INSIGHTS INTO THE GENOMIC AND METABOLIC ADAPTATIONS OF TERMITE ASSOCIATED *VERRCUCOMICROBIA* STRAINS USING FUNCTIONAL AND COMPARATIVE GENOMICS

by

Malini Kotak

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
DECEMBER 2016

Acknowledgements

I would like to express my sincere gratitude to Dr. Jorge L. M. Rodrigues for being my mentor and providing opportunities, encouragement, guidance and support throughout my doctoral studies. I thank him for believing in me and pushing me beyond what I thought were my limits. Next, I would like to thank Dr. Demuth for agreeing to be my supervisor, and giving me an opportunity to finish my dissertation. I especially thank him for patiently answering all my questions regarding evolutionary and computational biology.

I would like to thank Dr. Chrz for letting me use his lab, and for thought provoking conversations about microbial physiology. I am also thankful to Dr. Grover and Dr. Christensen for finding time amidst their busy schedule to guide me. Needless to mention, science is a collaborative endeavor and I would like to express my immense gratitude to all the erstwhile scientists who have worked hard to bring human knowledge to where it stands. I am humbled to have had the chance to be able to contribute to this collaborative effort.

Completing my Ph. D. studies would have been very difficult without support form the Biology Department personnel including Linda, Gloria, Paulette, Sherri, Sufera, Kim, Melissa and Anya. I am very grateful to Jantiya, Lita, Nit, Noom, Fon, and Luis for helping me with my research and being excellent labmates. A special thanks to Wadud for everything, starting from letting me use his pipettes to listeining to my endless rants. Thanks are due to Shweta, Debanjana, Utpal, Catherine, Diwash, Elisa, Fabio, Cori and my Micro team for making my stay in the US even more memorable.

I would like to thank Rasika, Prashanth, Bhushan, Dhvani and Dhananjay for being my family in the US and helping me make great lifetime memories. A special thanks to Prashanth for long insightful conversations on everything science/non-science, and for pushing me to be a better person. Finally, I would like to thank my parents, sisters and Adwait for their unconditional love and support throughout my graduate studies.

November 15, 2016

Abstract

INSIGHTS INTO THE GENOMIC AND METABOLIC ADAPTATIONS OF TERMITE ASSOCIATED *VERRCUCOMICROBIA* STRAINS USING FUNCTIONAL AND COMPARATIVE GENOMICS

Malini Kotak, Ph. D.

University of Texas at Arlington, 2016

Supervising Professor: Dr. Jeff Demuth

The symbiotic gut organisms of termites play an important role in lignocellulose digestion and other related metabolic processes. Previous studies on *Diplosphaera colitermitum* TAV2, a *Verrucomicrobium* isolated from the hindgut of *Reticulitermes flavipes,* revealed its microaerophilic nature and involvement in important ecophysiological processes in the termite gut. In the present study, a detailed account of genomic sequences of 3 additional strains is provided, and the genomes of all 5 TAV strains isolated so far are compared, to gain a comprehensive understanding of these organisms. Our comparative genomic study shows that these strains represent two distinct species, and despite significant number of genes being conserved across the genomes, synteny of genomes between species is lost. We found important ecological attributes like microaerophilic physiology and the ability to digest lignocellulose to be conserved at the genomic level in these strains, suggesting their common roles in O_2 scavenging and lignocellulose digestion. However, high numbers of genes associated with functional category transcription were present in species-specific genomes, suggesting differential regulation of gene expression to be a potential primary feature distinguishing these otherwise highly similar genomes. Our study identified signatures of adaptive variation in enzymes of the electron transport chain, which might be linked to microaerophilic physiology. Additionally, several membrane proteins/enzymes involved in recombination and repair, nitrogen metabolism and lignocellulose degradation were found undergoing adaptive evolution, reflecting the evolutionary mechanisms used by TAV strains to survive in the termite gut. Furthermore, the transcriptome of strain TAV2 under aerobic (20% O_2) and hypoxic conditions (2% O_2) was studied to understand its microaerophilic physiology. Putative CAZymes that degrade lignocellulose and other polysaccharides like pectin represent the majority of up-regulated genes. Genes associated with functional categories like oxidative stress protection, motility, amino acid metabolism, coenzyme metabolism were down-regulated. Additionally, changes in electron transport chain and putative oxygen sensing signal transducing proteins in the TAV2 strain were identified. In addition to corroborating our previous findings, these results reveal the metabolic adaptations used by TAV2 to survive in hypoxic conditions, and will serve as a foundation for future studies to understand the microaerophilic physiology of TAV strains.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Termites

Termites are descendants of the wood-feeding cockroach *Cryptocercus*, and researchers study them to discover lignocellulases which could aid in biofuel production, to understand the coevolution of the host and symbionts, and to study the structure and function of a complex gut ecosystem (1). Based on the presence or absence of hindgut protozoan flagellates, termites can be divided into two main groups, lower or higher termites, respectively. Lower termites contain a wide diversity of flagellate protozoa that help them in the digestion of wood whereas higher termites have been reported not to contain such symbiotic protists (2).

