdia_1246*) and *moaDE* (*Gdia_0049-Gdia_0050)* displayed large fitness losses (Figure 10). These genes, though not located within the main *nif* cluster, are still evidently critical for nitrogenase assembly and BNF, and may be important for FeMoCo assembly in this diazotroph.

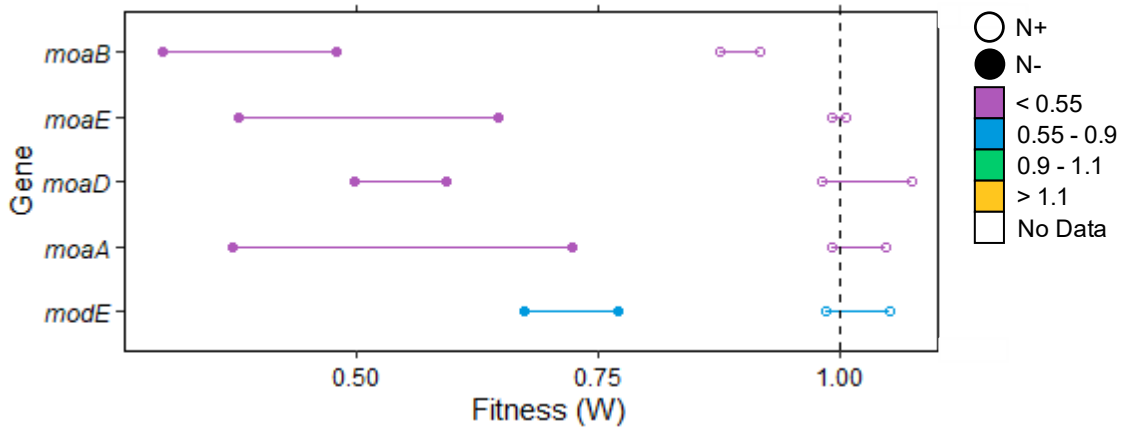

Figure 10. Select *moa* and *mod* genes associated with molybdenum transport and accumulation. Replicates shown per condition. Closed circles shown are without nitrogen (N-), while open circles are with nitrogen provided as ammonium sulfate (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order.

There are several genes within *G. diazotrophicus* that are associated with electron transport to nitrogenase. BNF is an energetically expensive process, requiring 16 mol ATP and 8 electrons per reduction of 1 mol N_2 (Equation 2). The primary electron donors to *nifHDK* are reduced ferredoxin and flavodoxin (85). Contained within the main cluster of nitrogen fixation genes, *G. diazotrophicus* displays two putative ferredoxin proteins: the 4Fe-4S ferredoxin iron-sulfur binding domain protein *fdxN* (*Gdia_1575*) and the nif-specific ferredoxin III protein *fdxB* (*Gdia_1562*). Both proteins displayed differential fitness under nitrogen fixation conditions. Other ferredoxin proteins outside of the major *nif* cluster are found throughout the genome. These include the 4Fe-4Fs ferredoxin iron-sulfur binding domain protein *fdxA* (*Gdia_0255*), and the 2Fe-2S ferredoxin protein *fdxE* (*Gdia_0615*). Though both *fdxA* and *fdxE* displayed high variability between replicates so

we are unable to draw conclusions regarding their fitness, these ferredoxins are both commonly found in diazotrophs (85).

The genome of *G. diazotrophicus* additionally contains two putative *fix* clusters, one within the main cluster of *nif* genes, and one elsewhere in the genome: *fixABCX* (*Gdia_1552-Gdia_1555*) and *fixABC* (*Gdia_1988-Gdia_1990*). This is one of the few instances of functional redundancy in the genes related to nitrogen fixation in *G. diazotrophicus*. While their function is not explicitly known, *fix* genes are thought to provide electrons to drive nitrogen fixation through an electron transfer chain. It has been previously proposed that the diazotrophic taxa acquiring the *fix* complex are able to generate reduced ferredoxin from NADH/NADPH (86). The *fixABCX* cluster found in the major *nif* cluster displayed significant fitness detriments under BNF conditions, however the second *fix* homolog did not likewise display differential fitness. Like the *fixABCX* cluster, this second *fixABC* homolog could be associated with electron transport to support nitrogen fixation. However, the second *fixABC* did not exhibit fitness loss in any of the three genes, although *fixB*2 was not sampled by the transposon library. It may be that the secondary *fixABC* cluster plays a role in nitrogen fixation, but is insufficient for the needs of the bacterium which are otherwise met by activity of the *fixABCX* genes, Alternatively, given the fitness values, it may be that this second *fix* cluster does not participate in providing electrons to power BNF.

We theorize that this lack of fitness differential displayed in this second *fix* cluster could be due to the absence of *fixX* as present in the first homolog. The *fixABCX* proteins are thought to operate in concert, where the NADH electrons are bifurcated by the protein products of *fixAB*, splitting the electrons to both ferredoxin and the *fixCX* proteins in the respiratory chain (86). The lack of *fixX* could deem this cluster unusable by the genome. Further, given the distance of the second *fix* cluster from the major *nif* cluster within the genome, it seems to follow that these genes would be less directly involved in BNF. The secondary *fix* genes would likely not be directly transcribed by *nifA* along with the other *nif* genes. Additionally, while other diazotrophs such as *A. vinelandii* contain the similar *rnf* nitrogen fixation complex, another set of genes thought to be involved in providing electrons for BNF (87), *G. diazotrophicus* contains no *rnf* homologs. Similarly, *G.*

diazotrophicus contains no *nifJ* homologs. *nifJ* is another gene involved in electron transport to nitrogenase and found in the diazotroph *Klebsiella pneumoniae* (88, 89). While two *fix* homologs are indeed present in *G. diazotrophicus*, only the *fixABCX* cluster located in the main *nif* cluster displayed differential fitness under BNF. This further displays the streamlined nitrogen fixation genetic system present in *G. diazotrophicus*.

Several genes with no assigned function in the major *nif* cluster displayed differential fitness under BNF conditions. Genes *nifT* (Gdia_1572) and a conserved hypothetical protein (*Gdia_1571*) both displayed moderate-to-large fitness defects. A few genes within this major cluster also displayed no change in fitness, including a conserved hypothetical protein (*Gdia_1578*), and the protein of unknown function DUF683 (*Gdia_1563*). As hypothetical proteins, their functions have not yet been defined, and it is possible their function is divergent from BNF, which would explain the lack of fitness differential presented here. Additionally, the genes *nifX* and *nifQ*, the carrier of FeMo-co precursors and the processor of Mo prior to its incorporation into FeMo-co respectively (46), were represented in the transposon library here, though their results were too variable from which to reliably draw conclusions.

Nitrogen assimilation and regulation genes

Central to nitrogen fixation, beyond the genes within the major *nif* cluster, are the proteins associated with nitrogen assimilation and regulation. These genes range in function from ammonium uptake into the cell to activation of *nif* gene transcription and are not arranged in one genetic cluster. Rather, these genes are found throughout the genome, some organized in operons and some independently transcribed. The fitness values of these genes associated with nitrogen assimilation and regulation are presented in Figure 11. Like the major *nif* cluster, these genes were also assigned to the same fitness categories indicating the change to the growth of the mutant with these genes silenced.

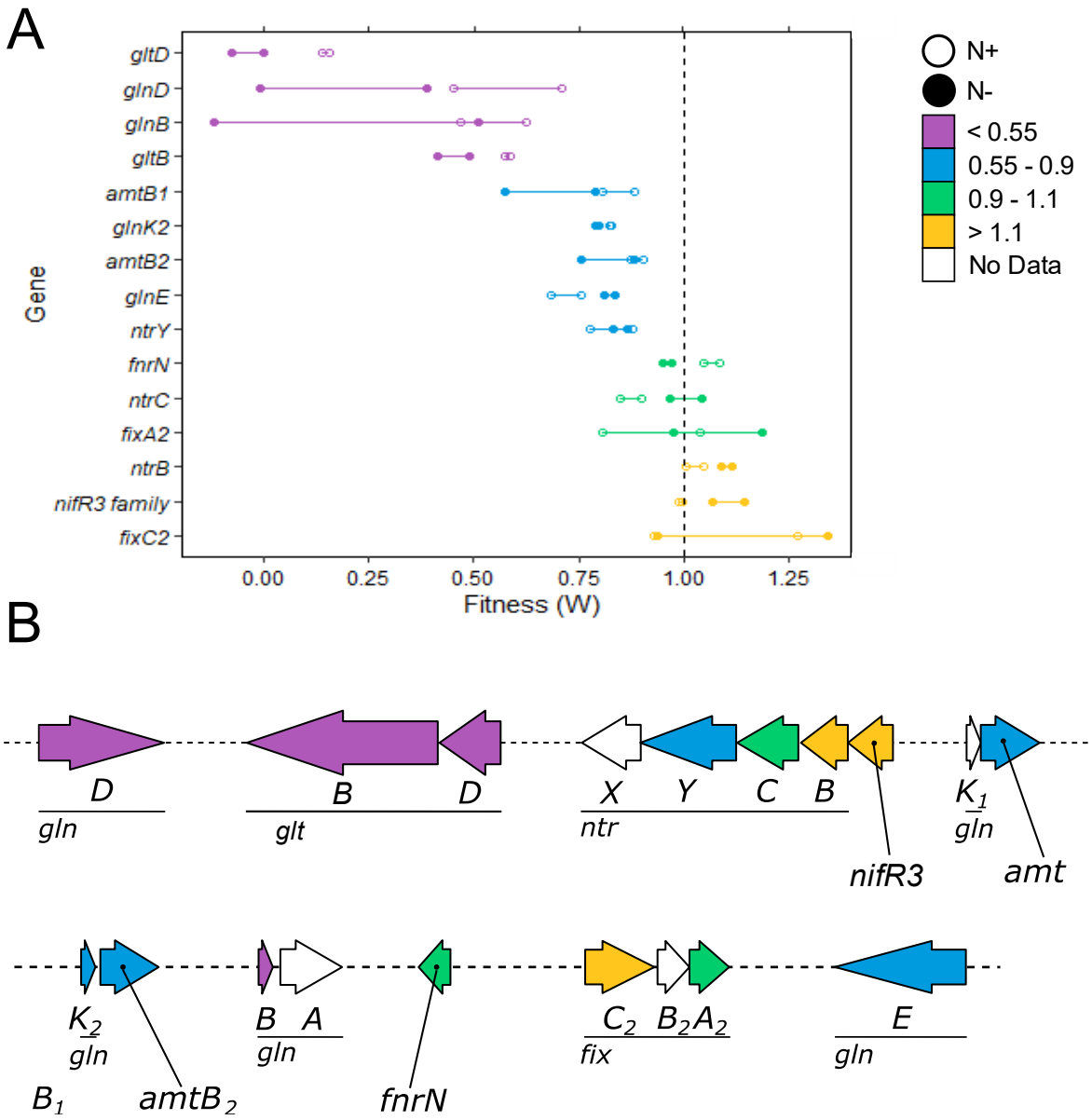

Figure 11. Select *nif*-associated genes. **A)** Gene fitness. Replicates shown per condition. Closed circles shown are without nitrogen (N-), while open circles are with nitrogen provided as ammonium sulfate (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order. Fitness of genes *ntrX*, $glnK_1$ and *glnA* omitted due to high variability between replicates; $fixB_2$ omitted as was not covered by the library. **B)** Gene schematic of genes shown in A.

As mentioned above, *G. diazotrophicus* largely contains little functional redundancy in genes related to nitrogen fixation. However, this microbe does display

some genetic homologs with redundancy in genes related to nitrogen regulation. *G. diazotrophicus* contains two homologs of the ammonium transporter *amtB*, termed *amtB1* (*Gdia_0598*) and *amtB2* (*Gdia_1303*), which bring ammonia into the cell (44). While in separate genomic regions, both genes displayed moderate fitness defects under BNF conditions, indicating that both homologs are active under such an environment. Additionally, both *amtB* genes indicated moderate fitness defects in medium with ammonium present, suggesting a general fitness independent of diazotrophic growth associated with these genes. This is not what was noted regarding the other prominent nitrogen fixation genes that displayed functional redundancy, the dual *fix* clusters. This indicates that both *amtB* genetic products have some essentiality for BNF or under conditions where BNF produces the primary source of reduced nitrogen. However, the potential fitness defects by the interruption of one homolog may be mitigated by the presence and activity of the other. Deletion of both genes simultaneously may result in a stronger fitness defect, which we were unable to observe with this dataset. Further construction of a dual *amtB* deletion *G. diazotrophicus* strain through targeted gene deletion allowed for observation of strain growth under diazotrophic conditions (Chapter 4).

Directly upstream of the two ammonium transporters are two of the three P_{II} protein homologs found in *G. diazotrophicus*, *glnK1* (*Gdia_0597*) and *glnK2* (*Gdia_1302*). The third P_{II} protein homolog is *glnB* (*Gdia_1481*), which lies upstream of the glutamate synthase *glnA* (*Gdia_1482*). Single, double and triple mutants of these genes in *G. diazotrophicus* have previously been constructed (52). The function of each of these three P_{II} protein products were described as nonessential for *nif* gene expression (52), where the *glnK2* protein product was found to act as the *nif* inhibitor and the proteins encoded by *glnK1* and *glnB* controlled *nif* expression in response to ammonium availability. These proteins likely directly interact with the protein product of *nifA*, unlike in other systems (Chapter 1). While these genes are considered homologs, their functions appear divergent from one another. In this dataset, the fitness of *glnK1* was too highly variable to reliably report. However, under diazotrophicus conditions, *glnK2* and *glnB* both exhibited moderate and large fitness differentials respectively. As *glnB* exhibited a

higher fitness defect, it is possible that this protein product is more essential to controlling *nif* expression than *glnK1*, though we cannot confirm that comparison given the data presented here. Additionally, the fitness deficit seen in *glnK2* is only moderate, so it is possible that additional *nif* inhibitors could be present.

Nitrogen assimilation in *G. diazotrophicus*, as in other diazotrophs, is carried out by the glutamine synthetase/glutamate synthetase pathway (GS-GOGAT) (28, 53). While the transposon library covered glutamate synthase (*glnA*), the resulting fitness values under growth without nitrogen were highly variable between the two replicates and as such are not presented here. However, *glnA* is regulated by an adenyltransferase (*glnE*), which displayed a moderate fitness deficit under nitrogen fixation conditions. Additionally, a genomic analysis of *G. diazotrophicus* suggested that the microbe contains alternative routes of ammonia incorporation beyond the GS-GOGAT system (46). These potential pathways include four putative proteins: aminomethyltransferase *gcvT* (*Gdia_0534*), NAD-synthase *nadE* (*Gdia_1103*), histidine ammonia-lyase *hutH (Gdia_1458*), and D-amino acid dehydrogenase *dadA* (*Gdia_2257*). Both *nadE* and *gcvT* demonstrated fitness deficits in both nitrogen fixation conditions and nitrogen-sufficient conditions, with greater fitness deficits under BNF, while *dadA* and *hutH* did not demonstrate fitness differentials. This indicates that *nadE* and *gcvT* may indeed be involved in the proposed alternative ammonia incorporation routes, though the lack of fitness differential for *dadA* and *hutH* suggests that these two genes are not critical.

Depending on nitrogen concentration within the cell, the structure and resulting function of the P_{II} proteins are modified by uridylylation. The uridylyltransferase or uridylyl-removing enzyme *glnD* (*Gdia_0300*), previously identified in *G. diazotrophicus* (50), is a critical protein important in the GO-GOGAT regulatory pathway. This gene displayed a large fitness defect under nitrogen fixation conditions. Prior work indicated that *glnD* alongside the regulatory P_{II} proteins regulate expression of *nifA* (52), the transcriptional regulator of nitrogen fixation (36).

Glutamine synthase is not the only gene product supporting nitrogen fixation in *G. diazotrophicus*. Four proteins are thought to regulate the GS-GOGAT system in other diazotrophs. Together, these proteins were originally termed the global nitrogen

regulatory system (50). In other organisms, the two-component regulatory system NtrBC, encoded by the genes *ntrBC*, is thought to control activation and repression of a variety of genes involved in nitrogen regulation. In this system under nitrogen limitation, NtrC phosphorylates NtrB, activating transcription of genes involved in both nitrogen metabolism and fixation, such as the transcriptional activator *nifA* (44, 51).

In *G. diazotrophicus*, the operon containing *ntrBC* (*Gdia_0486; Gdia_0485*) further contains the similar two-component *ntrXY* (*Gdia_0483; Gdia_0484*) and a *nifR3* family protein (*Gdia_0487*). This operon largely did not exhibit any fitness defect under nitrogen fixation conditions. Of these, *ntrY* had the greatest fitness deficit, which was only moderate, while *ntrB* and the *nifR3* protein indicated slight growth promotion upon their silencing. Previous research on this system in *G. diazotrophicus* indicated that these genes are nonessential for nitrogen fixation (51), a finding that is supported by the fitness data presented here. This data supports the previously suggested reasoning that NifA is instead directly targeted by the P_{II} proteins (52) and the NtrBC system is bypassed in *G. diazotrophicus* (see Chapter 1).

Aside from the GS-GOGAT system and related genes, we examined the one-component regulator *fnrN*. This gene is a common regulator of aerobic respiration found across numerous nitrogen-fixing Alphaproteobacteria (90). The *fnrN* protein product is thought to directly sense oxygen, as it contains an iron-sulfur cluster (90). It did not exhibit a significant fitness differential under nitrogen fixation conditions. As we grew the library under limited aerobic conditions, this is a reasonable result to expect, as little oxygen was present in the BNF environment. However, under the conditions in which *G. diazotrophicus* exhibited nitrogen fixation here, *fnrN* was nonessential. Other than the previously described regulators *rpoN* and *nifA*, which are found within the major *nif* gene cluster, *G. diazotrophicus* does not appear to exhibit further transcriptional regulators relating to nitrogen fixation, such as the two-component regulatory system *fixLJ* or other homologs (90). As *fnrN* did not exhibit a fitness differential, it appears that transcription regulation of *nif* genes is confined to *rpoN* and *nifA*.

Nitrogenase oxygen protection mechanisms

Both the Fe and MoFe proteins that make up nitrogenase are highly sensitive to oxygen, becoming irreversibly damaged and inactivated by its presence. To avoid nitrogenase inactivation, diazotrophs employ a variety of oxygen protectant mechanisms (91–93). For instance, *A. vinelandii* is believed to protect nitrogenase through increased respiration to decrease the oxygen concentration around nitrogenase, as well as through conformational protection of the protein through use of a the 2Fe-2S (FeSII) ferredoxin protein (94). Like *A. vinelandii*, *G. diazotrophicus* is thought to also employ these mechanisms.

Under aerobic growth, *G. diazotrophicus* utilizes a respiratory protection mechanism, including ubiquinone and various cytochromes (95). The complete respiratory chain complex of *G. diazotrophicus* has been identified (46), spanning from *nuoA* (*Gdia_0718*) to *nuoM* (*Gdia_0706*). Here, the *nuo* genes largely displayed no fitness differential under nitrogen fixation conditions (Figure 12). As this respiratory protection mechanism is thought to be displayed only under highly aerobic environments, it follows that these genes would display unessential characteristics, as the library was grown under limited oxygen when under BNF conditions, in an environment containing only 2.5% oxygen.

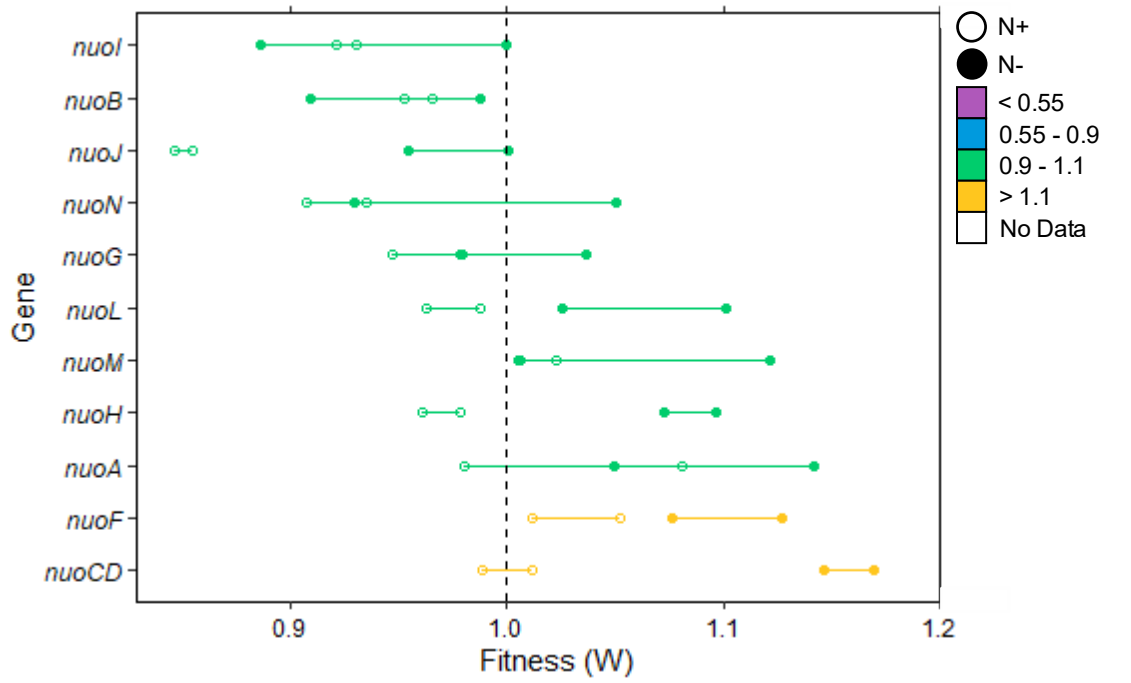

Figure 12. Select *nuo* genes associated with respiration. Replicates shown per condition. Closed circles shown are without nitrogen (N-), while open circles are with nitrogen provided as ammonium sulfate (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order.

