Functional characterization and bioinformatic analysis of glutamine synthetases of *Ochrobactrum anthropi*

by

Saud Alzahrani

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2020

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Glutamine synthetase (GS) is an essential enzyme required for the conversion of ammonium (NH^{4+}) into glutamine. It is commonly classified into GSI, GSII, and GSIII based on molecular size, number of subunits, underlying regulation and enzyme structure. GSI is further subdivided into GSI-α and GSI-β. The latter form contains a conserved motif (NLYDLP) for the adenylation of a tyrosine residue near the active site and insertion of a specific 25-amino acid residue domain. Five ORFs were described to have GS activity in the genome of *Ochrobactrum anthropi* ATCC 49188. However, there are no studies regarding their functional significance and bioinformatic analysis. Therefore, with thorough bioinformatic analysis, I identified and classified these five putative distantly related GSs. Moreover, I modified the previously reported conserved motif (NLYDLP) for adenylation of tyrosine at the N-terminus of GSI-β to N/D-LYDLP. Using this modified motif as criterion as well as insertion of specific 25 amino acids, I identified the chromosome I GS (Oant_2087) of *O. anthropi* as GSI-β. Since those features were absent in the GS from pONAT01(Oant_4491) and the two GS of chromosome II (Oant_3936 and Oant_3881), they were identified as GSI-α. Chromosome II GS (Oant_4157) is GSII type. Further, my results from the bioinformatic analysis strongly indicate that GS on pONAT01 was acquired through horizontal gene transfer from either *Ensifer adhaerens* plasmid or *Ensifer adhaerens* chromosome 1. Interestingly, this transferred enzyme was found to be functional in *O. anthropi* as knocking it out from pONAT01 of *O. anthropi* resulted in 50% reduction in enzyme activity. All these findings will provide an insight the underlying mechanism of regulation for the five GSs present in *O. anthropi* and could serve as the basis for further investigation into the molecular functions of these five GSs, and the plasmid based one in particular.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation and gratitude to my supervisor Dr. Trevor Charles for accepting me to do this study in his laboratory as well as sharing his excellent knowledge and rich experience in the field of bacterial genetics and analyzing genetic material throughout my stay in his lab. Dr. Charles is more than a mentor for me, and his pretty friendly approach is unforgettable. A great debt of gratitude is my mother who has been providing me unconditional love and encouragement though out my study. I appreciate all my family for their continued support during my stay in Canada. I am also grateful to my colleagues, Teshager Kefale and Dr. Sumera Perveen, for providing me guidance and technical support on my experiments throughout my stay in the lab and enriching me with their great lab experience. Teshager and Sumera are such a wonderful co-worker and had so much fun working with them. I am very thankful to my supervisory committee members, Dr. McConkey and Dr. Doxey, for their great supervision and contributions towards the successful completion of this study. Finally, I would like to extend my gratitude to all Dr. Charles lab members and all friends in Canada. Last but not least, I also thank Elaine Lincoln who made my stay in Canada comfortable.

DEDICATION

This MSc thesis is dedicated to my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DNA	deoxyribonucleic acid
µ	Micro
oriT	origin of transfer
sacB	levansucrase gene
bp	base pair
oriV	origin of replication
GS	Glutamine synthetase
G	Gram
GC	gas chromatography
Gm	Gentamicin
kb	kilobase pair
Mb	megabase pair
LB	lysogeny broth
OD	optical density
OD600	optical density at 600 nm
PCR	polymerase chain reaction
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SOB	super optimal broth
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside
dnaA	based replication system
EDTA	ethylenediaminetetraacetic acid
BBII	Binding Buffer II
M	Molar
m	Milli
mM	Millimolar
nm	Nanometer

1. INTRODUCTION

Nitrogen is an essential element required for biosynthesis of nucleic acids, amino-sugars and proteins (Reitzer, 1987) and many bacteria meet their nitrogen need though assimilation of the preferred nitrogen source, ammonium (NH^{+4}), into organic compounds using an enzyme called glutamine synthetase (GS) (Pesole, Gissi, Lanave, & Saccone, 1995). Because of its vital role, this enzyme is ubiquitous and well conserved from unicellular organisms to mammals (Pesole et al., 1995) and the overall chemical reaction this enzyme catalyze is:

$$\text{Glutamate} + NH_3 + \text{ATP} \longrightarrow \text{Glutamine} + \text{ADP} + P_i$$

GS is commonly classified into GSI, GSII, and GSIII based on molecular size, number of subunits, underlying regulation and enzyme structure (Brown, Masuchi, Robb, & Doolittlel, 1994; Liu et al., 2018). GSI and GSIII typically exist in bacteria and archaea in a dodecamer form, consisting of two back-to-back hexameric rings, while GSII predominantly present in eukaryotes with decamer form, which is comprised of two back-to-back pentameric rings (Patel, 2015). In terms of size, GSIII is the largest of all followed by GSII (Brown, Masuchi, Robb, & Doolittlel, 1994; Liu et al., 2018).

GSI-α is mainly found in Gram-positive bacteria and thermophilic bacteria whereas GSI-β are found in other bacteria species (Fisher, 1999; Shapiro, Kingdon, & Stadtman, 1967). In addition, a conserved motif (NLYDLP) for the adenylation of a tyrosine residue near the active site and insertion of a specific 25-amino acid residues motif (residues 146–170 AA in *E. coli*) distinguishes GSI-β from GSI-α form (Brown et al., 1994; Joo et al., 2018). As a result, they have different regulatory mechanisms to modulate their activity

through posttranslational adenylation and feedback inhibition by nitrogenous compounds (Goss, Perez-Matos, & Bender, 2001).

Although GS is one of the most extensively studied enzymes and many have been characterized to date (Chen & Silflow, 1996; Deuel, Ginsburg, Yeh, Shelton, & Stadtman, 1970; Joo et al., 2018; Liu et al., 2018; Lu et al., 2019; Reinecke, Zarka, Leu, & Boussiba, 2016), it is still of paramount importance to identify more efficient ones. Given that most characterized GSs are usually not free for commercial use (Abad, Coffin, & Goldman, 2015), finding alternatives which have higher enzyme activities and minimum regulation could be useful for synthetic biology applications.

There are five putative GSs present in the genome of *O. anthropi* ATCC 49188. However, functional properties of these GSs have not been studied. In depth bioinformatic analysis for these enzymes is also lacking. Therefore, this thesis project first focused on functionally characterizing the GS present on pOANT01 via gene manipulation, growth assay and enzyme activity study, followed by a through bioinformatic analysis that enabled to identify these potential GSs into their respective forms, evolutionary relationships of these GSs and the possible origin of the GS contained in plasmid on pOANT01.

To date, the putative GSs of *O. anthropi* have not been identified into GSI-α and GSI-β. In this study, using predicted 3D structures, biochemical analysis, amino acids alignments and conserved motif analysis (i. e N/D-LYDLP), chromosome I GS (Oant_2087) was classified as GSI-β whereas the GS from pONAT01(Oant_4491) and the two GS of chromosome II (Oant_3936 and Oant_3881) were classified as GSI-α. Moreover, using multiple genome alignment tools and analysis of a 10-kb conserved sequences (containing GS plus 8 other flanking genes), *Ensifer adhaerens* strain Casida A plasmid pCasidaAB was found to have common origin for the GS on pOANT01. My bioinformatics analysis strongly

suggests that the GS on pOANT01 was acquired by neither gene duplication from its own chromosomal GS nor vertical gene transfer but rather by horizontal gene transfer from a plasmid. These results will be discussed in detail in this thesis work and, in general, the findings from this work will certainly provide an insight for underlying mechanism of regulation for the five GSs present in *O. anthropi* and could serve as the basis for the further investigation into the molecular functions of these five GSs, particularly for the plasmid based one.

2. LITERATURE REVIEW

2.1 Bacterial plasmids

Bacterial plasmids can be defined as a class of circular extrachromosomal DNA that exist and replicate independently from the chromosomal DNA. They have the ability to be transferred between different hosts and mostly contain non-essential genes (Skippington & Ragan, 2011). They vary in length from a few to several hundred kilobase pairs and encode genes that confer accessory functions such as antibiotic resistance, heavy metal resistance and utilization of toxic compounds (Allen et al., 2010; Bruins, Kapil, & Oehme, 2000). These features enable bacteria to adapt and survive in various environmental niches (Gyles & Boerlin, 2014). Their classification is challenging due to very diverse characteristics. Some plasmids are circular in configuration, while others are linear, their copy number varies from one to several hundred per cell, and various methods of replication have been identified. Understanding their strategies for replication, which affect their copy number, host-range and dependence, and response to environmental conditions (Bengtsson-Palme, Kristiansson, & Larsson, 2017) could provide insight into the role of plasmids.

Bacterial plasmids have modular structures, making it possible to group them into several functional genetic modules (Fig. 1). Plasmid replication and its control are usually located in a region called the basic replicon. The genes encoding Rep proteins are required for replication, and often participate in its control. The plasmid backbone is composed of a set of conserved modules, coding for replication, stability and conjugal transfer functions, which are crucial for plasmid maintenance and spread (Dziewit et al., 2015). This genetic information is not important to determine the host viability, but it plays a crucial role in the adaptability of bacteria to diverse environments (Heuer & Smalla, 2012). Studies have proven that the most common phenotypic modules present in bacterial plasmids are heavy metal resistance genes. Furthermore, metal resistance genes are sometime contained together with antibiotic resistance genes on the plasmids, and they are frequently present within transposable and integrative mobile elements (Fang et al., 2016). They contain genes that are essential for their maintenance and functions, for example the initiation and control of replication. Some contain genes that control traits ensuring stable inheritance, such as equipartitioning during cell division or conjugal transfer (Carattoli, 2003). Many plasmids contain genes that are useful not only to themselves, but also to their host, examples of such are genes controlling drug resistance, degradation of organic compounds, and virulence factors, including the production of toxins (Couturier, Bex, Bergquist, & Maas, 1988). These kinds of genes are often found within transposons, and this has led to a great deal of variation and flexibility in the composition of plasmids (Heuer & Smalla, 2012).

Plasmids containing heavy metal resistance systems are typically related to chromosomal-encoded determinants found in other bacteria, this demonstrates the natural flow of genes among bacteria (Dziewit et al., 2015). Horizontal gene transfer among bacteria promote diversity and adaptability in the host cell, plasmids contribute significantly to this

process. Many plasmids are self-transferable replicons, that can be transmitted from one host to another, together with other genes embedded within their transposons, that are of great importance during selection process. This process is highly valuable to bacteria and, help advancing the process of evolution (Aminov, 2011).

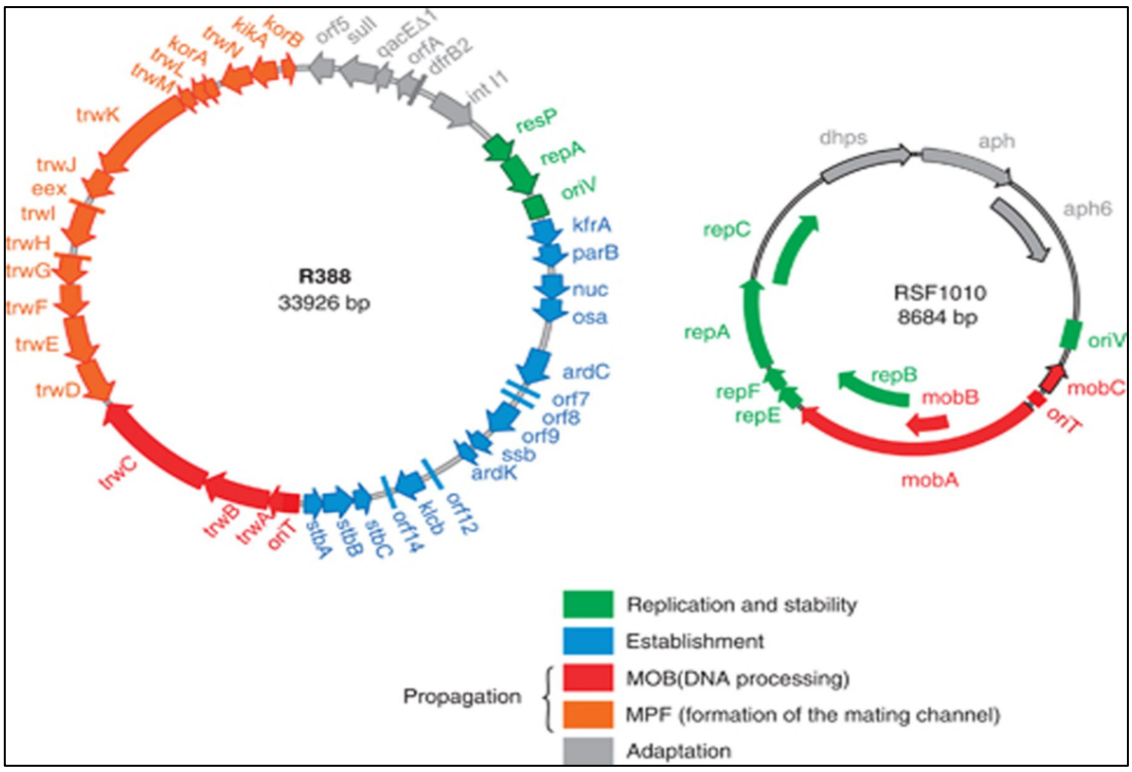

Fig 1. Modular genetic organization of two plasmids. The genes are represented in different colors according to their functional module. Adapted from Garcillán-Barcia, Alvarado, and de la Cruz (2011).

2.2 RepABC plasmid family

Members of the *repABC* plasmid family have several properties in common. One of the noticeable features is that all elements essential for plasmid partitioning and replication, with those responsible for incompatibility, are located in the same operon (Żebracki, Koper, Marczak, Skorupska, & Mazur, 2015). The *repABC* operons sequenced to date share some general characteristics. All of them contain at least three protein-encoding genes: *repA*, *repB* and *repC*. The first two genes encode proteins involved in plasmid segregation while *repC* encodes protein crucial for replication (Cevallos, Cervantes-Rivera, & Gutiérrez-Ríos, 2008; Pinto, Pappas, & Winans, 2012). The genetic organization of the *repABC* cassette is well conserved: *repA* is upstream of *repB*, and both precede *repC* (Fig. 2). The origin of replication maps within the *repC* gene. Despite their apparent structural homogeneity, *repABC* operons have diverse DNA sequences (MacLellan, Zaheer, Sartor, MacLean, & Finan, 2006; Żebracki et al., 2015). There are variations based on the presence of peptide-encoding minigenes, the numbers and class of the regulatory elements involved in operon transcription, and the numbers and positions of centromeric *parS* sequences (Cevallos et al., 2008). The structural diversity of *repABC* operons resulting from their complex and independent evolution tend to affect the regulation and function of certain replication elements.

The *repABC* replicons are found in large plasmids of low copy number and chromosomes (Cervantes-Rivera, Pedraza-López, Pérez-Segura, & Cevallos, 2011; Ramírez-Romero et al., 2000). They have been identified in some α-proteobacteria such as *Rhizobium*, *Mesorhizobium*, *Sinorhizobium*, *Agrobacterium*, *Rhodobacter*, *Ruegeria*, *Paracoccus*,

Agrobacterium, *Ochrobactrum* and *Brucella*. Some strains of these bacterial genera contain multiple *repABC* replicons indicating that this plasmid family includes several incompatibility groups. Native plasmids could be suitable candidates for the construction of vectors for genetic manipulations of biotechnologically important bacteria. However, this requires insight into the biology and function of plasmids of interest.

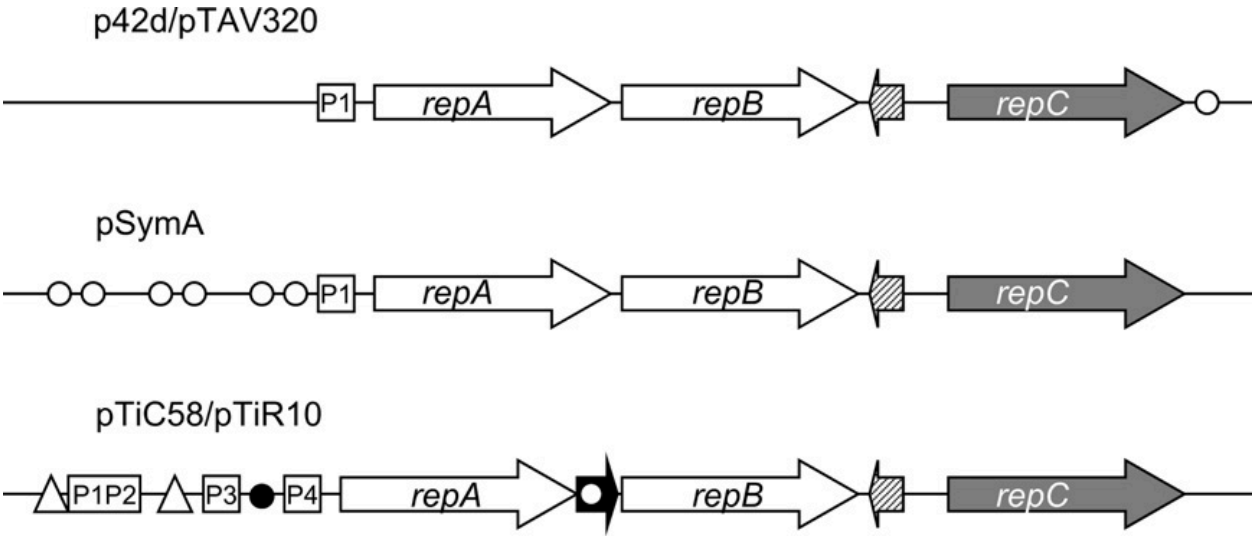

Fig 2. The diversity of the genetic organization of *repABC* replicons in different plasmids. Adapted from Cevallos et al. (2008).

The replication, partitioning and stable maintenance of these replicons depend on the presence of a *repABC* operon. The *repABC* replicons contain an operon encoding the initiator protein (RepC) and partitioning proteins (RepA and RepB). RepA and RepB encoding genes belong to the families of ParA and ParB which are partitioning proteins respectively. RepA and RepB play dual roles in plasmid maintenance. In conjunction with *parS* centromere-like sites, they participate in partitioning and in the negative transcriptional regulation of their

own operons (Pinto et al., 2012; Soberón, Venkova-Canova, Ramírez-Romero, Téllez-Sosa, & Cevallos, 2004). This is required for efficient maintenance of the low copy DNA replicons in a dividing population of cells (Mazur & Koper, 2012). The third protein-encoding gene of the operon, *repC*, is essential for plasmid replication. RepC is an initiator protein, function by binding to the origin of replication *oriV* located within its own coding sequence, close to or inside of a large A+T region. The *repC* gene is the minimal region sufficient for replication when inserted into a non-replicating vector (Mazur & Koper, 2012). Mazur, Majewska, Stasiak, Wielbo, and Skorupska (2011) reported that the genome of *Rhizobium leguminosarum* bv. *Trifolii* TA1 (RtTA1) consists of five replicons and that all these RtTA1 plasmids are equipped with functional *repABC* genes.

2.3 *Ochrobactrum anthropi*

Ochrobactrum spp. belongs to the *Brucellaceae* and is known to be isolated from *Leguminosae* nodules. Its name is derived from the Greek *ochros*, meaning pale yellow; this is the characteristic color of *Ochrobactrum* colonies on plates (Cevallos et al., 2008). This genus was first described by Holmes in 1988 and 19 species have been identified to date, which include *O. anthropi*, *O. intermedium* and *O. pseudintermedium* that have been reported in clinical samples (Hagiya, Ohnishi, Maki, Watanabe, & Murase, 2013; Teyssier et al., 2007). *Ochrobactrum anthropi* is a non-fastidious, strictly aerobic, motile (with peritrichous flagella), oxidase-positive, non-fermenting, gram-negative bacillus with strong urease activity. It is a soil bacterium that colonizes a wide variety of habitats and is being increasingly recognized as an opportunistic human pathogen (Chain et al., 2011). Potentially life-threatening infections, such as endocarditis, are included in the list of reported infections caused by *O. anthropi* (Thoma et al., 2009). This opportunistic pathogen has exhibited

resistance to β-lactam antibiotics. In addition, the *O. anthropi* genome contains four plasmids pOAN01, pOAN02, pOAN03 and pOAN04 which are found to be linked to antibiotic resistance exhibited by this pathogen (Chain et al., 2011).