Reticulitermes flavipes, the Eastern subterranean lower termite, harbors a tripartite symbiosis in its hindgut consisting of protozoal, bacterial and archaeal symbionts (3). The degradation of lignocellulose by *R. flavipes* is possible due to the above-mentioned symbionts, which aid in digesting wood particles and in providing acetate as a nutrient for the host. *R. flavipes* is a eusocial insect and its colonies have a caste system made up of juveniles, workers, soldiers, and reproductives (3). The workers forage for food and return to the nest to feed other members. Therefore, the worker caste is reported to have the highest number and diversity of symbionts in their gut as compared to other colony members (1). This makes them a model system for the study of symbiosis and a potential source of lignocellulases. The present research focuses on microorganisms isolated from lower termites; consequently all future information and discussion pertain to lower termites unless mentioned otherwise.

1.2 Plant structural polysaccharides

Plant structures are made up of cellulose, hemicellulose and lignin. Cellulose is a linear polysaccharide formed by units of β-1,4-linked D-glucopyranosyl units and is the major component of plant material (20–40%), being the most abundant biomass component on earth (4). Hemicellulose is the second most common polysaccharide on earth and chains of hemicellulose are composed of xylose, or glucose/mannose often acetylated with arabinose, galactose or other acidic sugars (4). Lignin, on the other hand, has a complex and recalcitrant structure consisting of p-hydroxyphenylpropane monomer units joined by ether or C-C bonds (4).

Cellulose microfibrils are connected to hemicelluloses via hydrogen bonds crosslinking them into a very robust network. This network like structure hinders the access of the enzymes to glycosidic bonds of the cellulose fibrils. In addition, the interfibrillar space is filled with lignin. Together, cellulose, hemicellulose and lignin form a highly recalcitrant interconnected structure requiring a battery of enzymes for its digestion (2).

1.3 Termite gut structure and digestive strategy

To digest the highly recalcitrant plant polysaccharides, termites take advantage of a dual system that combines enzymatic activities of both the host and their gut symbionts (Fig 1.1). The termite gut is divided into the foregut, midgut and hindgut regions. Wood is converted to small particles by the powerful mandibles, and these particles are passed on to the foregut where they are mixed with enzymes of the termite salivary glands, initiating the hydrolysis of cellulose. Particles are further ground by muscular gizzard and transported to the midgut where they are digested by enzymes secreted from the host gut epithelium (5). The digestion products released in the midgut are resorbed by the epithelium and the partially digested particles are sent to the hindgut, which is a major site of microbial colonization. The wood particles are phagocytized by flagellates, which hydrolyze them using cellulases and hemicellulases (2, 5). The released oligosaccharides are fermented into short chain fatty acids, H_2 and CO_2. The host absorbs the microbial fermentation products like acetate, while fatty acids and lignin rich residues are excreted (2).

1.4 Symbionts and their lignocellulose-degrading activities in hindgut.

Lower termites have microorganisms in their hindguts that span through the three domains of life: archaea, protozoa and bacteria. Flagellate protists, which fill up a majority of the hindgut, belong to two phyla – the *Parabasalia* and the *Preaxostyla*. These flagellates participate in lignocellulose hydrolysis, and are also responsible for the generation of the bulk of the fermentation products which are subsequently resorbed by the termite (2). Other characteristic symbionts of the hindgut are bacteria called spirochaetes. These bacteria can mostly be found free in the hindgut fluid, but they are also found in association with flagellates. Lastly, the majority of small bacteria and archaea are found in the cytoplasm and the external surface of the flagellates as endosymbionts of flagellates or associated with the hindgut cuticle (2).

Flagellates that have a large variety of celluloses and hemicelluloses in their digestive vacuoles carry out the majority of lignocellulose digestion in lower termites. The bacterial symbionts of lower termites are believed to have no major role in fiber digestion because the flagellates immediately sequester all wood particles that enter the hindgut. The detailed mode of action of various lignocellulose-digesting enzymes is given below. Most of the cellulases and hemicellulases are modular proteins and contain catalytic modules that are usually glycoside hydrolases (GH), or in the case of hemicellulases they can also be carbohydrate esterases (CE). In addition, they also have carbohydrate-binding modules, which help in targeting enzymes to the polysaccharide. These modules are assigned into families marked by numbers based on the homology of sequences (e.g. GH1 for glycoside hydrolase family 1).