Like the nitrogenase oxygen protection displayed in *A. vinelandiii*, this phenomenon has similarly been observed in *G. diazotrophicus*. A previous study provided strong evidence for conformational nitrogenase protection in *G. diazotrophicus* by an FeSII protein, where the FeSII protein forms a complex with nitrogenase, protecting it from oxygen damage and rendering it inactive (94). This FeSII protein was later suggested to be the putative 2Fe-2S ferredoxin protein *fdxE* (*Gdia_0615*) (96). While *fdxE* was hit within our library, the resulting fitness values were flagged in this dataset for growth under BNF conditions due to high variability. Further, as *fdxE* did not demonstrate a fitness differential under non-BNF conditions, we are unable conclude the essentiality of *fdxE* for BNF growth given our dataset, though we would not expect to see a fitness differential here given the oxygen-poor environment of BNF growth.

Polysaccharides

While we did not grow *G. diazotrophicus* under endophytic conditions here, it is worthwhile to mention a few genes associated with an endophyte lifestyle that have been

identified in this organism. There are several stages to the process of endophytic infection and colonization of a host plant that has been studied in detail in other endophytes (1, 97). Initially studied in *G. diazotrophicus* through colonization of micropropogated sugarcane (98), those authors proposed that this bacterium colonized the roots first and then used existing plant structure to colonize the internal tissues. This colonization process has similarly been identified in other endophytes (1). Central to the initial attachment of bacterial cells in the colonization process colonization process is polysaccharides (1). As an endophyte, *G. diazotrophicus* contains genes encoding proteins that may be important for host-bacterial interactions, including capsular polysaccharides (CPS), lipopolysaccharides (LPS), and exopolysaccharides (EPS) (46). Of these polysaccharides, exopolysaccharides may assist with bacterial cell attachment early in endophytic colonization (1), as was reported for *G. diazotrophicus* rice root surface attachment (64), while lipopolysaccharides have been found to be necessary for bacterial attachment within the host plant in other endophytes (99).

Due to its status as an endophyte, it comes as no surprise that within *G. diazotrophicus*, many genes related to polysaccharides have been identified. Within its genome, *G. diazotrophicus* was found to contain two key gene clusters related to CPS biosynthesis and export (*Gdia_0642*-*Gdia-0646*; *Gdia_0660*-*Gdia_0664*) (46). Of note, two CPS genes displayed large fitness deficits when grown under BNF conditions: *ctrB* (*Gdia_0662*) and *ctrD* (*Gdia_0660*) (Figure 13). Other genes such as *Gdia_0643* and *Gdia_0645* instead displayed fitness gains, indicating that the deletion of these genes was beneficial for the growth of the mutant (Figure 13). CPS have not been studied extensively in *G. diazotrophicus* (46, 100), so the interactions of these specific protein products have yet to be determined.

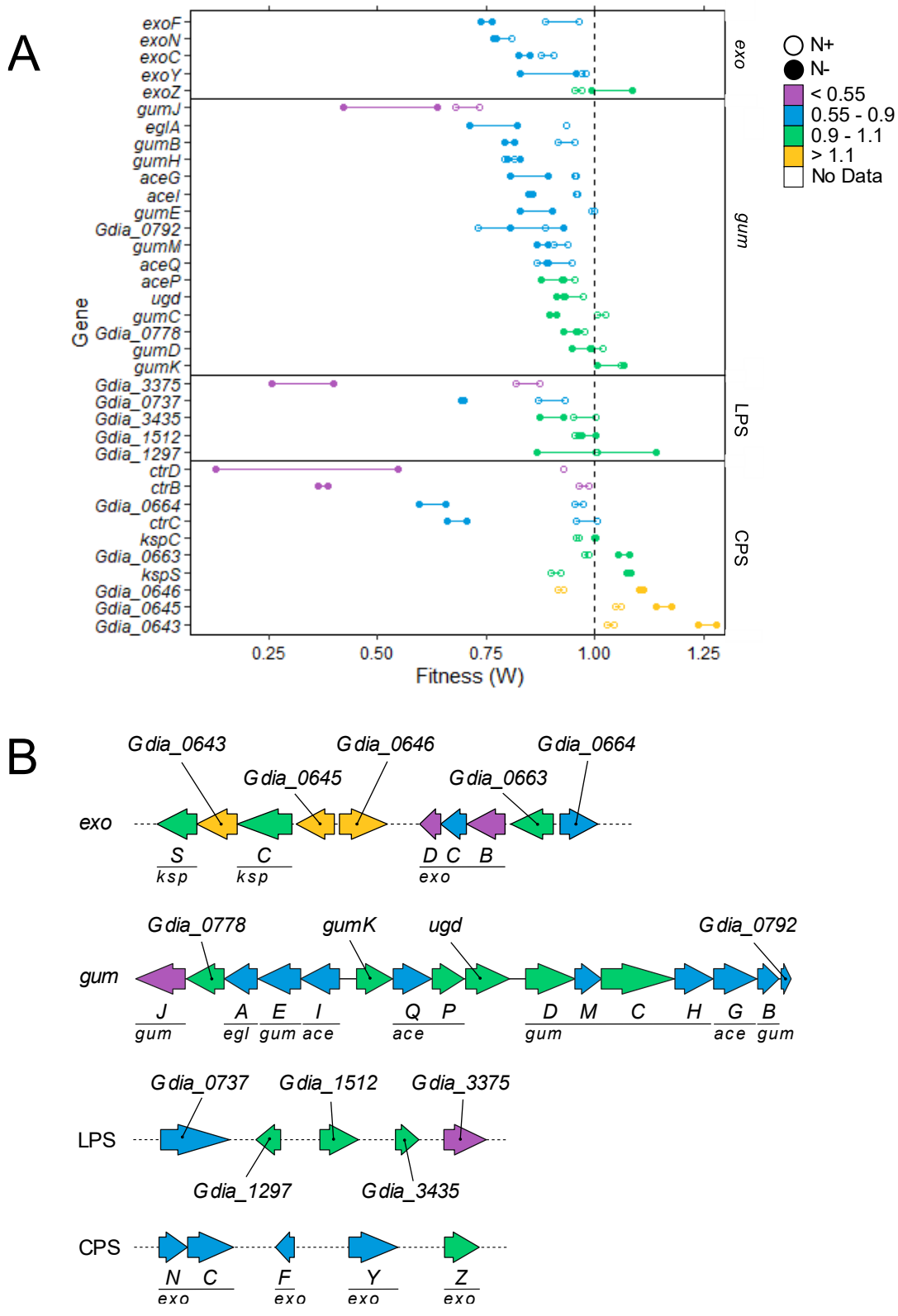

Figure 13. Select genes associated with polysaccharide formation. **A)** Replicates shown per condition. Closed circles shown are without nitrogen (N-), while open circles are with

nitrogen provided as ammonium sulfate (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order. From top to bottom, genes presented are in the following groups: *exo* cluster, *gum* cluster, LPS, CPS. **B**) Gene schematic of genes shown in A.

In addition, several genes that were scattered throughout the genome were identified to encode several proteins related to LPS biosynthesis (46) that were explored in this study (Figure 13). Of those genes that were also contained within the mutant library of this study and not otherwise excluded due to high variability, two genes coding for glycosyltransferases (*Gdia_3435*, *Gdia_1512*) demonstrated no fitness differential as did the gene coding for a nucleotidyl transferase (*Gdia_1297*). However, an O-antigen polymerase (*Gdia_3375*) and a protein product related to lipopolysaccharide transport (*Gdia_0737*) demonstrated moderate-to-significant fitness detriment under BNF.

As previously described (46), *G. diazotrophicus* possesses a large *gum* gene cluster spanning from *gumJ* (*Gdia_0777*) to a conserved hypothetical protein (*Gdia_0792*) that is important in exopolysaccharide production and critical for successful endophytic association (64). Under BNF conditions, the majority of these genes did exhibit moderate differential fitness, though not all (Figure 13). The polysaccharide biosynthesis protein *gumJ* displayed the largest fitness deficit when grown without nitrogen, with the next largest deficit displayed by cell-wall-degrading endogluconase *elgA (Gdia_0779*) (46). Additionally, *gumD*, thought to be the first step in exopolysaccharide biosynthesis (64), did not display a fitness differential. In addition to this more well-studied cluster, several other genes related to the synthesis of other polysaccharides are found scattered throughout the genome (46). These so-called *exo* genes (*Gdia_0121, Gdia_0122, Gdia_0734, Gdia_0736* and *Gdia_3379*) displayed minimal fitness differentials under BNF conditions, though a moderate defect could be observed for the majority of genes (Figure 13)..

It is yet unknown why these genes important for endophytic growth displayed these greater fitness differentials when grown under these conditions. Exopolysaccharide production by the *gum* genes could be tangentially related to nitrogenase oxygen protection, as a prior study on *G. diazotrophicus* suggested that colony mucilage

formation and the position of *G. diazotrophicus* therein was related to the protection of nitrogenase (101). It has further been speculated that the *gum* gene cluster is related to the localization of hydrolytic enzymes produced by the bacterium (46). While these genes did not all display critical changes under BNF, it is important to establish baseline fitness values of these genes important for endophytic lifestyle for future studies in *G. diazotrophicus*.

Other genes

In addition to filling in the gaps of knowledge in known systems, global analysis studies such as this study have the additional potential to reveal gene essentiality outside of the expected. Beyond the expected *nif* and *nif*-associated genes, we additionally found unexpected genes that yielded interesting fitness values under nitrogen fixation conditions. Divided into two groups, these anomalies included genes that were important to growth under nitrogen fixation conditions with large fitness losses when grown without reduced nitrogen, and those whose removal was beneficial for growth without reduced nitrogen and had large fitness gains (Figure 14).

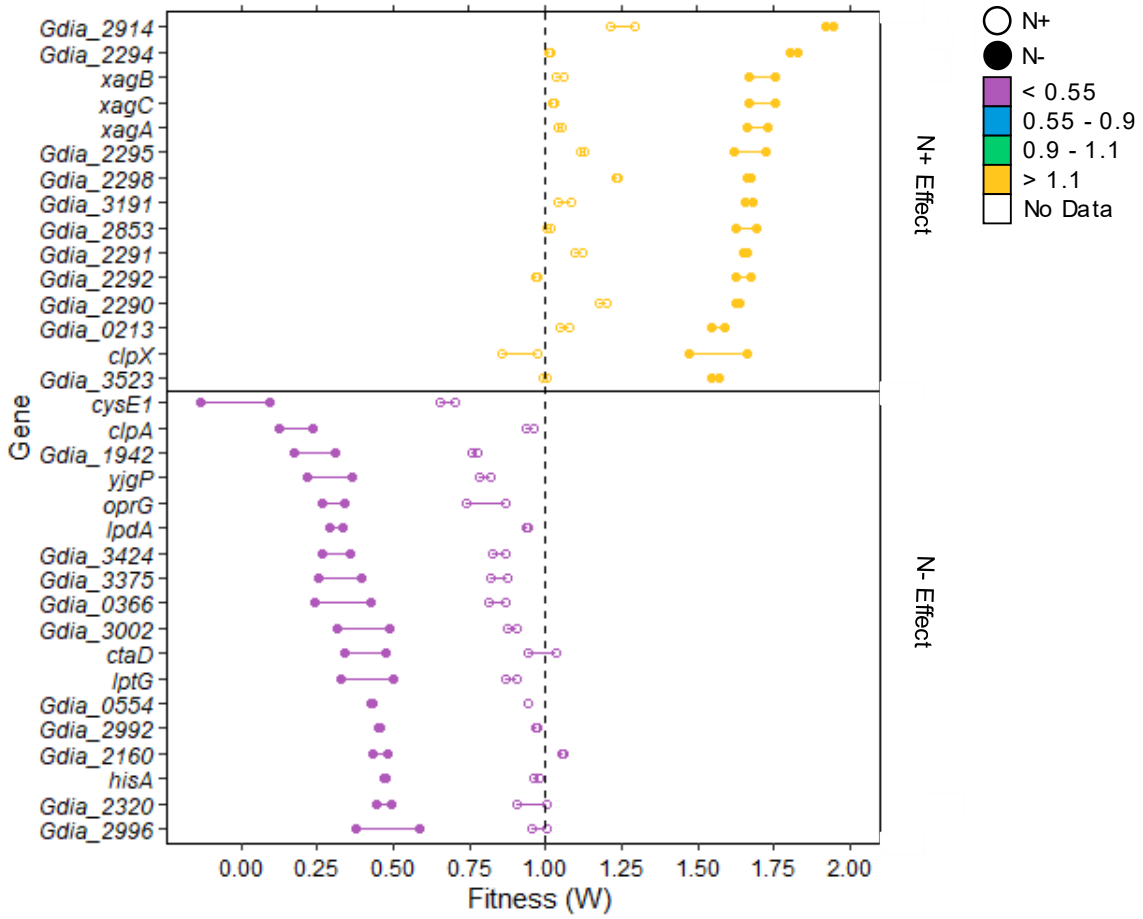

Figure 14. Top genes with the largest fitness differential between samples grown with and without nitrogen. Genes divided into two categories: those grown without nitrogen with higher fitness (top) and those with lower fitness (bottom). Closed circles shown are without nitrogen (N-) while open circles are with nitrogen (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order.

Regarding the genes that were important to growth under nitrogen fixation conditions, several genes were found to display large fitness losses that were not entirely unexpected. These genes, such as *cysE1* (Figure 14) and the *moa* genes (Figure 10), are related to nitrogen fixation in other organisms. Cysteine is critical for the Fe cluster in nitrogenase and could be brought to the system by serine O-acetyltransferase, encoded by *cysE1* (*Gdia_0595*). However, the bulk of the genes reported here displayed unexpected fitness (Figure 14). The first category, those essential to BNF, largely included proteins potentially found on the outer surface of the membrane. These proteins included an O-

antigen polymerase (*Gdia_3375*), the dihydrolipoamide dehydrogenase *lpdA* (*Gdia-_0164*), which is involved in sugar production, the polysaccharide export inner-membrane protein *ctrB* (*Gdia_0662*), the permease *lgtG* (*Gdia_3416*), and the OmpW family protein *oprG* (*Gdia_1985*). Regarding genes that displayed fitness gains, these included the glycosyltransferase proteins *xagABC* (*Gdia_2854-Gdia_2856*) and the ATP-dependent protease *clpX* (*Gdia_3325*). In other organisms, the *xagABC* genes are involved in exopolysaccharide production (102). Reduction in polysaccharide production in *G. diazotrophicus* could be related to oxygen balance due to the low-oxygen growth nitrogen fixation growth conditions. The latter gene, *clpX*, was not the only ATP-binding protease subunit to display an unusual fitness value. The subunit *clpA* displayed the opposite effect to the similar *clpX* and appeared essential to BNF. In *E. coli*, the gene products of *clpA* and *clpX* are ATP-dependent unfoldases that promote different protein degradation through interaction with the protease, ClpP (103, 104).

Conclusion

This work provides insight into genes essential for BNF in *G. diazotrophicus*. To our knowledge, studies exploring the fitness of *G. diazotrophicus* genes through the tools provided by Tn-seq have not before been reported. Our data suggests minimal functional redundancy in genes essential for BNF and presents *G. diazotrophicus* as a streamlined diazotrophic system. Beyond the genes central to BNF, our data also suggested the conditional essentiality of several genetic mechanisms related to other microbial characteristics under BNF. Understanding into genes critical for BNF in this microbe may allow for its optimization as a biofertilizer through future strain engineering. Beyond potential use as a biofertilizer, this work forms a basis for future studies on this organism in an endophytic environment to determine genes essential for plant associations. Armed with the knowledge of the genes critical for diazotrophic growth in *G. diazotrophicus*, we can look towards answering bolder questions about endophytic nitrogen fixation.

Chapter 3: Exploring Glucose, Sucrose, Galactose and Asparagine Catabolism of *G. diazotrophicus*

Introduction

Study of the metabolic capabilities of plant endophytes may reveal potential lifestyle adaptations. Soil- and rhizosphere-dwelling microbes encounter significant competition for resources, while endophytic bacteria do not face that level of competition (105). Plant endophytes have instead adapted to grow in their host plants and utilize readily available resources with fewer competitors (105). This specialized environment lends itself to study of the metabolism of different compounds to elucidate understanding of plant-endophytic relationships.

Beyond nitrogen fixation, as a plant endophyte, *G. diazotrophicus* has other interesting features to be studied. While *G. diazotrophicus* is natively found in sugarcane and can thus grow well in high sucrose concentrations (106), it is capable of growth on nutrients beyond sucrose. It appears to be able to catabolize a variety of carbon and nitrogen sources for energy due to its robust growth under laboratory conditions on compounds such as galactose, arabinose and glutamic acid (12, 27, 107, 108). Further, *G. diazotrophicus* is capable of colonizing plants other than its native host (13), indicating its metabolic versatility. The wide range of growth conditions for *G. diazotrophicus* makes it a remarkable organism for study to further understand the plant endophytic environment.

The sugar metabolism of *G. diazotrophicus* has previously been studied. As in other *Acetobacter* and *Gluconacetobacter* strains examined that do not synthesize cellulose, it lacks the Entner-Douderoff metabolic pathway (109). Indeed, the enzymes of this pathway have not been detected in this organism (110, 111). Therefore, the oxidative pentose phosphate pathway (PPP) is thought to be the main route of carbon catabolism in this organism (28, 108, 110–112). This pathway originates with periplasmic glucose oxidation (112), the principal pathway of glucose metabolism in *G. diazotrophicus* (13). The utilization of the PPP instead of other catabolic pathways is in agreement with other species within the *Acetobacteriaceae*, where the PPP has likewise been previously described as the main sugar pathway (113). While the enzymes of the Emden-Meyerhoff-

Parnas pathway were originally undetected (110), a protein expression profile of the organism suggested that the lower half of this pathway may play a role in catabolism or function in gluconeogenesis (111). Additionally, the TCA cycle is complete in *G. diazotrophicus* (28, 111), unlike strains of the genus *Gluconobacter* (113).

The metabolic capabilities of this organism have received little attention in recent years and have not yet been studied through the means of a Tn-seq experiment. As previously discussed (Chapter 2), Tn-seq allows for rapid characterization of gene function spanning the entire genome of an organism. Using this method, we can quickly examine genes thought to be essential for central metabolism. Further, this method can potentially reveal new, not yet studied genes likewise critical for growth under specific conditions. It is a versatile method that is well-suited to build understanding of the metabolism of *G. diazotrophicus*.

In this extension of the Tn-seq study on *G. diazotrophicus* (Chapter 2), we examined fitness within genes relevant to the catabolism of several carbon and nitrogen sources. This method of study provides rapid insight into the metabolic differences of this organism under different growth conditions beyond nitrogen fixation. Few global analysis studies have been performed on this organism, and particularly regarding its metabolic function, this study can inform future work with *G. diazotrophicus* in novel endophytic environments. Here, we present an analysis of fitness in *G. diazotrophicus* for those genes related to growth on several carbon and nitrogen sources including sucrose, glucose, and asparagine.

Methods

Bacterial strains and growth conditions

Gluconacetobacter diazotrophicus PA1 5 (ATCC 49037) was obtained from Cedric Owens and was grown aerobically at 30ºC on GAD basal medium and modified Burk's (B) medium (72) unless otherwise specified. The composition of the GAD and B media were previously noted (Chapter 2), with slight alterations. The same medium was supplemented instead with various carbon and nitrogen sources (see Table 3). All media was additionally supplemented with 100 mg/L tetracycline unless otherwise noted.

Growth conditions of *Escherichia coli* WM3064 were previously noted (Chapter 2).

Genetic constructs and library construction

All methods regarding the construction the Tn-seq library of *G. diazotrophicus* were as previously described (Chapter 2) as this study was performed simultaneously with the previously noted Tn-seq. The identical Tn-seq library was used here as in Chapter 2.

Library growth conditions

Growth conditions of the initial growth of the library (t_1) and preparation of the t_1 library for inoculation were as described (Chapter 2). The cells of the initial growth of the library were instead inoculated in duplicate into 250 mL GAD medium or modified B medium (see Table 3). The inoculated media were incubated at 28ºC with 160 rpm shaking under air until 7 generations of growth had been achieved. Conditions for without nitrogen was previously described (Chapter 2). The OD_{600} values of each culture and time of harvest are provided in Table A1.

Sequencing and analysis of Tn-seq data

All further preparation of samples, Illumina sequencing and subsequent analysis of Tn-seq were as previously described (Chapter 2) and performed in concert with the previously described Tn-seq.

Table 3. B media composition.

Carbon (20 g/L)	**Nitrogen**
Galactose	10 mM $NH_4(SO_4)_2$
Sucrose	10 mM $NH_4(SO_4)_2$
Sucrose	10 mM asparagine
Sucrose	-

Results

Library statistics

The statistics of the t_1 library were as previously noted (Chapter 2). The same genes flagged as potentially misleading in the t_1 dataset were again not examined within the t_2 dataset as the insertions in the t_1 dataset are indicative of the entire library depth.

We additionally also identified genes with a coefficient of variance, also known as relative standard deviation, of greater than 70 as specified previously (Chapter 2).

Glucose catabolism: PPP and glucose dehydrogenase

Glucose is the main carbon source of *G. diazotrophicus*. Through previous studies, the main pathway of *G. diazotrophicus* carbon catabolism has been proposed to be the pentose phosphate pathway (PPP) (28, 110, 112). Enzymes of the Embden-Meyerhof-Parnas and Entner-Doudoroff pathways, such as 6-phopshofructokinase and 6-phosphogluconate dehydratase, have not been detected in *G. diazotrophicus* (28). The use of the PPP as the primary glucose catabolism pathway is a finding in agreement with other acetic acid bacteria such as *Gluconacetobacter oxydans* (109, 114).