2.4 Molecular content of *O. anthropi*

A classical genome content consists of one or more central chromosome(s) made up of housekeeping genes for fundamental metabolic function and plasmid(s) coding for secondary features. The primary chromosome possesses *dnaA*-based replication system and controls crucial genes involved in cellular processing such as transcription, translation, DNA replication, and energy metabolism (Landeta et al., 2011), whereas the secondary chromosomes contain certain vital genes that are absent in the primary chromosome. Most of these secondary chromosomes possess duplicates of these genes which contribute towards cell survival (Harrison, Lower, Kim, & Young, 2010). The whole genome of *O. anthropi* has been sequenced and deposited in public databases. Chain et al. (2011) reported the draft sequenced of *O. anthropi* ATCC 49188T. The whole genome consists of 4.8 Mb of circular chromosomes with *repABC* origin. The molecular composition is made up of 56% GC content, 4,424 protein-coding genes (87% coding), along with 31 pseudogenes and 73 structural RNAs (rRNA, tRNA, and small RNA) (Chain et al., 2011).

2.5 Functions of plasmids in *O. anthropi*

RepABC plasmids are commonly found in alpha-proteobacteria species and particularly among the order *Rhizobiales*. In some species the *repABC* plasmids constitute about 35% of the genomic content, containing several incompatibility groups i.e. more than

one type of *repABC* plasmid can be found in the same bacterial species (Castillo-Ramírez, Vázquez-Castellanos, González, & Cevallos, 2009). The substantial amounts of plasmids that make up the genome of the organism, the conjugative ability of these plasmids, and multiplicity of the plasmids serve as a platform for researchers to have an insight into the structure and functions of plasmids?. The plasmids are composed of replicons with various sizes and diversity in their conjugative systems (Cevallos et al., 2002). The phenotypes of the plasmids are wide-ranging and vary from replicon to replicon, and from species to species.

These replicons are defined by the presence of a *repABC* operon, whereas the replicon itself is a large plasmid (ranging in size from roughly 50 kb to 1.35 Mb) (Cevallos et al., 2008). The *repABC* operon is highly unusual in that it arranges the replication and partitioning functions under the control of a single promoter. RepA and RepB form a partitioning cassette with ATPase domain that provides the energy needed for the partitioning machinery (Pinto et al., 2012). It has been investigated that RepB is responsible for recognition and binding to the centromere-like site on the replicon as well as binding RepA, and also acting as an adaptor between the replicon and the partition motor (Yip, Ding, & Hynes, 2015). Studies suggested that RepB binds only with low affinity to the centromere-like site, meanwhile the binding of RepA greatly increases the affinity for this interaction (Pappas & Winans, 2003). Research has shown that RepC is the limiting factor in replication and it has been found to be the initiator protein (Cervantes-Rivera et al., 2011; Lilly & Camps, 2015).

Chain et al. (2011) reported that the genome of *O. anthropi* contains four plasmids (pOAN01, pOAN02, pOAN03 and pOAN04) and three of these plasmids contain one or

more RepABC or replication-partition systems (Chain et al., 2011). They are also found to contain genes responsible for stabilization (Pinto et al., 2012). The first three plasmids have RepABC and/or RepC including a large number of genes encoding transposases and integrases from different families, whereas pOAN04 was found to lack such genes.There is a set of complete type IV secretion system that indicates it is self-transmissible, and pOAN02 is predicted to be mobilizable (coding for antirestriction and mobilization proteins). These four plasmids are predicted to contribute to the stability of the organism, as they contain several transporters, although metabolic genes are found only in pOAN01, pOAN03, and pOAN04 (Chain et al., 2011).

2.6 Host range of plasmid RepABC

The transfer of plasmids between bacteria confers pathogenically and environmentally relevant features that promotes evolution and adaptation essential for their survival. Plasmids are usually transmitted by conjugation; this is one of the most effective machineries to distribute genetic materials among bacteria. It is important to understand the relationship between plasmid traits and host taxonomy in order to comprehend the proliferation of plasmids among microbes. Information about the host range of these plasmids is critical to effectively use them as genetic tools for microbial engineering (Shintani, Sanchez, & Kimbara, 2015). Several plasmids in *Alphaproteobacteria* carry genes encoding RepABC proteins (Cevallos et al., 2008).

Bartosik, Baj, and Wlodarczyk (1998) reported the molecular and functional analysis of pTAV320, a *repABC*-type replicon in *Paracoccus versutus,* previously *Thiobacillus versutus,* which a Gram-negative, facultatively chemolithoautotrophic soil bacterium

belonging to the *Proteobacteria* subclass. *P. versutus* UW1 harbours a large (107 kb), low-copy-number, cryptic plasmid, pTAVl (Bednarska *et al.,* 1983), which is stably maintained in its bacterial host. The replication system appears to be related to several plasmids commonly found in Gram- negative soil bacteria. It was discovered that the putative translation products of pTAV320 show a significant and very similar level of homology to previously reported RepA, RepB and RepC proteins (Ramirez-Romero et al., 1997; Sarah L. Turner & Young, 1995).

2.7 Incompatibility of RepABC plasmid

Ramírez-Romero et al. (2000) defined plasmid incompatibility as the inability of two different plasmids to reside in the same host cell as independent replicons in the absence of selective pressure because of sharing similar replication and/or partition systems. A DNA fragment was considered to exhibit incompatibility if its introduction into host cell (i) caused the displacement of the symbiotic plasmid or (ii) induced the cointegration of the symbiotic plasmid with another cognate plasmid (Ramírez-Romero et al., 2000). Plasmid incompatibility is a limiting factor in the acquisition of new plasmids and plays a crucial role in modeling bacterial evolution. Perez-Oseguera & Cevallos (2013) indicated that plasmid incompatibility is as a result of functional interference between the replication systems and/or partitioning mechanism of the plasmids involved. In all *repABC* loci that have been characterized, an apparently untranslated intergenic region between the *repB* and *repC* genes encodes a strong incompatibility determinant (referred to as incα). Yip et al. (2015) reported that *repABC* operons of the three largest plasmids of VF39SM were found to have strong incompatibility determinants in the non-protein coding regions. It was discovered that in the *repABC* operons, the intergenic region between *repB* and *repC* was the strongest

incompatibility factor. This region for many *repABC* plasmids encoded a counter-transcribed RNA (ctRNA) which regulates RepC abundance, making it to also control the rate of initiation of replication. They were able to establish that ctRNA controls replication and incompatibility.

2.8 Toxin-antitoxin of RepABC

A majority of bacteria contain toxin coding genes which could inhibit cell growth by targeting a central molecule in a crucial cellular process such as DNA replication, mRNA stability, protein synthesis and several other significant biosynthesis (Yamaguchi, Park, & Inouye, 2011). These toxins are co-transcribed and co-translated with their cognate antitoxins from an operon called a toxin-antitoxin (TA) operon.

2.9 Glutamine synthetase (GS)

Glutamine synthetase (GS) is a ubiquitous enzyme present in all organisms and commonly classified into GSI, GSII, and GSIII based on molecular size, number of subunits, underlying regulation and enzyme structure (Brown et al., 1994; Liu et al., 2018; Shatters, Liu, & Kahn, 1993; Somerville, Shatters, & Kahn, 1989). GSI and GSIII typically exist in bacteria and archaea in a dodecamer form (consisting of two back-to-back hexameric rings) whereas GSII predominantly present in eukaryotes with decamer form, comprised of two back-to-back pentameric rings (Patel, 2015).

GSI is encoded by the *glnA* gene and the best characterized among all glutamine synthetases (Gill, Pfluegl, & Eisenberg, 2002). Most enteric bacteria including *E. coli* contain one *glnA* gene and mutants deficient in this gene are glutamine auxotrophs (L J Reitzer et al., 1987; Wei & Kustu, 1981). GSII is encoded by *glnII* gene whose protein is smaller than GSI (~370 residues average length) and has been less studied than its prokaryotic counterpart,

both in functional and structural terms (Gill et al., 2002). Most bacteria in the Rhizobiales order possess both GSI and GSII (R L Fuchs & D L Keister, 1980; R. L. Fuchs & D. L. Keister, 1980) and a single enzyme encoded by one of these genes was reported to be sufficient for glutamine prototrophy in *Rhizobium meliloti* (Somerville et al., 1989). In addition, an *R. meliloti* double mutant was found to be not a strict glutamine auxotroph as it could grow on media supplemented with glutamate and ammonia (Somerville et al., 1989). This observation later led to the discovery of GSIII in this bacterium (Shatters et al., 1993).

GSI is further subdivided into GSI-α and GSI-β (Brown et al., 1994; Joo et al., 2018). GSI-α is mainly found in gram-positive bacteria and thermophilic bacteria whereas GSI-β are found in other bacteria species (Fisher, 1999; Shapiro et al., 1967). In addition, a conserved motif (NLYDLP) for the adenylation of a tyrosine residue near the active site and insertion of a specific 25-amino acid residues (i.e. residues 146–170 AA in *E. coli*) distinguishes GSI-β from GSI-α form (Brown et al., 1994; Joo et al., 2018). As a result, they have different regulatory mechanisms to modulate their activity through posttranslational adenylation and feedback inhibition by nitrogenous compounds (Goss et al., 2001).

2.10 GS enzymatic assays

GS enzymatic activity has mainly been quantified using either biosynthetic assay or transferase assay (Bressler & Ahmed, 1984; Dharmawardene, Haystead, & Stewart, 1973; Gawronski & Benson, 2004). In the biosynthetic assay, GS activity is measured spectrophotometrically by determining the amount of inorganic phosphate released from its physiological substrates. To set up this assay reactions, 1.0 M Imidazole-HC1 buffer (pH 7.0), 10.0 mM NH_4Cl, 60.0 mM ATP (pH 7.0), 1.67 M $MgC1_2$. 6H20, 1.0 M Na-glutamate

(pH 7.0) and cell-free extract are mixed. The reactions are then initiated by incubating for 30 min at 25°C and stopped by addition of $FeSO_4.7H_20$ solution (0.8 % w/v in 0.015N H_2S0_4). Color develops when ammonium molybdate (6.6 % w/v in 7.5 N H_2S0_4) is added and the amount of liberated inorganic phosphate is measured by spectrophotometer at 850 nm (Bressler & Ahmed, 1984; Dharmawardene et al., 1973; Gawronski & Benson, 2004). To set up reactions for transferase assay, 1M Imidazole-HC1 buffer (pH 7.3), 0.1M glutamine, 0.01M $MnCl_2$, 0.01M ADP (pH 7.3), 1M K-arsenate, 2M Hydroxylamine and cell-free extract are added. The reaction is then stopped after 30 min by the addition a mixture of 10 % $FeCl_3$, 24 % trichloroacetic acid, 6N HCl and 6.5 m1 double-distilled water. Reading for this assay is done by measuring absorbance at 540 nm using spectrophotometer(Bressler & Ahmed, 1984; Dharmawardene et al., 1973; Gawronski & Benson, 2004).

2.11 Glutamine synthetases (GSs) in *O. anthropi*

Five GS open reading frames are present in *O. anthropi.* The ORFs encoding enzymes have not been well characterized and identified, particularly the one from the plasmid pOANT01. Therefore, in depth functional and bioinformatic study could broaden our current understanding of these GSs functions and regulations.

3. HYPOTHESES AND OBJECTIVES

3.1 Hypotheses

- GS present in pOANT01 plays a role in nitrogen assimilation in *O. anthropi*
- The chromosomal GSs and the GS present on plasmid pOANT01 of *O. anthropi* are different in terms of evolution and they are distantly related
- The five ORFs predicted to have GS activity in the genome of *O. anthropi* are classified into different GS forms

3.2 Objectives

- To investigate the function/phenotype of glutamine synthetase, which is present in pOANT01.
- To identify and classify the five putative GSs present in *O. anthropi*
- To study the evolution of the five putative GSs present in *O. anthropi*
- Identify the possible origin of the GS from plasmid pOANT01 of *O. anthropi*

4. MATERIALS AND METHODS

4.1 Total DNA extraction

For extraction of total DNA, pellets were harvested in 2 ml tubes from 3 ml overnight cultures and resuspended in 200 µl of 0.2 M NaCl. Into these solutions, 20 µl of 0.2M EDTA, 10 µl of 10% SDS, and 5 µl of 10 mg/ml pronase were added sequentially and the contents were then heated up to 70^{o}C for 1 h. After cooling down the solutions briefly, 300 µl of chloroform was added into the tubes, vortexed and centrifuged at maximum speed for 10 min for phase separation. The aqueous layers were then removed and dispensed into new tubes. To precipitate DNA, 450 µl of ice cold 95% ethanol was added, tubes were vortexed and centrifuged at maximum speed for 15 min followed by removal of the supernatant. Further washing of DNA pellets were done with 1 ml of 70% ethanol. Eventually, after drying the DNA pellets for about 10-20 minutes, 50 µl of 2 mM Tris were added to dissolve the pellets.

4.2 PCR amplification

Standard PCR was performed under the following conditions: 98°C for 10 s, 53°C for 30 s, and 68°C for 1 min and 35 cycles using Taq DNA polymerase for the amplification of 343 bp of GS from pONAT01 plasmid. 2 µL genomic DNA was used as the template for PCR amplification. 1 µl of GS-FOR and GS-REV primers (10 µM) plus 25 µL of 2X Taq master mix from Thermo Scientific/Fisher were added to the template. Then, mixing the PCR reaction was done by flicking the tube and spin down the liquid from the sides of the tube before running it in a thermocycler machine.

4.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyze the size and condition of DNA fragments from a variety of processes. Routine agarose gel electrophoresis was carried out using a 0.8-1% agarose gel in 1X Tris-Acetate-EDTA (TAE) buffer. The running buffer was also 1X TAE. The protocol was adapted from Sambrook and Russell (2001).

4.4 Gel extraction

The desired fragment was cut and placed in an Eppendorf tube containing 4 volumes of Binding Buffer II (BBII). The content was incubated at 65ºC for 5 min to melt the gel. After mixing the solution by inverting the tube few times, the solution was then loaded on a spin column and eluted by centrifuging at 10,000 rpm for 1 min. The flow through was discarded. The column was washed twice with 500 µl of Wash Buffer and centrifuge for 1 min at 10,000 rpm. Finally, the DNA was eluted into a new microcentrifuge tube with 50 µL of Elution Buffer and quantified using Nanodrop.

4.5 Plasmid construction

To construct pK19mobsacB-GS-343 plasmid for disrupting GS by single crossover and remove pONAT01, 343 bp GS fragment was amplified using primers (GS-FOR and GS-REV, Table 1) and digested with HindIII and PstI. The vector, pK19mobsacB, was also digested with the same enzymes. Then, the purified insert and vector were placed in a tube in 3:1 molar ratio along with 10x ligase buffer, 0.5 µl of T4 DNA ligase, and ddH_20. The content was flicked to mix, quickly centrifuged to collect the liquid at the bottom of the tube and incubated at room temperature for 30 min to allow ligation to occur.

Table 1. Bacterial strains, plasmids, and primers used in this study

Strain, plasmid, or primer	Relevant characteristic(s) or sequence
Ochrobacterum strain Oa49188	WT isolate
Plasmids	
pK19mobsacB	Suicide vector, sacB, mobRK2, oriR6K, Nm-Km
pRK600	Mobilizer plasmid for conjugal transfer, Cm Nm-Km
E. coli	
DH5α	F– φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1
Primers:	
GS-SC-FOR-	CGCGAAGCTTTCCGGCCTGACGGAACTTGCGAC
GS-SC-REV-	CGGCCTGCAGCGAACTGGGAGAGCGTCGCCTTCAG
KanR_F1	ATGATTGAACAAGATGGATTGCACG
KanR_R1:	TCAGAAGAACTCGTCAAGAAGGC
pK19mobsacB.FOR	GTCTCGATATGAATTACGCCAAGCTTGCATGC
p15A-888.FOR	GTTCTTTCCGCTCCAAGCACTAGTAACAACTTATAT CG.
888-p15A.REV	TACTAGTGCTTGGAGCGGAAAGAACCGCAGC
982-p15A.FOR	CCTAGGTATAAAATCGTGCTCTTCACCGAC
982-GmR.REV	CATGGAGATAAGCGCTGTTCCCCTTTGAAGCGAT
SucB-GmR.FOR	AGCGTCAGACCCCGGGACTCTGGGGTTCG
GmR-SucB.REV	CCCCAGAGTCCCGGGGTCTGACGCTCAGTGGA
888-SucB.FOR	GTTAACAAATAATGACCTCGACCAAGACCA
SucB-888.REV	TTGGTCGAGGTCATTATTTGTTAACTGTTAATTGTC CTTGTTC

4.6 Transformation of bacteria

E. coli (DH5α) cells were made chemically competent and transformation of these cells was performed with the ligation reaction mix mentioned in section 4.5 via heat shock method following the protocol established by (Sambrook & Russell, 2001). Positive DH5α colonies were screened with LB plates containing X-Gal and kanamycin. Afterwards, the plasmid (pK19mobsacB-GS-343) was isolated from a positive colony and transformation of this

plasmid to *O. anthropi* was done by electroporation after confirmation with restriction enzyme.

4.7 Growth assay

For the growth assay, wild type and mutant *O. anthropi* were cultured overnight in LB media with selection. The next day, cells were diluted to 0.1OD_{600} in defined media containing the following nitrogen sources: no nitrogen, 1mM glutamate, 1mM glutamine, 1g/L NH_4Cl, and 1g/L NH_4Cl plus 1mM glutamate. The defined media was composed of M9 salts (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4 and 5 g/L NaCl in the absence of NH_4Cl), trace elements (2.86 g/L H_3Bo_3, 1.81 g/L $MnCl_2.4H_2O$, 0.222 g/L $ZnSo_4.7H_2O$, 0.39 g/L $Na_2MoO_4.2H_2O$, 79 µg/L $CuCl_2.2H_2O$ and 49.4 µg/L $CoSo_4.7H_2O$) and 20 g/L glucose as a carbon source. OD_{600} (cell density) was measured using spectrophotometer up to 48 hours.

4.8 GS biosynthetic enzyme assay

To measure GS activity via biosynthetic method, wild type and mutant cells were grown overnight in LB media with selection. In the following day, cultures were diluted to 0.1 OD_{600} and further grown in super broth with selection. Cell were then harvested at 0.8 OD_{600} by centrifuging at 8000 rpm for 5 min. Before cell lysates preparation, cells were washed 3 times with the widely used S30 buffer (10 mM Tris-acetate buffer (pH 8.2), 14 mM Magnesium acetate, 60 mM potassium acetate, and 1 mM DTT) following previously established protocol (Adachi et al., 2019). Cells were resuspended in S30 having Halt™ protease inhibitor cocktail, sonicated, and lysates obtained by centrifuging at 25,000 g for 30 min at 4°C.

GS biosynthetic enzyme reaction was set up as described in (Bressler & Ahmed, 1984), which contained the following:

200 µL Imidazole-HC1 buffer 1.0 M, pH 7.0

10 µL NH_4Cl 10.0 mM

50 µL ATP 60.0 mM, pH 7.0

50 µL $MgC1_2$. 6H20 1.67 M

50 µL Na-glutamate 1.0 M, pH 7.0

100 µL cell-free extract

460 µL total volume of reaction mixture

The reactions were allowed to take place by incubating for 30 min at 25°C and stopped by addition of 1.8 mL of $FeSO_4.7H_20$ solution (0.8 % w/v in 0.015N H_2S0_4). Thereafter, samples were vortexed and placed on ice followed by addition of 0.15 mL of the color-forming reagent ammonium molybdate (6.6 % w/v in 7.5 N H_2S0_4). The samples were vortexed vigorously again and placed on ice immediately. Reading then was carried out using spectrophotometer at 850 nm. An internal control, which is minus substrates, was included in the experiment and values were deducted to obtain the actual activity of GS. This enzymatic assay was done in duplicates.