1.4.1 Cellulases

Cellulase is a general terminology used for cellulose digesting enzymes. Based on enzyme action and substrate specificities these cellulases can be classified in three categories (i) endoglucanases, (ii) exoglucanases and (iii) β-glucosidases. Endo- β-1,4-glucanases randomly cleave amorphous sites of cellulose chains. Exoglucanases (1,4-β-d-glucan cellobiohydrolases or 1,4-β-d-glucan glucohydrolases) act on nonreducing or reducing termini of cellulose fibers and release cellobiose or glucose depending on the type of exoglucanase. β-glucosidases digest cellobiose or cello-oligomers to glucose from the nonreducing ends (Fig 1.2). Exoglucanases generally produce cellobiose whereas most endoglucanases produce cellobiose and smaller amounts of glucose and cellotriose (6). In the lower termites endoglucanases from families GH5, GH7 and CBM 7 are present, and they act as core enzymes conserved among all protistans. In addition, endoglucanases from family GH45 and β-glucosidase from family GH3 might supplement the core cellulases (4).

1.4.2 Hemicellulases

Hemicelluloses are a heterogeneous group of branched and linear polysaccharides. The most common type of hemicellulose is xylan which is composed of D-xylopyranosyl units linked by β-1,4-glycosidic bonds. The xylan backbone is modified with various side chains like 4-O-methyl-D-glucuronic acid linked to the xylose units via a-1,2-glycosidic bonds, and acetic acid that esterifies the xylose units (7). In non-acetylated xylans, in addition to uronic acids, there are

L-arabinofuranose residues attached to the main chain by glycosidic linkages. In some varieties, phenolic substitutions of p-cumaric acid and ferulic acid esterify the arabinofuranoses (7). The abundance and linkage types of these substitutions vary between xylans from different sources. Fig 1.3 shows the various basic structural components of hemicellulose and the hemicellulases that act on these structures.

The variable structure and organization of hemicellulose requires the synergistic action of many enzymes for its complete degradation. These enzymes include xylanases, β-xylosidases, α-glucuronidase, α-arabinofuranosidase and acetylxylan esterase. Xylanases and β-xylosidases cleave the backbone of xylan and release xylose. Metatranscriptomic studies done in lower termites show that xylanases from families GH10 and GH11 are abundant in the hindgut. In addition, β-xylosidases from family GH1 have also been reported (4).

1.4.3 Lignin modifying enzymes

Lignin is the major constituent of the faeces of wood-feeding termites suggesting that the termites are not able to degrade lignin completely. But structural modifications of lignin that improve the accessibility of polysaccharides to glycoside hydrolases help in improving the efficiency of lignocellulose digestion. There is a possibility that endogenous enzymes secreted by salivary glands and midgut of the termite cause such modifications of lignin. In addition, the mechanical grinding by mandible and gizzard cleaves some linkages in lignin which help in increasing the efficiency of lignocellulose digestion (2).

The traditional view that the bacterial microbiota of lower termites has no major role in lignocellulose digestion and that the digestion is exclusively carried out by gut flagellates has been contradicted by recent studies carried out in lower termites. In this study several genes and potential plant fiber degrading enzymes were identified as having a bacterial origin (8). In addition, the presence of lignocellulose genes has been reported for all TAV strains sequenced thus far (9, 10). Therefore, currently there exist conflicting views regarding the role of bacteria in lignocellulose digestion in lower termites, which warrants further investigation.

1.5 Other activities in the hindgut

In addition to lignocellulose digestion, the hindgut microorganisms participate in a variety of other processes described below.

1.5.1 Fermentation

The gut flagellates and other microorganisms ferment the wood particles into short-chain fatty acids, which are later absorbed by the host. The other by-products of wood fermentation by hindgut organisms are acetate, lactate, formate, H_2 and CO_2. All the by-products are either readily absorbed or converted to other molecules except H_2, which is accumulated. The H_2 is highly accumulated in the gut center, but most of it is consumed before escaping the gut (2).

1.5.2 Reductive acetogenesis

Hydrogen is a key intermediate in the gut community, which drives the reduction of CO_2 and yields additional acetate. Reductive acetogenesis, carried out by spirochaetes using H_2 and CO_2, is the major hydrogen sink in lower termites. Evidence of the important role of spirochaetes in reductive acetogenesis is provided by the fact that marker genes of Wood-Ljungdahl pathway and several hydorgenases identified in the lower termites have been assigned to spirochaetes (2).