The first committed step in the PPP is glucose oxidation, and this is the main route of glucose metabolism in *G. diazotrophicus* (28, 110). In the periplasmic space, pyrroloquinoline quinone-linked glucose dehydrogenase (PQQ-GDH) carries out this first step by oxidizing glucose to gluconic acid. PQQ-GDH is thought to be synthesized constitutively (112). The growth conditions on GAD medium were the only conditions containing glucose that our transposon library was challenged against. Under these growth conditions, no fitness defects were observed for any of the three genes coding for PQQ-GDH (*Gdia_1469*, *Gdia_2382*, and *Gdia_3089*) (Figure 15). It appears that any one gene coding for PQQ-GDH is not individually essential. However, the importance of this protein function is emphasized by the functional redundancy of these genes in *G. diazotrophicus*. Due to the presence of this gene in triplicate, it can be gathered that inactivation of one gene through silencing does not completely remove the capability to carry out this reaction. Silencing of all three PQQ-GDH genes simultaneously and combinations thereof may result in a stronger fitness defect, or otherwise inform which genes are more critical for this process.

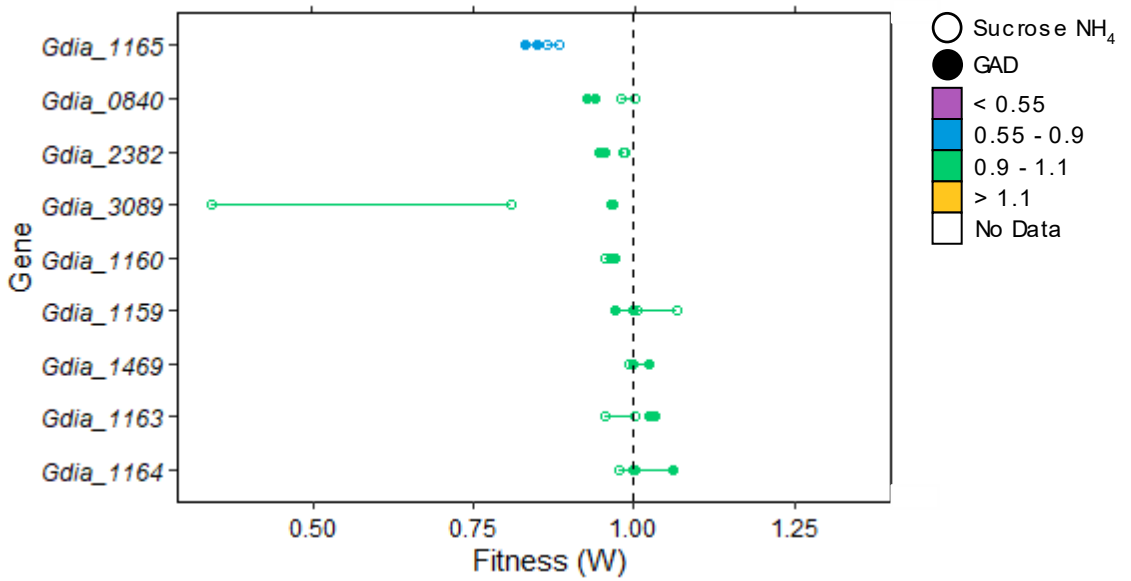

Figure 15. Select genes of the pentose phosphate pathway. Select genes of the pentose phosphate pathway. Replicates shown per condition. Closed circles shown are growth on GAD while open circles are growth with sucrose ammonium. Genes color coded as denoted by the average GAD fitness value and presented in ascending order.

It has been further found that PQQ-GDH activity is greater under BNF conditions, where it provides an abundance of energy for the bacterium (112, 115). Under BNF conditions, one gene coding for PQQ-GDH (*Gdia_3089*) displayed a moderate fitness differential, while the remaining two genes did not (Figure 16). This fitness differential implies that this one gene is more critical for providing energy under BNF conditions than the remaining two homologs, and its functioning cannot be as easily duplicated. However, this gene displayed variability between replicates, so we cannot confidently draw conclusions from this data. Even if this gene only expressed a moderate fitness deficit, we speculate that the activity of the remaining PQQ-GDH homologs were able to meet the metabolic needs of the cell. Further genetic manipulation of the three PQQ-GDH genes under these conditions would allow for greater understanding of their conditional essentiality. As PQQ-GDH is primarily synthesized in the early stages of growth under diazotrophic conditions to meet the additional energy demands of the organism during BNF (112, 115), these results are not unexpected. When grown under diazotrophic conditions, it is additionally important to note that sucrose was the sole

carbon source. This sugar would have been further catabolized in advance of interaction with PQQ-GDH. To understand the essentiality of PQQ-GDH more clearly under diazotrophic conditions, growth of the library on glucose-only media would be required.

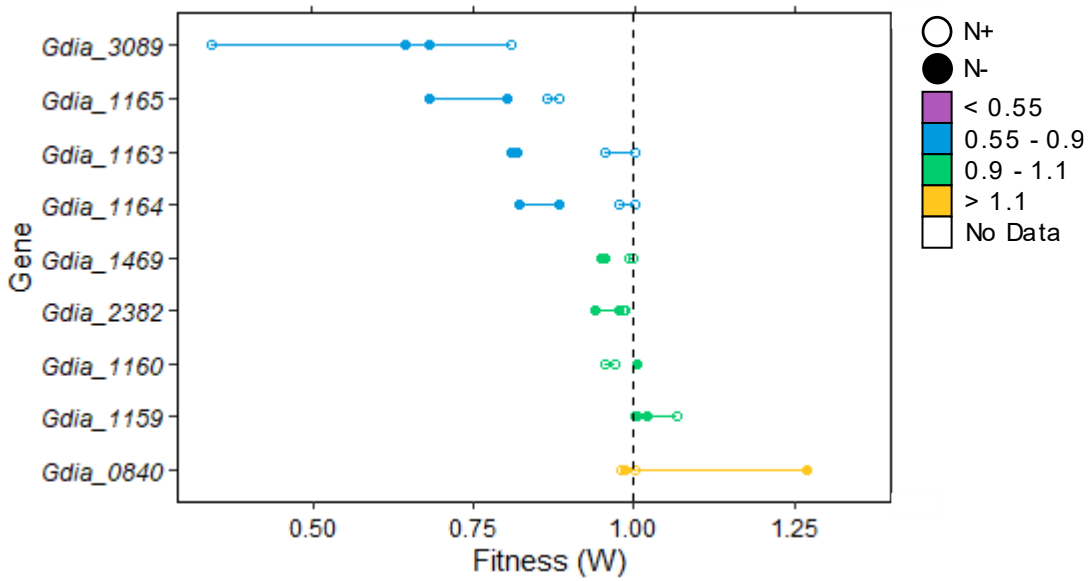

Figure 16. Select genes of the pentose phosphate pathway, BNF conditions. Replicates shown per condition. Closed circles shown are without nitrogen (N-), while open circles are with nitrogen provided as ammonium sulfate (N+). Genes color coded as denoted by the average N- fitness value and presented in ascending order.

While PQQ-GDH is the main pathway of glucose metabolism, an alternative intracellular glucose metabolic pathway can additionally be followed. When under conditions of glucose excess, the nicotinamide adenine dinucleotide-linked glucose dehydrogenase (NAD-GDH) additionally promotes glucose oxidation (28, 112). This gene (*Gdia_0840*) did not display a fitness defect under GAD growth conditions (Figure 15). As NAD-GDH has been suggested to be an alternative route for glucose oxidation with PQQ-GDH serving as the principal pathway, it follows that this gene would be nonessential. It appears that here, as with PQQ-GDH, the remaining glucose dehydrogenase homologs are sufficient for metabolic function without reduction in growth.

It has been proposed that complete periplasmic oxidation of glucose is carried out by several additional enzymes. The periplasmic oxidation of gluconate to 2-ketogluconic

acid occurs through the action of gluconate-2-dehydrogenase, encoded by three genes (*Gdia_1163, Gdia_1164*, and *Gdia_1165*) (46). While two of these genes did not exhibit major fitness deficits when grown on GAD, *Gdia_1165* demonstrated moderate fitness loss (Figure 15). Minimal fitness differentials were observed for these genes when compared to growth on sucrose with ammonium, indicating that this reaction is not critical for growth under either growth condition. However, moderate fitness differentials were observed for diazotrophic growth, supporting the idea that sugar metabolism by PQQ-GDH is more critical to meet energy demands when nitrogenase is active.

Further catabolism of gluconate in *G. diazotrophicus* is carried out by various other enzymes. A 5-ketogluconate reductase was originally proposed for the conversion of gluconate to 5-ketogluconate in strain Pal3 (28). Complete genome sequencing revealed this reaction may be performed by a glucose/methanol/choline oxidoreductase (*Gdia_1160*) (46). Further, complete sequencing uncovered that the production of 2,5-di-ketogluconate may be performed by a putative gluconate 2-dehydrogenase acceptor (*Gdia_1159*). Like the genes coding to produce 2-ketoglutonic acid, these genes demonstrated no fitness loss under GAD conditions (Figure 15), nor under BNF (Figure 16). Together, the lack of essentiality of these genes suggests that this pathway is largely inconsequential in glucose metabolism, or at least under these specific conditions. However, it is additionally important to note that the GAD medium was an undefined rich medium also containing yeast extract, tryptone and glutamic acid. Due to the presence of carbon sources other than glucose in this medium, these sugars could have additionally been metabolized under these conditions. Alternatively, the presence of these additional carbon sources could have allowed *G. diazotrophicus* to bypass these steps in glucose metabolism.

Sucrose catabolism: levansucrase

G. diazotrophicus grows well with high sucrose concentration. Growth of this organism was previously indicated under conditions of up to 30% sucrose, while 10% sucrose was found to serve as ideal growth conditions (12). As sugarcane, a native host of *G. diazotrophicus*, contains an abundance of sucrose (106), it should come as no surprise that *G. diazotrophicus* has adapted to grow under high sucrose conditions. However, *G.*

diazotrophicus is unable to transport nor directly metabolize sucrose (28). Instead, it relies upon the action of the levansucrase *lsdA* (sucrose-2,6-β-D-fructan 6-β-D-fructosyltransferase; *Gdia_1531*) (116, 117). This levansucrase converts the sucrose to beta-1,2-ligofructans and levan (13, 28, 46, 118), sugars that *G. diazotrophicus* is then able to metabolize. Levansucrase is largely thought to be critical for the growth of this organism on sucrose (13). One early study on levansucrase found mutants in four strains of *G. diazotrophicus* with inactivated *lsdA* via insertion lost LsdA protein activity and were also unable to grow in medium with sucrose as the sole source of carbon (117). More recently, the construction of a *lsdA*-defective mutant likewise by insertion in *G. diazotrophicus* Pal 5 R, a rifampicin resistant strain, indicated an inability to grow in media containing sucrose as the sole carbon source (118). Given these previous findings, we would expect that a functioning *lsdA* is critical for growth of *G. diazotrophicus* on sucrose. However, here, when grown with sucrose as the sole carbon source, *lsdA* did not exhibit a fitness differential when compared to growth on rich medium containing glucose (Figure 17). This gene appeared to have a high tolerance for transposon insertions throughout the gene (Figure 18) and further did not demonstrate conditional essentiality under any conditions of sucrose as the only carbon source. It appears that mutants within the library with *lsdA* silenced were capable of growth on sucrose-only medium. These results indicate the possibility of other undisclosed protein homologs in the genome carrying out the same process as levansucrase, or the utilization of alternative nutritional components in the medium for energy. It is important to note that function of levansucrase has been more extensively studied in *G. diazotrophicus* strain SRT4, and the exact strain PA1 5 studied here was not among those previously described with *lsdA* mutants. The lack of conditional essentiality of *lsdA* may be unique to this strain.

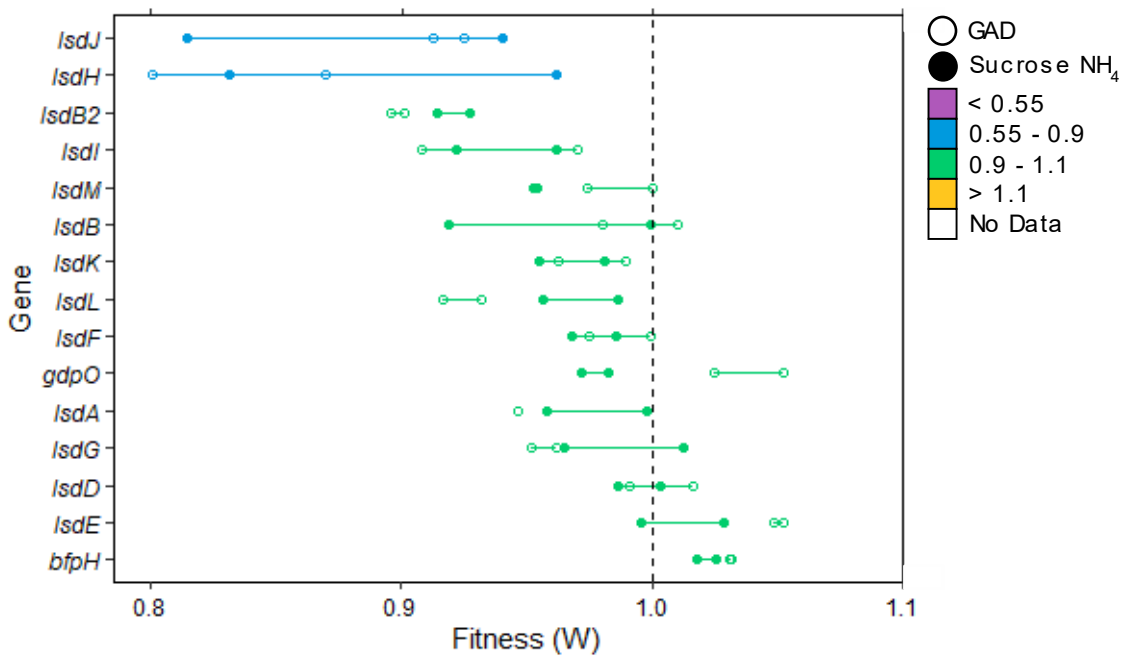

Figure 17. Select genes associated with levansucrase activity. Replicates shown per condition. Closed circles shown are growth with sucrose ammonium while open circles are growth on GAD. Genes color coded as denoted by the average sucrose ammonium fitness value and presented in ascending order.

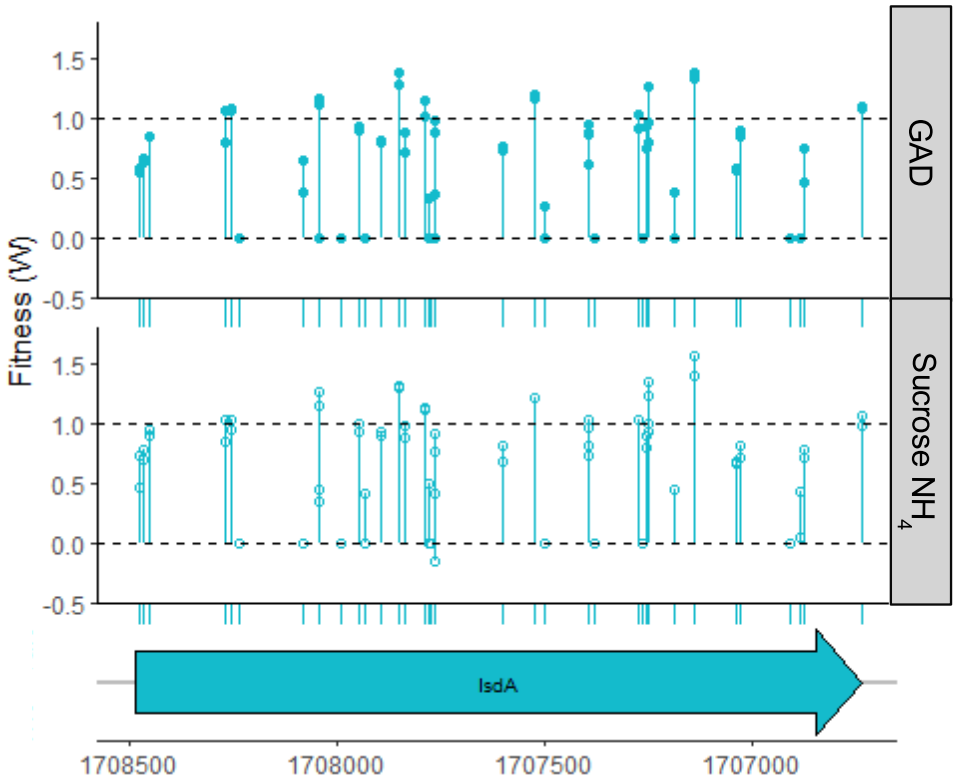

Figure 18. Fitness of each TA site in levansucrase. *lsdA* (*Gdia_1531*) shown for growth on GAD medium and sucrose ammonium. Two replicates shown per condition.

Levansucrase in *G. diazotrophicus* is thought to require the activity of several other proteins to break down sugars. These genes are all located adjacent to *lsdA* in the genome. Levansucrase is secreted to the periplasm of the cell through the action of a type II protein secretion pathway encoded by various *lsd* genes in one gene cluster (*Gida_1517-Gdia_1528*) (119). Through this transport process, the secretion of levansucrase accounts for more 70% of all proteins secreted in *G. diazotrophicus* (116). However, these genes associated with type II secretion did not display a fitness differential or fitness defects upon transposon insertion under sucrose-only growth (Figure 17). This supports the idea presented above that an alternative protein is carrying out this process, or even that *lsdA* is not required in this strain of *G. diazotrophicus* for sucrose breakdown. Additionally within this operon is the exolevanase *lsdB* (*Gdia_1530*), a protein that converts polyfructans outside of the cell into the metabolically useful fructose (46). Similarly, *lsdB* did not display a fitness defect when grown with sucrose as

the sole carbon source. However, *G. diazotrophicus* has a levanase homolog, *lsdB2* (*Gdia_2613*). The lack of fitness differential for this gene could be indicative of functional redundancy. However, these results are overall unexpected, as the genes associated with levansucrase would be thought to be similarly critical for the growth and survival of the bacterium.

Sucrose catabolism: other genes

Other genes beyond levansucrase are involved in the degradation of sugars before their entrance into central carbon metabolism. In the Kyoto Encyclopedia of Genes and Genomes (KEGG), 17 additional genes are noted to be present in *G. diazotrophicus* in the starch and sucrose metabolic pathways. Their exact functions are varied, but all involve complex sugar catabolism. When examining the fitness of these genes, few exhibited a fitness differential when comparing growth on sucrose ammonium medium and GAD medium (Figure 19). Other than the 1,4-alpha glucan branching enzyme encoded by *glgA* (*Gdia_1464*), which showed a moderate conditional fitness loss, few other genes displayed a significant fitness differential. As few genes displayed redundancy, this was surprising. However, the fitness values of several genes are not presented here due to high variability between replicates. These genes may be of greater essentiality for starch and sugar metabolic breakdown, but we cannot reliably report those values here. Of note, *Gdia_3234* is not displayed as it was found to be an essential gene in the t_1 library. Due to this gene essentiality, the conditional essentiality of *glgA* and the high variability, aspects of these pathways can be considered essential to sucrose catabolism.

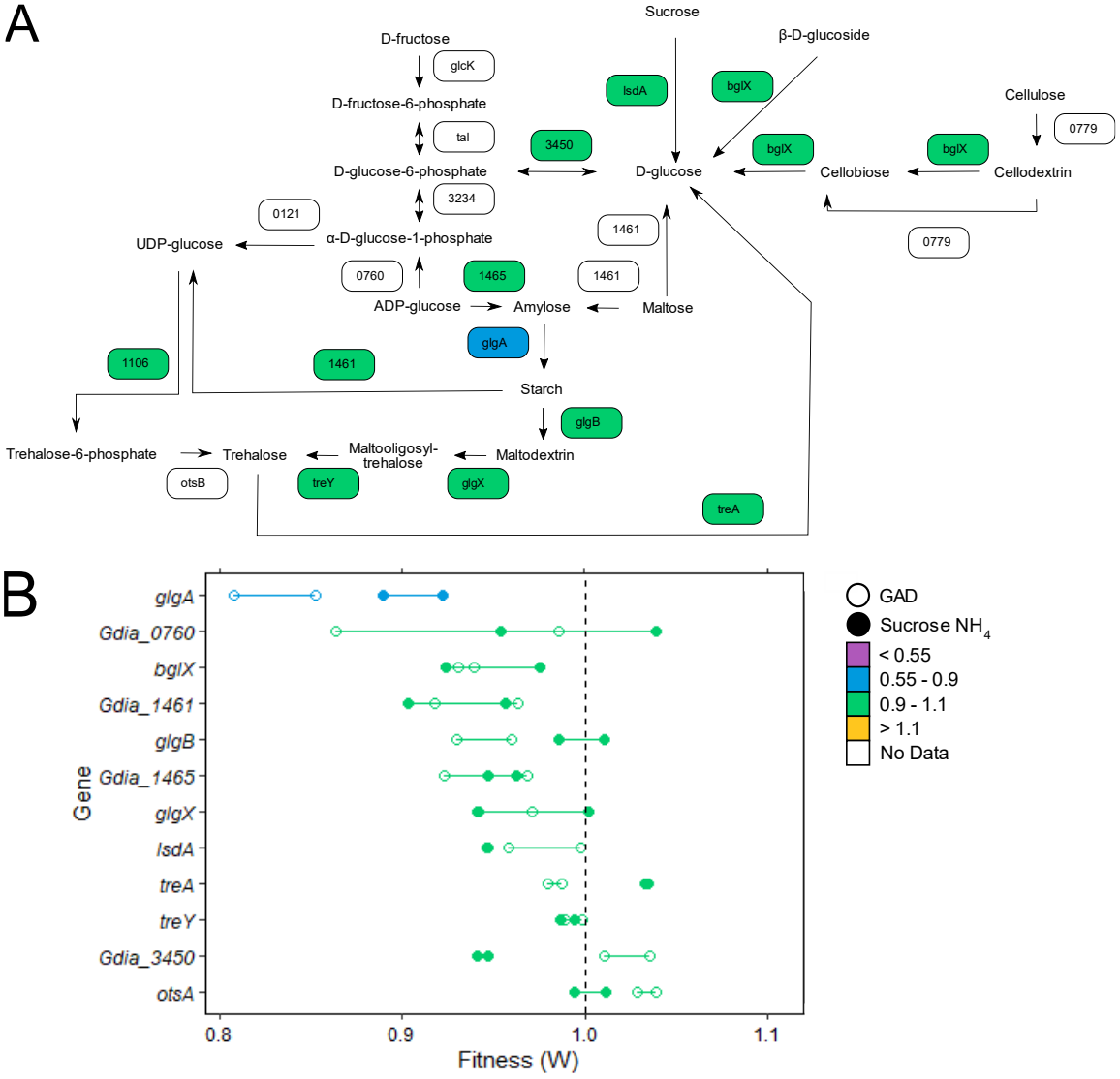

Figure 19. Genes associated with sucrose catabolism. **A**) Proposed sucrose metabolic pathway. Genes of the proteins that catalyze the reactions of the pathway are shown in boxes. Gene color coded as denoted by the average sucrose ammonium fitness value. Adapted from KEGG database. **B**) Gene fitness of sucrose pathway. Closed circles shown are with sucrose ammonium while open circles are with GAD. Genes color coded as denoted by the average sucrose ammonium fitness value and presented in ascending order.