4.9 Bioinformatics analysis

To identify the five putative GSs from the genome of *Ochrobactrum anthropi* ATCC 49188, I used phmmer (https://www.ebi.ac.uk/Tools/hmmer/search/phmmer). To do so, I used the amino acids sequences of the catalytic domain of chromosome I GS (Appendix I). Two approaches where utilized to generate the 3D structures of the GSs: 1) built using SWISS-MODEL (https://swissmodel.expasy.org/) by using the amino acid sequences of the putative GSs and 2) acquired from PDB website after obtaining the PDB identifier number for the top hit of the PDB search in NCBI for each putative GS. Then, the result of these two methods were compared side by side for visualization. Amino acids alignments of GS were performed using seaview (http://pbil.univ-lyon1.fr/software/seaview3) (Muscle algorithm) and MEGA10 software. Sequence logo for the amino acid alignments were made using WebLogo 3(https://weblogo.berkeley.edu/logo.cgi). NCBI blastn and blastp were utilized to analyze nucleotide and amino acids, respectively, and gene tree was constructed with Geneious Prime. Criteria for species selection to build the phylogenetic tree were reference sequences, top 5 hits from sequence search result and plus additional sequences to increase species diversification. Multiple genome alignments were carried out using Mauve platform and YASS genomic similarity search tool.

5. RESULTS

5.1 Disruption of glutamine synthetase by single crossover

To disrupt the glutamine synthetase from pONAT01 plasmid of *O.* anthropi by single crossover, I generated the pK19mobsacB-GS-343 plasmid shown in Fig 3A. Transforming this plasmid to *O. anthropi* by electroporation, positive colonies were initially screened on LB plates with 100 μg/ml of G418. The success of single crossover in these colonies is then further investigated by PCR using different sets of primers as shown in Fig 3A. Using set of primers (GS-SC-FOR and GS-SC-REV) outside of recombination site (Fig 3A), genomic DNA of wild type and G418 positive strains (mutants) and Phusion polymerase, the integration of the whole pK19mobsacB-GS-343 into the targeted recombination site was checked via PCR. However, I did not observe the expected 7000 bp band in the mutants (Fig. 3B). Although Phusion polymerase can amplify up to 20 kb, I thought amplifying a shorter fragment using forward primer from the vector (pK19_FOR) and a reverse primer (GS-SC-REV) from pONAT01 could be a better strategy to confirm the success of the single crossover. With this set of primers, a 1000 bp band was expected. As shown in Fig 3C, a band of this size was detected in both wild type and mutants. The presence of this band in the wild type was confusing. Because I thought this could happen from unspecific binding of the primers to other sites, I blasted the primer sequences in *O.* anthropi genome. The blast result showed there are high similar sequences for both primers in plasmid pONAT02 and chromosome I and that could be the reason why the same size band was observed in all samples.

An alternative strategy to verify the integration of the pK19mobsacB-GS-343 into pONAT01 is to amplify the gene coding kanamycin resistance by PCR. For this purpose, I

performed PCR with forward primer (KanR_For) and reverse primer (KanR_Rev), which are highlighted in green boxes in Fig 3A. Unfortunately, KanR gene was detected in both wild type and mutants (Fig 3C). This observation made my effort to confirm the incorporation of pK19mobsacB-GS-343 into pONAT01 by single crossover more complicated.

In general, my attempt to confirm the success of single crossover by PCR was not conclusive for the reasons discussed above. Given that wild type and mutants displayed different band patterns (Fig 3B and C) and only the mutants were resistant to 100 μg/ml of G418, the result somehow indicate that the single crossover could occur in the mutants. To gain more insight regarding the success of the single crossover as well as studying the effect of the disruption on GS function, I performed growth assay under limited nitrogen condition.

5.2 Functional analysis of the loss of glutamine synthetase

5.2.1 Growth assay

Glutamine synthetase is important for the synthesis of glutamine and assimilation of nitrogen and disruption of the gene encoding this enzyme from pONAT01 of *O. anthropi* may cause dependence on exogenous supply of nitrogen. If the gene is disrupted, then the mutant could display a difference in phenotype. However, conformation for disruption of GS by PCR was not convincing enough as discussed in the previous section and I thought functional assay such as growth assay under limited nitrogen condition could shed insight on the difference between wild type and G418 resistant mutant. Therefore, I performed growth assay using minimum media under five conditions (i.e. no nitrogen, glutamate, glutamine, NH_4CL and NH_4CL plus glutamate).

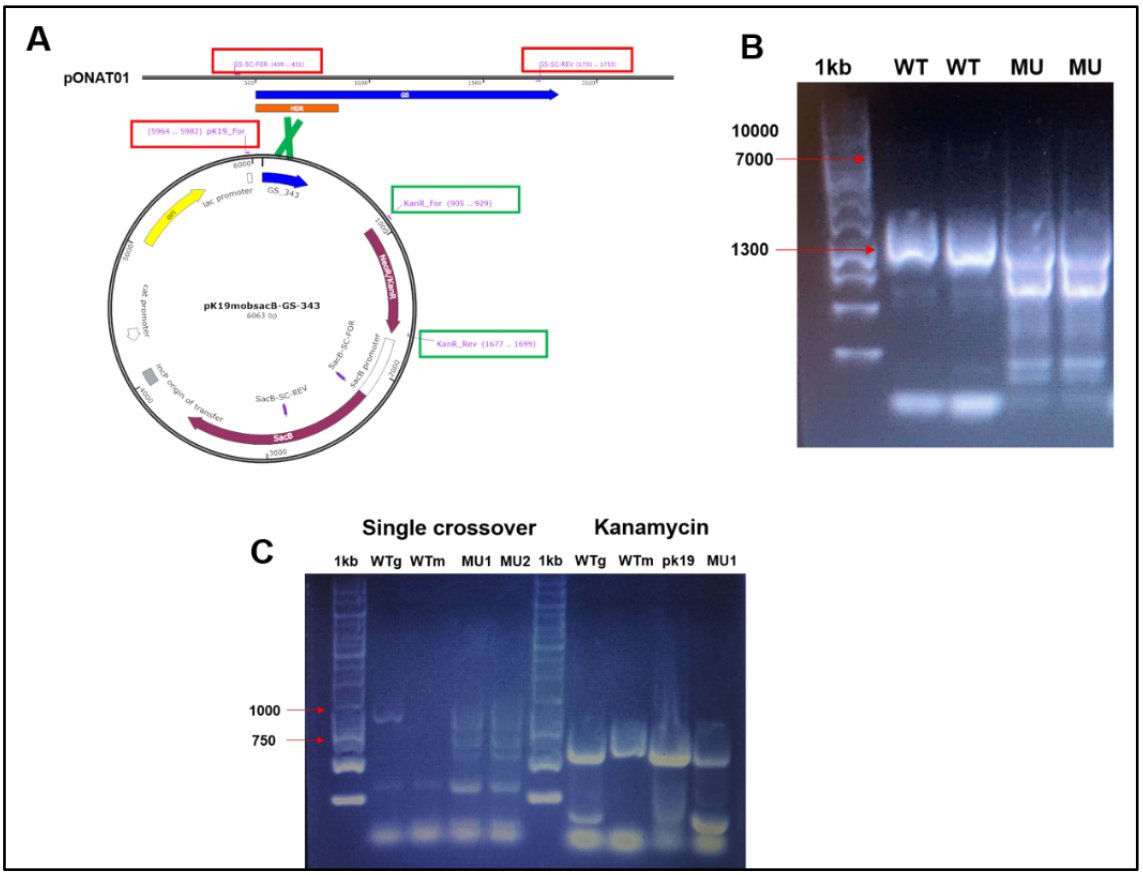

Fig 3. Disruption of glutamine synthetase (GS) by single crossover. A) Map of the plasmid used to disrupt GS and confirmation strategies for single crossover success. Primers highlighted in red boxes were used to validate the integration of pK19mobsacB into pONAT01. Primers highlighted in green boxes used to check the presence of kanamycin cassette. B) Confirmation of single crossover by PCR. C) Confirmation of single crossover and kanamycin by PCR. MU is mutant developed while WT represents wild type strain.

As expected, the lowest growth for both wild type and mutant was under no nitrogen sources in 48 hours of cultivation (Fig 4). Supplementation of NH_4Cl plus glutamate resulted in the highest growth regardless of the strain type followed by NH_4Cl. Under these two conditions (NH_4Cl and NH_4Cl plus glutamate), the mutant grew at much higher rate compared to the wild type. This observation contradicts with my hypothesis. I expected that the disruption of the GS from pONAT01could result in the reduction of growth rate. The

possible explanation for this unexpected growth rate of the mutant on all nitrogen sources tested could be the activation of the other glutamine synthetase present in the chromosome I.

All in all, the result from the growth assay suggested that the single crossover could occur as the wild type and mutant growth rates were highly variable and a biosynthetic enzyme assay was done to further obtain supporting data on functional differences between the wild type and G418 resistant mutant.

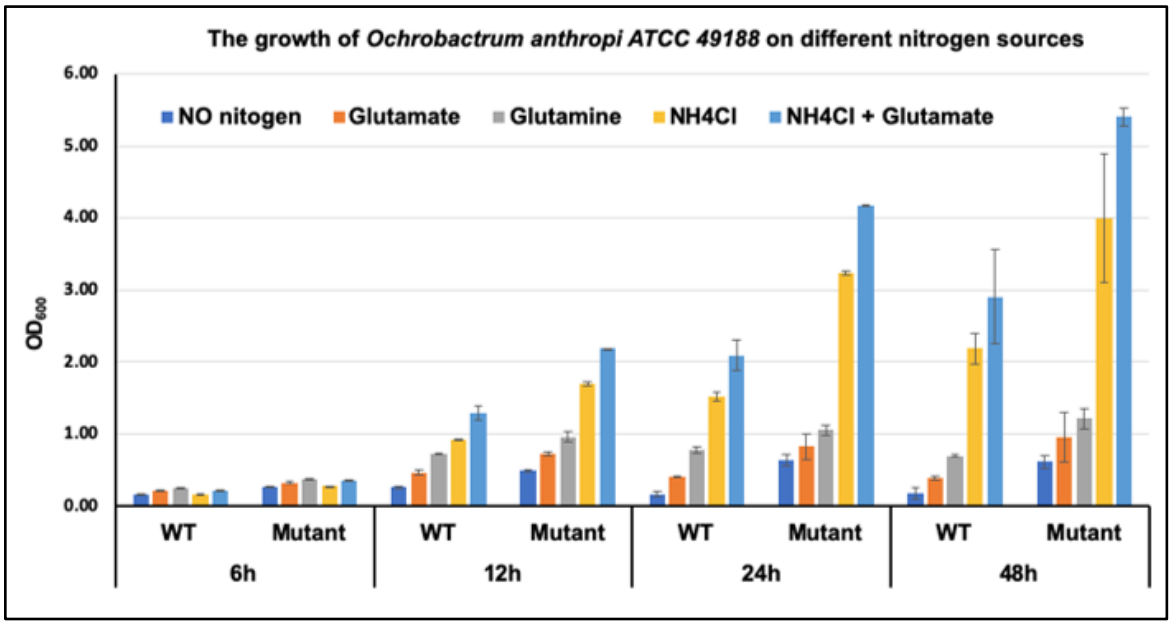

Fig 4. The effect of GS loss on the growth of *O. anthropi* under limited nitrogen sources. Cultures were diluted to 0.1OD_{600} from overnight cultures grown on LB plus selection and further cultivated in minimum media having the indicated nitrogen sources. The experiment was done in duplicates and error bars represent standard deviation.

5.2.2 GS biosynthetic enzyme assay

Using cell lysate from mutants generated by single crossover, I conducted an enzymatic assay for GS via biosynthetic method as described in (Bressler & Ahmed, 1984). Lysates from both wild type and mutant were prepared with the commonly used S30 buffer. This assay enables to determine the activity of GS indirectly based on the released phosphate from the reaction. If the GS in pONAT01 is knocked out, then lower activity of GS is expected. Fig 5 shows that 50% reduction in GS activity in the mutant occurred compared to the activity of the wild type.

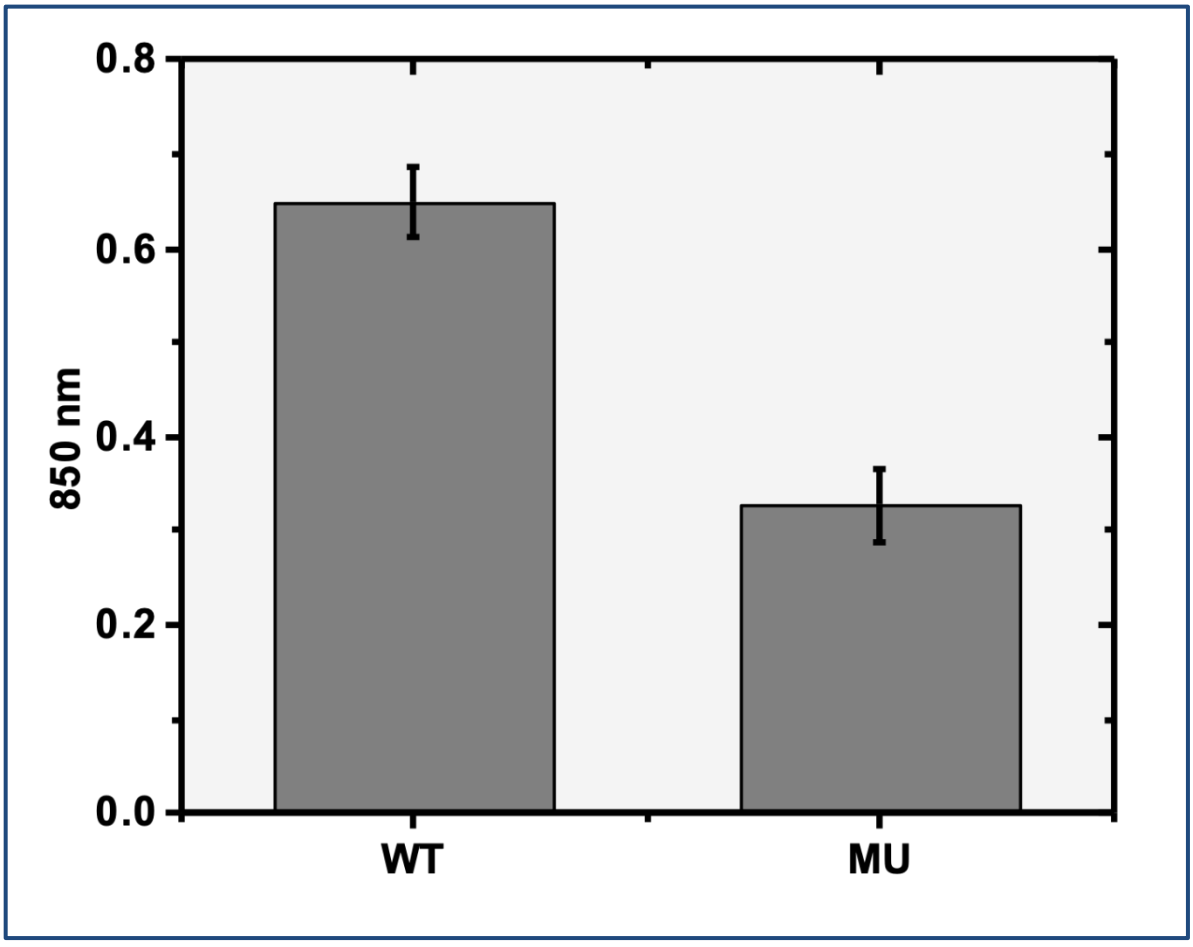

Fig 5. GS biosynthetic enzyme assay. Normalized OD_{850} of mutant (MU) and wild type (WT) compared in this assay in duplicates. Error bars indicate standard deviations. To normalize the data, the formula: $X_{normalized} = (X-X_{min}) / (X_{max}-X_{min})$ was used. After normalization of the data, the corresponding internal controls were subtracted.

5.3 Bioinformatic analysis of glutamine synthetases of *O. anthropi*

5.3.1 Identification and characterization of glutamine synthetases of *O. anthropi*

There are five ORFs described in the reference sequence of *Ochrobactrum anthropi* ATCC 49188 that are predicted to have GS activity. Thus, to identify them and undertake comprehensive bioinformatic analysis of GSs present in this strain, I performed a search via phmmer using the amino acid sequence of the catalytic domain (Gln_synth_cat_dom) of the GS present in chromosome I (Appendix I). As shown in Table 2, the search resulted in 5 potential GSs having similar catalytic domains as the chromosome I GS catalytic domain. To further validate if the 5 putative enzymes are actually GS, I performed in-depth assessment on them in terms of GS signature domains, which were acquired from InterPro. As presented in Table 3, all of them contains the catalytic domain (Gln_synth_cat_dom), the N-terminal domain (Gln_synt_b-grasp), and Gln_synth_gly_rich_site domain, which are a characteristic features of GS.

Moreover, I utilized criteria such as predicted structures (structure and number of subunits), number of amino acid residues, insertion of 25 amino acids (in case of GSI-β) and the presence of adenylation site to identify and classify these potential GSs.

Structurally, GS exist in either a dodecamer or decamer forms where GSI and GSIII typically exist in bacteria and archaea in a dodecamer form (Patel, 2015). Therefore, I opted to build the structures of these putative GS presented in Table 2 using SWISS-MODEL (https://swissmodel.expasy.org/) to determine if they possess such structures (Fig 6). I also compared the structures build by SWISS-MODEL to the top structures obtained from PDB search to see if the two methods match (Fig 6). Based on these structural analyses, all the

putative GS except GS (Oant_4157) appeared to be type GSI or GSIII as they retained dodecamer structures as shown in Fig 6. GS (Oant_4157) possesses a predicted decamer structure (Fig 6), which is a common feature of GSII that is found mostly in eukaryotes (Patel, 2015). Taking number of amino acids into account, I classified all the GS except GS (Oant_4157) as GSI (Table 3). GSIII is larger in size and contains over 700 amino acid residues as a result they cannot be classified to GSIII. GS (Oant_4157) possesses 352 amino acids and was classified as GSII (Fig 6).

Table 2. Putative glutamine synthetases in *Ochrobactrum anthropi* ATCC 49188 as identified using phmmer.

Target	Locus	Description	E-value
A6X0P9_OCHA4	Oant_2087	Glutamine synthetase	8.4e-136
A6X5Y7_OCHA4	Oant_3936	Glutamate--putrescine ligase	9.1e-106
A6X5T2_OCHA4	Oant_3881	Glutamate--ammonia ligase	1.7e-105
A6X7I3_OCHA4	Oant_4491	Glutamine synthetase catalytic region	2.0e-71
A6X6K3_OCHA4	Oant_4157	Glutamine synthetase	1.1e-12

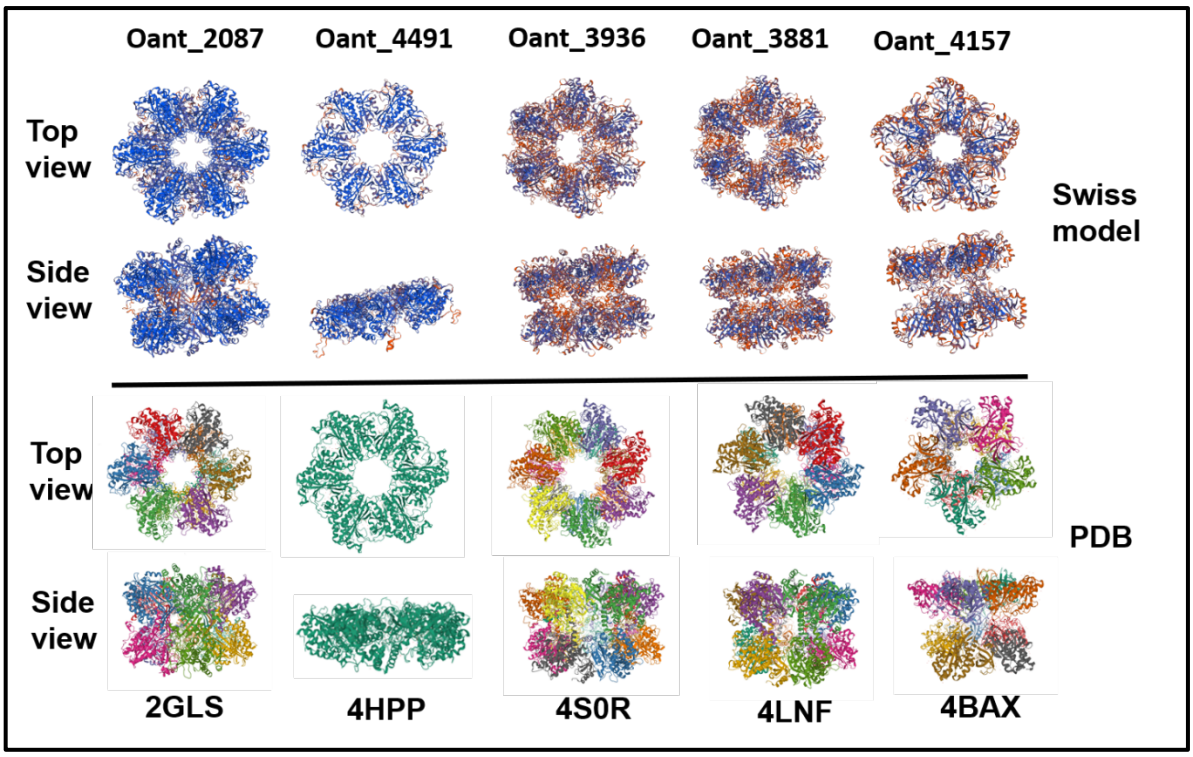

Fig 6. Structures of putative glutamine synthetases of *Ochrobactrum anthropi* ATCC 49188. The structures in the upper panel were built using SWISS-MODEL and structures in the lower panel were obtained from PDB website. The PDB identifiers in the lower panel indicate the top hits in the PDB search for the corresponding putative GSs in the upper panel.