1.5.3 Methanogenesis

Methanogens use H_2 and CO_2 to form methane and thereby directly compete with acetogens for H_2. Although thermodynamically methanogenesis is favored, acetogenesis is predominant in the gut environment. This can be because while acetogens are found in the gut lumen, the methanogens typically colonize the hindgut wall, which places them downstream in the hydrogen gradient (2). The gut periphery seems like a very conducive environment for aerobic methane oxidation because of counter gradients of methane and oxygen, but so far no evidence of methanotrophic bacteria or their activity has been reported within the termite gut.

1.5.4 Nitrogen metabolism

Termites thrive on lignocellulose, which is very low in nitrogen and hardly contains any amino acids or vitamins. Therefore, termites depend heavily on their gut symbionts for dietary supplements of nitrogen and amino acids. Several gut microorganisms have been reported to have the ability to fix nitrogen and assimilate ammonia (product of nitrogen fixation) into various amino acids. Moreover, gut symbionts are also known to recycle the nitrogenous waste of the host in order to increase the nitrogen budget (11).

Thus, the termite hindgut symbionts participate in various activities like lignocellulose digestion, fermentation, host nutrition etc. And this symbiotic relationship between host and symbionts drives their co-speciation (2).

1.6 Effect of oxygen on the hindgut activity

The termite’s 1ul-volume hindgut harbors a rich diversity of microbial populations of up to 250 species. The hindgut environment, in addition to being highly diverse, is also very structured. Different microbial populations and their activities are organized at a microscale level in the hindgut. The respiratory activities of microorganisms maintain a steep gradient of O_2, creating a microaerophilic zone surrounding a completely anoxic core where important processes like methanogenesis and acetogenesis occur (12) (Fig 1.4). This structured environment is very important for maintaining the bacterial community structure and its associated functions. However, there is a constant influx of O_2 in the hindgut, which can jeopardize the gut symbiosis by disturbing the anoxic core. Therefore, the microbial community that colonizes the hindgut wall is essential for O_2 removal and maintenance of the anoxic core. So far, very limited attention has been paid to the understanding of microaerophilic microorganisms, their metabolism and physiology.

1.7 Termite Associated *Verrucomicrobia* (TAV) strains

Verrucomicrobia is a phylum under domain Bacteria and members of this phylum are found in quite diverse environments like soil, marine waters, hot springs etc. Recently, this phylum has been found related to *Planctomycetes* and *Chlamydiae*, and this finding is supported by 16S and 23S rRNA sequence analysis (13) leading to proposition of the PVC (*Planctomycetes-Verrucomicrobia-Chlamydiale)* superphylum. Based on the 16S rRNA phylogeny phylum *Verrucomicrobia* has been divided into seven subdivisions out of which subdivisions1 (*Verrucomicrobiae*), 2 (*Spartobacteria*), 3 and 4 (*Opitutae*) are the most common. While subdivision 2 is dominant in soil, subdivisions 1, 3 and 4 are present in lower numbers, and Subdivisions 2 and 4 are abundant in freshwater (14). The members of this phylum have been reported to possess very interesting features. For example, one of the members has a homolog of eukaryotic tubulin, *Verrucomicrobium spinosum* possesses FtsZ divergent from that present in other phyla of domain Bacteria and the only extreme acidophilic methanotroph know so far belongs to this phylum (14). These findings suggest that knowledge of this phylum will enhance our understanding about microbial evolution and ecology.

Previously, five isolates TAV1-5 belonging to phylum *Verrucomicrobia* were isolated from the termite gut under constant hypoxia (2% O_2). Their faster growth rate at 2% O_2 as compared to 20% O_2 confirmed that they are microaerophilic and might be involved in oxygen removal in the termite gut (15). Moreover, members of this obscure phylum have been consistently detected in lower and higher termites in studies based on 16S rRNA gene sequence, but due to a difficulty in culturing their roles in the termite gut are still unclear. Recently, a few members of this phylum have been reported to have a potential role in complex polysaccharide degradation, hinting towards their importance in the biogeochemical cycling of carbon (15). Various genomic and physiological studies carried out with the TAV strains so far revealed important characteristics and possible roles of these strains and also posed interesting questions, which led to the current study.

1.7.1 Preliminary Genomic Characterization

The genomic sequences of strains TAV1 and TAV2 revealed important functional attributes related to the microaerophilic physiology and other capabilities such as nitrogen fixation, amino acid production and lignocellulose degradation (9, 10).

1.7.1.1 Biological nitrogen fixation, amino acid production and urea metabolism

The genome sequences indicated the presence of the molybdenum-containing dinitrogenase reductase gene (*nifH*), and other core genes for nitrogen fixation such as *nifD* and *nifK*. Other auxiliary *nif* genes, *nifN* and *nifE* and genes for molybdenum ABC transport system were also identified. They also have genes for the production of several amino acids essential and non-essential to the termite, and genes for recycling uric acid excreted by the termite (9).