We additionally explored genes that were unexpected to be essential for growth on sucrose. Through examining genes displaying fitness differentials for all three sucrose-only growth conditions compared with growth on GAD, we present a select list of genes that may otherwise be essential for growth on sucrose (Figure 20). The gene

with highest level of conditional essentiality for growth on any sucrose media was the large subunit of acetolactate synthase (*Gdia_1779*). This protein is involved in L-isoleucine and L-valine biosynthesis and catalyzes the reaction of pyruvate into either α-acetolactate or α-ketobutyrate. Other genes with high conditional essentiality on all sucrose-containing media included an ATP phosphoribosyltransferase (*Gdia_2141*), an GTP diphosphokinase (*Gdia_0376*), and a 3-isopropylmalate dehydratase (*Gdia_3280*). However, as these were additionally required for growth on galactose as the sole carbon source, we can gather that these genes are required for the biosynthesis of amino acids that were otherwise provided by the undefined GAD medium.

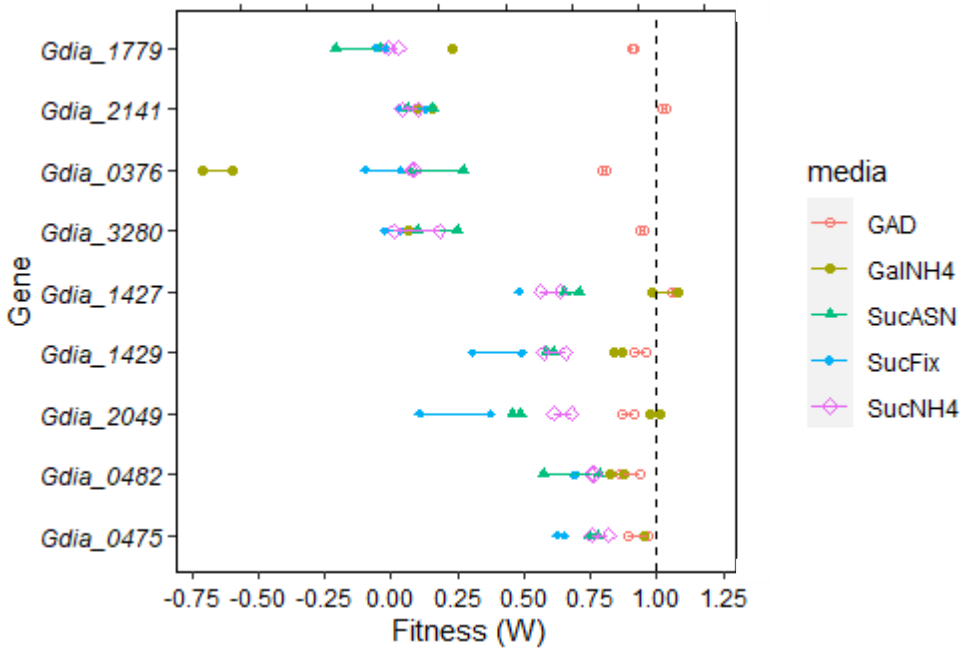

Figure 20. Select genes related to sucrose metabolism. Replicates shown per condition. Five different media conditions shown as indicated. Abbreviations: GalNH4, galactose ammonium; SucASN, sucrose asparagine; SucFix, sucrose without reduced nitrogen; SucNH4, sucrose ammonium.

With this in mind, we further examined genes displaying fitness differentials for the sucrose-only conditions when compared to both GAD and galactose-only media. The gene coding for a nicotinate-nucleotide pyrophosphorylase (*Gdia_1427*) and the A subunit of the quinolinate synthase complex (*Gdia_1429*) displayed low fitness values for the sucrose-only media. These two genes along with *Gdia_1428* are involved in the NAD

biosynthetic pathway from aspartate to nicotinate D-ribonucleotide. Other genes potentially involved in sucrose catabolism included a gene coding for an MFS drug resistance transporter in the Bcr subfamily (*Gdia_2049*) and a host factor-I protein (*Gdia_0482*) involved in quorum sensing and RNA degradation. We also noted the fitness of the protein adenylosuccinate synthase (*Gdia_0475*), which catalyzes the reaction of inosinic acid to adenylosuccinate. These gene all displayed moderate-to-significant fitness differentials under sucrose-only conditions compared to both galactose-only and GAD media. These genes appear to be related to sucrose catabolism in *G. diazotrophicus*.

Other sugars: galactose catabolism

Beyond sucrose and glucose, *G. diazotrophicus* can grow upon a wide variety of additional carbon sources. Among these substrates, robust growth has previously been demonstrated for D-galactose as the sole carbon source (10, 12, 107). Given this additional metabolic capability, we challenged our mutant library to growth on galactose to further explore the carbon metabolism of *G. diazotrophicus*. Unlike many other organisms that convert galactose to glucose via the well-studied Leloir pathway (120), *G. diazotrophicus* may instead utilize the alternative DeLey-Doudoroff pathway. The LeLoir pathway more directly converts D-galactose to D-glucose-6-P, and in the KEGG database, *G. diazotrophicus* appears to lack several genes required for the Leloir pathway. However, the DeLey-Doudoroff pathway appears largely intact. In this pathway, D-galactose is first converted to D-galactonate before undergoing further processing to pyruvate and D-glyceraldehyde-3-phosphate (121). Other diazotrophs such as *A. vinelandii* also rely upon the DeLey-Doudoroff pathway (121). However, the metabolism of other sugars such as galactose in *G. diazotrophicus* has not been extensively studied.

When examining the transposon library when grown with galactose as the sole carbon source, the three key genes along the DeLey-Doudoroff pathway stood out as having the highest fitness differential as compared to growth with sucrose as the sole carbon source (Figure 21). The 2-dehydro-3- deoxyphosphogalactonate aldolase encoded by *dgoK* (*Gdia_0391*), galactonate dehydratase *dgoD* (*Gdia_0390*), and 2-dehydro-3-

deoxygalactonokinase *dgoA* (*Gdia_0392*) displayed large fitness deficits, indicating their conditional essentiality. These genes were contained within a gene cluster additionally containing genes coding for a galactose-hydrogen symporter (*Gdia_0388*), D-xylose 1-dehydrogenase (*Gdia_0389*), and a transcriptional regulator *dgoR* (*Gdia_0393*). While *Gdia_0388* displayed no fitness differential when grown under galactose as compared to sucrose, both latter genes displayed a moderate fitness defect.

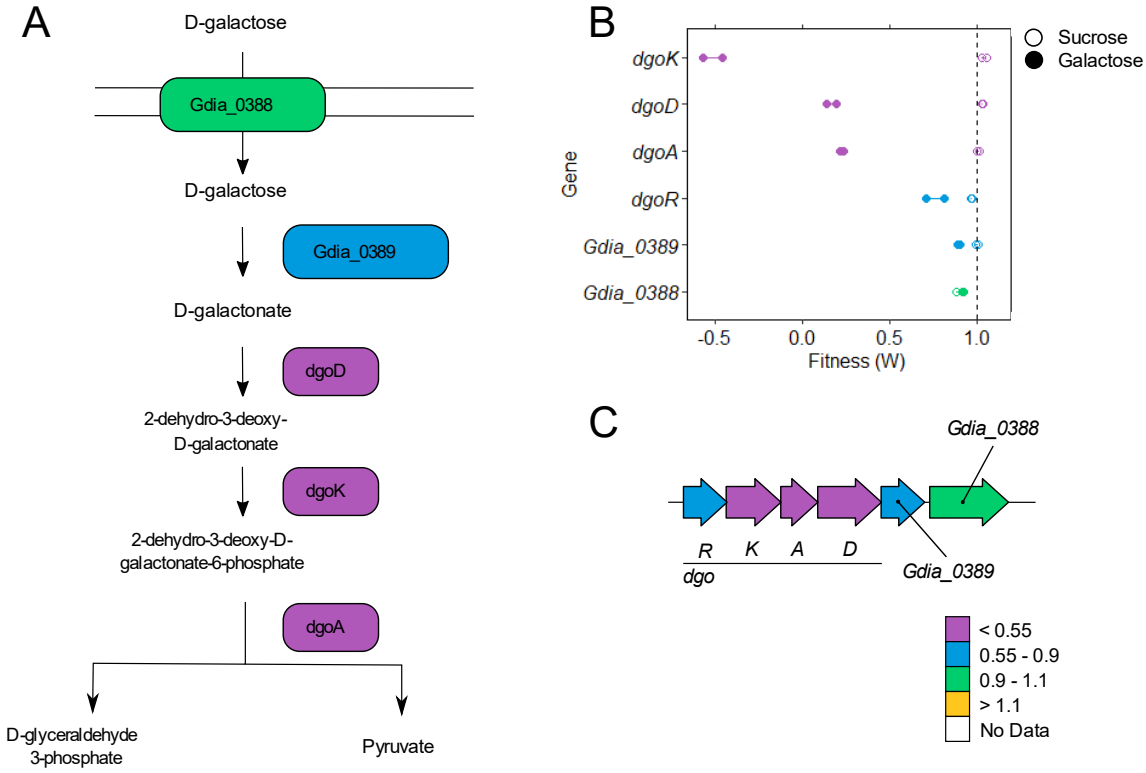

Figure 21. Select genes associated with galactose catabolism. **A**) Proposed galactose catabolic pathway. Genes of the proteins that catalyze the reactions of the pathway are shown in boxes. Genes color coded as denoted by the average galactose fitness value. Follows the DeLey-Doudoroff pathway, with Gdia_0389 serving as the D-galactose dehydrogenase and Gdia_0388 as the D-galactose transporter. **B**) Gene fitness of galactonate pathway. Closed circles shown are with galactose while open circles are with sucrose. Genes color coded as denoted by the average galactose fitness value and presented in ascending order. **C**) Gene schematic of galactose gene cluster shown in B.

While these genes make up the DeLey-Doudoroff pathway, this metabolic pathway has not fully been described in *G. diazotrophicus*. The three key genes identified

above—*dgoK, dgoD,* and *dgoA*—are related to the metabolism of D-galactonate, as opposed to the sole carbon source provided in this context, D-galactose, leaving the question of the conversion of D-galactose to D-galactonate unanswered. We thus tentatively propose that the role of this missing D-galactose dehydrogenase in this pathway is filled by the dehydrogenase already present in this gene cluster, D-xylose 1-dehydrogenase (*Gdia_0389*). Due to similar function, its proximity to the other *dgo* genes, as well as its moderate fitness differential, we put forward that this dehydrogenase has additional galactose utilization capability. This is not the first such instance of incompletely described galactose metabolism. A prior study on differential carbohydrate utilization in a microbial mat found the bacterium *Halomonas* spp. to be lacking in a galactose dehydrogenase within the DeLey-Doudoroff pathway, and it was similarly proposed that the D-xylose dehydrogenase had specificity for multiple substrates, such as galactose and xylose (122). Further characterization of this gene in *G. diazotrophicus* and the reaction its protein product catalyzes is necessary to elucidate if this dual specificity is the case.

An intriguing aspect of Tn-seq studies are the potential to reveal genes that were otherwise unexpected to be essential for growth under different conditions. A few further genes potentially related to the growth of *G. diazotrophicus* on galactose as the main carbon source displayed significant fitness differentials when compared to growth on sucrose (Figure 22). Notably, the gene encoding for fructose-bisphosphate aldolase (*Gdia_1516*) displayed a large fitness differential. This protein is involved in central carbon metabolism and catalyzes the reversible reaction of fructose 1,6-bisphosphate into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. As one of the key products of the DeLey-Doudoroff pathway has been indicated to be glyceraldehyde 3-phosphate (Figure 21), the conditional essentiality of this gene here indicates that its action is critical for the incorporation of galactose. While not in close proximity in the genome to the other genes involved in galactose catabolism, given the large fitness differential exhibited by *Gdia_1516*, it is plausible that this gene encodes the next step in galactose catabolism after *dgoA*.

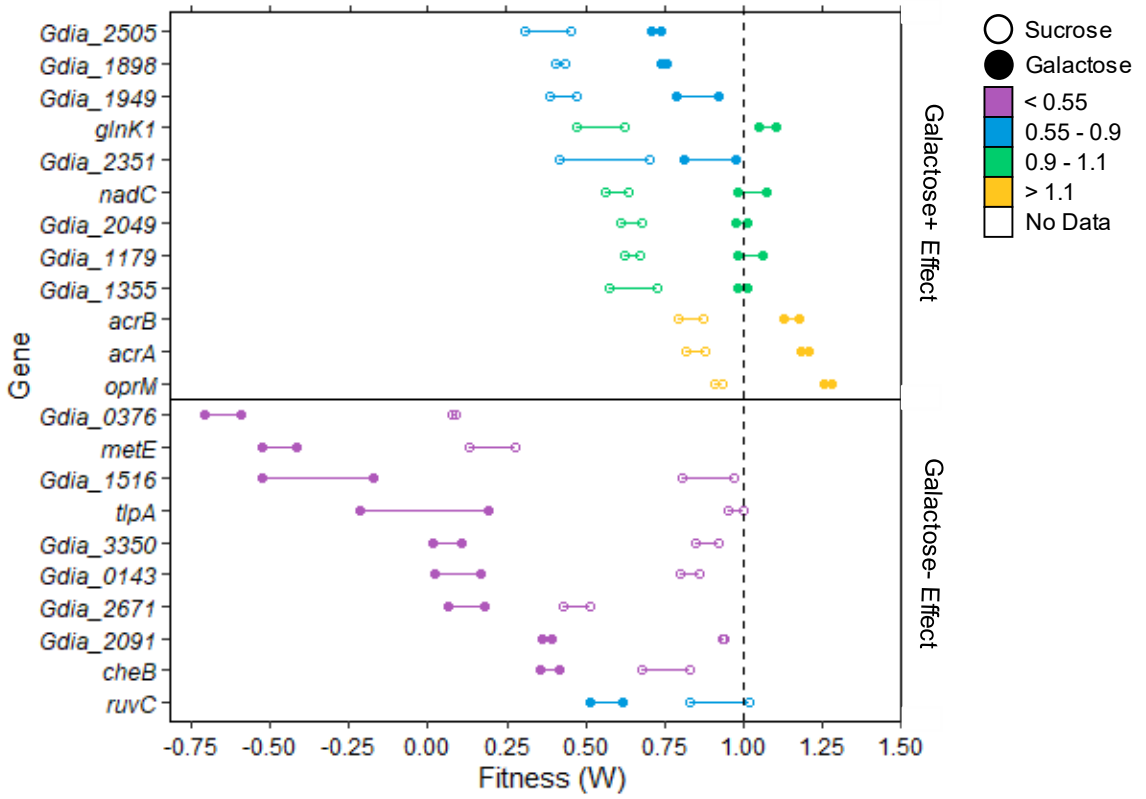

Figure 22. Top genes with the largest fitness differential between samples grown with galactose and sucrose as the only carbon source. Genes divided into two categories: those grown with galactose with higher fitness (top) and those with lower fitness (bottom). Closed circles shown are with galactose while open circles are with sucrose. Genes color coded as denoted by the average galactose fitness value and presented in ascending order.

Other genes were indicated to be essential for growth on galactose due to large fitness differentials. One of these genes was *tlpA* (*Gdia_3077*), encoding a methyl accepting chemotaxis protein. Chemotaxis proteins are largely involved in environmental and sensory signals to alter the activity of the histidine kinase CheA (123). The sole presence of galactose could result in greater chemotaxis activity than other more desirable sugars such as glucose. Further genes with significant fitness detriments included *Gdia_3350*, encoding the A subunit of DNA topoisomerase IV, and a pimeloyl-acyl carrier protein methyl ester esterase (*Gdia_0143*). The reason behind the large fitness deficits of these genes under these conditions are yet unknown.

Additionally, several genes displayed a significant relative fitness gain upon growth on galactose. The fitness of a small number of these genes indicated that their

silencing led to increased mutant growth and that their absence is beneficial. Of note, three outer membrane proteins demonstrated fitness gain upon their deletion. These genes are in an operon encoded by *acrA* (*Gdia_1994*), *acrB* (*Gdia_1995*) and *oprM* (*Gdia-_1996*) and are all related to a multidrug efflux system. Efflux pumps are thought to be involved in antibiotic resistance, but have additionally been suggested to be involved in biofilm formation (124). The efflux pump system involving *acrAB*—the RND-type efflux pumps—transport substrates from within the cell to outside, and play a larger role in biofilm formation than other similar systems (124). As the Tn-seq library here was exposed to the same antibiotic concentration regardless of media, it seems probable that these genes are involved in biofilm formation as opposed to antibiotic resistance in *G. diazotrophicus*. However, as the fitness data indicates that these genes are less necessary in *G. diazotrophicus* when grown in the presence of galactose, is appears that the presence of the other biofilm-related genes in this organism are sufficient for growth on galactose.

Asparagine metabolism

As a plant endophyte, the growth of *G. diazotrophicus* on various nitrogen sources has been previously studied. It was found that this microbe is capable of robust growth sustained only by amino acid-provided nitrogen such as via glutamine or asparagine (108). However, this growth is complicated due to its diazotrophic characteristics. Nitrogenase activity in *G. diazotrophicus* is significantly diminished in the presence of free asparagine (108, 125). To further explore the metabolism of other nutrients in G. diazotrophicus, we subjected the libraries to growth on asparagine as the sole nitrogen source, with sucrose as the carbon source. While we did not study this growth additionally under nitrogen fixation conditions, understanding the metabolism and accumulation of asparagine in this diazotroph remains of interest. Asparagine is a commonly understood amino acid that takes on the role of nitrogen storage in plant tissues (126). In sugarcane specifically, a native host of *G. diazotrophicus*, asparagine is abundant (126, 127). It follows that this bacterium may then have metabolic methods for the breakdown of this amino acid.

Perhaps surprisingly, here few protein products displayed significant fitness differentials. Of note, the putative L-asparaginase II protein (*Gdia_1961*) had a high fitness differential. Asparaginase catalyzes the reaction of L-asparagine to L-aspartate, using 1 mol H_2O and producing 1 mol NH_3 as a byproduct (Figure 23). Upon completion of this reaction, the resulting amino acid is free to enter the TCA cycle and the ammonia is assimilated into the cell. This gene has previously been suggested, along with various genes coding for putative aspartate aminotransferases, to represent a potential evolutionary advantage to growth in these amino-acid rich endophytic conditions (126). One such aspartate aminotransferase (*Gdia_3243*) did display a significant fitness differential when grown with asparagine as the nitrogen source. While these two genes were flagged for high variability between the two replicates grown with asparagine, it is still worthwhile to report here due to this potential endophytic lifestyle adaptation.

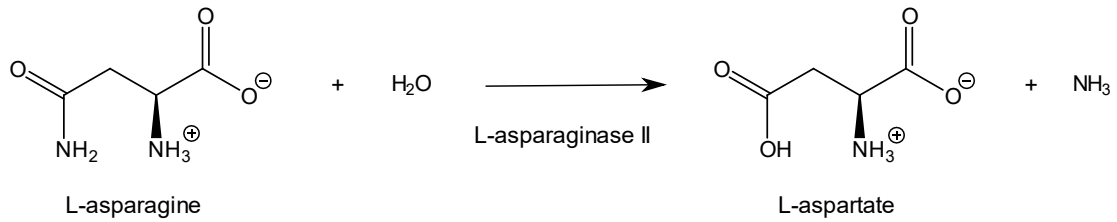

Figure 23. Reaction carried out by L-asparaginase II.