Table 3. Analyses and identification of the putative GS types present in *Ochrobactrum anthropi* ATCC 49188.

Locus	Location	# of AA	Structure	Domains				25 AA insertion	Type
				Gln_synth_cat_dom	Gln_synt_b-grasp	Gln_synth_gly_rich_site	Gln_synth_I_adenylation_site		
Oant_2087	Chromosome I	469	Dodecamer	Yes	Yes	Yes	Yes	Yes	**GSI-β**
Oant_4491	Plasmid I	444	Dodecamer	Yes	Yes	Yes	No	No	**GSI-α**
Oant_3936	Chromosome II	476	Dodecamer	Yes	Yes	Yes	No	No	**GSI-α**
Oant_3881	Chromosome II	456	Dodecamer	Yes	Yes	Yes	No	No	**GSI-α**
Oant_4157	Chromosome II	352	Decamer	Yes	Yes	Yes	No	No	**GSII**

With the paraments described above, I could classify ORF Oant_2087, Oant_4491, Oant_3936 and Oant_3881 were classified as GSI. GSI is further subdivided into GSI-α and GSI-β (Brown et al., 1994; Joo et al., 2018). This division is based on a conserved motif (NLYDLP) for the adenylation of a tyrosine residue near the active site and insertion of a specific 25-amino acid residues (i.e. residues 146–170 AA in *E. coli*) (Brown et al., 1994; Joo et al., 2018). GSI-β have these distinguishing futures while GSI-α lacks them. However, such classification for the GSs of *O. anthropi* is lacking.

Before classifying the four putative GSs as GSI-α and GSI-β forms, first, I chose to implement such criteria to study chromosome I GS. To do so, I did an amino acids alignment for the top 20 hits (Appendix 2) from blastp search using the amino acid sequences of chromosome I GS (Oant_2087). Using this alignment, I generated amino acid sequence logo (Fig 7). This logo indeed indicated the insertion of a specific 25-amino acid residues in the chromosome I GS amino acids (Fig 7A). However, half of the GSs used in the alignment contained the conserved motif (NLYDLP) for the adenylation of a tyrosine residue near the active site (Fig 7B). The rest half contained DLYDLP motif (Fig 7B), suggesting that the motif for adenylation of tyrosine could be modified to D/NLYDLP.

The conserved motif analyses I performed using InterPro indicate that only chromosome I GS (Oant_2087) contain the motif for tyrosine adenylation (Gln_synth_I_adenylation_site). I further confirmed this by looking into the alignment of amino acids sequences of the four putative GSI (Appendix 3). The insertion of the specific 25-amino acid residues occurred only in the chromosome I GS. Therefore, based on these observations, I classified the GSs with locus number Oant_4491, Oant_3936 and Oant_3881 as GSI-α while Oant_2087 as GSI-β (Table 3).

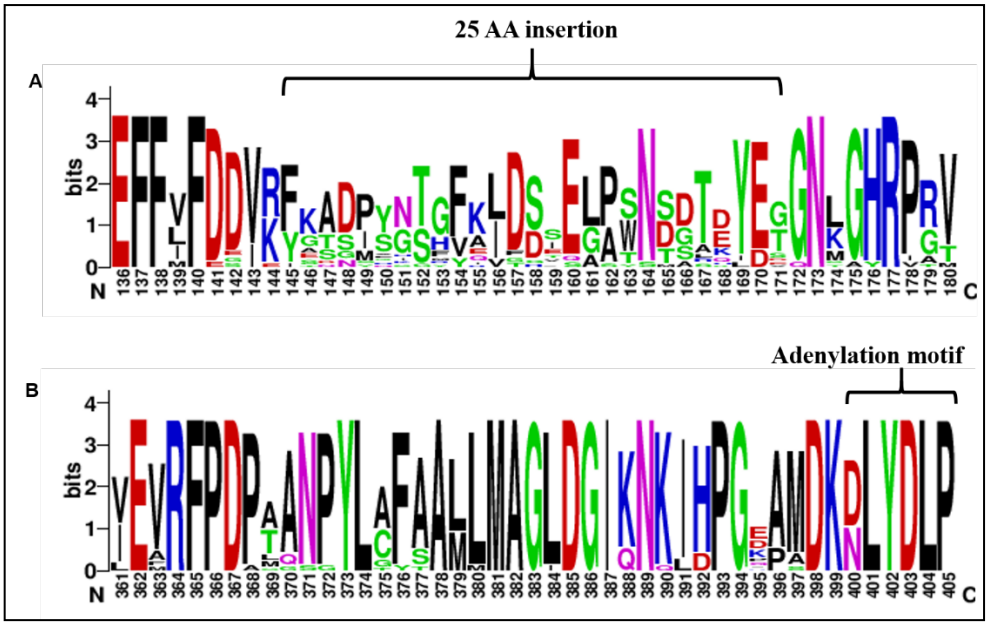

Fig 7. Motif logo. Logo of the amino acid sequences alignment for the top 20 hits obtained from blastp search using chromosome I GS. Regions containing the 25 AA insertion and the Adenylation motif are shown.

5.3.2 Evolution of glutamine synthetases of *O. anthropi*

To investigate the evolutionary relationship of the five putative GSs of *O. anthropi* ATCC 49188, I constructed a gene tree using Geneious Prime. Top five hits with > 70% GS amino acid identity were included from blastp search results for each putative GS. As shown in the gene tree (Fig 8), each putative GS were found in a separate clade. Chromosomal GS of *Ensifer adhaerens* OV14 and the GS from *O. anthropi* ATCC 49188 plasmid 1 (Oant_4491) appeared in the same clade in the gene tree. The tree also indicated that the GSII present in chromosome II (Oant_4157) and the GS from plasmid 1 (Oant_4491) diverged from recent common ancestor, which was supported by high boot strap value (Fig 8). Chromosome II GS (Oant_3936) was found to be the least evolved compared to the other four GSs.

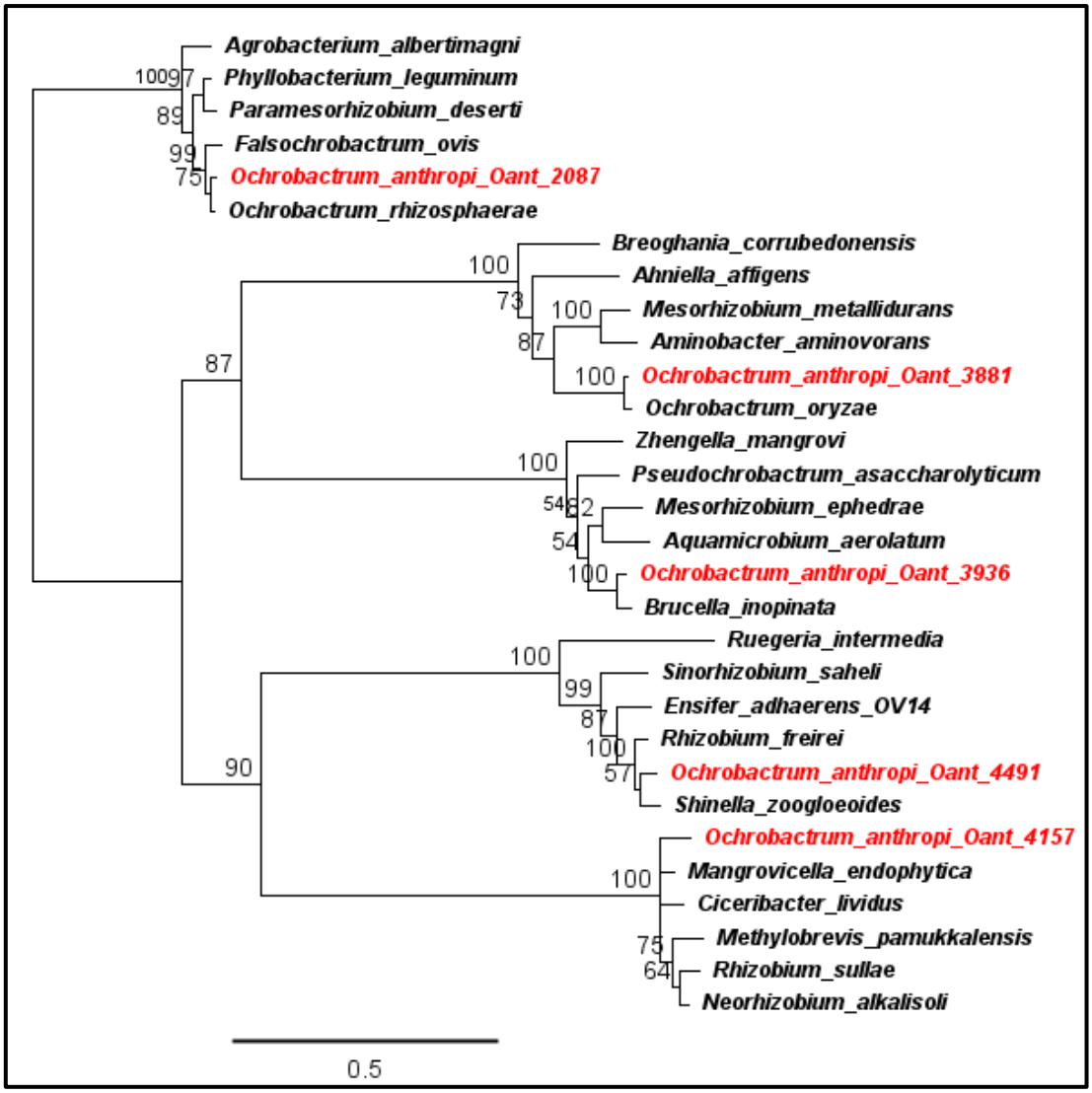

Fig 8. Evolutionary relationship of five putative GSs of *O. anthropi* ATCC 49188. For selecting sequences to build this tree, top protein identity of the species and diversity were considered. The complete amino acid sequences of all the 30 GSs are presented in Appendix 3.

5.3.3 Origin of the glutamine synthetase present in plasmid (pOANT01) of *O. anthropi*

The chromosomal GS of *O. anthropi* is the one which has been evolved with the strain and vertically inherited whereas the one in plasmid pOANT01 could be acquired via horizontal gene transfer at some point. To address this question, I blasted the amino acid sequence of the GS from pOANT01 across the genus *Ochrobactrum*. As shown in Table 4, the majority of the bacteria in this genus have distantly related GS with protein identity <55% except the GS from *O. rhizosphaerae*, *O. haematopilum*. *O. sp* P6B-III and *O. lupini* where the protein identities were 89.62%, 79.37%, 78.92% and 79.37%, respectively.

Table 4. Protein identity of GS from O. anthropi plasmid pOANT01 to the GSs across the genus *Ochrobactrum*.

Description	MAX Score	Total Score	Query cover	Per. Ident	Accession
Glutamine synthetase *[Ochrobactrum anthropi]*	905	905	100%	100%	WP_011982907.1
Glutamine synthetase *[Ochrobactrum rhizosphaerae]*	827	827	99%	89.62%	WP_024899879.1
Gglutamine synthetase *[Ochrobactrum haematophilum]*	735	735	100%	79.37%	WP_138784552.1
Gglutamine synthetase *[Ochrobactrum sp.P6BS-III}*	734	734	100%	78.92%	Wp_078337376.1
Glutamine synthetase *[Ochrobactrum lupini]*	731	731	99%	79.37%	WP_094513496.1
Glutamine synthetase *[Ochrobactrum oryzae]*	485	485	97%	53.79%	WP_104756786.1
Glutamine synthetase *[Ochrobactrum sp. LP_5_YM]*	461	461	98%	50.91%	WP_134157980.1
Glutamine synthetase *[Ochrobactrum pecoris]*	455	455	97%	50.12%	WP_140022293.1
Glutamine synthetase *[Ochrobactrum rhizosphaerae]*	455	455	97%	50.35%	WP_094575837.1
Glutamine synthetase *[Ochrobactrum grignonense]*	447	447	97%	49.65%	WP_094542705.1

In addition, the phylogenetic tree in Fig 8 elucidated that the GS from plasmid pOANT01 has more resemblance to GSs outside of the *Ochrobactrum* genus. This result somehow indicated that the GS on pOANT01 could be acquired through horizontal gene transfer. To further validate this observation and identify where it was transferred from, I performed nucleotide alignments of the whole plasmid pOANT01 with whole genome of the species (mainly in Rhizobiales order) that were shown to have high GS homology (>80% protein identity). The species tested for this purpose includes: *Rhizobium freirei*, *Shinella zoogloeoides*, *Ensifer adhaerens* strain Casida A, *Ensifer adhaerens* strain OV14, *Sinorhizobium saheli* and *Ensifer alkalisoli*. The alignments were performed using both Mauve multiple genome alignment software and YASS a web-based genomic similarity search tool. From all genome alignments I carried out, only *O. anthropi* plasmid pOANT01,

Ensifer adhaerens strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome I possess a unique 10-Kb fragment (the green boxes) as shown in Fig 9.

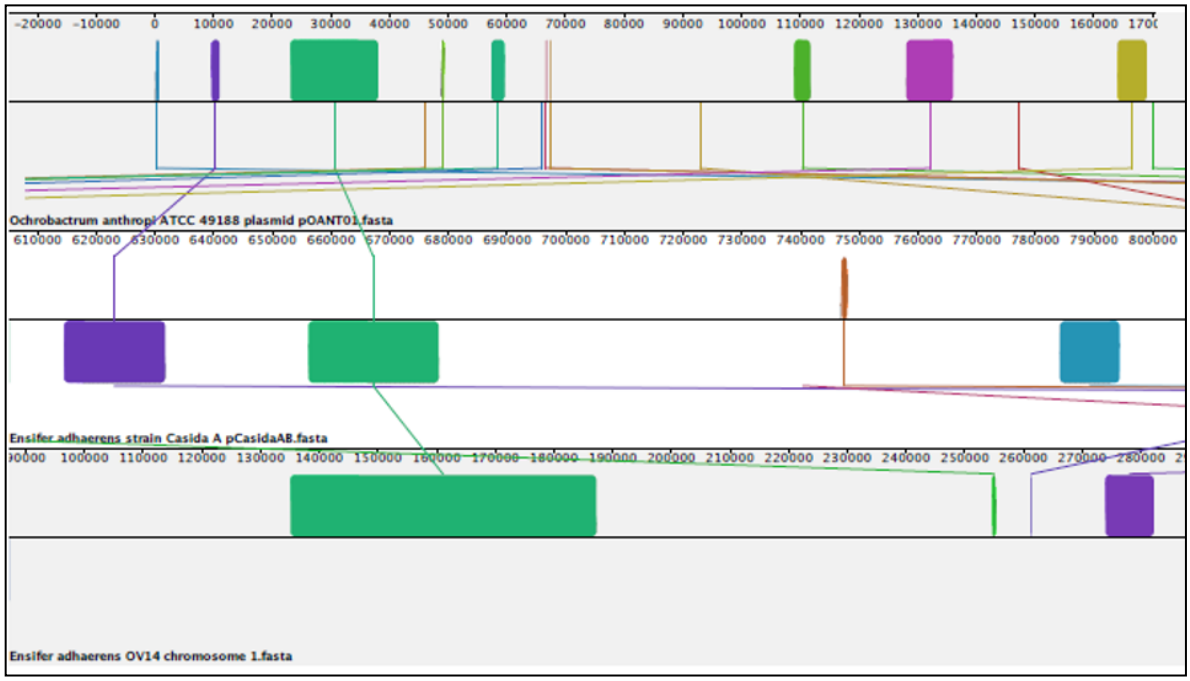

Fig 9. Multiple sequence alignment. Nucleotides alignment of pOANTO1 of *O. anthropi* with pCasidaAB of *E. adhaerens* strain Casida A and chromosome 1 of *E. adhaerens* OV14 was performed with Mauve multiple genome alignment tool. The green boxes are sequences shared among the GS sources.

Table 5. Comparison of pOANT01 and potential GS sources

Name	pOANT01	pCasidaAB	OV14 chromosome 1
Strain origin	*Ochrobactrum anthropi* ATCC 49188	*Ensifer adhaerens* strain Casida A	*Ensifer adhaerens* strain OV14
Accession	CP000760	CP015882	NZ_CP007236
Plasmid size	170,351 bp	1,459,374 bp	3,956,045 bp
Plasmid type	RepABC	RepABC	not RepABC
Total Genes	172	1330	7326
Pseudogenes	27	74	229
Total proteins	147	1250	7017
pOANT01 GS identity	100%	86.91% (99% QC)	87.58% (99% QC)
ori	replication protein C	replication protein C	No protein C
ori position	75312..76592	3003..4217	N/A

To confirm if GS open reading frame (ORF) is contained in the 10-kb fragment from *O. anthropi* plasmid pOANT01, *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1, I blasted the nucleotides of each fragments. Interestingly, all these fragments have the GS ORF (Fig 10). Locus tags: Oant_4491, FA04_30675 and OV14_0164, highlighted with red boxes in Fig 10, are for GS from *O. anthropi* plasmid pOANT01, *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1, respectively. Also, as presented in Table 5, the GS from pOANT01 shares 86.91% (99% query coverage) and 87.57% (99% query coverage) homology with the GS from pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1, respectively. Both pOANT01 (170,351 bp) and pCasidaAB (1,459,374 bp) contain RepC replication origins while this origin is absent in *Ensifer adhaerens* strain OV14 chromosome 1, which is of 3,956,045 bp in size (Table 5).

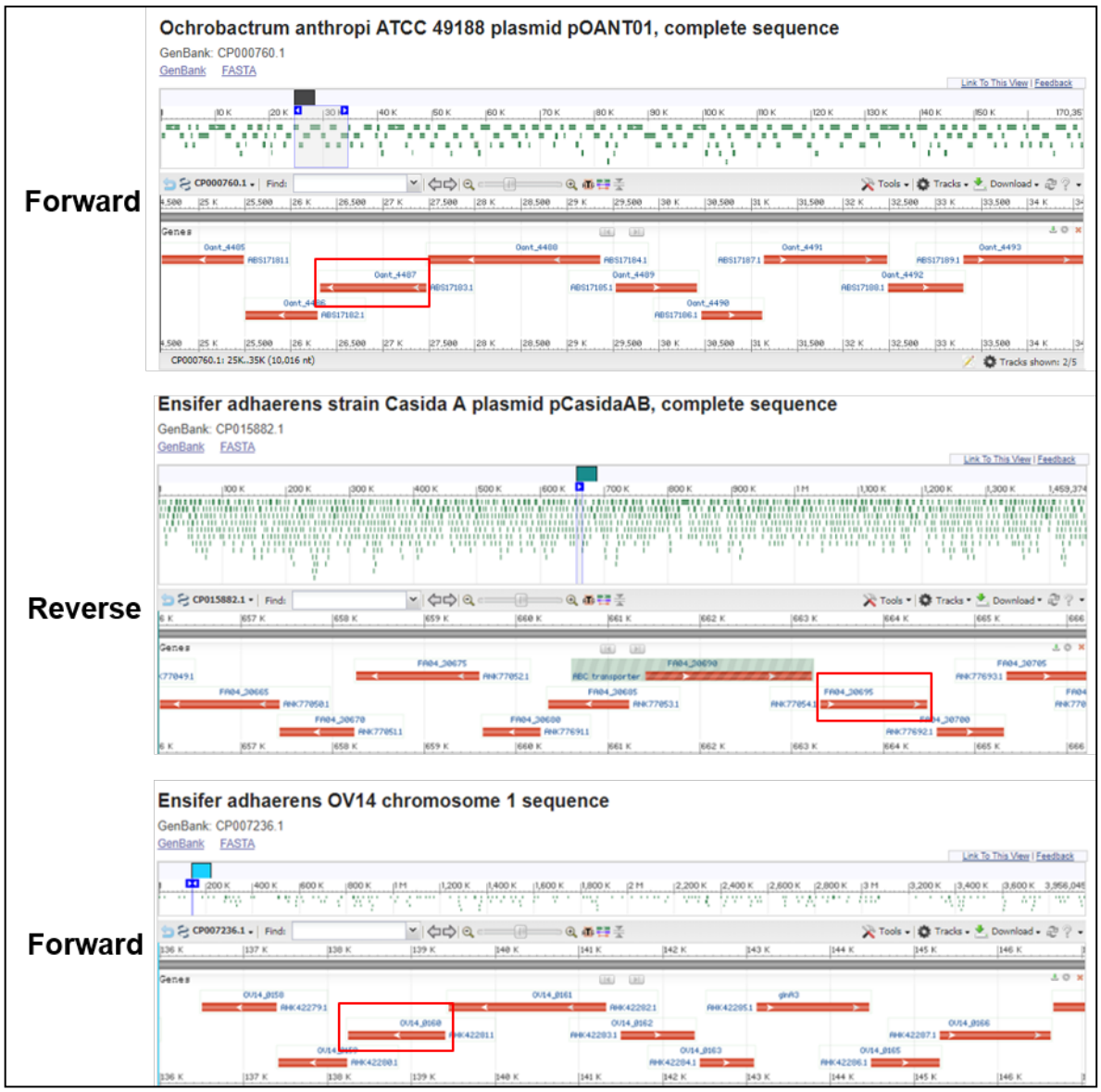

Fig 10. GS and genes flanking GS in the identified 10kb DNA fragments from *O. anthropi* plasmid pOANT01, *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1.