1.7.1.2 O_2 consumption

TAV genomes also have all genes necessary for *ccb*$_3$-type cytochrome oxidase. This enzyme is characterized by its strong affinity to capture oxygen at very low concentrations, and has been identified in several human pathogens and plant symbionts, which colonize O_2-restricted tissues. This finding represents the first identification of the *ccb*$_3$-type cytochrome oxidase encoding genes within the phylum *Verrucomicrobia* (9, 15).

1.7.1.3 Lignocellulose degradation

Their genomes have a substantial amount of genes encoding for enzymes like cellulases, xylanases, acetyl esterases and 1,4-β-glucanases which are involved in the degradation of cellulosic material (9, 10, 15).

1.7. 2 Transcriptomic and proteomic studies with strain TAV2

In addition to investigating the genomic content of TAV2, a customized microarray targeting approximately 50% genes and a liquid chromatography-tandem mass spectrometry (LC-MS/MS) was carried out on TAV2 cells grown under 2% and 20% oxygen concentrations (15). Forty-nine genes were up-regulated under 2% O_2, and these genes belonged to functional categories like energy production and conversion, cell cycle control, replication, cellular trafficking and secretion. A hundred and fifty five unique proteins were detected in cells grown under 2% oxygen concentration while 70 unique proteins were detected under 20% O_2

concentration. Among the unique proteins, an antioxidant super oxide dismutase was detected only when cells were grown under 20% O_2 concentration, while enzymes such as *cbb₃* cytochrome oxidase, acetyl xylan esterase and NifH were observed only under 2% O_2 concentration (15). These results helped to infer a possible role for TAV2 in biological nitrogen fixation, amino-acid production, hemicellulose degradation and consumption of O_2 in the termite hindgut.

1.7.3 Phylogenetic Characterization

The 16s rRNA sequence analysis of these five TAV strains revealed that these strains are evolutionary related. TAV3 and TAV4 are very closely related with 99.9% sequence identity. TAV2 shares 99.6% identity with TAV3 and TAV4, which makes them strains of the same species. TAV5 is moderately distant, sharing 97.5% identity with the three aforementioned strains. Lastly, TAV1 is the most distant, sharing 95.3-95.8% identity with the four preceding TAV strains. Due to the presence of such an evolutionary gradient, these TAV strains can be used to study intra- and interspecies variations.

1.8 Aims of current investigation

Aim 1: Comparative genomic study of TAV strains.

Based on 16S rRNA identity and preliminary physiological analyses, TAV2 was determined as a model organism to draw conclusions regarding ecological and physiological roles of TAV strains in the termite gut (9). However, several studies conducted on closely related strains show that even subtle variations in genetic structure could lead to drastic differences in ecological strategies adopted by such strains. These studies prove that 16S rRNA relatedness can be deceiving, and we need to study the genetic similarity and differences in the TAV strains to understand their ecological preferences and their specific roles. We don't know how much genetic variation is present in TAV genomes and if the conclusions derived for one strain e.g. TAV2 can be extended to rest of the TAV strains. To answer these questions, genomes of strains TAV3, TAV4 and TAV5 were sequenced, and Chapter 2 and Chapter 3 describe the sequencing and in-silico characterization of these genomes. Once the genomic sequences of all 5 strains were available, a comparative study was performed to identify the core genome that gave us insights into the metabolic machinery common in all TAV strains and also identify the strain variability that might have implication in their ecological roles. Additionally, we also attempt to identify the evolutionary adaptations developed by these strains to survive in termite gut. Chapter 4 gives a detailed methodology and findings of the comparative genomic study.

Aim 2: Understand the metabolic pathways and genes involved in the microaerophilic phenotype of TAV strains.

To address this aim, transcriptomic analysis of strain TAV2 was performed at 2% and 20% O_2 concentration where 2% O_2 concentration was used to mimic conditions in the termite gut periphery. The methodology and results of this RNA-seq analysis are given in Chapter 5. Transcriptome analysis has proven to be well suited to determine the genetic response of a bacterium to change in environmental conditions. By characterizing a stimulon i.e. the group of genes that change expression in response to an environmental stimulus it is possible to understand the physiological strategy applied by that organism to cope with that environmental change. It was predicted that a fine-scale change in expression of TAV2 transcriptome will be observed in response to different oxygen concentration and statistical quantification using Illumina reads will be able to capture this change.