A few further genes potentially related to the breakdown of asparagine by *G. diazotrophicus* displayed significant fitness differentials when compared to ammonium-provided nitrogen (Figure 24). Of those with significant fitness defects, the protein products of two genes were involved in carbon metabolism: prephenate dehydrogenase *pheA (Gdia_0102*), converting between prephenate and phenyl-pyruvate in the shikimate pathway, and malate dehydrogenase (*Gdia_1833*), which catalyzes the oxidation of malate to oxaloacetate. As both conditions were provided the same carbon source, it follows that these genes would be involved in the metabolism related to asparagine. Regarding other genes that displayed large fitness defects, such as the genes for the bifunctional enzyme sulfate adenylyltransferase and adenylyl-sulfate kinase (*Gdia_3126*) and the ubiquinol oxidase (*Gdia_0259*), the reason for their conditional essentiality is yet

unknown. Under these conditions, few genes displayed a significant relative fitness gain, or where the silencing of the gene led to increased mutant growth.

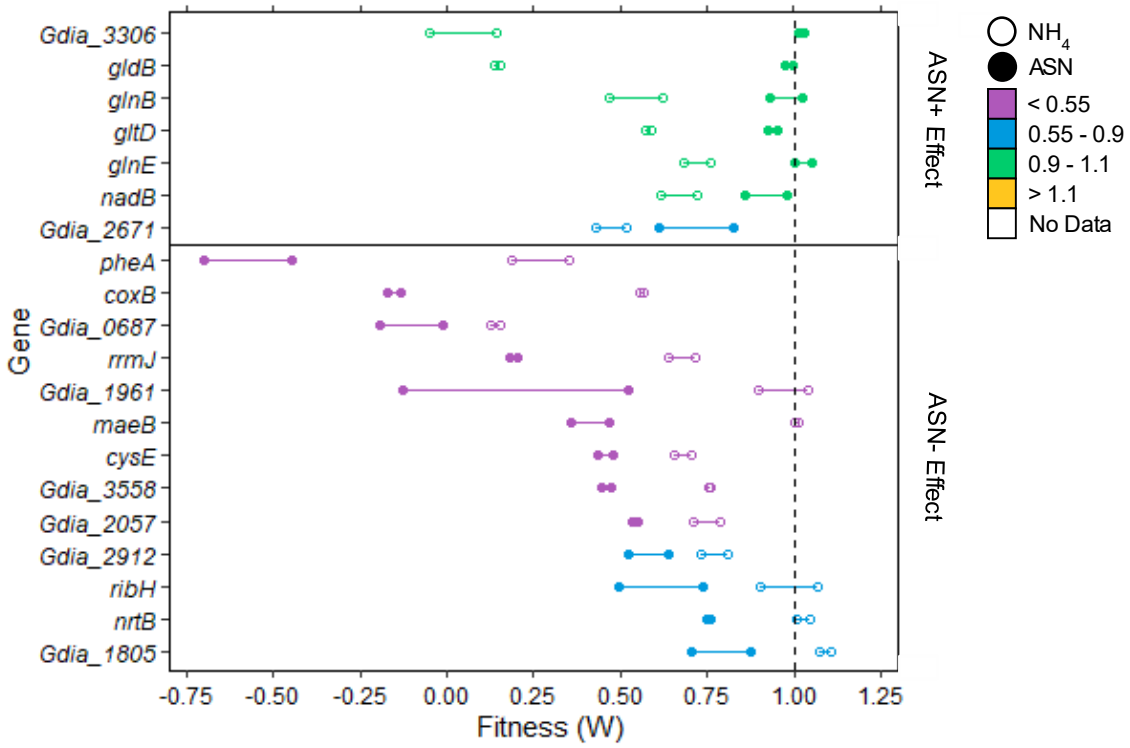

Figure 24. Top genes with the largest fitness differential between samples grown with ammonium and asparagine (ASN) as the only nitrogen source. All samples grown on sucrose as the sole carbon source. Genes divided into two categories: those grown with asparagine with higher fitness (top) and those with lower fitness (bottom). Closed circles shown are with asparagine while open circles are with ammonium. Genes color coded as denoted by the average asparagine fitness value and presented in ascending order.

Conclusion

This work expands upon the results of Chapter 2 to provide insight into genes critical for carbon and nitrogen metabolism in *G. diazotrophicus*. Our data suggests *G. diazotrophicus* has varied and fascinating metabolic capabilities. This diversity in catabolism of different carbon and nitrogen sources for energy suggests the occurrence of endophytic lifestyle adaptations in this strain. Our data also informs potential future characterization of the metabolism of this strain. Several genes previously indicated to be essential for sugar catabolism, such as levansucrase, demonstrated nonessentially,

suggesting possible functional redundancy with other yet-to-be-determined genes. Further, this work suggests new complete pathways of previously regarded metabolic routes, such as galactonate catabolism. This work allows for a renewed appreciation of the catabolic capabilities of this organism. Simultaneously, our data serves as a basis for the future application of this work in understanding its metabolism in novel endophytic environments. Understanding of the genes essential for nutrient breakdown permits us to look towards future studies with a greater holistic understanding of *G. diazotrophicus*.

Chapter 4: Optimizing Ammonium Production in *G. diazotrophicus* Through Genetic Manipulation

Introduction

Sufficient soil nutrients are required to stimulate robust plant growth. However, as previously discussed (Chapter 1), bioavailable or reduced nitrogen is a key limiting nutrient for plant growth. Given that reduced nitrogen is not readily available in soil, plants must acquire this form of nitrogen through alternative means. Often, this comes in the form of synthetic nitrogen. Approximately half of the total nitrogen input to agriculture is produced through the Haber-Bosch process (8). Through burning fossil fuels to generate N_2 gas and the necessary temperatures to carry out the reaction, this process is highly energy-intensive (8). Nonetheless, great amounts of Haber-Bosch derived nitrogen are still applied to agricultural fields to increase crop yields; in some instances, more than 200 kg N $ha^{-1}yr^{-1}$ (6). Only slightly over half of applied nitrogen-containing fertilizer is actually used by plants (128, 129), and the rest is lost to the environment through surface runoff or leaching. This contributes to the low nitrogen use efficiency of many agricultural crops (130). The combination of both fossil fuel use and the harmful environmental consequences of groundwater contamination and eutrophication brings into question the sustainability of synthetic fertilizer use. There is a need to reduce our reliance upon synthetic fertilization and look towards utilizing BNF to meet the world's nitrogen demands.

Depending upon the diazotroph, BNF provides some nitrogen to agricultural crops. For instance, legumes engage in root nodule symbiosis with diazotrophs like rhizobia, and these microbes provide upwards of 50 kg N $ha^{-1}yr^{-1}$ to their host plant (6). However, other crops are not so lucky. Around 60% of synthetic nitrogen fertilizers are used in growth of cereal crops such as wheat and corn (128), as these crops do not form these symbiotic relationships with rhizobia. Diazotrophs still provide fixed nitrogen through endosymbiotic or associative means to cereal crops, though the amount is not sufficient to withhold supplementation with synthetic fertilizer. It has therefore been a long-standing dream of the agronomic community to improve BNF in non-leguminous crops (6, 131).

Nitrogen-fixing plant endophytes provide a unique opportunity for study as biofertilizers due to their wide-ranging PGP characteristics. While other biofertilizers,

typically nutrient-providing microbial inoculants (132), must compete with soil bacteria, endophytes such as *G. diazotrophicus* do not have this challenge. Present in large numbers in the intercellular spaces of the host plant (12), *G. diazotrophicus* has been clearly established as an endophyte. Due to its demonstrated characteristics of plant growth and yield benefits when present in a host plant, as well as its host versatility as an endophyte (see Chapter 1), we explored the potential of *G. diazotrophicus* to serve as a biofertilizer-like strain through performing novel genetic manipulations within the strain.

Manipulations within the Q-linker region (Chapter 1) of *nifA* have been carried out previously. Deletions or insertions in strains such as *Bradyrhiobium japonicum* or *K. pneumoniae* did not affect the activation function of the modified NifA protein (45). Manipulations of *nifA* in these bacterial species did not affect the activation of NifA. This same modification within the Q-linker has resulted in constitutive nitrogenase gene expression in other species. In the purple non-sulfur bacterium *Rhodopseudomonas palustris*, previously described *nifA** mutants were non-stable single point mutations in the Q-linker region that resulted in an active conformation of NifA and constitutively expressed nitrogenase even in the presence of ammonium (133). Further work was carried out in *R. palustris* to circumvent potential reversion to the wild type phenotype through the deletion of 48 basepairs within the Q-linker region of *nifA* (134). This deletion resulted in the desired stable *nifA** phenotype (134). Likewise, a deletion of the similar amino acid sequence within the *Rhodobacter sphaeroides* has since been performed to similar results (135). Here, we carried out similar genetic manipulations in *G. diazotrophicus* to build a *nifA** mutant that would constitutively express nitrogenase (Figure 25).

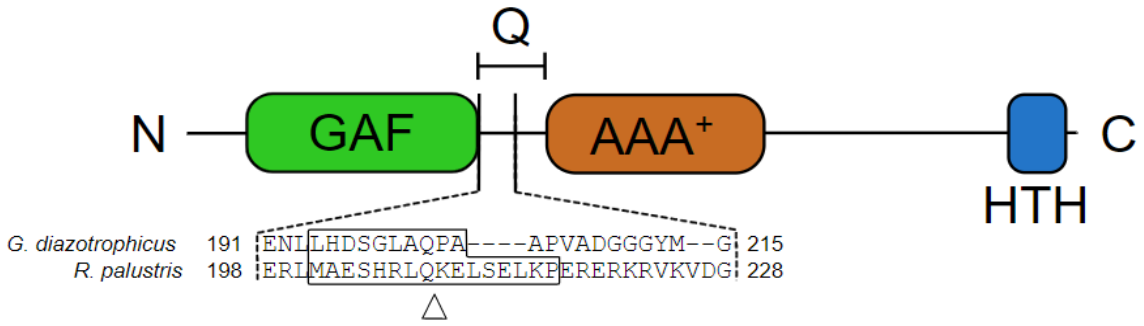

Figure 25. Genetic manipulations in the *nifA* of *G. diazotrophicus* to produce the *nifA** mutant as was done in *R. palustris*. Indicated are the N- and C-termini of the protein, the locations of the GAF, AAA+ and HTH domains, and the Q-linker interdomain region. The amino acids of the Q-linker for *G. diazotrophicus* and *R. palustris* are indicated. The 10-aa deleted Q-linker region in *G. diazotrophicus* to construct the *nifA** mutant is shown (Δ) with the corresponding 16-aa deleted in *R. palustris*.

Other genetic manipulations within diazotrophs have been previously found to increase extracellular nitrogen. Like *nifA*, previous studies have been performed regarding the deletion of genes coding for the ammonium transporter in diazotrophs. Prior work in *A. vinelandii*, a free-living diazotroph, disrupted *amtB*, the gene coding for the ammonium transporter (136). This resulted in the slow release of low levels of ammonium into the media (136). This modified strain also supported the growth of a coculture of the nitrogen-requiring green alga *Chlorella sorokiniana* in nitrogen-free medium (136). Similarly, deletion of the dual homologs *amtB1* and *amtB1* in the nitrogen-fixing *Pseudomonas stutzeri* also resulted in ammonium excretion under nitrogen fixation conditions (137). Like *P. stutzeri*, *G. diazotrophicus* has two genes coding for the ammonium transporter, *amtB1* and *amtB2* (Chapter 1). In this work, we constructed a similar mutant strain of *G. diazotrophicus* lacking both *amtB* homologs (Figure 26).

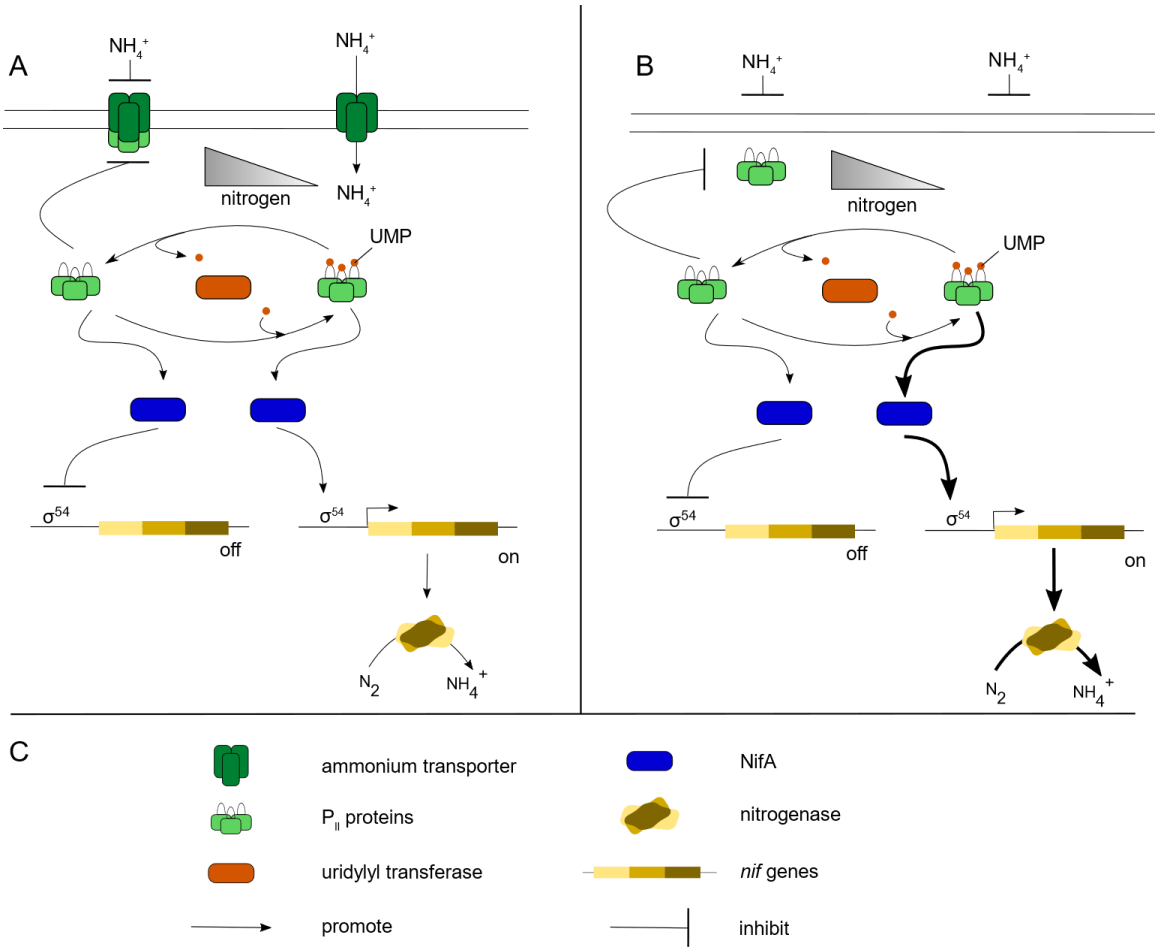

Figure 26. *nif* regulon. **A)** Regulation of nitrogen fixation in *G. diazotrophicus*, as previously described (Figure 2). **B**) Proposed regulation of nitrogen fixation in a *G. diazotrophicus* strain lacking the dual *amtB* homologs. Bold lines indicate proposed increase of nitrogen fixation expression in mutant. **C**) Figure legend indicating icons used.

Here, we performed novel genetic manipulations in *G. diazotrophicus* with the intent to increase extracellular nitrogen production. These manipulations include the deletion of the two genes coding for the ammonium transporter, *amtB1* and *amtB2*, and a deletion within the interdomain Q-linker region of *nifA*, resulting in the *nifA** mutant. In prior studies in other diazotrophs, these manipulations resulted in increased extracellular nitrogen production, a result we looked to replicate in *G. diazotrophicus*. As this bacterium is already an important endophyte that provides fixed nitrogen to its host plants, a mutant strain of *G. diazotrophicus* that expels greater amounts of extracellular nitrogen would have

the potential to serve as a better biofertilizer. Providing a greater share of the nitrogen requirements of the host plant, application of this new strain would reduce the amount of synthetic fertilizer required for prosperous crop growth. A potential new biofertilizer strain such as one from *G. diazotrophicus* would take us one step further in the direction of a more sustainable agronomic future.

Methods

Bacterial strains and growth conditions

G. diazotrophicus PA1 5 (ATCC 49037) was obtained from Cedric Owens and was grown aerobically at 30ºC (12) on GADN medium unless otherwise specified. GADN was adapted from DYGS basal medium (35) (g L^{-1}: glucose, 2; yeast extract, 2; tryptone, 1.5; $MgSO_4$, 0.5; glutamic acid, 1.5; adjusted to pH 6.2 with NaOH before sterilization). When noted, GADN was supplemented with antibiotics at the following concentrations: 100 mg/L tetracycline, 150 mg/L chloramphenicol. *G. diazotrophicus* was also grown on GADB(N-)2 solid media, adapted from Burk's medium (72) (50 mM citrate, g L^{-1}: sucrose, 20; $MgSO_4$, 0.2; $CaCl_2$, 0.02; Na_2MoO_4, 0.005; $FeSO_4$, 0.01g; 2 mL 100X phosphate buffer composed of 20 g KH_2PO_4 and 80 g K_2HPO_4 in 1 L dH2O; adjusted to pH 6.2 with NaOH before sterilization), and grown under nitrogen fixation conditions at 28.9ºC in an controlled atmosphere of 0.030 L/min nitrogen within a sealed chamber. *E. coli* JM109 was acquired from New England Biolabs (Ipswich, MA) and used to construct the plasmids for *G. diazotrophicus*.

Genetic constructs

GABB004 and GABB027 were the targeted strains of *G. diazotrophicus* for construction, with GABB026 as the biosensor strain. These strains, as well as other related strains used for their assembly, were constructed as described (Table 4) using the different plasmids detailed (Table 5). Primers used for the construction of these genetic constructs as well as their confirmation are also described (Table 6).

All plasmids were transformed to *G. diazotrophicus* through electroporation, a method that has been previously described (47) and was adapted for use here. In brief, following two days of growth of *G. diazotrophicus* on GADN, ~50 µL of cells were inoculated into 50 mL GADN in a 125 mL base/acid washed Erlenmeyer flask. The

culture of *G. diazotrophicus* was grown overnight at 30ºC with 180 rpm shaking. The following morning, 1 mL of the overnight culture was transferred into fresh 50 mL GADN medium and grown for several hours or until the culture reached a density indicative of exponential growth. Then, 8 mL of the 50 mL culture was pelleted at 3,500 x *g* for 1.5 minutes. The supernatant was removed, and the resulting cell pellet was resuspended with 1 mL filter-sterilized 10% glycerol in water (v/v). This was repeated once, washing with 1 mL 10% glycerol before pelleting again. The supernatant was removed, and the pellet was washed through resuspending twice more with 500 µL 10% glycerol. Following the final wash, the pellet was resuspended in 150 µL 10% glycerol in an electroporation cuvette as was 3 µL of the desired plasmid. The cell culture and plasmid were electroporated on a BioRad pulser apparatus at the following conditions: 12.5 kV cm^{-1}, 25 uF capacitance and 600 Ω resistance with a desired 12 ms pulse length. The electroporated cell culture was immediately added to 50 mL GADN medium and incubated overnight at 30ºC with 180 rpm shaking. Following overnight growth, 1 mL of cells were pelleted, resuspended in 100 µL GADN medium and plated onto GADN medium with the corresponding antibiotics. Plates were incubated at 30ºC for several days until formation of colonies.

Coculture of strains with biosensor

To make a qualitative assessment of the potential of strains GABB004 and GABB027 to support the growth of the biosensor GABB026, a minimal amount of cells from each strain (10 µL of each strain mixed) were spotted onto GADB(N-)2 plates and grown under nitrogen fixation conditions within the atmospheric chamber. Strains were first grown overnight in 50 mL GADN medium at overnight at 30ºC with 180 rpm shaking before 1 mL of each culture was pelleted and the supernatant was removed to minimize any supplemented nitrogen present. Equivalent amounts of cells were added from each culture.

Table 4. Key strains used in genetic manipulations.

***G. diazotrophicus* strain**	**Genetic features[a]**	**Plasmid**	**Parent strain**
PA1 5 (ATCC 49037)	Wild-type *G. diazotrophicus*	None	None
GABB002	*amtB1*::Tet^r	pPCRERIN8	PA1 5
GABB003	*amtB2*::Chlor^r	pPCRERIN13	PA1 5

GABB004	*amtB1*::Tet^r; *amtB2*:: Chlor^r	pPCRERIN8	GABB003
GABB011	*Gdia*_2010::*vioABCE* transposon (R6K origin, Kan^r, *vioABCE*)	pBB327	PA1 5
GABB019	*nifA*::Tet^R	pPCRGNIF5	PA1 5
GABB026	*nifHDK*::Tet^R and *Gdia*_2010::*vioABCE* transposon	pPCRGNIF11	GABB011
GABB027	*nifA**: Q-linker deletion	pPCRGNIF3	PA1 5

[a] Tet^r, tetracycline resistance; Chlor^r, chloramphenicol resistance.

Table 5. Key plasmids used in genetic manipulations.