The previous bioinformatic analysis indicated that the GSs in *O. anthropi* are distantly related (Fig 8) and the GS from *O. anthropi* plasmid pOANT01 has not been vertically inherited across *Ochrobactrum* (Table 4). The GS in the 10-kb conserved fragments from *O. anthropi* plasmid pOANT01*, Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome I share high protein identity (>87%) (Table 4). All these observations led me to do more in-depth study on the content of the 10-kb conserved fragments to gain insight regarding the possible origin of the GS on *O. anthropi* plasmid pOANT01. As summarized in Table 5, the conserved 10-kb fragment from *O. anthropi* plasmid pOANT01 contains the following 9 genes in order: inner-membrane translocator (Oant_4485, 295 AA), ABC transporter (Oant_4486, 260 AA), extracellular ligand-binding receptor (Oant_4487, 384 AA), ABC transporter (Oant_4488, 624 AA), RpIR family transcriptional regulator (Oant_4489, 293 AA), isocorismataase (Oant_4490, 220 AA), glutamine synthetase (Oant_4491, 444 AA), N-formylglutamate amidohydrolase (Oant_4492, 272 AA) and aminotransferase (Oant_4493, 436 AA). These 9 genes are conserved in the same order in the 10-kb fragments from *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1 (Table 5). Although there is a GS in chromosome 1 of each of strain considered in the analysis, this 10-kb fragment with the 9 genes is only found in *O. anthropi* plasmid pOANT01*, Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome I. The fragment was also not observed in all bacteria from Rhizobiales order that have GS with high protein identity to the GS on pOANT01 (Fig 8).

Table 6. Summary of proteins flanking GS in the identified 10kb DNA fragments from *O. anthropi* plasmid pOANT01, *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome I.

***Ochrobactrum anthropi* ATCC 49188 plasmid pOANT01 ACCESSION: CP000760 REGION:24599..34614**			***Ensifer adhaerens strain* Casida A plasmid pCasidaAB ACCESSION:CP015882 REGION:656120..666218**			**Ensifer adhaerens OV14 chromosome 1 sequence ACCESSION: CP007236 REGION: 135999..147047**		
Locus_tag	**AA number**	**Function**	**Locus_tag**	**AA number**	**Function**	**Locus_tag**	**AA number**	**Function**
Oant_4485	295	Inner-membrane translocator	FA04_30705	293	Branched-chain amino acid ABC transporter	OV14_0158	295	Putative transmembrane component of ABC
Oant_4486	260	ABC transporter	FA04_30700	243	ABC transporter	OV14_0159	272	Putative ATP-binding component of ABC
Oant_4487	384	Extracellular ligand-binding receptor	FA04_30695	387	ABC transporter substrate-binding protein	OV14_0160	387	Substrate-binding
Oant_4488	624	ABC transporter	FA04_30690	604	ABC transporter	OV14_0161	627	ABC transporter
Oant_4489	293	Transcriptional regulator, RpiR family	FA04_30685	293	RpiR family transcriptional regulator	OV14_0162	293	Transcriptional regulator, RpiR family
Oant_4490	220	Isochorismatase hydrolase	FA04_30680	208	Isochorismatase	OV14_0163	215	Isochorismatase/putative cysteine hydrolase
Oant_4491	**444**	**Glutamine synthetase**	**FA04_30675**	**447**	**Glutamine synthetase**	**OV14_0164**	**448**	**Glutamine synthetase**
Oant_4492	272	N-formylglutamate amidohydrolase	FA04_30670	272	N-formylglutamate amidohydrolase	OV14_0165	274	N-formylglutamate amidohydrolase
Oant_4493	436	Aminotransferase	FA04_30665	434	Aminotransferase	OV14_0166	439	Aminotransferase

6. DISCUSSION

Well conserved from unicellular organisms to mammals, GS is an enzyme involved in condensation of glutamate and ammonia to produce the essential amino acid, glutamine (Pesole et al., 1995). This ubiquitous enzyme is commonly classified into GSI, GSII, and GSIII based on molecular size, number of subunits, underlying regulation and enzyme structure (Brown et al., 1994; Liu et al., 2018). Although there are five ORFs described to have possible GS activity in *O. anthropi* ATCC 49188 genome, functional characterization and bioinformatic analysis have been lacking for all them. Therefore, here in this section, I discuss my bioinformatics analysis and experimental results.

Based on structures built using SWISS-MODEL and from predicted structures of top PDB hits for each putative GSs, all the putative GS except GS (Oant_4157) displayed dodecamer structures and can be categorised as type GSI or GSIII (Fig 6). Exceptionally, both methods generated a GS with hexamer structure for the plasmid 1 GS (Oant_4491), missing the other biological hexamer to give it the complete dodecamer form. GS (Oant_4157) possesses a decamer structure (Fig 6), which is a common feature of GSII that is found mostly in eukaryotes (Patel, 2015). As a result, I classified it as GSII. Taking number of amino acids into account, I classified all the GS except GS (Oant_4157) as GSI (Table 3). Because GSIII is larger in size and contains over 700 amino acid residues, all the putative GSs cannot fall into this type.

GSI is further subdivided into GSI-α and GSI-β (Brown et al., 1994; Joo et al., 2018). GSI-α is mainly found in gram-positive bacteria and thermophilic bacteria whereas GSI-β is found in other bacteria species (Fisher, 1999; Shapiro et al., 1967). One of the main biochemical features that distinguish these two forms is the presence of a conserved motif

(NLYDLP) for the adenylation of a tyrosine residue near the active site GS (Brown et al., 1994; Joo et al., 2018). This motif exhibited only by GSI-β form. Given that the amino acids aspartate (D) and asparagine (N) have similar structures, the DLYDLP motif could alternatively serve as a site for adenylation of a tyrosine residue near the active site of GSI depending on the species. I further verified this observation through in-depth amino acid sequences analysis presented with sequence logo (Fig 7) developed using 20 GS top hit sequences for chromosome I GS (Appendix 2). According to the sequence log (Fig 7), 50% of the GSs had NLYDLP motif while the rest had DLYDLP motif, indicating that species can use these motifs for adenylation of tyrosine alternatively. Therefore, although it is required to validate this finding experimentally, the motif can be expanded to accommodate both aspartate (D) and asparagine (N) alternatively and modified as N/D-LYDLP. Based on this modified motif, the chromosome I GS (Oant_2087) of *O. anthropi* can be classified as GSI-β whereas the GS from pOANT01 (Oant_4491) and the other GSI from chromosome II (Oant_3936 and Oant_3881) as GSI-α because this motif is missing from them (Appendix 3). The insertion of the specific 25-amino acid residues was observed only in chromosome I GS (Fig 7A and appendix 3). This further confirm that only chromosome I GS is GSI-β. In general, my bioinformatic analysis highly suggest that only the chromosome I GS (Oant_2087) can be regulated though adenylation of a tyrosine residue.

As an essential enzyme, the chromosomal GSs could provide an essential function to *O. anthropi.* Most likely, they have evolved with the strain and vertically inherited. Despite this, the origin of the GS in plasmid pOANT0 (Oant_4491) has not been known. Possibly, it could be a result of gene duplication or was acquired via horizontal gene transfer from other species at some point. If this GS were encoded by duplicate genes, it would share high protein

identify with one of the putative GSs of *O. anthropi* unless they diverged independently to their current form through evolution. As a result, it should fall in the same clade with one of the putative GSs in the gene tree (Fig 8). Rather, this GS has high protein similarity with the GS of *Rhizobium freirei*, *Shinella zoogloeoides*, *Ensifer adhaerens* OV14 and *Sinorhizobium Saheli* (Fig 8). This result affirmed that the GS in plasmid pOANT0 (Oant_4491) was not a result of gene duplication. Moreover, majority of the bacteria in *Ochrobactrum* genus have distantly related GS compared to the GS in pOANT01 of *O. anthropi* (Table 4). The highest matching GS are from *O. rhizosphaerae*, *O. haematopilum*. *O. sp* P6B-III and *O. lupini* where the resemblance accounted 89.62%, 79.37%, 78.92% and 79.37%, respectively. The rest of the species have GSs with protein identity <55% GS (Table 4), implying that the GS on pOANT01 could be acquired through horizontal gene transfer from outside *Ochrobactrum* genus.

To elucidate if the GS in pOANT01 of *O. anthropi* was acquired via horizontal gene transfer and identify its potential origin, I performed multiple genome alignments and sequence analyses. According to the result from the phylogeny analysis in Fig 8, the top matches to the GS in pOANT01 of *O. anthropi* are from Rhizobiales order. With >80% protein identity, the species considered for the genome alignments includes:*Rhizobium freirei*, *Shinella zoogloeoides*, *Ensifer adhaerens* strain Casida A, *Ensifer adhaerens* strain OV14, *Sinorhizobium saheli* and *Ensifer alkalisoli*. Surprisingly, from all genome alignments I carried out, only *O. anthropi* plasmid pOANT01*, Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1 possess a unique 10-kb fragment (the green boxes) as shown in Fig 9. This fragment contained GS open reading frame (ORF) with locus tags: Oant_4491(*O. anthropi* plasmid pOANT01), FA04_30675

(*Ensifer adhaerens* strain Casida A plasmid pCasidaAB) and OV14_0164 (*Ensifer adhaerens* strain OV14 chromosome 1) (Fig 9). In addition, as presented in Table 5, the GS from pOANT01 shares 86.91% (99% query coverage) and 87.57% (99% query coverage) homology with the GS from pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1, respectively. These observations strongly suggest the GS on pOANT01 of *O.* anthropi could be acquired by horizontal gene transfer from either *Ensifer adhaerens* strain Casida A plasmid pCasidaAB or *Ensifer adhaerens* strain OV14 chromosome 1. With further sequence analysis, the10-kb fragment was found to contain 9 genes, which are conserved in the same order in *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1 (Table 6). Three of these genes which are in proximity: glutamine synthetase (Oant_4491, 444 AA), N-formylglutamate amidohydrolase (Oant_4492, 272 AA) and aminotransferase (Oant_4493, 436 AA) are involved in ammonium assimilation pathway. The fact that this fragment with its unique gene content did not exist in the other bacteria from Rhizobiales order except on *Ensifer adhaerens* strain Casida A plasmid pCasidaAB and *Ensifer adhaerens* strain OV14 chromosome 1 further strengthen the concept that the GS on pOANT01 of *O. anthropi* could be acquired from those sources through horizontal gene transfer. With this fragment transfer, *O. anthropi* might have gained added capabilities in terms of nitrogen assimilation as the transferred GS could have different regulation and activity level.

Horizontal gene transfer can cause abrupt alterations in the structure and organization of genomes of bacteria to generate variants of bacterial strains with new capabilities (Dutta & Pan, 2002). However, in some cases, the transferred gene can be non-functional in the recipient bacteria (Dutta & Pan, 2002). Therefore, it is important to know if the GS on

pOANT01 is functional in *O. anthropi.* As shown in Fig 5, knocking out the GS from pONAT01 resulted in 50% reduction in GS activity compared to the activity of the wild type, indicating that the transferred GS is functional in *O. anthropi.* The mutant contained the GS encoded from the chromosome 1 and the activity of the two GSs were not assessed separately. As their activities can vary, I recommend further investigation to figure out the contribution of each of the GS toward nitrogen assimilation in *O. anthropi*. The nature of their underlying mechanism of regulation can have a role towards their activity. The chromosomal GS is GSI-β type as it contains the motif (DLYDLP) for adenylation of tyrosine. The function of this type of GS is modulated through posttranslational adenylation of tyrosine present in this motif by the adenylyltransferase enzyme (is to transfer adenyly group to tyrosine to deactivate GS when there is too much Glutamine in the cell). (ATase), which is encoded by *glnE* from the same operon (Schulz, Collett, & Reid, 2001). The GS on pONAT01 of *O. anthropi* is GSI-α type and does not have the motif, as a result, GlnE does not regulate it (Garner, Fulkerson, & Mobley, 1998). This difference in regulation can cause variation between the five putative GSs of *O. anthropi.*

7. CONCLUSION

With thorough bioinformatic analysis, I identified and classified five putative GSs from the genome of *O. anthropi*, which were found to be distantly related. Also, I modified the previously reported conserved motif (NLYDLP) for adenylation of tyrosine at the N-terminal of GSI-β to N/D-LYDLP. Using this modified criteria and insertion of specific 25 amino acids, I identified the chromosomal GS of *O. anthropi* as GSI-β. Since those features were not present in the GS from pONAT01 and the other GS from chromosome II of *O. anthropi*, I identified them as GSI-α. *O. anthropi* also contains GSII in chromosome II. This classification would provide insight towards the underlying mechanism of regulations of these GSs in *O. anthropi*. My analysis also led to the conclusion that the GS on the pONAT01 of *O. anthropi* was acquired through horizontal gene transfer from either *Ensifer adhaerens* strain Casida A plasmid pCasidaAB or *Ensifer adhaerens* strain OV14 chromosome 1. Knocking out this transferred GS from pONAT01 of *O. anthropi* resulted in 50% reduction in enzyme activity. Thus, the enzyme is functional in *O. anthropi*.

REFERENCES

Abad, M. S., Coffin, M., & Goldman, B. S. (2015). Genes encoding glutamine synthetase and uses for plant improvement: Google Patents.

Adachi, J., Katsura, K., Seki, E., Takemoto, C., Shirouzu, M., Terada, T., . . . Yokoyama, S. (2019). Cell-Free Protein Synthesis Using S30 Extracts from Escherichia coli RFzero Strains for Efficient Incorporation of Non-Natural Amino Acids into Proteins. *International journal of molecular sciences, 20*(3), 492. doi: 10.3390/ijms20030492

Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J., & Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology, 8*(4), 251.

Aminov, R. I. (2011). Horizontal gene exchange in environmental microbiota. *Frontiers in Microbiology, 2*, 158.

Bartosik, D., Baj, J., & Wlodarczyk, M. (1998). Molecular and functional analysis of pTAV320, a repABC-type replicon of the Paracoccus versutus composite plasmid pTAV1. *Microbiology, 144*(11), 3149-3157.

Bengtsson-Palme, J., Kristiansson, E., & Larsson, D. J. (2017). Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiology Reviews, 42*(1), doi.org/10.1093/femsre/fux1053.

Bressler, S. L., & Ahmed, S. I. (1984). Detection of glutamine synthetase activity in marine phytoplankton: optimization of the biosynthetic assay. *Mar. Ecol. Prog. Ser, 14*, 207-217.

Brown, J., Masuchi, Y., Robb, F., & Doolittlel, W. (1994). Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. *Journal of molecular evolution, 38*(6), 566-576.

Bruins, M. R., Kapil, S., & Oehme, F. W. (2000). Microbial resistance to metals in the environment. *Ecotoxicology and Environmental Safety, 45*(3), 198-207.

Carattoli, A. (2003). Plasmid-mediated antimicrobial resistance in Salmonella enterica. *Current Issues in Molecular Biology, 5*(4), 113-122.

Castillo-Ramírez, S., Vázquez-Castellanos, J. F., González, V., & Cevallos, M. A. (2009). Horizontal gene transfer and diverse functional constrains within a common replication-partitioning system in Alphaproteobacteria: the repABC operon. *BMC Genomics, 10*(1), doi:10.1186/1471-2164-1110-1536.

Cervantes-Rivera, R., Pedraza-López, F., Pérez-Segura, G., & Cevallos, M. A. (2011). The replication origin of a repABC plasmid. *BMC Microbiology, 11*(1), 158.

Cevallos, M. A., Cervantes-Rivera, R., & Gutiérrez-Ríos, R. M. (2008). The repABC plasmid family. *Plasmid, 60*(1), 19-37.

Cevallos, M. A., Porta, H., Izquierdo, J., Tun-Garrido, C., Garcıa-de-los-Santos, A., Dávila, G., & Brom, S. (2002). *Rhizobium etl*i CFN42 contains at least three plasmids of the repABC family: a structural and evolutionary analysis. *Plasmid, 48*(2), 104-116.

Chain, P. S., Lang, D. M., Comerci, D. J., Malfatti, S. A., Vergez, L. M., Shin, M., . . . Tolmasky, M. E. (2011). Genome of *Ochrobactrum anthropi* ATCC 49188T, a versatile opportunistic pathogen and symbiont of several eukaryotic hosts. *Journal of Bacteriology, 193*(16), 4274-4275.

Chen, Q., & Silflow, C. D. (1996). Isolation and characterization of glutamine synthetase genes in Chlamydomonas reinhardtii. *Plant Physiology, 112*(3), 987-996.

Couturier, M., Bex, F., Bergquist, P. L., & Maas, W. K. (1988). Identification and classification of bacterial plasmids. *Microbiological Reviews, 52*(3), 375-395

Deuel, T. F., Ginsburg, A., Yeh, J., Shelton, E., & Stadtman, E. (1970). Bacillus subtilis glutamine synthetase purification and physical characterization. *Journal of Biological Chemistry, 245*(20), 5195-5205.

Dharmawardene, M., Haystead, A., & Stewart, W. (1973). Glutamine synthetase of the nitrogen-fixing alga Anabaena cylindrica. *Archiv für Mikrobiologie, 90*(4), 281-295.

Dutta, C., & Pan, A. (2002). Horizontal gene transfer and bacterial diversity. *Journal of biosciences, 27*(1), 27-33.

Dziewit, L., Pyzik, A., Szuplewska, M., Matlakowska, R., Mielnicki, S., Wibberg, D., . . . Bartosik, D. (2015). Diversity and role of plasmids in adaptation of bacteria inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals. *Frontiers in Microbiology, 6*, https://doi.org/10.3389/fmicb.2015.00152.

Fang, L., Li, X., Li, L., Li, S., Liao, X., Sun, J., & Liu, Y. (2016). Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from *E. coli* isolates of food-producing animals. *Scientific Reports, 6*, doi:10.1038/srep25312.

Fisher, S. H. (1999). Regulation of nitrogen metabolism in Bacillus subtilis: vive la difference! *Molecular microbiology, 32*(2), 223-232.

Fuchs, R. L., & Keister, D. L. (1980). Comparative properties of glutamine synthetases I and II in Rhizobium and Agrobacterium spp. *Journal of Bacteriology, 144*(2), 641-648.

Fuchs, R. L., & Keister, D. L. (1980). Identification of two glutamine synthetases in Agrobacterium. *Journal of Bacteriology, 141*(2), 996-998. doi: 10.1128/JB.141.2.996-998.1980

Garcillán-Barcia, M. P., Alvarado, A., & de la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiology Reviews, 35*(5), 936-956.