1.9 References

1. **Benjamino J**, **Graf J**. 2016. Characterization of the core and caste-specific microbiota in the termite, Reticulitermes flavipes. Front. Microbiol. **7**.
2. **Brune A**. 2014. Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. **12**:168–80.
3. **Ohkuma M**. 2003. Termite symbiotic systems: efficient bio-recycling of lignocellulose. Appl. Microbiol. Biotechnol. **61**:1–9.
4. **Ni J**, **Tokuda G**. 2013. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol. Adv. **31**:838–50.
5. **Brune A**, **Dietrich C**. 2015. The Gut Microbiota of Termites : Digesting the Diversity in the Light of Ecology and Evolution. Annu. Rev. Microbiol. **69**:145–166.
6. **Watanabe H**, **Tokuda G**. 2010. Cellulolytic systems in insects. Annu. Rev. Entomol. **55**:609–32.
7. **Shallom D**, **Shoham Y**. 2003. Microbial hemicellulases. Curr. Opin. Microbiol.
8. **Tokuda G**, **Tsuboi Y**, **Kihara K**, **Saitou S**, **Moriya S**, **Lo N**, **Kikuchi J**. 2014. Metabolomic profiling of 13C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. Proc. Biol. Sci. **281**:20140990.
9. **Wertz JT**, **Kim E**, **Breznak JA**, **Schmidt TM**, **Rodrigues JLM**. 2012. Genomic and physiological characterization of the Verrucomicrobia isolate Diplosphaera colitermitum gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. Appl. Environ.

Microbiol. **78**:1544–1555.

10. **Isanapong J**, **Goodwin L**, **Bruce D**, **Chen A**, **Detter C**, **Han J**, **Han CS**, **Held B**, **Huntemann M**, **Ivanova N**, **Land ML**, **Mavromatis K**, **Nolan M**, **Pati A**, **Pennacchio L**, **Pitluck S**, **Szeto E**, **Tapia R**, **Woyke T**, **Rodrigues JLM**. 2012. High-quality draft genome sequence of the Opitutaceae bacterium strain TAV1, a symbiont of the wood-feeding termite Reticulitermes flavipes. J. Bacteriol. **194**:2744–5.
11. **Hongoh Y**, **Sharma VK**, **Prakash T**, **Noda S**, **Taylor TD**, **Kudo T**, **Sakaki Y**, **Toyoda A**, **Hattori M**, **Ohkuma M**. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proc. Natl. Acad. Sci. U. S. A. **105**:5555–5560.
12. **Brune a.**, **Emerson D**, **Breznak J a.** 1995. The termite gut microflora as an oxygen sink: Microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol. **61**:2681–2687.
13. **Lee KC**, **Webb RI**, **Janssen PH**, **Sangwan P**, **Romeo T**, **Staley JT**, **Fuerst JA**. 2009. Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC Microbiol **9**:5.
14. **Freitas S**, **Hatosy S**, **Fuhrman JA**, **Huse SM**, **Welch DBM**, **Sogin ML**, **Martiny AC**. 2012. Global distribution and diversity of marine Verrucomicrobia. ISME J. **6**:1499–505.
15. **Isanapong J**, **Sealy Hambright W**, **Willis AG**, **Boonmee A**, **Callister SJ**, **Burnum KE**, **Paša-Tolić L**, **Nicora CD**, **Wertz JT**, **Schmidt TM**, **Rodrigues JL**. 2013. Development of an ecophysiological model for Diplosphaera colotermitum TAV2, a termite hindgut Verrucomicrobium. ISME J. **7**:1803–13.