Plasmid	**Relevant cloned gene(s) or manipulated plasmids[a]**	**Parent vector**	**Source or reference**
pBB284	Backbone for *amtb1* genetic manipulations	pBB227	This study
pBB295	pTnMm1 with addition of BamHI sites	pTnMme1	Chapter 2
pBB308	Cloned transposon region from pBB295 via site-specific mutagenesis	pBB295	This study
pBB309	Digested pBB308 with KpnI and ligated	pBB308	This study
pBB327	Inserted *vioABCE* transposon from pBB309 into pBB284; contains Kan^r	pBB284	This study
pBBTET6	pUC19 with Tet^r with BamHI sites	pBBTET2	Chapter 2
pPCRERIN2	Cloned *amtB1* and flanking regions from *G. diazotrophicus* into pBB284	pBB284	This study
pPCRERIN7	Removed *amtB1* from pPCRERIN2, leaving flanking regions	pPCRERIN2	This study
pPCRERIN8	Inserted Tet^r into *amtB1* flanking regions in pPCRERIN7	pPCRERIN7	This study
pPCRERIN5	Cloned *amtB2* and flanking regions from *G. diazotrophicus* into pBB284	pBB284	This study
pPCRERIN10	Removed *amtB2* from pPCRERIN5, leaving flanking regions	pPCRERIN5	This study
pPCRERIN13	Inserted Chlor^r into *amtB2* flanking regions in pPCRERIN10.	pPCRERIN10	This study
pPCRGNIF1	Cloned *nifA* and flanking regions *from G. diazotrophicus* into pBB284	pBB284	This study
pPCRGNIF2	Removed *nifA* from pPCRGNIF1, leaving flanking regions	pPCRGNIF1	This study
pPCRGNIF3	Removed Q-linker region of *nifA* by failsafe vent PCR of pPCRGNIF1	pPCRGNIF1	This study

pPCRGNIF4	Removed Q-linker region of *nifA* by failsafe vent PCR of pPCRGNIF2	pPCRGNIF2	This study
pPCRGNIF5	Inserted Tet^r from pBBTET6 into *nifA* flanking regions in pPCRGNIF2	pPCRGNIF2	This study
pPCRGNIF8	Cloned *nifHDK* from G. *diazotrophicus* into pBB284	pBB284	This study
pPCRGNIF9	Cloned *nifHDK* flanking regions from *G. diazotrophicus* into pBB284	pBB284	This study
pPCRGNIF10	Inserted *nifHDK* from pPCRGNIF8 into flanking regions in pPCRGNIF9	pPCRGNIF9	This study
pPCRGNIF11	Inserted Tet^r from pBBTET6 between *nifHDK* flanking regions in pPCRGNIF10	pPCRGNIF10	This study

[a] Tet^r, tetracycline resistance; Kan^r, kanamycin resistance.

Table 6. Key primers used in genetic manipulations.

Primer	**Sequence (5'-3')**	**Purpose**
BBP 3158	NNNGAATTC GTGCTGCCAGCCCATAAATCAGG	*amtB1* gene and region cloning
BBP 3159	NNNTCTAGA CGACTCGCTGGAATATTTGGGCGACAC	*amtB1* gene and region cloning
BBP 3160	NNNGGATCCGAATTCTGTGCATATG GCGTCTTCTCCCGTCTCGC	*amtB1* gene deletion
BBP 3161	NNNGGATCC CTGATCCATTTCACCGAACAACAGACAGG	*amtB1* gene deletion
BBP 3162	NNNGAATT CTGATGACGACATGGTTGTCGTCATCC	*amtB2* gene and region cloning
BBP 3163	NNNAAGCTT GATCTGGTATTCGGCCTTCGTCGGGTCGATG	*amtB2* gene and region cloning
BBP 3164	NNNGGATCCGAATTCGTTCATATG GCAATCTCCCCGGTTTCATTCGTACGGATAC	*amtB2* gene deletion
BBP 3165	NNNGGATCC TGATCGTTCCGGAAACGGGC	*amtB2* gene deletion
BBP 3184	NNNGAATT CCTTTTTTATTTCACA ATCGATGCAACCAGCGTTTCC	*amtB1* gene region cloning
BBP 3185	NNNAAGCTT CTTCGTCAGCCATGGCAGGTGGGTATC	*amtB1* gene region cloning
BBP 3186	NNNGAATTCTTTAAGGAGGTAAAACATATGA AACATTCTTCCGATATCTGCATTGTTGG	violacein genes from pBB284 cloning
BBP 3187	NNNGAATTCGTAATCATGT CATAGCTGTTTCCTGTGTG	violacein genes from pBB284 cloning
BBP 3188	GATCGTCTCATGGGCACTCAGTGAC	*amtB1* gene deletion confirmation

BBP 3189	GTGTATCAGACCGTTCCGCTGCAG	*amtB1* gene deletion confirmation
BBP 3192	GAATGACGAGACGTACAAGGTCGCCATGAC	*amtB2* gene deletion confirmation
BBP 3193	CGATAGATTTCGCCGCGATAATTGAAGG	*amtB2* gene deletion confirmation
BBP 3269	NNNTCTAGA CTGGTGAGGTGGAGGAGGCAAGTAG	*nifA* gene and flanking region cloning
BBP 3270	GGTCAGCGACGTCCGGTTGGTTTCG	*nifA* gene and flanking region cloning
BBP 3271	CAGGTTTTCCCGGTCCCGCTGCAC	*nifA* Q-linker genetic region deletion
BBP 3272	GCGCCGGTTGCCGATGGCGG	*nifA* Q-linker genetic region deletion
BBP 3273	NNNGGATCC CAGGTTTTCCCGGTCCCGCTGCAC	*nifA* Q-linker genetic region deletion
BBP 3274	NNNGGATCC GCGCCGGTTGCCGATGGCGG	*nifA* Q-linker genetic region deletion
BBP 3329	GTGTGCGAGAACATCCCGATCGTC	*nifHDK* gene deletion confirmation
BBP 3330	CAGAATCCGCGCTGCCTTGGCCTGCAC	*nifHDK* gene deletion confirmation
BBP 3346	NNNGGATCCATATG GGCGTCCTGTTTTCCGATAGATC	*nifHDK* genes flanking region cloning
BBP 3347	NNNGAATTC GAACTACATCATCGCGGATTCCGAC	*nifHDK* genes flanking region cloning
BBP 3348	CATGATGCCGGATCC GAGGGTTGGAGAGGACGATGAG	*nifHDK* genes flanking region cloning
BBP 3349	NNNAAGCTT CATCGTCCCGATCACATAATCCAGCAG	*nifHDK* genes flanking region cloning

Results and Discussion

Constructed strains

We constructed several strains of *G. diazotrophicus* through genetic manipulations as identified in Table 4 (Figure 27). The dual ammonium transporter gene *amtB* deletion strain, GABB004, was one key target. Previous deletions of *amtB* in *A. vinelandii* resulted in the slow release of ammonium in sufficient amounts to support the growth of a coculture nitrogen-dependent green algal strain (136). As *G. diazotrophicus* has two *amtB* homologs, it was necessary to delete both genes to remove all identified genes coding for ammonium transporters. After electroporation of the desired plasmid,

the replacement of each gene with the antibiotic resistance cassettes was confirmed in the constructed strain through both colony PCR and sequencing. Both methods indicated successful deletion of both *amtB* homologs and replacement with antibiotic resistance cassettes.

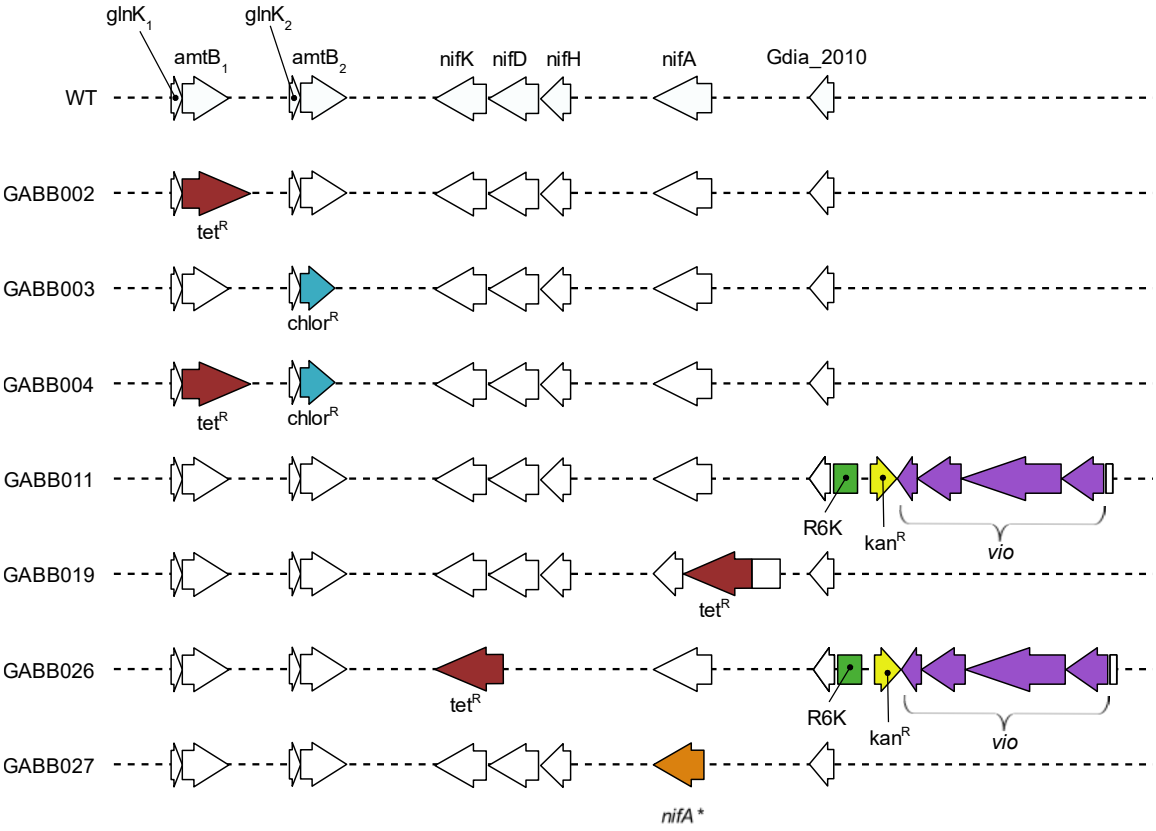

Figure 27. Depictions of strains constructed in this work. Shown are simple representations of the modifications made to the various regions of the *G. diazotrophicus* genome in the strains.

In addition to construction of the dual *amtB* deletion strain, we additionally aimed to carry out precise basepair deletions within the Q-linker region of *nifA* in *G. diazotrophicus* to construct a strain with deregulated nitrogen fixation. As the master regulator of all *nif* genes, *nifA* is critical for the activation of BNF. Prior work in the diazotroph *R. palustris* indicated that the *nifA** mutant, missing 48 bp of the Q-linker domain, resulted in *nifA* constitutive expression (134). We identified the similar amino acid sequence in *nifA* in *G. diazotrophicus* and subsequently constructed the *nifA** mutant, GABB027. We additionally built a strain of *G. diazotrophicus* with *nifA* replaced

with a tetracycline resistant cassette (GABB019). The genotype of each strain was confirmed through colony PCR and subsequent sequencing, as was done for the *amtB* dual deletion strain.

The construction of a *nif*- sensor strain was the final component of the *G. diazotrophicus* strain engineering. Built to test the ammonium excretion capabilities of the previously described strains quantitatively, this strain (GABB026) lacked *nifHDK*, the genes encoding for molybdenum-dependent nitrogenase. Without these genes, this strain would be unable to perform BNF and unable to survive under diazotrophic growth. Due to a random transposon insertion of pBB327 into a putative glutamine amidotransferase (*Gdia_2010*), GABB026 additionally possessed the *vioABCE* genes coding for production of violacein, giving the strain its distinctive purple color (Figure 28). Thus, this strain would be able to act as a biosensor since any growth of this strain under diazotrophic conditions would be easy to detect due to this vibrant pigmentation. Together, the lack of *nifHDK*, the insertion of the genes coding for violacein production, and the requirement of a nitrogen source for growth of GABB026 under diazotrophic conditions yielded these biosensor characteristics. This final engineered strain was again confirmed through the above methods.

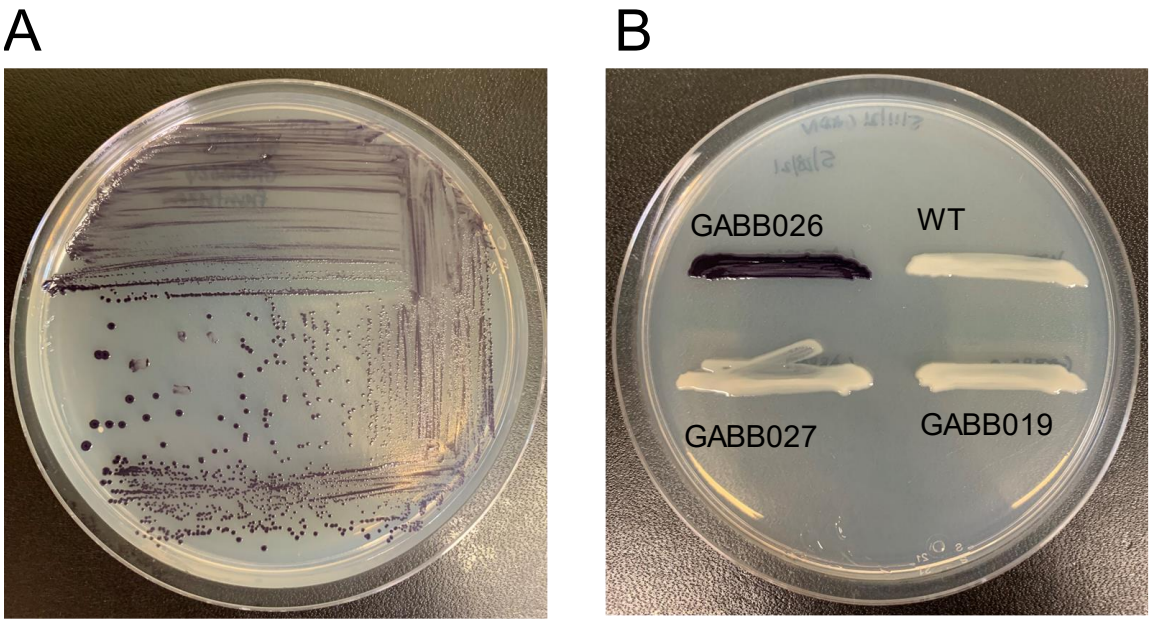

Figure 28. Plate images of various constructed strains. **A**) Image of GABB026 strain grown on GADN medium at 30ºC in air. Purple pigmentation is visible. **B**) Image of four strains grown on GADN medium at 30ºC in air.

Biosensor coculture

Under the conditions in the atmospheric chamber, neither the wild type strain, GABB004 nor GABB027 indicated support of growth of the biosensor GABB026 (Figure 29). Purple pigmentation of the coculture cells under diazotrophic growth, as indicated for the strains under nitrogen-rich conditions on the nitrogen-free medium, would have indicated support of growth of GABB026. As GABB026 lacks *nifHDK*, this strain should not have been capable of independent growth under nitrogen fixation conditions. We expected wild type *G. diazotrophicus* to provide sufficient exogenous nitrogen to support growth of GABB026, and to view the purple phenotype. We further expected either GABB004 or GABB027 to provide a greater amount of extracellular nitrogen than the wild type. None of the strains appeared to support growth of the biosensor strain. This lack of observed support of the biosensor may be due to a lack of sufficient nitrogen provided to the biosensor strain for violacein production. However, phenotypes of these strains observed may provide additional answers. Further, GABB026 exhibited a level of translucent independent growth under diazotrophic conditions. The performances of these individual strains are further discussed below.

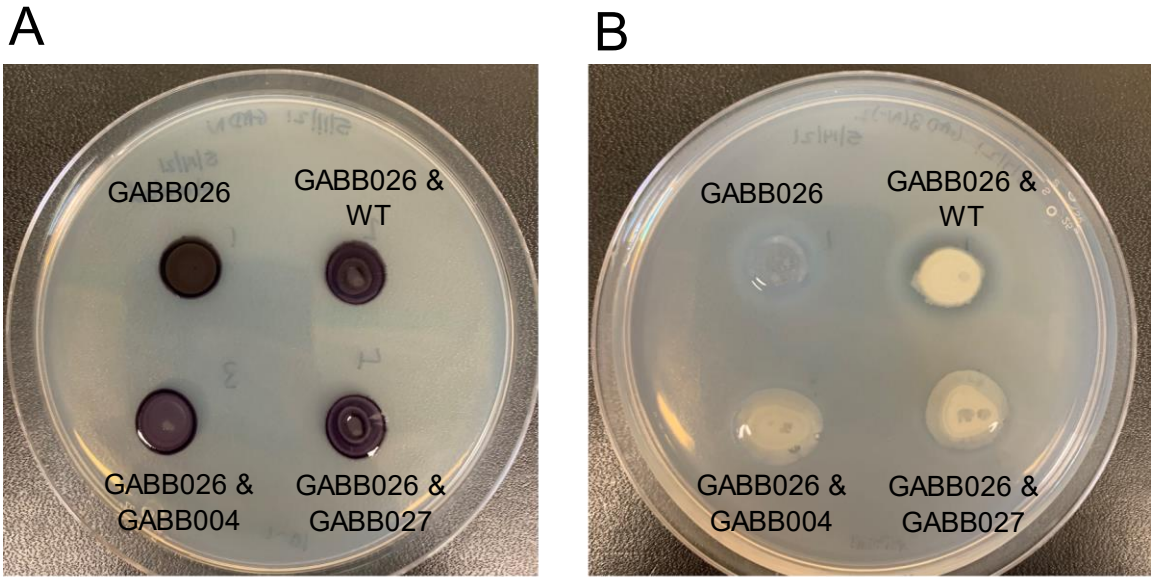

Figure 29. Plate images of coculture of strains with biosensor. **A**) Growth of all four coculture spots grown on GADN medium at 30ºC in air. **B**) Growth of all four coculture spots on GADB(N-)2 medium at 28.9ºC under diazotrophic conditions.

Under the diazotrophic conditions tested here, all strains regardless of pigmentation exhibited a mucoid phenotype. As this unique characteristic was not observed under nitrogen sufficient growth conditions, this phenotype appears to be associated with BNF. This mucoid phenotype may be related to exopolysaccharide production. *G. diazotrophicus* contains a variety of genes related to exopolysaccharides, as previously described (46). The production of these polysaccharides have been largely attributed to the colonization process of the bacterium into host plants (1). Key among these genes are those within the large *gum* gene cluster for exopolysaccharide biosynthesis. These genes are essential for plant colonization as well as biofilm formation by this bacterium (64). Prior work demonstrated that a mutant lacking a *gumD* homolog, the gene thought to be the first committed step of exopolysaccharide biosynthesis, was unable to produce exopolysaccharides, though continued normal growth (64). Additionally, most of the *gum* genes exhibited a fitness differential under diazotrophic conditions (Chapter 2). Given this prior work regarding the *gum* genes and this fitness differential, we thus hypothesize that the *gum* genes are largely responsible for this mucoid phenotype. Deletion of the entire *gum* gene cluster or perhaps solely *gumD*, may likewise result in minimal exopolysaccharide production in *G. diazotrophicus*. Through reduction of exopolysaccharide production, we hypothesize that *G. diazotrophicus* will no longer exhibit this mucoid phenotype under nitrogen fixation conditions. Alternatively, deletion of other genes associated with polysaccharide production, such as those involved in LPS or CPS biosynthesis (Chapter 2) may likewise reveal answers surrounding this phenotype. Upon removing the ability of the strain to exhibit this mucoid phenotype, we hypothesize that the strains may be able to then support biosensor growth.

This mucoid phenotype exhibited here may additionally be involved in regulating oxygen provided the bacterium. Prior work indicated that *G. diazotrophicus* exhibits unique nitrogenase oxygen protectant colony morphology mechanisms (101). It has been

suggested that *G. diazotrophicus* uses bacterial positioning within its mucoid colony to achieve the proper oxygen conditions for nitrogenase activity (101). Given the diazotrophic environment of growth, this oxygen protection mechanism may occur here. The phenotype may be produced to achieve the proper oxygen flow and provide the bacterial cells with the proper positioning within the colony for diazotrophic growth. As this optimal bacterial positioning of *G. diazotrophicus* was previously observed under non-diazotrophic conditions, the production of the mucoid phenotype could play a role in the ability of these strains to adapt to growth under diazotrophic conditions. Given this potential mechanism, it is plausible that removal of the mucoid phenotype, which we posit may be possible through removal of the *gum* genes, may instead be detrimental to diazotrophic growth.

Further quantitative and qualitative tests of these strains would be worthwhile to carry out to further characterize these strains. As the original coculture experiment with the *nif*- strain GABB026 yielded inconclusive results, growth of these strains with instead an algal strain could reveal extracellular nitrogen production. The previous *amtB* deletion strain of *A. vinelandii*, which this work was based upon, supported growth of the green algae *C. sorokiniana* (136). While growth of *G. diazotrophicus* with an algal strain presents additional challenges, such as light-dark cycles and algal growth under the low oxygen environment essential for diazotrophic growth, coculture with an algal strain instead of a bacterial strain may allow for a more clear understanding of ammonium excretion from strains deregulated for nitrogen fixation. It may be that these strains are producing sufficient extracellular nitrogen to support the growth of a biosensor strain, but perhaps it is not entering the media nor being taken up by the biosensor strain due to the presence of the excreted polysaccharides and tough membrane of *G. diazotrophicus*. Growth with algae may reduce these issues.