Garner, R. M., Fulkerson, J., & Mobley, H. L. T. (1998). Helicobacter pylori Glutamine Synthetase Lacks Features Associated with Transcriptional and Posttranslational Regulation. *Infection and Immunity, 66*(5), 1839-1847. doi: 10.1128/iai.66.5.1839-1847.1998

Gawronski, J. D., & Benson, D. R. (2004). Microtiter assay for glutamine synthetase biosynthetic activity using inorganic phosphate detection. *Analytical biochemistry, 327*(1), 114-118.

Gill, H. S., Pfluegl, G. M. U., & Eisenberg, D. (2002). Multicopy Crystallographic Refinement of a Relaxed Glutamine Synthetase from Mycobacterium tuberculosis Highlights Flexible Loops in the Enzymatic Mechanism and Its Regulation. *Biochemistry, 41*(31), 9863-9872. doi: 10.1021/bi020254s

Goss, T. J., Perez-Matos, A., & Bender, R. A. (2001). Roles of Glutamate Synthase, gltBD, and gltF in nitrogen metabolism of *Escherichia coli* and *Klebsiella aerogenes*. *Journal of Bacteriology, 183*(22), 6607-6619. doi: 10.1128/jb.183.22.6607-6619.2001

Gyles, C., & Boerlin, P. (2014). Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Veterinary Pathology, 51*(2), 328-340.

Hagiya, H., Ohnishi, K., Maki, M., Watanabe, N., & Murase, T. (2013). Clinical characteristics of *Ochrobactrum anthropi* bacteremia. *Journal of Clinical Microbiology, 51*(4), 1330-1333.

Harrison, P. W., Lower, R. P., Kim, N. K., & Young, J. P. W. (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in Microbiology, 18*(4), 141-148.

Heuer, H., & Smalla, K. (2012). Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiology Reviews, 36*(6), 1083-1104.

Joo, H. K., Park, Y. W., Jang, Y. Y., & Lee, J. Y. (2018). Structural Analysis of Glutamine Synthetase from Helicobacter pylori. *Scientific Reports, 8*(1), 11657. doi: 10.1038/s41598-018-30191-5

Landeta, C., Dávalos, A., Cevallos, M. Á., Geiger, O., Brom, S., & Romero, D. (2011). Plasmids with a chromosome-like role in rhizobia. *Journal of Bacteriology, 193*(6), 1317-1326.

Lilly, J., & Camps, M. (2015). Mechanisms of theta plasmid replication. *Microbiology Spectrum, 3*(1), PLAS-0029-2014.

Liu, X., Zhang, Q., Huan, Z., Zhong, M., Chen, W., & Du, H. (2018). Identification and characterization of glutamine synthetase isozymes in Gracilaria lemaneiformis. *Aquatic Botany, 146*, 23-30.

Lu, Z., Qin, Z., Sarath Babu, V., Ye, C., Su, G., Li, J., . . . Lin, L. (2019). Expression and functional characterization of glutamine synthetase from giant freshwater prawn (Macrobrachium rosenbergii) under osmotic stress. *Aquaculture Research, 50*(9), 2635-2645.

MacLellan, S. R., Zaheer, R., Sartor, A. L., MacLean, A. M., & Finan, T. M. (2006). Identification of a megaplasmid centromere reveals genetic structural diversity within the repABC family of basic replicons. *Molecular Microbiology, 59*(5), 1559-1575.

Mazur, A., & Koper, P. (2012). Rhizobial plasmids—replication, structure and biological role. *Central European Journal of Biology, 7*(4), 571-586.

Mazur, A., Majewska, B., Stasiak, G., Wielbo, J., & Skorupska, A. (2011). repABC-based replication systems of *Rhizobium leguminosarum* bv. *trifoli*i TA1 plasmids: incompatibility and evolutionary analyses. *Plasmid, 66*(2), 53-66.

Pappas, K. M., & Winans, S. C. (2003). The RepA and RepB autorepressors and TraR play opposing roles in the regulation of a Ti plasmid repABC operon. *Molecular Microbiology, 49*(2), 441-455.

Patel, S. I. (2015). *Soluble expression of plasmodium falciparum glutamine synthetase and three-dimensional structure by single particle reconstruction.* University of Cape Town.

Pérez-Oseguera, Á., & Cevallos, M. A. (2013). RepA and RepB exert plasmid incompatibility repressing the transcription of the repABC operon. *Plasmid, 70*(3), 362-376.

Pesole, G., Gissi, C., Lanave, C., & Saccone, C. (1995). Glutamine synthetase gene evolution in bacteria. *Molecular Biology and Evolution, 12*(2), 189-197. doi: 10.1093/oxfordjournals.molbev.a040197

Pinto, U. M., Pappas, K. M., & Winans, S. C. (2012). The ABCs of plasmid replication and segregation. *Nature Reviews Microbiology, 10*(11), 755–765.

Ramirez-Romero, M. A., Bustos, P., Girard, L., Rodríguez, O., Cevallos, M. A., & Dávila, G. (1997). Sequence, localization and characteristics of the replicator region of the symbiotic plasmid of *Rhizobium etli*. *Microbiology, 143*(8), 2825-2831.

Ramírez-Romero, M. A., Soberón, N., Pérez-Oseguera, A., Téllez-Sosa, J., & Cevallos, M. A. (2000). Structural elements required for replication and incompatibility of the *Rhizobium etli* symbiotic plasmid. *Journal of Bacteriology, 182*(11), 3117-3124.

Reinecke, D. L., Zarka, A., Leu, S., & Boussiba, S. (2016). Cloning, molecular characterization, and phylogeny of two evolutionary distinct glutamine synthetase isoforms in the green microalga Haematococcus pluvialis (Chlorophyceae). *Journal of phycology, 52*(6), 961-972.

Reitzer, L. J. (1987). Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine and D-alanine. *In Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology 2*, 302-320.

Reitzer, L. J., Bueno, R., Cheng, W. D., Abrams, S. A., Rothstein, D. M., Hunt, T. P., . . . Magasanik, B. (1987). Mutations that create new promoters suppress the sigma 54 dependence of glnA transcription in Escherichia coli. *Journal of Bacteriology, 169*(9), 4279-4284. doi: 10.1128/jb.169.9.4279-4284.1987

Sambrook, J., & Russell, D. W. (2001). Molecular cloning: a laboratory manual. 2001: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sarah L. Turner, & Young, J. P. W. (1995). The replicator region of the Rhizobium leguminosarum cryptic plasmid pRL8JI. *FEMS Microbiology Letters, 133*(1-2), 53-58.

Schulz, A. A., Collett, H. J., & Reid, S. J. (2001). Nitrogen and carbon regulation of glutamine synthetase and glutamate synthase in Corynebacterium glutamicum ATCC 13032. *FEMS microbiology letters, 205*(2), 361-367.

Shapiro, B. M., Kingdon, H. S., & Stadtman, E. (1967). Regulation of glutamine synthetase. VII. Adenylyl glutamine synthetase: a new form of the enzyme with altered

regulatory and kinetic properties. *Proceedings of the National Academy of Sciences of the United States of America, 58*(2), 642.

Shatters, R., Liu, Y., & Kahn, M. (1993). Isolation and characterization of a novel glutamine synthetase from Rhizobium meliloti. *Journal of Biological Chemistry, 268*(1), 469-475.

Shintani, M., Sanchez, Z. K., & Kimbara, K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology, 6*, 242.

Skippington, E., & Ragan, M. A. (2011). Lateral genetic transfer and the construction of genetic exchange communities. *FEMS Microbiology Reviews, 35*(5), 707-735.

Smalla, K., Jechalke, S., & Top, E. M. (2015). Plasmid detection, characterization and ecology. *Microbiology Spectrum, 3*(1), doi:10.1128/microbiolspec.PLAS-0038-2014.

Soberón, N., Venkova-Canova, T., Ramírez-Romero, M. A., Téllez-Sosa, J., & Cevallos, M. A. (2004). Incompatibility and the partitioning site of the repABC basic replicon of the symbiotic plasmid from *Rhizobium etli*. *Plasmid, 51*(3), 203-216.

Somerville, J., Shatters, R., & Kahn, M. (1989). Isolation, characterization, and complementation of Rhizobium meliloti 104A14 mutants that lack glutamine synthetase II activity. *Journal of Bacteriology, 171*(9), 5079-5086.

Teyssier, C., Marchandin, H., Jean-Pierre, H., Masnou, A., Dusart, G., & Jumas-Bilak, E. (2007). *Ochrobactrum pseudintermedium* sp. nov., a novel member of the family Brucellaceae, isolated from human clinical samples. *International Journal of Systematic and Evolutionary Microbiology, 57*(5), 1007-1013.

Thoma, B., Straube, E., Scholz, H. C., Al Dahouk, S., Zöller, L., Pfeffer, M., . . . Tomaso, H. (2009). Identification and antimicrobial susceptibilities of *Ochrobactrum spp*. *International Journal of Medical Microbiology, 299*(3), 209-220.

Wei, G. R., & Kustu, S. (1981). Glutamine auxotrophs with mutations in a nitrogen regulatory gene, ntrC, that is near glnA. *Molecular and General Genetics MGG, 183*(2), 392-399. doi: 10.1007/BF00270646

Yamaguchi, Y., Park, J.-H., & Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annual Review of Genetics, 45*, 61-79.

Yip, C. B., Ding, H., & Hynes, M. F. (2015). Counter-transcribed RNAs of *Rhizobium leguminosarum* repABC plasmids exert incompatibility effects only when highly expressed. *Plasmid, 78*, 37-47.

Żebracki, K., Koper, P., Marczak, M., Skorupska, A., & Mazur, A. (2015). Plasmid-encoded RepA proteins specifically autorepress individual repABC operons in the multipartite *Rhizobium leguminosarum* bv. Trifolii genome. *PloS One, 10*(7), doi: 10.1371/journal.pone.0131907.

APPENDICES

Appendix 1. Catalytic domain of chromosome I GS (Oant_2087)

GRDPRTTAKKAEAYMKSLGIGDTVYVGPEAEFFVFDDVKYKADPYNTGFKLDSTELPSNDDTEYET
GNLGHRPRVKGGYFPVPPIDSAQDMRSEMLTVLTEMGVTVEKHHHEVASAQHELGVKFDTLVRNA
DKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHVHFSIWKEGKPTFAGNEYAGLSENCLYFI
GGVIKHAKAVNAFTNPSTNSYKRLVPGYEAPVLLAYSARNRSASCRIPFGNSPKSKRLEVRFPDPAA
NPYLCFAALLMAGLDGIKNKIHPGQAMDKDLYDLPAKELKKIPTVCGSLREALQSLDKDREFLKAG
GVFDDDQIDSFIELKMAEVMRYETTPHPVEYDMY

Appendix 2. Amino acid sequences of 20 GSI used to generate Logo in Fig 7.

Ochrobactrum_anthropi (Oant_2087)

MTTANDILKQIKDNDIKFVDLRFTDPKGKLQHVTMDIGLVDEEMFVDGVMFDGSS
IAGWKAINESDMVLMPDPTTAHIDPFFAQSTLVILCDILDPISGEAYGRDPRTTAKK
AEAYMKSLGIGDTVYVGPEAEFFVFDDVKYKADPYNTGFKLDSTELPSNDDTEYE
TGNLGHRPRVKGGYFPVPPIDSAQDMRSEMLTVLTEMGVTVEKHHHEVASAQHE
LGVKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHVHFSI
WKEGRPTFAGNEYAGLSENCLYFIGGVIKHAKAVNAFTNPSTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPFGNSPKSKRLEVRFPDPAANPYLCFAALLMAGLDGIKNKI
HPGQAMDKDLYDLPAKELKKIPTVCGSLREALQSLDKDREFLKAGGVFDDDQIDS
FIELKMAEVMRYETTPHPVEYDMYYSV

Brucella_ovis

MTTANDILKQIKDNDVKFVDLRFTDPKGKLHHVTMDVGLVDEDMFIDGVMFDGS
SIAGWKAINESDMVLMPDPETAHIDPFFAQSTLVILCDILEPLSGESYSRDPRTTAKK
AEAYLKSLGIGDTVYVGPEAEFFVFDDVKFKADPFNTGFKLDSAELPSNDDTDYET
GNLGHRPRVKGGYFPVPPIDSLQDMRSEMLTVLSEMGVTVEKHHHEVASAQHEL
GTKFDTLVRNADKMQIQKYVVHQVANAYGKTATFMPKPVFGDNGSGMHVHFSI
WKDGKPTFAGNEYAGLSENCLYFIGGVIKHAKAVNAFTNPSTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPFGSSPKSKRLEVRFPDPSANPYLCFSALLMAGLDGIKNKI
HPGQAMDKDLYDLPPKELKQIPTVCGSLREALQSLDKDREFLKAGGVFSDDQIDSF
IELKMAEVMRFEMTPHPVEFDMYYSV

Sinorhizobium_meliloti

MTTANEVLKQIKENDVKFVDLRFTDPKGKLQHVTMDVVCVDEDMFADGVMFDG
SSIGGWKAINESDMVLMPDPETAHMDPFFAQSTMVIFCDILDPVSGEAYNRDPRGT
AKKAEAYLKASGIGDTVFVGPEAEFFVFDDVKYKADPYNTGFKLDSSELPSNDDT
DYETGNLGHRPRVKGGYFPVPPVDSSQDMRSEMLTVLSEMGVTVEKHHHEVAAA
QHELGVKFDALVRNADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHV
HLSIWKDGKPTFAGDEYAGLSESCLYFIGGIIKHAKALNAFTNPSTNSYKRLVPGYE
APVLLAYSARNRSASCRIPFGTNPKAKRVEVRFPDPTANPYLAFAAMLMAGLDGIK
NKLHPGKAMDKDLYDLPPKELKKIPTVCGSLREALESLDKDRKFLTAGGVFDDDQ
IDSFIELKMQEVMRFEMTPHPVEYDMYYSV

Agrobacterium_tumefaciens

MTTANDILKQIKDNDIKFVDLRFTDPKGKLQHVTMDVVCVDEDMFADGVMFDGS
SIAGWKAINESDMVLMPDPATAHIDPFFAQSTLVVLCDILDPVSGEAYNRDPRGTA
KKAEAYLKASGIGDTVFVGPEPEFFVFDDVKYKADPYNTGFKLDSSELPSNDDTDY
ETGNLGHRPRVKGGYFPVPPIDSLQDMRSEMLTVLAEMGVVVEKHHHEVAAAQH
ELGVKFDTLVSSADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHVHQS
IWKGGKPTFAGDEYAGLSENCLYYIGGIIKHAKAINAFTNPTTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPFGSNPKAKRVEVRFPDPTQNPYLGFAAMLMAGLDGIKN
KIHPGKPMDKDLYDLPAKELKKIPTVCGSLREALESLDKDRKFLTTGGVFDDDQID
SFIELKMQEVMRFEMTPHPVEFDMYYSA

Rhizobium_leguminosarum

MATASEILKQIQENDVKFVDLRFTDPKGKLQHVTMDVACVDEDMFADGVMFDGS
SIGGWKAINESDMVLMPDTETVHMDPFFAQSTMVIVCDILDPVSGEAYNRDPRGT
AKKAEAYLKASGIGDTIFVGPEAEFFVFDDVKYKADPYNTGFKLDSTELPSNDDTD
YETGNLGHRPRVKGGYFPVPPVDSAQDMRSEMLTVLSEMGVVVEKHHHEVAAA
QHELGIKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHVH
QSIWKGGKPTFAGDEYAGLSETCLYYIGGIIKHAKAINAFTNPSTNSYKRLVPGYEA
PVLLAYSARNRSASCRIPFGSNPKAKRVEVRFPDPTANPYLAFAAMLMAGLDGIKN
KIHPGKAMDKDLYDLPPKELKKIPTVCGSLREALESLDKDRKFLTAGGVFDDDQID
AFIELKMAEVMRFEMTPHPVEYDMYYSA

Hoeflea_olei

MTTATEILKQIKDNDVKFVDLRFTDPKGKLQHVTMDVSAVDEDLFADGVMFDGS
SIGGWKAINESDMVLMPDTATAHMDPFFAQSTMVLICDILDPITGEAYNRDPRTTA
KKAEAYLQASGLGDTVYIGPEPEFFIFDDVKYKADPYNTGFKLDSSELPSNDDTDY
ETGNMGHRPRVKGGYFPVPPVDSCQDMRSEMLTVLAEMGLTVEKHHHEVASAQ
HELCMVFDTLTRQADKTQIYKYGVHQVANAYGKTATFMPKPVFGDNGSGMHVH
QSIWKGGKPTFAGDEYAGLSETCLFYIGGIIKHAKAINAFTNPSTNSYKRLVPGYEA
PVLLAYSARNRSASCRIPFGSSPKAKRVEVRFPDPTANPYLAFAAMLMAGIDGIKN
KIHPGKAMDKDLYDLPPKELKKIPTVSGSLREALENLDKDRKFLTAGGVFDDDQID
AFIELKMVEVMRYEMTPHPVEFDMYYSA

Bartonella_vinsonii

MTTASDIIKQIADNEIRFVDLRFTDPRGKLHHVTMDIAEISEDTFSDGVMFDGSSISG
WKTINESDMVLMPDPETAHIDPFFAQSTLVIFCDVLDPVSGEFYRRDPRSIAKRAEV
YMKSLGIGDTINVGPEAEFFIFDDVRYKTDPYNTGFKLDSSELPSNDDTEYEVGNL
GHRPRMKGGYLPVPPIDSCQDMRSEMLTALKDMGVRVEKHHHEVAAGQHELGIR
FDTLVREADKMQIFKYVVHQIANSYGKTATFMPKPVFGDNGSGMHVHISIWKDGK
PIFAGNEYAGLSETCLFFIGGVIKHAKALNAFTNPSTNSYKRLVPGYEAPVLLAYSA
RNRSASCRIPMSSSPNSKRVEVRFPDPTANPYLAFAALLMAGLDGIKNKIHPGHAM
DKDLYDLPLKERKEIPTVSGSLREALEALDKDRSFLKAGDVFDDDQINSFIQVKMQ
EVLCYETTPHPVEFDMYYSV

Labrenzia_marina

MTTAAEVLKEIQEKDVKFVDLRFTDPRGKMQHVTMDVALVDEDMFAEGVAFDG
SSIAGWKAINESDMMLILDPESAHIDPFFAQSTLAIFCDIVDPITGEGYNRDPRMTAK
KAEAYVKSGGFGDTIYIGPEAEFFMFDDVRFTADPYNTGFILDGDELPSNMGSEYE
TGNLGHRPRTKGGYFPVPPIDSAQDIRSEMLSVMGEMGVPTEKHHHEVAAAQHEL
GMKFDHLTRCADNMQVYKYVVHQVAHAYGKTATFMPKPVFGDNGTGMHCHLSI
WNKGEPVFAGNQYADLSETCLYFIGGILKHAKALNAFTNPSTNSYKRLVPGYEAP
VLLAYSSRNRSASCRIPFTASPKSKRVEVRFPDPTANPYLCFSALLMAGLDGIKNKL
HPGDAMDKNLYDLPPEELSEIPTVCGSLREALEAVDEDREFLKAGGVFDDDQIDAY
IELKMEEVERYEMTPHPVEFDMYYSV

Nitrobacter_hamburgensis

MKTASDVLKSIKDNDVKYVDLRFTDPRGKWQHVTFDVSMIDEDIFTEGTMFDGSS
IAGWKAINESDMMLMLDPATAAIDPFFAETTMVITCDIMEPSTGEPYNRDPRGIAK
KAEAMVKSMGIGDTVYIGPEAEFFVFDDVRFSADPYNTGFKLDSSELPTNSATEYE
GGNLGHRIRTKGGYFPVPPQDSVQDMRSEMLGAMAKMGVKVEKHHHEVASAQH
ELGMKFDTLTYMADQMQVYKYCIHQVAHIYGKTATFMPKPIFGDNGSGMHVHQS
IWKDGKPTFAGNKYADLSETCLHYIGGIIKHAKAINAFTNPSTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPYTTSPKAKRVEVRFPDPMANPYLAFAAMLMAGLDGIKN
KIDPGPAMDKDLYDLPKEELKQIPTVCGSLREALESLDKDRAFLKNGGVFDDDFIN
AYIELKMTEVERFDMTPHPVEFDMYYSY