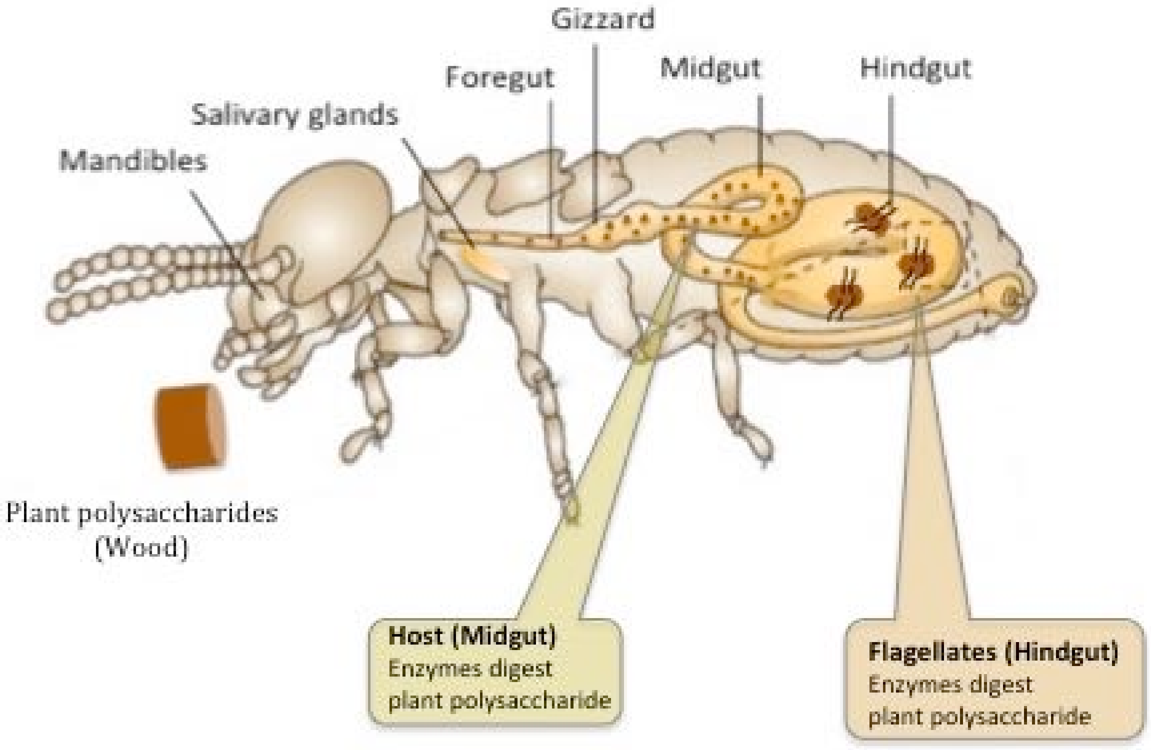


Figure 1.1 Anatomy of termite gut and an over view of plant polysaccharide digestion by lower termites. For detailed explaination refer section 1.3 in text. Adapted from Reference 2.

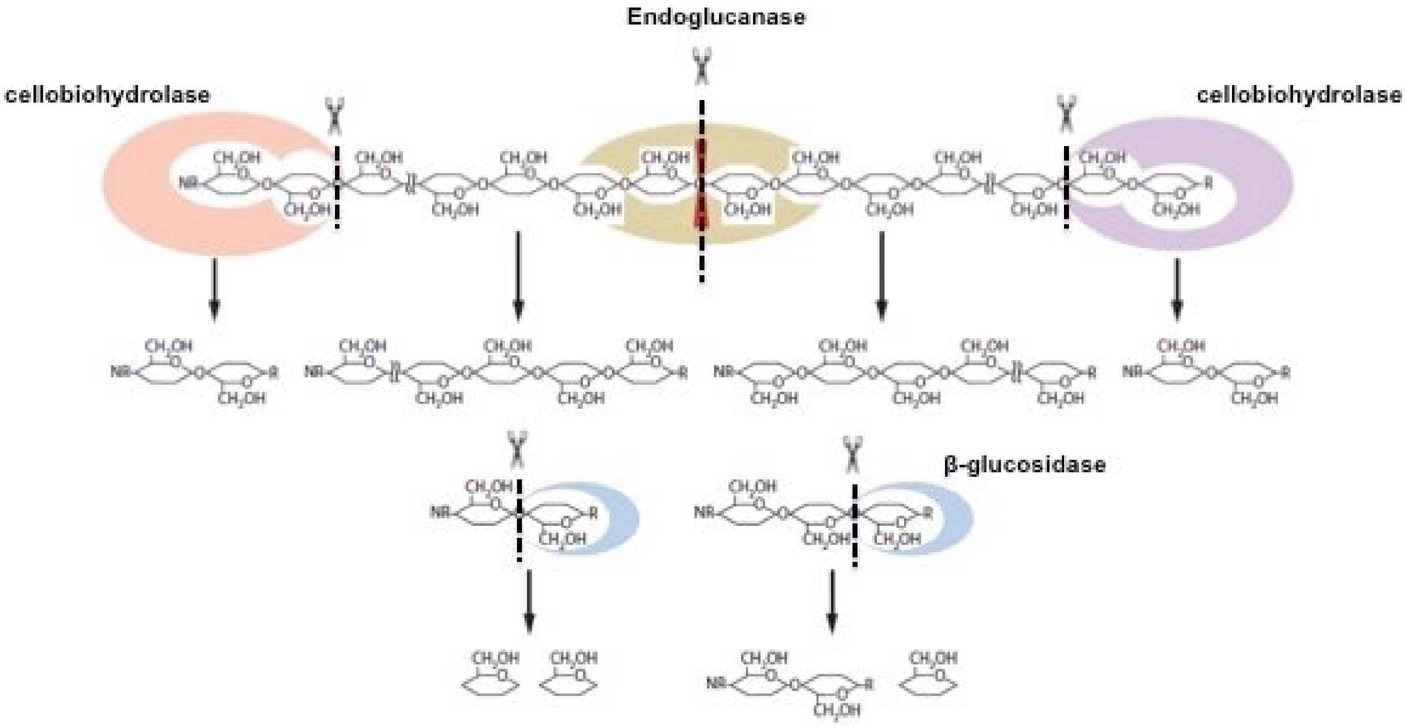


Figure 1.2 Different types of cellulases and their mode of action to digest cellulose chains.