Deletion of* amtB *genes

Deletion of the genes encoding for putative ammonium transporters, *amtB1* and *amtB2*, in strain GABB004 did not appear to result in a marked difference from the wild type strain regarding ammonium excretion to the medium based on the observed phenotype of the biosensor strain. While the ammonium excretion was not quantified

here due to time restraints and it is thus possible that both strains did accumulate low levels of ammonium in the extracellular space, neither indicated this characteristic in the qualitative assessment. We thus conclude from this assessment that neither strain was capable of supporting growth of the non-diazotrophic biosensor strain. This was unexpected, given that previous work in the diazotrophs *A. vinelandii* (136) and *P. stutzeri* (137) indicated that deletion of *amtB* homologs resulted in ammonium secretion under nitrogen fixation conditions. However, deletion of both *amtB* homologs simultaneously in *G. diazotrophicus* did not appear to drastically affect the growth or health of this strain under diazotrophic conditions. Given that both *amtB1* and *amtB2* displayed moderate conditional essentiality due to fitness deficits under BNF growth in the Tn-seq performed on this organism (Chapter 2), it is suspicious that a stronger fitness defect was not detected in the double mutant. While further characterization of this strain is necessary, it appears that neither *amtB* homolog is strictly required for diazotrophic growth.

The trimeric ammonium transporter proteins are thought to be channel proteins that acquire ammonium from the extracellular space. While it is commonly assumed that these proteins are not essential for growth on ammonium, as under sufficient conditions ammonium is capable of diffusing across the cell membrane (138), ammonium transporters are nonetheless regulated physically by P_{II} proteins such as GlnK. In this model, GlnK negatively regulates AmtB by binding to it under conditions of nitrogen sufficiency, negatively regulating the transport capabilities of the protein (Chapter 1). As *amtB* and *glnK* are thought to have historically evolved together (138), AmtB is the predominant target for P_{II} proteins in this model. However, the ammonium transporter is not the sole target of P_{II} proteins across all prokaryotes. In the native rice endophyte and diazotroph *Azoarcus* sp. BH72, GlnK is capable of membrane association independent of AmtB (139). Here, interaction of GlnK with the membrane-bound protein complex encoded by the *rnf1* genes, Rnf1, controls electron transfer to nitrogenase (140). This is a novel regulatory feature of Rnf1 (140). While *G. diazotrophicus* does not possess *rnf* genes, it has two *fix* homologs, one of which is conditionally essential under diazotrophic conditions (Chapter 2). As diazotrophs with *rnf* typically do not additionally have *fix*

genes, though *A. vinelandii* is an exception (141), this is not unexpected. Prior studies found the FixABCX complex to be membrane-associated in other diazotrophs (141, 142). Given the similar electron transfer role of these two systems and that both are membrane-associated, it is plausible that the FixABCX complex is instead a target of P_{II} proteins for nitrogen regulation in *G. diazotrophicus*. We expected that deletion of the ammonium transporters would result in constitutive *nif* expression and higher external ammonium excretion due to the inability of the strain to inactivate or "switch off" nitrogenase activity. The proposed *nif* regulon for *G. diazotrophicus* indicates that ammonium uptake is largely related to the activity of the ammonium transporters (Figure 26). From this model, we expected that deletion of the ammonium transporters would lower intracellular nitrogen and result in constitutive *nif* expression as the strain would otherwise be unable to acquire sufficient reduced nitrogen. If instead the FixABCX complex is the alternative target of the P_{II} proteins, this would explain the unexpected performance of the *amtB* dual mutant strain regarding its observed ammonium release. If there is a similar process occurring in *G. diazotrophicus* as in *Azoarcus* sp. BH72, the deletion of the *amtB* genes would be largely inconsequential for nitrogen sensing and the eventual inactivation of nitrogenase activity. Further alternative P_{II} protein targets have been found to interact with other membrane proteins, such as PamA in *Synechocystis* and a component of the nitrate utilization pathway in cyanobacteria (138). Alternative targets of P_{II} proteins in *G. diazotrophicus* are plausible. Further study of the *glnK-amtB* system in this organism is required.

Q-linker deletion

In addition to the *amtB* deletion strain, the *nifA** mutant strain GABB027 did not demonstrate a difference to the wild-type *G. diazotrophicus* in terms of ammonium excretion to the medium, as neither appeared to support growth of the biosensor strain. We expected constitutive *nif* gene expression in this *G. diazotrophicus* strain as prior *nifA** mutant strains in *R. palustris* (134) and *R. sphaeroides* (135) had revealed this result. As nitrogenase activity was not directly quantified here, and instead its activity was indirectly indicated through a lack of support of the biosensor strain, we cannot accurately claim that the *nif* genes were not upregulated in some way by this deletion.

Due to the growth of this strain and the moderate essentiality of *nifA* under diazotrophic conditions (Chapter 2), it appeared that this *nifA* deletion was not detrimental to the growth of the strain. However, our qualitative assessment of the strain implies that the *nifA** mutant in *G. diazotrophicus* did not act as expected. No additional ammonium in an amount necessary to support the growth of the biosensor was excreted into the medium by the action of a deregulated nitrogenase.

While the deletion of the Q-linker interdomain region resulted in constitutive *nif* expression in *R. palustris* (134) and *R. sphaeroides* (135) as mentioned previously, manipulations in *nifA* in other diazotrophs have revealed conflicting results regarding the region critical for nitrogen sensitivity. In the plant endophyte and diazotroph *Herbaspirillum seropedicae*, which additionally does not possess a *nifL* homolog, a point mutation in the N-terminal GAF domain resulted in an active conformation of *nifA* that did not require regulation by GlnK (143). Further, full removal of the GAF domain resulted in the active *nif* conformation even under conditions of excess nitrogen (143, 144). This mutation occurred elsewhere in the gene and still resulted in a similar active conformation as seen by the Q-linker deletion in *R. palustris*. This suggests variety in the region of *nifA* required for nitrogen sensing across diazotrophs. Even though *R. palustris* and *R. sphaeroides* lack a *nifL* homolog as does *G. diazotrophicus*, these bacteria are purple non-sulfur photosynthetic bacteria, and so the function of the *nifA* domains may have evolutionarily diverged.

More in-depth study of this strain is necessary to understand the possible constitutive *nif* expression occurring in *G. diazotrophicus* following deletion of the Q-linker interdomain region of *nifA*. An acetylene reduction assay measures nitrogenase activity, and this was done for the previous *nifA** mutants of *R. palustris* (133, 134, 145). Measuring the activity of the nitrogenase of GABB027 under nitrogen rich as well as diazotrophic conditions would yield important information regarding its expression without the Q-linker. Further, the *R. palustris* mutant strain produced H_2 gas via nitrogenase in nitrogen-rich medium (133, 134). Carrying out these same measurements would additionally yield relevant information about GABB027 and its possible constitutive expression of *nif* genes. Further exploration of the effect of the *nifA** mutant

on nitrogenase activity will allow deeper understanding of the Q-linker region in *nifA* in *G. diazotrophicus*. If quantitative assessment of nitrogenase activity reveals it is unchanged from the wild-type *nifA* by the deletion of the Q-linker region, it would be worthwhile to pursue further manipulations within *nifA* to result in the constitutive expression of *nif* genes.

Biosensor strain

While the *nif*- strain GABB026 did have purple pigmentation from the presence of the *vio* genes, it additionally indicated an amount of independent growth under nitrogen fixation conditions (Figure 29). As this strain lacks *nifHDK*, this was unexpected. *G. diazotrophicus* does not possess alternative nitrogenases, so this strain should not be able to carry out BNF without the capability of synthesis of its molybdenum-dependent nitrogenase. However, we hypothesize that this growth seen may be excessive extracellular polysaccharide production by the strain. As discussed above, this strain displayed the mucoid phenotype as observed in the other constructed strains. However, it exhibited translucent growth, unlike the growth of other strains, potentially indicating an overproduction of exopolysaccharides instead of sustained growth. This exopolysaccharide production may be a protective mechanism by the strain for production of a non-diazotrophic environment. To understand this possibility, deletion of the *gum* gene cluster would be critical. While unlikely, it is also possible that insertion of the genes encoding violacein is related to its continued ability to grow under diazotrophic conditions despite lacking the genes encoding nitrogenase, or that it was an unstable deletion and the strain reverted to a wild type genotype.

Future work

Further genetic manipulations are planned in these strains. First, we plan to perform the deletion of the *gum* gene cluster to potentially eliminate the mucoid phenotype observed under diazotrophic conditions. We propose that this genetic deletion will be essential in the GABB026 strain for its future use as a biosensor strain. Its current level of polysaccharide production under diazotrophic conditions is potentially inhibitory for use concurrently with other *G. diazotrophicus* strains. Additionally, we plan to perform markerless deletion of the *amtB* dual homologs. The strain GABB004 contains

two antibiotic resistance cassettes, whose use may result in a fitness cost to the growth of the organism due to increased metabolic load (146). Further, if this strain was used outside of a laboratory setting, the potential for spread of antibiotic drug resistance could be concerning (147). Deletion of these genes through markerless means via *sacB*, as previously described for *G. diazotrophicus* (148), or *pyrF* (146) is a worthwhile future goal for this strain. We additionally plan to examine the growth rates of all constructed strains to further understand potential fitness costs of these deletions.

Strains GABB004 and GABB027 did not appear to sustain growth of the biosensor strain GABB026 under diazotrophic conditions. However, the issue may be that the manipulations within each strain did not result in a sufficient level of ammonium excretion to support GABB026 growth independently. We plan to present their genetic manipulations within a single strain of *G. diazotrophicus*: both the *nifA** mutation and deletion of the dual *amtB* homologs. We hypothesize that this further construction may result in a deregulated nitrogenase with constitutive expression and a high level of ammonium excretion to the medium.

Given the performances of strains GABB004 and GABB027, targeting instead different genes or genetic regions within these strains may yield the desired results. Further manipulations within *nifA* may be necessary, such as targeting the *nifA* GAF domain instead of the Q-linker region for production of a strain deregulated for BNF, as suggested above. Additionally, as the *amtB* deletion strain did not perform as expected, it would be worthwhile to identify if the ammonium transporters are the true target of the P_{II} proteins, or if instead the P_{II} proteins target another membrane-bound protein, such as in the Rnf1 system in *Azoarcus* sp. BH72 (140).

Finally, if growth of these strains with a bacterial or algal coculture yields an ammonium-producing strain, we plan to inoculate the strain into a plant host. Given the PGP characteristics displayed by *G. diazotrophicus* (Chapter 1), it is worthwhile to directly explore potential of an ammonium-producing engineered strain for use as a biofertilizer. Improving BNF for agronomic use in an endophytic strain such as *G. diazotrophicus* puts us one step closer to sustainable global nitrogen production.

Conclusion

Our results indicate construction of several agriculturally interesting strains of *G. diazotrophicus*. Successful deletion of the two genes coding for the ammonium transporters as well as removal of the *nifA* Q-linker region demonstrate the potential of this strain for further modification for improved ammonium production. Further study is necessary to characterize these strains in greater detail and optimize their modified BNF pathways for high nitrogen output. We nonetheless present these constructed strains as a preliminary study into modification of *G. diazotrophicus* for improved extracellular ammonium production.

Bibliography

1. S. Kandel, P. Joubert, S. Doty, Bacterial Endophyte Colonization and Distribution within Plants. *Microorganisms* **5**, 77 (2017).
2. M. Rosenblueth, E. Martínez-Romero, Bacterial endophytes and their interactions with hosts. *Mol. Plant-Microbe Interact.* **19**, 827–837 (2006).
3. P. R. Hardoim, *et al.*, The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiol. Mol. Biol. Rev.* **79**, 293–320 (2015).
4. A. Pinski, A. Betekhtin, K. Hupert-Kocurek, L. A. J. Mur, R. Hasterok, Defining the genetic basis of plant–endophytic bacteria interactions. *Int. J. Mol. Sci.* **20** (2019).
5. H. Liu, *et al.*, Inner plant values: Diversity, colonization and benefits from endophytic bacteria. *Front. Microbiol.* **8**, 1–17 (2017).
6. V. C. S. Pankievicz, T. B. Irving, L. G. S. Maia, J. M. Ané, Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biol.* **17**, 1–17 (2019).
7. J. W. Erisman, M. A. Sutton, J. Galloway, Z. Klimont, W. Winiwarter, How a century of ammonia synthesis changed the world. *Nat. Geosci.* **1**, 636–639 (2008).
8. B. E. Smith, Nitrogenase Reveals Its Inner Secrets. *Science (80-.).* **109**, 1654–1655 (2002).
9. F. Mus, *et al.*, Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl. Environ. Microbiol.* **82**, 3698–3710 (2016).
10. M. Gillis, *et al.*, *Acetobacter diazotrophicus* sp. nov., a nitrogen-fixing acetic acid bacterium associated with sugarcane. *Int. J. Syst. Bacteriol.* **39**, 361–364 (1989).
11. Y. Yamada, K. I. Hoshino, T. Ishikawa, The Phylogeny of Acetic Acid Bacteria Based on the Partial Sequences of 16S Ribosomal RNA: The Elevation of the Subgenus Gluconoacetobacter to the Generic Level. *Biosci. Biotechnol. Biochem.* **61**, 1244–1251 (1997).
12. V. A. Cavalcante, J. Dobereiner, A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. *Plant Soil* **108**, 23–31 (1988).
13. N. Eskin, K. Vessey, L. Tian, Research progress and perspectives of nitrogen fixing Bacterium, *Gluconacetobacter diazotrophicus*, in monocot plants. *Int. J. Agron.* **2014** (2014).
14. M. A. Paula, V. M. Reis, J. Döbereiner, Interactions of *Glomus clarum* with *Acetobacter diazotrophicus* in infection of sweet potato (*Ipomoea batatas*), sugarcane (*Saccharum* spp.), and sweet sorghum (*Sorghum vulgare*). *Biol. Fertil. Soils* **11**, 111–115 (1991).
15. A. Tapia-Hernández, M. R. Bustillos-Cristales, T. Jiménez-Salgado, J. Caballero-Mellado, L. E. Fuentes-Ramírez, Natural endophytic occurrence of *Acetobacter diazotrophicus* in pineapple plants. *Microb. Ecol.* **39**, 49–55 (2000).
16. T. Jimenez-Salgado, *et al.*, *Coffea arabica* L., a new host plant for *Acetobacter diazotrophicus*, and isolation of other nitrogen-fixing acetobacteria. *Appl. Environ. Microbiol.* **63**, 3676–3683 (1997).
17. E. C. Cocking, P. J. Stone, M. R. Davey, Intracellular colonization of roots of *Arabidopsis* and crop plants by *Gluconacetobacter diazotrophicus*. *Vitr. Cell. Dev.*

Biol. - Plant **42**, 74–82 (2006).
18. G. Tian, P. Pauls, Z. Dong, L. M. Reid, L. Tian, Colonization of the nitrogen-fixing bacterium *Gluconacetobacter diazotrophicus* in a large number of Canadian corn plants. *Can. J. Plant Sci.* **89**, 1009–1016 (2009).
19. M. V. Rodriguez, *et al.*, Anatomical and Biochemical Changes Induced by *Gluconacetobacter diazotrophicus* Stand Up for *Arabidopsis thaliana* Seedlings From *Ralstonia solanacearum* Infection. *Front. Plant Sci.* **10**, 1–20 (2019).
20. M. Sevilla, A. De Oliveira, I. Baldani, C. Kennedy, Contributions of the bacterial endophyte *Acetobacter diazotrophicus* to sugarcane nutrition: A preliminary study. *Symbiosis* **25**, 181–191 (1998).
21. E. Lima, R. M. Boddey, J. Döbereiner, Quantification of biological nitrogen fixation associated with sugar cane using a 15N aided nitrogen balance. *Soil Biol. Biochem.* **19**, 165–170 (1987).
22. R. M. Boddey, J. C. Polidoro, A. S. Resende, B. J. R. Alves, S. Urquiaga, Use of the 15N natural abundance technique for the quantification of the contribution of N2 fixation to sugar cane and other grasses. *Aust. J. Plant Physiol.* **28**, 889–895 (2001).
23. M. F. Luna, J. Aprea, J. M. Crespo, J. L. Boiardi, Colonization and yield promotion of tomato by *Gluconacetobacter diazotrophicus*. *Appl. Soil Ecol.* **61**, 225–229 (2012).
24. L. Filgueiras, *et al.*, *Gluconacetobacter diazotrophicus* mitigates drought stress in Oryza sativa L. *Plant Soil* **451**, 57–73 (2020).
25. P. Delaporte-Quintana, N. C. Lovaisa, V. A. Rapisarda, R. O. Pedraza, The plant growth promoting bacteria *Gluconacetobacter diazotrophicus* and Azospirillum brasilense contribute to the iron nutrition of strawberry plants through siderophores production. *Plant Growth Regul.* **91**, 185–199 (2020).
26. R. Muthukumarasamy, G. Revathi, S. Seshadri, C. Lakshminarasimhan, *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. *Curr. Sci.* **83**, 137–145 (2002).
27. M. P. Stephan, M. Oliveira, K. R. S. Teixeira, G. Martinez-Drets, J. Döbereiner, Physiology and dinitrogen fixation of *Acetobacter diazotrophicus*. *FEMS Microbiol. Lett.* **77**, 67–72 (1991).
28. B. Alvarez, G. Martinez-Drets, Metabolic characterization of *Acetobacter diazotrophicus*. *Can. J. Microbiol.* **41**, 918–924 (1995).
29. F. Bastián, *et al.*, Production of indole-3-acetic acid and gibberellins A1 and A3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regul.* **24**, 7–11 (1998).
30. L. E. Fuentes-Ramirez, T. Jimenez-Salgado, I. R. Abarca-Ocampo, J. Caballero-Mellado, *Acetobacter diazotrophicus*, an indoleacetic acid producing bacterium isolated from sugarcane cultivars of México. *Plant Soil* **154**, 145–150 (1993).
31. P. Logeshwaran, M. Thangaraju, K. Rajasundari, Hydroxamate siderophores of endophytic bacteria *Gluconacetobacter diazotrophicus* isolated from sugarcane roots. *Aust. J. Basic Appl. Sci.* **3**, 3564–3567 (2009).
32. A. C. Intorne, *et al.*, Identification and characterization of *Gluconacetobacter diazotrophicus* mutants defective in the solubilization of phosphorus and zinc.

Arch. Microbiol. **191**, 477–483 (2009).
33. A. J. Jasniewski, C. C. Lee, M. W. Ribbe, M. W. Ribbe, Y. Hu, Reactivity, Mechanism, and Assembly of the Alternative Nitrogenases. *Chem. Rev.* **120**, 5107–5157 (2020).
34. J. B. Howard, K. J. Kechris, D. C. Rees, A. N. Glazer, Multiple Amino Acid Sequence Alignment Nitrogenase Component 1: Insights into Phylogenetics and Structure-Function Relationships. *PLoS One* **8** (2013).
35. K. Fisher, W. E. Newton, Nitrogenase proteins from *Gluconacetobacter diazotrophicus*, a sugarcane-colonizing bacterium. *Biochim. Biophys. Acta - Proteins Proteomics* **1750**, 154–165 (2005).
36. S. Burén, E. Jiménez-Vicente, C. Echavarri-Erasun, L. M. Rubio, Biosynthesis of Nitrogenase Cofactors. *Chem. Rev.* **120**, 4921–4968 (2020).
37. J. R. Lecomte, The nitrogenase system from Azotobacter: Two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ADP Hydrolysis (1966).
38. P. C. Dos Santos, Z. Fang, S. W. Mason, J. C. Setubal, R. Dixon, Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13**, 1–12 (2012).
39. R. A. Dixon, J. R. Postgate, Genetic transfer of nitrogen fixation from K*lebsiella pneumoniae* to *Escherichia coli*. *Nature* **237**, 102–103 (1972).
40. K. Temme, D. Zhao, C. A. Voigt, Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7085–7090 (2012).
41. S. Lee, A. Reth, D. Meletzus, M. Sevilla, C. Kennedy, Characterization of a major cluster of nif, fix, and associated genes in a sugarcane endophyte, *Acetobacter diazotrophicus*. *J. Bacteriol.* **182**, 7088–7091 (2000).
42. R. Dixon, D. Kahn, Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2**, 621–631 (2004).
43. I. Martinez-Argudo, R. Little, N. Shearer, P. Johnson, R. Dixon, The NifL-NifA System: a Multidomain Transcriptional Regulatory Complex That Integrates Environmental Signals. **186**, 1–10 (2004).
44. M. B. Batista, R. Dixon, Manipulating nitrogen regulation in diazotrophic bacteria for agronomic benefit. *Biochem. Soc. Trans.* **47**, 603–614 (2019).
45. H.-M. Fischer, Genetic regulation of nitrogen fixation in Rhizobia. *Microbiol. Rev.* **58**, 352–386 (1994).
46. M. Bertalan, *et al.*, Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5. *BMC Genomics* **10**, 450 (2009).
47. K. R. S. Teixeira, *et al.*, Molecular analysis of the chromosomal region encoding the nifA and nifB genes of *Acetobacter diazotrophicus*. *FEMS Microbiol. Lett.* **176**, 301–309 (1999).
48. P. J. Rutten, P. S. Poole, *Oxygen regulatory mechanisms of nitrogen fixation in rhizobia* (Elsevier Ltd, 2019).
49. J. C. Wootton, M. H. Drummond, The q-linker: A class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng. Des. Sel.* **2**, 535–543 (1989).