Bradyrhizobium_japonicum

MKTAKDVLKSIKDNDVKYVDLRFTDPRGKWQHVTFDISMIDEDIFAEGTMFDGSSI
AGWKAINESDMCLMPDPVTATIDPFFAETTMVITCDVLEPTTGEPYNRDPRGIAKK
AEAMVKSMGVGDTVFVGPEAEFFVFDDVRFSANPYSTGFRLDSSELPTNSDTEYE
GGNLGHRVRTKGGYFPVPPQDSVQDMRSEMLGAMAKMGVKVEKHHHEVASAQ
HELGMKFDTLTLMADHMQIYKYCIHQVAHIYGKTATFMPKPVYGDNGSGMHVH
QSIWKDGKPVFAGNKYADLSETCLHYIGGIIKHAKAINAFTNPSTNSYKRLVPGYE
APVLLAYSARNRSASCRIPYTASPKAKRVEVRFPDPLANPYLGFAAMLMAGLDGIK
NKIDPGPAMDKDLYDLPKEELKQIPTVCGSLREALENLDKDRGFLKNGGVFDDDFI
DSYIELKMTEVERFEMTPHPVEFDMYYSG

Erythrobacter_luteus

MSKASDIIKRIKDEEIEWVDLRFTDPKGKWQHLSMVASALDEDQLEEGLMFDGSSI
AGWKAINESDMILKPDLEEVWMDPFSATPMMIVNCDIVEPSTGEGYGRDPRTTAK
RAEAYLKSTGIGDTVYVGPEAEFFMFDDVRFEDGYAGSGFQIDDVELPTNTGTDY
DSGNMAHRPRVKGGYFPVAPIDSAVDIRGEMVSTMLEMGLPCDKHHHEVAAAQH
ELGLTFGTLVQTADRMQIYKYVVHQVAHAYGKTATFMPKPIKDDNGSGMHTHMS
IWDGGKPTFAGNGYAGLSDNCLYYIGGVIKHAKALNAFTNPTTNSYKRLVPGFEA
PVLLAYSARNRSASCRIPYGSGEKAKRVEFRFPDAMANPYLAYAALLMAGLDGIQ
NKIHPGEAMDKNLYDLPPEELKEVPTVCGSLREALTELEKDHEFLLAGDVFTKDQI
DAYIELKWEEVMRVETTPSAVEFDLYYSM

Nisaea_denitrificans

MSDVNAVLSMIKEHDCKFVDLRFTDPRGKMQHVTQAIETIDAESLVEGFMFDGSSI
AGWKAINESDMSLKLDLSTARVDPFFAQPTLMILCDVVDPITGQPYERDPRSTAKA
ALNHLNSLGIGDTAFFGPEAEFFVFEDVKIKTGSNIGYYEVDHPEGPYNSARSYEEG
NMGHRPGVKGGYFPVPPVDSEQDLRSEMLAVMGEMGVDIEKHHHEVAPAQHEL
GMKFGTLIETADALQMYKYVVHNVAHAYGKTATFMPKPIAEDNGSGMHVHQSI
WKDGKPLFAGNGYADLSEMCLYYIGGIIKHAKALNAFTNPTTNSYKRLVPGFEAP
VLLAYSARNRSASCRIPFVNSPKGKRVEVRFPDPAGNPYLAFSAMLMAGLDGIQN
KIHPGDPMDKDLYDLPPEELADIPTVAGSLREALESLDADRSFLTQGDVFTNDQID
AYIELKMDEVIRFEQTPHPVEFEMYYSV

Escherichia_coli

MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSI
GGWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRA
EDYLRSTGIADTVLFGPEPEFFLFDDIRFGSSISGSHVAIDDIEGAWNSSTQYEGGNK
GHRPAVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAHHHEVATAGQNEVAT
RFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLSKN
GVNLFAGDKYAGLSEQALYYIGGIIKHAKAINALANPTTNSYKRLVPGYEAPVML
AYSARNRSASIRIPVVSSPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPG
EAMDKNLYDLPPEEAKEIPQVAGSLEEALNELDLDREFLKAGGVFTDEAIDAYIAL
RREEDDRVRMTPHPVEFELYYSV

Shigella_dysenteriae

MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSI
GGWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRA
EDYLRSTGIADTVLFGPEPEFFLFDDIRFGSSISGSHVAIDDIEGAWNSSTQYEGGNK
GHRPAVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAHHHEVATAGQNEVAT
RFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLSKN
GVNLFAGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVML
AYSARNRSASIRIPVVSSPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPG
EAMDKNLYDLPPEEAKELPQVAGSLEEALNELDLDREFLKAGGVFTDEAIDAYIAL
RREEDDRVRMTPHPVEFELYYSV

Salmonella_enterica

MSSEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSI
GGWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRA
EDYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKYEGGN
KGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAHHHEVATAGQNEV
ATRFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLA
KNGTNLFSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPV
MLAYSARNRSASIRIPVVASPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIH
PGEAMDKNLYDLPPEEAKEIPQVAGSLEEALNALDLDREFLKAGGVFTDEAIDAYI
ALRREEDDRVRMTPHPVEFELYYSV

Vibrio_cholerae

MSVENVLSLIQENEVKFVDLRFTDTKGKEQHISIPAHQIDADFFEDGKMFDGSSVA
GWKGINESDMVMMPDPSSAVLDPFTEDATLNIRCDILEPATMQGYDRDPRSIAKR
AEEYMRSTGIADTVLVGPEPEFFLFDDVKFATNMSGSFFKIDDVEAAWNTGTEYED
GNKGHRPGVKGGYFPVAPVDSSQDIRSAMCLIMEEMGLVVEAHHHEVATAGQNE
IATRFNTLTTKADEIQIYKYVVHNVAHAFGKTATFMPKPLVGDNGSGMHVHQSLA
KDGVNLFAGDKYGGLSEMALYYIGGVIKHARALNAITNPSTNSYKRLVPHYEAPV
MLAYSARNRSSSIRIPVVPSPKARRIEVRFPDPAANPYLAFAALLMAGLDGIKNKIH
PGEAMDKDLYDLPAEEAAEIPKVAESLQQALQYLDADREFLTAGGVFSDDFIDSYI
ELKTKDVERVNVAVHPLEFELYYSV

Klebsiella_variicola

MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPSHQVNAEFFEEGKMFDGSSI
GGWKGINESDMVLMPDASTAVIDPFYEEPTLIIRCDILEPGTLQGYDRDPRSIAKRA
EEYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKYEGGNK
GHRPGVKGGYFPVPPVDSSQDIRSTMCMIMEEMGLVVEAHHHEVATAGQNEVAT
RFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLAK
NGTNLFSGDKYAGLSEQALFYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVM
LAYSARNRSASIRIPVVTSPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHP
GEAMDKNLYDLPPEEAKEIPQVAGSLEEALLALDADREFLTAGGVFTNDAIDAYIA
LRLEENDRVRMTPHPVEFELYYSV

Pseudomonas_aeruginosa

MSYKSHQLIKDHDVKWVDLRFTDTKGKQQHVTMPARDALDDEFFEAGKMFDGS
SIAGWKGIEASDMILMPDDSTAVLDPFTEEPTLILVCDIIEPSTMQGYERDPRNIAKR
AEEYLKSTGIGDTVFVGPEPEFFIFDEVKFKSDISGSMFKIFSEQASWNTDADIESGN
KGHRPGVKGGYFPVPPVDHDHEIRTAMCNALEEMGLVVEVHHHEVATAGQNEIG
VKFNTLVAKADEVQTLKYCVHNVADAYGKTATFMPKPLYGDNGSGMHVHMSIS
KDGKNTFAGEGYAGLSETALYFIGGIIKHGKALNGFTNPSTNSYKRLVPGFEAPVM
LAYSARNRSASIRIPYVSSPKARRIEARFPDPAANPYLAFAALLMAGLDGIQNKIHP
GDAADKNLYDLPPEEAKEIPQVCGSLKEALEELDKGRAFLTKGGVFTDEFIDAYIE
LKSEEEIKVRTFVHPLEYDLYYSV

Xanthomonas_campestris

MSVENVEKLIKDNKVEFVDLRFVDMRGVQQHVTFPANIIEPALFEEGKMFDGSSIA
GWKGINESDMVLLPDAGTAYLDPFFADPTVVLTCDILDPATMQSYARDPRGIAKR
AEAYLKSSGIADQAFFGPEPEFFIFDSVRFANDMGHTFFQVGSEEAAWNTGAKYDG
GNSGYRPGVKGGYFPVPPTDTLHDLRAEMIKTLEQVGIETEVHHHEVATAGQCEIG
TKFSSLVQKADELLTMKYIIKNVAYRNGKTATFMPKPIVGDNGSGMHVHQSLAKG
GANLFSGDGYGGLSQLALWYIGGIFKHARAINAFANSGTNSYKRLVPGFEAPVML
AYSARNRSASCRIPWVSNPKARRIEMRFPDPLQSGYLTFTALMMAGLDGIKNQIDP
GAPSDKDLYDLPPEEEKLIPQVCSSLDQALEALDKDREFLKAGGVMSDDFIDGYIA
LKMQEVTKFRAATHPLEYQLYYGN

Neisseria_meningitidis

MSIKNAVKLIEESEARFVDLRFTDTKGKQHHFTVPVRIVLEDPEEWFENGQAFDGS
SIGGWKGIQASDMQLRPDASTAFVDPFYDDVTVVITCDVIDPADGQGYDRDPRSIA
RRAEAYLKSSGIGDTAYFGPEPEFFVFDGVEFETDMHKTRYEITSESGAWSSGLHL
DGQNTGHRPTVKGGYAPVAPIDCGQDLRSAMVNILEELGIKVEVHHSEVGTGSQM
EIGTRFATLVKRADQTQDMKYVIQNVAHNFGKTATFMPKPIMGDNGSGMHVHQSI
WKDGQNLFAGDGYAGLSDTALYYIGGIIKHAKALNAITNPSTNSYKRLVPHFEAPT
KLAYSAKNRSASIRIPSVNSSKARRIEARFPDPTANPYLAFAALLMAGLDGIQNKIH
PGDPADKNLYDLPPEEDALVPTVCASLEEALAALKADHEFLLRGGVFSKDWIDSYI
AFKEEDVRRIRMAPHPLEFEMYYSL

Appendix 3. Amino acid sequences of 30 GSs used to build the gene tree in Fig 8.

>***Ochrobactrum_anthropi_Oant_2087***

MTTANDILKQIKDNDIKFVDLRFTDPKGKLQHVTMDIGLVDEEMFVDGVMFDGSS
IAGWKAINESDMVLMPDPTTAHIDPFFAQSTLVILCDILDPISGEAYGRDPRTTAKK
AEAYMKSLGIGDTVYVGPEAEFFVFDDVKYKADPYNTGFKLDSTELPSNDDTEYE
TGNLGHRPRVKGGYFPVPPIDSAQDMRSEMLTVLTEMGVTVEKHHHEVASAQHE
LGVKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHVHFSI
WKEGKPTFAGNEYAGLSENCLYFIGGVIKHAKAVNAFTNPSTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPFGNSPKSKRLEVRFPDPAANPYLCFAALLMAGLDGIKNKI
HPGQAMDKDLYDLPAKELKKIPTVCGSLREALQSLDKDREFLKAGGVFDDDQIDS
FIELKMAEVMRYETTPHPVEYDMYYSV

>Ochrobactrum_anthropi_Oant_3936

MSADTTEKKVTRAPRRRTPAYVKSLRGVKNWKQATEWLAWRDIEDIECITPDQA
GVARGKMMPSKKFTSNTSLALPSAVFMTTISGDYPEDGYGFHYPEDDGDLKLVPD
LSTLSAVPWESDPTAQVICDLVYQDGRGVEFTPRNVLRNVVAAYSKRGLKPVVAP
EIEFYLVRKNPDPDYPLTPPVGRSGRAIGGGQGYSIAGVNEFDELIDDIYHFSEGQG
LEIDTLIHEEGAGQLEINLRHGDPVELADQVFMFKRTIREAALKHDMYATFMAKPI
QGQPGSAMHIHQSIVDKKTGRNIFTNEDGSESEAFRHFIGGMQRHVPNALVMFAPY
VNSYRRLTPDASAPVNVKWGYDNRTTAFRVPRSDPNGRRVENRIPSSDANPYLAL
AASLACGLIGLVNKIEAEQPATTSVNTKEIELPRGLIDAVELFEEDTELRNLFGSSFV
TTYAAIKRAEFETFMEVISPWEREFLLLNV
>Ochrobactrum_anthropi_Oant_3881

MAGQLTFDALKKAVANDEIDTVLACFVDMQGRLIGKRFYGQFFVESGYDETHGC
NYLLADDIDMEPVPGYEAAGWDKGYGDFVIKPDLSTLRVATWLEKTAIVLCDVLD
HHDHQDLAHSPRAILKKQLARLHERGYRAYFASELEFYLFDETYKTARAKHWQD
METASPYVQDYVIHLTTKEEPVLRAMRNQLAAAGIPVENSKGEWGPGQEELNVRY
AEALEMADRHVIMKNAMKEIAEAHGKCITFMAKYDYGKAGSSSHVHNSIWSADG
KEPLFFDPKAPYTMTPLMRSWVAGQLKYATDYTYFLAPYINSYKRFQAGTFAPTKI
MWSQDNRTAGFRLCGEGTKGIRIECRIGGADLNPYLAFAALIASGLQGIDEQLELD
EPFVGDAYSAVKLKEIPYTLREAAQALKNSSFLKEAFGDAVVNHYVHTAHWEQIE
YDRRVTDWELHRGFERY

>Ochrobactrum_anthropi_Oant_4491

MNDKTLSGLTELATFVTTDIAGITRGRSFAASEIEDYLRKGVGWVPANLALTPFDLI
ADPNPWGSAGDLRLMADPASKARVTCLPDETPLHFYHSDITDLKGEPWDCCVRSF
LKATLSQFEKEAGLKVVSAVEQEFQVLGADWPAAPAFGLRAQRRAEPFGCLLMT
ALKEAGAEPEMFLPEYGKDQFEVTCRPAPALVAADRGATIRAVTREVAALFGWH
ASFAPKTDPNGVGNGVHLHVSFTDLDGNPVTFDAARPGRLSRVAGAFAAGVIKHL
PALVAFTAPSVLSYMRLVPHHWSAAYTCLGEKNREATLRICPTLDLPGSNPAKQFN
MEYRAADACASPHLSLAVLLRAGLEGIRAGLEQPPLINSDPSAFSSEEQVRLGIRRL
PSSLAEALDTLAADEVVTGWFPKDFLDCYFAMKRKEIEIVEGLSPEALCARYAAVY

>Ochrobactrum_anthropi_Oant_4157

MTERLAMTKYKLEYIWLDGCTPAAGLRGKTQIKEFDAFPTLEQLPLWGFDGSSTM
QAEGRSSDCVLKPVAIYPDPARKNGVLVMCEVMMPDGITPHPSNSRATVLEDDDA
WFGFEQEYFFYKDGRPLGFPEQGFPAPQGPYYTGVGYKNVGSIARKIVEEHLDLCL
DAGINHEGINAEVAKGQWEFQIFGKGSKNAADQIWMARYLLLRLCEQYEIDIEFHC
KPLGDTDWNGSGMHCNFSTKYMREVGGKDYFEALMAQFDKNLQDHIDVYGPDN
HMRLTGKHETAPWNKFSYGVADRGASIRVPHAFVRDGYRGYLEDRRPNSQGCPY
QIASQVLKTISEVPTSEDEALAA

>***Ochrobactrum_rhizosphaerae***

MTTANDKKDNDKVDRTDKGKHVTMDGVDMVDGVMDGSSAGWKANSDMVMD
DTAHDASTVCDDSGAYGRDRTTAKKAAYMKSGGDTVYVGAVDDVKYKVDNTG
KDSTSNDDTDYTGNGHRRVKGGYVDSADMRSMTVTMGVTVKHHHVASAHGVK
DTVRNADKMYKYVVHVANAYGKTATMKVGDNGSGMHVHSWKGKTAGNYAGS
NCGGVKHAKAVNATNSTNSYKRVGYAVAYSARNRSASCRGSSKSKRVRDAANY
CAAMAGDGKNKHGAMDKDYDAKKKTVCGSRASDKDRKAGGVDDDSKMAVMR
YTTHVDMYYSV

>Falsochrobactrum_ovis

MTTAADILKQIKDLDIKFVDLRFTDPKGKLQHVTMDIGLVDEDMFIDGVMFDGSSI
GGWKAINESDMVLMPDPETAHIDPFFAQSTLVILCDILDPVSGEAYSRDPRTTAKK
AEAYMRSLGIGDTVFVGPEAEFFVFDDVKYKVDPFNTGFKLDSTELPSNDDTDYET
GNLGHRPRMKGGYFPVPPIDSAQDMRSEMLTVLTEMGVTVEKHHHEVASAQHEL
GVKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPVFGDNGSGMHVHFSI
WKDGKPTFAGNEYAGLSETCLYFIGGVIKHAKAVNAFTNPTTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPFGSSPKSKRLEVRFPDPSANPYLCFAALLMAGLDGIKNKI
HPGQAMDKDLYDLPAKELKEIPTVCGSLREALQALDKDREFLKVGGVFEDDQIDS
FIELKMAEVMRYETTPHPIEFDMYYSV

>Phyllobacterium_leguminum

MTTANDILKQIKDNDVKFVDLRFTDPKGKLQHVTMDVGMVDEEMFTDGVMFDG
SSIAGWKAINESDMVLMPDVETAHIDPFFAQSTMVVICDILDPVSGEAYNRDPRTT
AKKAEAYLKSLGIGDTIFVGPEAEFFVFDDVKYKVDPYNTGFKLDSTELPSNDDTD
YETGNLGHRPRVKGGYFPVPPVDSAQDMRSEMLTVLTEMGVTVEKHHHEVAAA
QHELGLKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHV
HMSIWKDGKPTFAGNEYAGLSENCLFFIGGVIKHAKAINAFTNPSTNSYKRLVPGY
EAPVLLAYSARNRSASCRIPFGSSPKSKRVEVRFPDPAANPYLGFAALLMAGLDGI
KNKIHPGQPMDKDLYDLPAKELKKIPTVCGSLREALQSLDKDRGFLKAGGVFDDD
QIDAFIELKMAETMRYETTPHPVEYDMYYSV

>Paramesorhizobium_deserti

MTTANDILKQIKDNDVKFVDLRFTDPKGKLQHVTMDVGVVDEDMFADGVMFDG
SSIAGWKAINESDMVLMPDVETAHIDPFFAQSTMVVMCDILDPISGEAYNRDPRGT
AKKAEAYLKSLGIGDTIFVGPEAEFFVFDDVKYKADPYNTGFKLDSTELPSNDDTD
YETGNLGHRPRIKGGYFPVPPVDSAQDMRSEMLTVLTEMGVTVEKHHHEVAAAQ
HELGLKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPVFGDNGSGMHVH
MSIWKDGKPTFAGNEYAGLSEACLFFIGGIIKHAKAINAFTNPSTNAYKRLVPGYE
APVLLAYSARNRSASCRIPFGTSPKSKRVEIRFPDPSANPYLGFAALLMAGLDGIKN
KIHPGQAMDKDLYDLPAKELKKIPTVCGSLREALQSLDKDRGFLKAGGVFDDDQI
DAFIELKMAETLRYETTPHPVEYDMYYSV

>Agrobacterium_albertimagni

MTTANDILKQIKDNDIKFVDLRFTDPKGKLQHVTMDIACVDEDMFADGVMFDGSS
IAGWKAINESDMVLMPDPATVHMDPFFAQSTMVILCDILDPVSGEAYNRDPRGTA
KKAEAYLKASGIGDTVFVGPEAEFFVFDDVKYKADPYNTGFKLDSTELPSNDDTE
YETGNLGHRPRVKGGYFPVPPIDSLQDMRSEMLTVLGEMGVVVEKHHHEVAAAQ
HELGIKFDTLVRNADKMQIYKYVVHQVANAYGKTATFMPKPIFGDNGSGMHVHQ
SIWKDGKPTFAGDEYAGLSESCLYYIGGIIKHAKALNAFTNPSTNSYKRLVPGYEAP
VLLAYSARNRSASCRIPFGTNPKAKRVEVRFPDPMANPYLAFAAMLMAGLDGIKN
KIHPGKAMDKDLYDLPPKELKKIPTVCASLREALESLDKDRKFLTAGGVFDDDQID
SFIELKMQEVMRFEMTPHPVEYDMYYSA

>Brucella_inopinata

MSADTTEKKATRTPRRRTPAYVKSLRGVKNWKQATEWLAWRDIEDIECITPDQAG
VARGKMMPSKKFTSNTSLALPSAVFMTTISGAYPEDGYGFHYPEDDGDLKLLPDL
TTLSAVPWETDSTAQVICDLVYQDGRAVEFTPRNVLRNVIAAYSKRGLKPVVAPEI
EFYLVRKNPDPDYPLTPPVGRSGRAIGGGQGYSIAGVNEFDELIDDIYHFSEGQGLE
IDTLIHEEGAGQLEINLRHGDPVELADQVFMFKRTIREAALKHDMYATFMAKPIQG
QPGSAMHIHQSIVDKKTGRNIFTNEDGSESQAFHHFLGGMQRHVPNALVMFAPYV
NSYRRLTPDASAPVNVKWGYDNRTTAFRVPRSDPSARRVENRIPSSDTNPYLALAA
SLACGLIGLVNKIEPEQPATTSVNTKEVELPRGLIDAVELFEEDAELRNLFGSSFMTT
YAAIKRAEFETFMEVISPWEREFLLLNV

>Pseudochrobactrum_asaccharolyticum

MPVTPTEKKKPRKAVRRGTPAYVKSLRGVKNWKEASEWLAWRDIEDIECITPDQA
GVARGKMMPSKKFISNSTLALPSAVFMVTISGDYPDDGHGFVYPEDDGDLRLVAD
LSTLTVVPWESDPTAQVLCDLVYQDGRVAEFTPRNVLKRVVDEYAKLGLKPVVA
PEIEFYLVRKNPDPDYPLTPPVGRSGRAIGGGQGYSIAGVNEFDELIDDIYHFSEAQG
LEIDTLIHEEGAGQLEINLRHGDPIELADQAFMFKRTIREAALKHDMYATFMAKPIQ
GQPGSAMHIHQSIIDKKTGHNIFSNEDGTESEAFYHFIGGMQRHIPNALVMFAPYVN
SYRRLTPDASAPVNVKWGYDNRTTALRVPRSDPQARRVENRIPSSDANPYLALAA
SLACGLIGLKNKIMPDAPVATSVNRNEVELPRGLIEAVSLFEQDSELKALLGESFAT
TFAAIKRAEFETFMEVISPWEREYLLLNV

>Zhengella_mangrovi

MASEKRDVRNTSAQRARVPKFVKNLRGVKNWKEVNAWLDWRGIEDIECITPDQA
GVARGKMMPSKKFTSNTSLALPSAIFMTTISGDYPEESETFQYPEDDGDLKLMPDL
STLAVVPWESDPTAQVICDMVHQDGREVEFTPRNVLKRVIRAYDKLGLKPVVAPE
IEFYLVSKNTDPDYPLVPPLGRSGRPIGGGQGYSIAGINEFDELIDDIYHFSEGQGLEI
DTLIHEEGAGQLEINLRHGDPVELADQVFMFKRTIREAALKHDMYATFMAKPIQG
QPGSAMHIHQSIIDKKTGRNIFTSDDGSESEAFFHFIGGMQKHVPNALVMFAPYVNS
YRRLTKAVTAPVNVEWGYDNRTTAFRIPRSDPAARRVENRIPSSDANPYLALAASL
ACGLIGMKQKIKPKEPAGHTANEADIELPRGLIEATSLFEADNDLIDMLGAGFVGT
YAAIKRGEFETFMQVISPWEREFLLLNV

>Mesorhizobium_ephedrae

MPPAKKEARPTGRAGRVRVPAYVKNLRGVTAWKQVSEWLEWRSIEDIECITPDQA
GVARGKMMPSKKFTSNTSLALPSAPFMMTISGDYPEDGNGFSYPEDDGDLKLVPD
LSTLSVVPWEEDPTAQVICDLVHQDGRNVEFTPRNVLKRVVAAYGKRGLRPVVAP
EIEFYLVRKNPDPDYPLVPPVGRSGRPIGGGAGYSIAGVNEFDELIDDIYHFSESQGL
EIDTLIHEEGAGQLEINLRHGDPVELADQVFMFKRTIREAALKHDTYATFMAKPIQ
GQPGSAMHIHQSIVDKKTGRNVFTDETGKETEAFRHFIGGMQKHVPNALVMFAPY
VNSYRRLTQAASAPVNTKWGYDNRTTAFRVPRSDPAARRVENRIPSSDANPYLAL
AASLACGLIGMTKNIEPDAPVGTTVNEDEIDLPRGLLEAVDLFENDEDLRDMLGSS
FASTYAAIKRAEFETFMEVISPWEREFLLLNV

>Aquamicrobium_aerolatum

MSAQKKEARSTGRGARVRTPAFLQSLRGVSDLKQATAWLESRGIEDIECITPDQAG
VARGKMMPSKKFTSNTSLALPSAVFMATISGGYPEDGHGFVYPEDDGDLKLVPDL
STLSIVPWESDPTAQVICDLVHQDGRSIEFTPRNVLRRILAAYAEHGLRPVVAPEIEF
YLVRKNPDPDYPLEPPVGRSGRAIGGGQGYSIAGVNEFDELIDDIYHFSEAQGLEID
TLIHEEGAGQLEINLRHGDPIELADQVFMFKRTIREAAFKHDTYATFMAKPIQGQPG
SAMHIHQSIVDLKTGNNVFTAIDGTETEAFRHFIGGMQRHVPSALVMFAPYVNSYR
RLTQSASAPVNTKWGYDNRTTAFRVPRSDPAGRRVENRIPSSDANPYLAIAASLAC
GLIGMLNKEHCDAPVGTSANEDEIDLPRGLLEAVELFEGDDALRDLLGSDFVTTYA
AIKKAEFETFMEVISPWEREYLLLNV

>Ochrobactrum_oryzae

MAGQLTFDALKKAVAENEIDTILACFVDMQGRLIGKRFYGQFFVESGYDETHGCN
YLLADDIDMEPVPGYEAAGWDKGYGDFVIKPDLSTLRVATWLEKTAIVLCDVLDH
HDHQDLAHSPRAILKKQLARLHERGYRAYFASELEFYLFDETYKTARAKHWQDM
ETASPYVQDYVIHLTTKEEPVLRAMRNQLAAAGIPVENSKGEWGPGQEELNVRYA
EALEMADRHVIMKNAMKEIAEAHGKWITFMAKYDYSKAGSSSHVHNSIWSADGK
EPLFFDPKAPYTMTPLMRSWVAGQLKYATDYTYFLAPYINSYKRFQAGTFAPTKI
MWSQDNRTAGFRLCGEGTKSIRIECRIGGADLNPYLAFAALIASGLQGIDEQLELDE
PFVGDAYSAVKLKEIPYTLREAAQALKNSEFLKEALGEAVVNHYVHTAHWEQIEY
DRRVTDWELHRGFERY

>Mesorhizobium_metallidurans

MAGNFSFDQLKKAVSNGEVDTVLACIVDMQGRLAGKRFLAQYFVDSAHGETHGC
NYLLAADIDMEPVPGYKAASWSKGYGDFVMKPDLSTLRRIPWLEKTALVICDVLD
HHDHEDLAHSPRAILKKQIKRLSDRGYIGYFASELEFYLFSETYNSARKKHWQGLD
TASPYIGDYQIGITTKEEGVMRRLRNEMEAAGIPIENSKGEWGPGQEEINVRYAEAL
DMADRHVILKNGAKEIAESEGKAISFMAKYNYALAGNSSHIHNSLWSADGKTPLF
YDKKANWTLSTLGQQWAAGQLKYAKEFTWFLAPYINSYKRFQAGTFAPTKIMWS
EDNRTAGFRLCGEGTKGIRMECRIGGADLNPYLAFAALIAAGLGGIDEKLELQKPF
VGDAYQASRLPEIPKTLRDATETLAKSRMLKQAFGEDVLEHYVHTARWEQFEYDR
RITDWELHRGFERY

>Aminobacter_aminovorans

MAGNLSFDQLKKAVAAGEIDTVLACAVDMQGRLVGKRFLAKYFVESAYDETHGC
NYLLANDIDMEPVPGYKAASWSKGYGDFVMKPDLNTIRNVPWLEKTALLLCDLK
DHHTHEDLAHSPRGILRKQVKRLQERGYLAYFASELEFYLFSETYDSARAKHWQN
LDTASPYIGDYLIGITTKEEGVMRRLRNEMEAAGIPIENSKGEWGPGQEEINVRYAE
VLEMADRHVILKNGAKEIAASEGKAISFMSKYNYGLAGNSSHIHNSLWSADGKTP
LFYDKGAEWTLSKLGQQWSAGQLKYAKEFTWFLAPYINSYKRFQSGTFAPTKIM
WSEDNRTAGFRLCGEGTKGIRMECRIGGADLNPYLAFAALIAAGLAGIDEKLELQK
PFVGDAYQAASLPEIPKTLRDATETLASSKMLREALGEEVVDHYVHTARWEQFEY
DRRITDWELHRGFERY

>Ahniella_affigens

MAAALSFEALKQAVAAGTIDTVLACMVDMQGRLVGKRFQAEYFVDSAYEETHCC
NYLLADDIDMEPVPGYQAASWSKGYGDFVLKPDLSTLRLTPWLEGTALVLCDVL
DHHTHADLPHSPRAMLKRQIARLTERGFLGMFASELEFYLFDESYESIREGHYANP
KTAGHYIEDYNILQTTREEGVMRAMRKGLQAAGIPVENSKGEWGPGQEEINIRYC
DALTMADRHAILKNACKEIAMLQGKAITFMAKWRYDLAGSSSHIHNSLWDLEGK
TSKFFDPNAPYGMSALMRSWVAGQIKYARDITWFLAPYINSYKRFQVGTFAPTRA
VWSRDNRTAGFRLCAEGSKGIRIECRIGGADLNPYLAYTGLIAAGLAGIDEQLELPA
PFEGDAYYGASLPEVPKTLRAAADALAGSSMLRQALGDEVVEHYHHTAEWEQFE
YDRRVTDWELKRGFERY

>Breoghania_corrubedonensis

MSGNLSFDDLKRAVDAGEIDTVLVCFTDMQGRLIGKRFHARFFVDGGHEETHGCD
YLLANDIDMEPVPGYAATNWATGYGDFVMKPDMATLRRIPWLEGTALVLCDVLD
HHHHAPLPHSPRGILKSQIDRLEEKGFRAFFASELEFYLFDESYESARQKRYGDLDT
AGRYIEDYHIFQTSKEEGVMRAIRNGLQEAGIPVENSKGEWGPGQEEINVRYTDAL
EMADRHVVLKNGIKEIAHLAGKAVTFMAKWDYELAGNSAHVHASLWDKAGKTP
LFYDKNAEFGMSALMKSFMAGQLKYAADMTVFLAPYINSYKRFQSGTFAPTKLV
WSRDNRTAGFRLCGEGTKAIRTECRIGGADLNPYLAFAALIATGLAGIEEGLELGSP
HVGDAYVGEDLREIPKTLREATAALDGSALMRATFGDAMVDHYVHTAKWEQLE
YDRRVTDWELKRGFERY

>Ensifer_adhaerens_OV14

MTETNNGKAHSALTELATFVTTDIAGITRGRSFAAAQIDDYLRKGVGWVPANLAL
TPFDQIAENNPWGSAGDLRLMADPASKARVTCLPDVTPLHFYHSDITDLKGDPWE
CCVRSLLKRTLEEFEREAGLKVISAVEQEFQLLGVDWPDAPSFGLRAQRRAEPFGP
LLMTALQEAGAEPEMFLPEYGKDQFEITCRPADALTAADRGATIRAITKEVAALFG
WNASFAPKTSANGVGNGVHLHVSFTDLDGNPVTFDASRPGRLSKVAGSFAAGVIR
HLPALTAFTAPSVLSYMRLVPHHWSAAYTCLGEKNREATLRICPTLDLPGSNPAKQ
FNMEYRAADACASPHLSLAVLLKAGLEGIRAGLEQPPLINSDPSDFSDADQKKLGI
RRLPASLPEALETLAEDKVVTGWFAKDFLDCYVAMKRKEIEIVDGLSPDELCARY
AAVY

>Rhizobium_freirei

MNDSSNGKALSDLTELATFVTTDIAGITRGRSFAASYIEDYLRKGVGWVPANLALT
PFDLIADPNPWGSAGDLRLMADPASKARVTCLPDVTPLHFYHSDITDLKGEPWDC
CVRSFLKATLADFEKEAGLKVLSAVEQEFQVLGADWPAAPAFGLRAQRRAEPFGS
LLMTALKEAGAEPEMFLPEYGKDQFEVTCRPASALVAADRGATIRAVTREVAALL
GWNASFAPKTDPNGVGNGVHLHVSFTDLDGNPVTFDASRPGRLSKIAGSFAAGVI
KHLPALVAFTAPSVLSYMRLVPHHWSAAYTCLGEKNREATLRICPTLDLPGSNPAK
QFNMEYRAADACASPHLSLAVVLRAGLEGIRAGLEQPPLINSDPSEFSPEEQKKLGI
RRLPSSLAEALDTLAADEVVTGWFPKDFLDCYFAMKRKEIEIVDGLSPEDLCARYA
TVY

>Shinella_zoogloeoides

MNDSSNGKALSDLTELATFVTTDIAGITRGRSFAATEIEDYLRKGVGWVPANLALT
PFDLIADPNPWGSAGDLRLMADPASKARVTCLPDETPLHFYHADITDLKGDPWEC
CVRGFLKKTLADFEKEAGLRVVSAVEQEFQIIGADWPAAPSFGLRAQRRAEPFGSL
LMTALKEAGAEPEMFLPEYGKDQFEVTCRPAPALVAADRGATIRAVTREVAALFG
WHASFAPKTEADGVGNGVHLHVSFTDLDGKPVTFDAARPGRLSKVAGSFAAGVI
KHLPALVAFTAPSVLSYMRLVPHHWSAAYTCLGEKNREATLRICPTLDLPGSNPAK
QFNMEYRAADACASPHLSLAVVLRAGLEGIRQGLEQPPLINSDPSDFPPEEQERLGI
RRLPSSLAEALDTLAADKVVTGWFPKDFLDCYFAMKRKEIEIVEGLSPEALCARYA
AVY

>Sinorhizobium_saheli

MTGTTKQRALSDLTELATFVTTDIAGITRGRSFAARHIDEYLRKGVGWVPANLALT
PFDAIAEANPWGSAGDLRLIADARSEARVTCLPDVTPLTFYHADITDLGGNPWECC
VRGFLKRTLDAFEREAGLKVVAAVEQEFQLIGVDWPEAPSFSLRAQRRADPFGPLL
MKALEEAGAEPEMFLPEYGRDQFEITCRATDALAAADRGATIRVVTKEVAALFGW
QASFAPKTSPDGVGNGVHLHVSFTDREGNPVTFDASRPGRLSEVAGAFAAGVIRHL
PALVAFTAPSVLSYLRLVPHHWSAAYTCLGEKNREATLRICPTLDLPGSNPARQFN
MEYRAADACASPHLALAVLLRAGLEGIKAGLEPPPLVNSDPSGLPAEEQERLGIRR
LPSSLGEALDTLAGDGVVTGWFARDFLDCYLAMKRKEIEIVAGLSPDELCARYAA
VYRGPVR

>Ruegeria_intermedia

MIKAEELVTFVTTDIAAITRGRSVAAANLPDALSKGVGWVPANLSLTPFDEIASPNP
FGSSGDLRLMPDPEAGVRIEGLGGRTPLHFYHSNITNLDGTPWEGCVRSMLKAAV
ADLEALGLRVVAAFEQEFQILGASWPLAPSFALSAQRRADPFGPMLMAALGAAGC
APECFLPEYGRDQFEIVCGPAGAVQAADRAVTIREVTRELAATMGWRASFCPKTD
PNGVGNGVHIHLSLTDLQGNPVTFDAARPGRLSAQAGAFAAGIVRHMAALTALAA
PSAVSYQRLKPHHWSASWNTLGEKDREATLRICPTSERPGHDPSRAFNMEFRAAD
ATASPHLALAMLIRAGIEGLKAGLSTPPIVKGDPEEMSADERARLGIRRLPTSLHEA
LAALEADTVVCGWMSPTFLDCWKGMRLKELEIVDGLDDAALCRRYAGVY

>Rhizobium_sullae

MTKYKLEYIWLDGYTPTPNLRGKTQVKEFAEFPTLEQLPLWGFDGSSTMQAEGRS
SDCVLKPVALYPDPTRTNGVLVMCEVMMPDGVTPHPSNSRATILDDEDAWFGFEQ
EYFFYEDGRPLGFPEQGYPAPQGPYYTGVGYKNVGSVAREIVEEHLDLCLAAGINH
EGINAEVAKGQWEFQIFGKGSKRAADQIWMARYLLLRLCEKYGIDIEFHCKPLGDT
DWNGSGMHCNFSTKFMREVGGKAYFEALMAQFDKNLQDHIDVYGPDNHLRLTG
KHETAPWNKFSYGVADRGASIRVPHSFVKNDYKGYLEDRRPNSQGCPYQIASQVL
KTISEVSTSGFASAAA

>Mangrovicella_endophytica

MTKYKLEYIWLDGYTPVPNLRGKTQIKEFDSFPTLEQLPLWGFDGSSTMQAEGRSS
DCVLKPVAVYPDPARTNGALVMCEVMMPDGVTPHASNSRATILDDADAWFGFEQ
EYFFYENGRPLGFPETGYPAPQGPYYTGVGYKNVGEVARTIVEEHLDQCLAAGIN
HEGINAEVAKGQWEFQIFGKGSKTAADQVWMARYLLLRLTEKYGIDIEFHCKPLG
DTDWNGSGMHCNFSTKYMREVGGKEYFEALMAQFEKNLEDHISVYGPDNHMRL
TGKHETAPWNKFSYGVADRGASIRVPHSFVNNGYKGYLEDRRPNSQGDPYQIASQ
VLKTISEVPTSASVSAAA

>Neorhizobium_alkalisoli

MTKYKLEYIWLDGYTPTPNLRGKTQIKEFDSFPTLEQLPLWGFDGSSTMQAEGRSS
DCVLKPVAIYPDPARTNGVLVMCEVMMPDGVTPHPSNSRATILDDEDAWFGFEQE
YFFYENGRPLGFPEQGYPAPQGPYYTGVGYSNVGSIAREIVEEHLDLCLAAGINHE
GINAEVAKGQWEFQIFGKGSKKAADQIWMARYLLQRLTEKYGIDIEYHCKPLGDT
DWNGSGMHCNFSTKFMREVGGKAYFEALMAQFDKNLMDHINVYGPDNDKRLTG
KHETAPWNKFSYGVADRGASIRVPHSFVKNDYKGYLEDRRPNSQGDPYQIASQVL
KTISEVPTSGFASAAA

>Ciceribacter_lividus

MTKYKLEYIWLDGYKPVANLRGKTQVKEFDNFPTLEQLPLWGFDGSSTMQAEGH
SSDCVLKPVAVYPDPARTNGVLVMCEVMMPDGVTPHASNTRATILDDEDAWFGF
EQEYFFYENGRPLGFPEQGFPAPQGPYYTGVGYKNVGAIAREIVEEHLDLCLAAGI
NHEGINAEVAKGQWEFQVFGKGSKKAADQIWMARYLLLRLCEKYGIDVEFHCKP
LGDTDWNGSGMHCNFSTKYMREVGGKEYFEALMAAFAKNWKEHIDVYGPDNHL
RLTGKHETAPWNKFSYGVADRGASIRVPHSFVNNGYRGYLEDRRPNSQGDPYAIA
SQVLKTISEVPLAASAAA

>Methylobrevis_pamukkalensis

MTKYKLEYIWLDGYKPVPNLRGKTQIKEFEAFPTLEQLPLWGFDGSSTMQAEGRS
SDCVLKPVALYPDPARTNGVLVMCEVMMPDGVTPHPTNARATILDDEGTWFGFE
QEYFFYKDGRPLGFPEHGYPAPQGPYYTGVGYSNVGDIAREIVEEHLDLCLEAGIN
HEGINAEVAKGQWEFQIFGKGSKKAADEIWMARYLLQRLTEKYGIDIEYHCKPLG
DTDWNGSGMHCNFSTKYMREVGGKEYFEALMAAFEKNLNEHIAVYGPDNHMRL
TGKHETAPWNKFSYGVADRGASIRVPHSFIKNDYKGYLEDRRPNSQGDPYQIASQ
VLKTVSEVPTGAEASAAA