50. O. Perlova, R. Nawroth, E. M. Zellermann, D. Meletzus, Isolation and characterization of the *glnD* gene of *Gluconacetobacter diazotrophicus*, encoding a putative uridylyltransferase/uridylyl-removing enzyme. *Gene* **297**, 159–168 (2002).
51. D. Meletzus, *et al.*, Characterization of genes involved in regulation of nitrogen fixation and ammonium sensing in *Acetobacter diazotrophicus*, an endophyte of sugarcane. *Biol. Nitrogen Fixat. 21st Century*, 125–126 (1998).
52. O. Perlova, A. Ureta, S. Nordlund, D. Meletzus, Identification of Three Genes Encoding PII-Like Proteins in *Gluconacetobacter diazotrophicus*: Studies of Their Role(s) in the Control of Nitrogen Fixation. *Society* **185**, 5854–5861 (2003).
53. M. J. Merrick, R. A. Edwards, Nitrogen control in bacteria. *Microbiol. Rev.* **59**, 604–622 (1995).
54. W. C. van Heeswijk, H. V. Westerhoff, F. C. Boogerd, Nitrogen Assimilation in *Escherichia coli*: Putting Molecular Data into a Systems Perspective. *Microbiol. Mol. Biol. Rev.* **77**, 628–695 (2013).
55. C. Leaungvutiviroj, P. Ruangphisarn, P. Hansanimitkul, H. Shinkawa, K. Sasaki, Development of a new biofertilizer with a high capacity for N2 fixation, phosphate and potassium solubilization and auxin production. *Biosci. Biotechnol. Biochem.* **74**, 1098–1101 (2010).
56. R. Jacoby, M. Peukert, A. Succurro, A. Koprivova, S. Kopriva, The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Front. Plant Sci.* **8**, 1–19 (2017).
57. E. H. Cojho, V. M. Reis, A. C. G. Schenberg, J. DÃ¶bereiner, Interactions of *Acetobacter diazotrophicus* with an amylolytic yeast in nitrogen-free batch culture . *FEMS Microbiol. Lett.* **106**, 341–346 (1993).
58. R. O. Pedraza, Recent advances in nitrogen-fixing acetic acid bacteria. *Int. J. Food Microbiol.* **125**, 25–35 (2008).
59. A. Giongo, H. L. Tyler, U. N. Zipperer, E. W. Triplett, Two genome sequences of the same bacterial strain, *Gluconacetobacter diazotrophicus* PAl 5, suggest a new standard in genome sequence submission. *Stand. Genomic Sci.* **2**, 309–317 (2010).
60. L. M. S. Lery, W. M. A. von Krüger, F. C. Viana, K. R. S. Teixeira, P. M. Bisch, A comparative proteomic analysis of *Gluconacetobacter diazotrophicus* PAL5 at exponential and stationary phases of cultures in the presence of high and low levels of inorganic nitrogen compound. *Biochim. Biophys. Acta - Proteins Proteomics* **1784**, 1578–1589 (2008).
61. L. M. S. Lery, A. S. Hemerly, E. M. Nogueira, W. M. A. Von Krüger, P. M. Bisch, Quantitative proteomic analysis of the interaction between the endophytic plant-growth-promoting bacterium *Gluconacetobacter diazotrophicus* and Sugarcane. *Mol. Plant-Microbe Interact.* **24**, 562–576 (2011).
62. T. Cruz, *et al.*, *Arabidopsis thaliana* exudates induce growth and proteomic changes in *Gluconacetobacter diazotrophicus* (2020) https:/doi.org/10.7717/peerj.9600.
63. M. F. dos Santos, V. L. Muniz de Pádua, E. de Matos Nogueira, A. S. Hemerly, G. B. Domont, Proteome of *Gluconacetobacter diazotrophicus* co-cultivated with sugarcane plantlets. *J. Proteomics* **73**, 917–931 (2010).

64. C. H. S. G. Meneses, L. F. M. Rouws, J. L. Simões-Araújo, M. S. Vidal, J. I. Baldani, Exopolysaccharide production is required for biofilm formation and plant colonization by the nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus*. *Mol. Plant-Microbe Interact.* **24**, 1448–1458 (2011).
65. T. van Opijnen, K. L. Bodi, A. Camilli, Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **6**, 767–772 (2009).
66. A. K. Cain, *et al.*, A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**, 526–540 (2020).
67. E. D. Brutinel, J. A. Gralnick, Anomalies of the anaerobic tricarboxylic acid cycle in *Shewanella oneidensis* revealed by Tn-seq. *Mol. Microbiol.* **86**, 273–283 (2012).
68. T. Van Opijnen, A. Camilli, A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res.* **22**, 2541–2551 (2012).
69. B. K. Fabian, S. G. Tetu, I. T. Paulsen, Application of Transposon Insertion Sequencing to Agricultural Science. *Front. Plant Sci.* **11** (2020).
70. K. B. Pechter, L. Gallagher, H. Pyles, C. S. Manoil, C. S. Harwood, Essential genome of the metabolically versatile alphaproteobacterium *Rhodopseudomonas palustris*. *J. Bacteriol.* **198**, 867–876 (2015).
71. F. Rosconi, S. P. W. de Vries, A. Baig, E. Fabiano, A. J. Grant, Essential genes for in vitro growth of the endophyte *Herbaspirillum seropedicae* SmR1 as revealed by transposon insertion site sequencing. *Appl. Environ. Microbiol.* **82**, 6664–6671 (2016).
72. P. C. Dos Santos, Molecular biology and genetic engineering in nitrogen fixation. *Methods Mol. Biol.* **766**, 81–92 (2011).
73. E. M. Lenneman, J. M. Ohlert, N. P. Palani, B. M. Barney, Fatty alcohols for wax esters in *Marinobacter aquaeolei* VT8: Two optional routes in the wax biosynthesis pathway. *Appl. Environ. Microbiol.* **79**, 7055–7062 (2013).
74. N. Palani, Transposon insertion sequencing (Tn-seq) library preparation protocol - includes UMI for PCR duplicate removal. 1–2 (2019).
75. B. E. Poulsen, *et al.*, Defining the core essential genome of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10072–10080 (2019).
76. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *J. EMBnet* **17**, 10–12 (2013).
77. L. A. Gallagher, J. Shendure, C. Manoil, Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *MBio* **2**, 1–8 (2011).
78. C. A. H. Iii, *et al.*, Global Transposon Mutagenesis and a Minimal Mycoplasma Genome. **286**, 2165–2170 (1999).
79. S. A. Lee, *et al.*, General and condition-specific essential functions of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5189–5194 (2015).
80. B. K. Fabian, *et al.*, Elucidating essential genes in plant-associated *Pseudomonas protegens* Pf-5 using transposon insertion sequencing . *J. Bacteriol.*, 1–40 (2020).
81. M. H. Ryu, *et al.*, Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* **5**, 314–330 (2020).
82. F. Mus, A. B. Alleman, N. Pence, L. C. Seefeldt, J. W. Peters, Exploring the alternatives of biological nitrogen fixation. *Metallomics* **10**, 523–538 (2018).

83. L. Demtröder, F. Narberhaus, B. Masepohl, Coordinated regulation of nitrogen fixation and molybdate transport by molybdenum. *Mol. Microbiol.* **111**, 17–30 (2019).
84. Y. Zhang, V. N. Gladyshev, Molybdoproteomes and Evolution of Molybdenum Utilization. *J. Mol. Biol.* **379**, 881–899 (2008).
85. S. Poudel, *et al.*, Electron transfer to nitrogenase in different genomic and metabolic backgrounds. *J. Bacteriol.* **200**, 1–19 (2018).
86. G. Herrmann, E. Jayamani, G. Mai, W. Buckel, Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *J. Bacteriol.* **190**, 784–791 (2008).
87. L. Curatti, C. S. Brown, P. W. Ludden, L. M. Rubio, Genes required for rapid expression of nitrogenase activity in *Azotobacter vinelandii*. **102**, 6291–6296 (2005).
88. S. Hill, E. P. Kavanagh, Roles of *nifF* and *nifJ* gene products in electron transport to nitrogenase in *Klebsiella pneumoniae. J. Bacteriol.* **141**, 470–475 (1980).
89. D. Nieva-Gomez, G. P. Roberts, S. Klevivkis, W. Brill, Electron transport to nitrogenase in *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 2555–2558 (1980).
90. O. V. Tsoy, D. A. Ravcheev, J. Čuklina, M. S. Gelfand, Nitrogen fixation and molecular oxygen: Comparative genomic reconstruction of transcription regulation in Alphaproteobacteria. *Front. Microbiol.* **7**, 1–14 (2016).
91. J. R. Gallon, Tansley Review No . 44 Reconciling the incompatible: N2 fixation and O2. *Rev. Lit. Arts Am.* **122**, 571–609 (1992).
92. J. Oelze, Respiratory protection of nitrogenase in *Azotobacter* species: Is a widely held hypothesis unequivocally supported by experimental evidence? *FEMS Microbiol. Rev.* **24**, 321–333 (2000).
93. R. L. Robson, Characterization of an oxygen-stable nitrogenase complex isolated from Azotobacter chroococcum. *Biochem. J.* **181**, 569–575 (1979).
94. A. Ureta, S. Nordlund, Evidence for Conformational Protection of Nitrogenase against Oxygen in *Gluconacetobacter diazotrophicus* by a Putative FeSII Protein. **184**, 5805–5809 (2002).
95. B. González, *et al.*, Respiratory system of *Gluconacetobacter diazotrophicus* PAL5 Evidence for a cyanide-sensitive cytochrome bb and cyanide-resistant cytochrome ba quinol oxidases. **1757**, 1614–1622 (2006).
96. L. M. S. Lery, M. Bitar, M. G. S. Costa, S. C. S. Rössle, P. M. Bisch, Unraveling the molecular mechanisms of nitrogenase conformational protection against oxygen in diazotrophic bacteria. **11**, 1–11 (2010).
97. E. C. Cocking, Endophytic colonization of plant roots by nitrogen-fixing bacteria. *Plant Soil* **252**, 169–175 (2003).
98. E. K. James, V. M. Reis, F. L. Olivares, J. I. Baldani, J. Dobereiner, Infection of sugar cane by the nitrogen-fixing bacterium *Acetobacter diazotrophicus*. **45**, 757–766 (1994).
99. Y. R. Saadat, A. Y. Khosroushahi, B. P. Gargari, A comprehensive review of anticancer, immunomodulatory and health beneficial effects of the lactic acid bacteria exopolysaccharides. *Carbohydr. Polym.* **217**, 79–89 (2019).
100. M. Penteado, L. Mendonca-Previato, K. A. R. Paulo, Differentiation of Capsular

Polysaccharides from *Acetobacter diazotrophicus* Strains Isolated from Sugarcane. **39**, 237–242 (1995).
101. Z. Dong, *et al.*, Evidence for protection of nitrogenase from O2 by colony structure in the aerobic diazotroph *Gluconacetobacter diazotrophicus*. *Microbiology* **148**, 2293–2298 (2002).
102. F. Tao, S. Swarup, L. H. Zhang, Quorum sensing modulation of a putative glycosyltransferase gene cluster essential for *Xanthomonas campestris* biofilm formation. *Environ. Microbiol.* **12**, 3159–3170 (2010).
103. S. K. Singh, *et al.*, Functional Domains of the ClpA and ClpX Molecular Chaperones Identified by Limited Proteolysis and Deletion Analysis. *J. Biol. Chem.* **276**, 29420–29429 (2001).
104. G. Martínez-Noël, L. Curatti2, J. A. Hernandez, L. M. Rubio, NifB and NifEN protein levels are regulated by ClpX2 under nitrogen fixation conditions in *Azotobacter vinelandii*. *Mol Microbio* **79**, 1182–1193 (2011).
105. G. Brader, S. Compant, B. Mitter, F. Trognitz, A. Sessitsch, Metabolic potential of endophytic bacteria. *Curr. Opin. Biotechnol.* **27**, 30–37 (2014).
106. Dong Zhongmin, *et al.*, A nitrogen-fixing endophyte of sugarcane stems. A new role for the apoplast. *Plant Physiol.* **105**, 1139–1147 (1994).
107. L. E. Fuentes-Ramírez, *et al.*, Novel nitrogen-fixing acetic acid bacteria, *Gluconacetobacter johannae* sp. nov. and *Gluconacetobacter azotocaptans* sp. nov., associated with coffee plants. *Int. J. Syst. Evol. Microbiol.* **51**, 1305–1314 (2001).
108. N. A. Tejera, E. Ortega, R. Rodés, C. Lluch, Influence of carbon and nitrogen sources on growth, nitrogenase activity, and carbon metabolism of *Gluconacetobacter diazotrophicus*. *Can. J. Microbiol.* **50**, 745–750 (2004).
109. D. Mamlouk, M. Gullo, Acetic Acid Bacteria: Physiology and Carbon Sources Oxidation. *Indian J. Microbiol.* **53**, 377–384 (2013).
110. M. M. Attwood, J. P. van Dijken, J. T. Pronk, Glucose metabolism and gluconic acid production by *Acetobacter diazotrophicus*. *J. Ferment. Bioeng.* **72**, 101–105 (1991).
111. L. M. S. Lery, *et al.*, Protein expression profile of *Gluconacetobacter diazotrophicus* PAL5, a sugarcane endophytic plant growth-promoting bacterium. *Proteomics* **8**, 1631–1644 (2008).
112. M. F. Luna, C. E. Bernardelli, M. L. Galar, J. L. Boiardi, Glucose metabolism in batch and continuous cultures of *Gluconacetobacter diazotrophicus* PAL 3. *Curr. Microbiol.* **52**, 163–168 (2006).
113. E. Rosenberg, *The prokaryotes: Alphaproteobacteria and betaproteobacteria* (2013) https:/doi.org/10.1007/978-3-642-30197-1.
114. K. Matsushita, H. Toyama, N. Tonouchi, A. Okamoto-Kainuma, Acetic acid bacteria: Ecology and physiology. *Acetic Acid Bact. Ecol. Physiol.*, 1–350 (2016).
115. M. L. Galar, J. L. Boiardi, Evidence for a membrane-bound pyrroloquinoline quinone-linked glucose dehydrogenase in *Acetobacter diazotrophicus*. *Appl. Microbiol. Biotechnol.* **43**, 713–716 (1995).
116. L. Hernandez, *et al.*, Isolation and enzymic properties of levansucrase secreted by *Acetobacter diazotrophicus* SRT4, a bacterium associated with sugar cane.

Biochem. J. **309**, 113–118 (1995).
117. L. Hernández, *et al.*, Structural levansucrase gene (lsdA) constitutes a functional locus conserved in the species Gluconacetobacter for diazotrophicus. *Arch. Microbiol.* **174**, 120–124 (2000).
118. M. L. Velázquez-Hernández, *et al.*, *Gluconacetobacter diazotrophicus* levansucrase is involved in tolerance to NaCl, sucrose and desiccation, and in biofilm formation. *Arch. Microbiol.* **193**, 137–149 (2011).
119. J. G. Arrieta, *et al.*, A type II protein secretory pathway required for levansucrase secretion by *Gluconacetobacter diazotrophicus*. *J. Bacteriol.* **186**, 5031–5039 (2004).
120. E. S. Maxwell, K. Kurahashi, H. M. Kalckar, [20] Enzymes of the Leloir pathway. *Methods Enzymol.* **5**, 174–189 (1962).
121. T. Y. Wong, X. T. Yao, The DeLey-Doudoroff pathway of galactose metabolism in Azotobacter vinelandii. *Appl. Environ. Microbiol.* **60**, 2065–2068 (1994).
122. S. A. Leyn, Y. Maezato, M. F. Romine, D. A. Rodionov, Genomic reconstruction of carbohydrate utilization capacities in microbial-mat derived consortia. *Front. Microbiol.* **8**, 1–17 (2017).
123. K. Wuichet, R. P. Alexander, I. B. Zhulin, Comparative Genomic and Protein Sequence Analyses of a Complex System Controlling Bacterial Chemotaxis. *Methods Enzymol.* **422**, 3 (2007).
124. I. Alav, J. M. Sutton, K. M. Rahman, Role of bacterial efflux pumps in biofilm formation. *J. Antimicrob. Chemother.* **73**, 2003–2020 (2018).
125. V. M. Reis, J. Döbereiner, Effect of high sugar concentration on nitrogenase activity of *Acetobacter diazotrophicus*. *Arch. Microbiol.* **171**, 13–18 (1998).
126. V. M. Reis, K. R. dos S. Teixeira, Nitrogen fixing bacteria in the family Acetobacteraceae and their role in agriculture. *J. Basic Microbiol.* **55**, 931–949 (2015).
127. N. Tejera, E. Ortega, R. Rodes, C. Lluch, Nitrogen compounds in the apoplastic sap of sugarcane stem: Some implications in the association with endophytes. *J. Plant Physiol.* **163**, 80–85 (2006).
128. C. Santi, D. Bogusz, C. Franche, Biological nitrogen fixation in non-legume plants. *Ann. Bot.* **111**, 743–767 (2013).
129. K. Mahmud, S. Makaju, R. Ibrahim, A. Missaoui, Current progress in nitrogen fixing plants and microbiome research. *Plants* **9**, 1–17 (2020).
130. L. K. Sharma, S. K. Bali, A review of methods to improve nitrogen use efficiency in agriculture. *Sustain.* **10**, 1–23 (2017).
131. M. Rosenblueth, *et al.*, Nitrogen fixation in cereals. *Front. Microbiol.* **9**, 1–13 (2018).
132. S. Umesha, P. K. Singh, R. P. Singh, *Microbial biotechnology and sustainable agriculture* (Elsevier Inc., 2017) https:/doi.org/10.1016/B978-0-12-812160-3.00006-4.
133. F. E. Rey, E. K. Heiniger, C. S. Harwood, Redirection of metabolism for biological hydrogen production. *Appl. Environ. Microbiol.* **73**, 1665–1671 (2007).
134. J. B. McKinlay, C. S. Harwood, Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **107**,

11669–11675 (2010).
135. T. Shimizu, H. Teramoto, M. Inui, Engineering the transcriptional activator NifA for the construction of Rhodobacter sphaeroides strains that produce hydrogen gas constitutively. *Appl. Microbiol. Biotechnol.* **103**, 9739–9749 (2019).
136. B. M. Barney, L. J. Eberhart, J. M. Ohlert, C. M. Knutson, M. H. Plunkett, Gene deletions resulting in increased nitrogen release by *Azotobacter vinelandii*: Application of a novel nitrogen biosensor. *Appl. Environ. Microbiol.* **81**, 4316–4328 (2015).
137. T. Zhang, *et al.*, Involvement of the ammonium transporter AmtB in nitrogenase regulation and ammonium excretion in *Pseudomonas stutzeri* A1501. *Res. Microbiol.* **163**, 332–339 (2012).
138. L. F. Huergo, G. Chandra, M. Merrick, PII signal transduction proteins: Nitrogen regulation and beyond. *FEMS Microbiol. Rev.* **37**, 251–283 (2013).
139. D. E. Martin, B. Reinhold-Hurek, Distinct Roles of PII-Like Signal Transmitter Proteins and amtB in Regulation of nif Gene Expression, Nitrogenase Activity, and Posttranslational Modification of NifH in *Azoarcus* sp. Strain BH72. *Microbiology* **184**, 2251–2259 (2002).
140. A. Sarkar, J. Köhler, T. Hurek, B. Reinhold-Hurek, A novel regulatory role of the Rnf complex of *Azoarcus* sp. strain BH72. *Mol. Microbiol.* **83**, 408–422 (2012).
141. R. N. Ledbetter, *et al.*, The Electron Bifurcating FixABCX Protein Complex from *Azotobacter vinelandii*: Generation of Low-Potential Reducing Equivalents for Nitrogenase Catalysis. *Biochemistry* **56**, 4177–4190 (2017).
142. T. Edgren, S. Nordlund, The *fixABCX* Genes in *Rhodospirillum rubrum* Encode a Putative Membrane Complex Participating in Electron Transfer to Nitrogenase. *J. Bacteriol.* **186**, 2052–2060 (2004).
143. B. Aquino, *et al.*, Effect of point mutations on *Herbaspirillum seropedicae* NifA activity. *Brazilian J. Med. Biol. Res.* **48**, 683–690 (2015).
144. R. A. Monteiro, *et al.*, Expression and functional analysis of an N-truncated NifA protein of *Herbaspirillum seropedicae*. *FEBS Lett.* **447**, 283–286 (1999).
145. Y. Oda, *et al.*, Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. *J. Bacteriol.* **187**, 7784–7794 (2005).
146. L. J. Eberhart, C. M. Knutson, B. M. Barney, A methodology for markerless genetic modifications in *Azotobacter vinelandii*. *J. Appl. Microbiol.* **120**, 1595–1604 (2016).
147. A. R. Hall, D. C. Angst, K. T. Schiessl, M. Ackermann, Costs of antibiotic resistance - Separating trait effects and selective effects. *Evol. Appl.* **8**, 261–272 (2015).
148. R. Miah, *et al.*, Major aldehyde dehydrogenase AldFGH of *Gluconacetobacter diazotrophicus* is independent of pyrroloquinoline quinone but dependent on molybdopterin for acetic acid fermentation. *Appl. Microbiol. Biotechnol.* **105**, 2341–2350 (2021).

Appendix

Table A1. Expansion factors of t2 samples (Tn-seq study, Chapters 2 and 3).

Selective condition	Replicate	OD600	Growth Time (h)
Sucrose NH4	1	1.86	42.38
Sucrose NH4	2	2.31	42.4
Sucrose N2 Fix	1	0.5	75.2
Sucrose N2 Fix	2	0.54	116.18
GAD	1	2.2	18.33
GAD	2	2.15	18.33
Sucrose ASN	1	2.5	42.45
Sucrose ASN	2	2.1	42.47
Galactose NH4	1	1.9	75.2
Galactose NH4	2	1.7	75.2

Table A2. Mapped reads (Tn-seq study, Chapters 2 and 3).

Strain	Selective condition	Time	Replicate 1	Replicate 2	Replicate 3
GDIA	NA	t1	22,387,551	18,951,263	20,632,937
GDIA	Sucrose NH4	t2	13,641,334	19,007,638	-
GDIA	Sucrose Nitrogen Fixation	t2	18,156,008	17,346,979	-
GDIA	GAD	t2	20,451,044	20,061,635	-
GDIA	Sucrose ASN	t2	20,728,852	25,562,162	-
GDIA	Galactose NH4	t2	27,242,591	30,735,334	-