Adapted from Reference 6

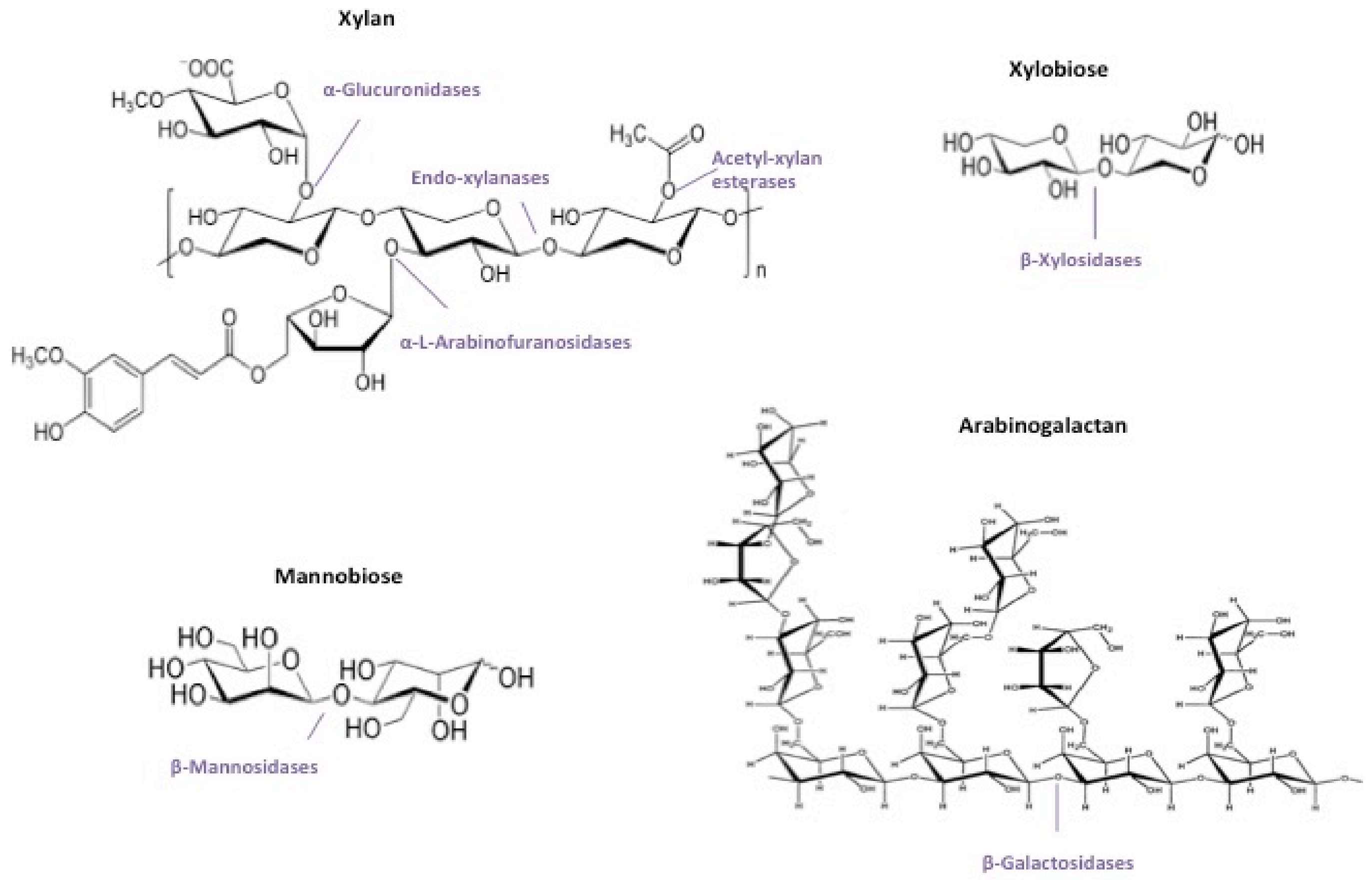


Figure 1.3 Basic structural components of hemicellulose and hemicellulases that act on these structures.

Adapted from Reference 7

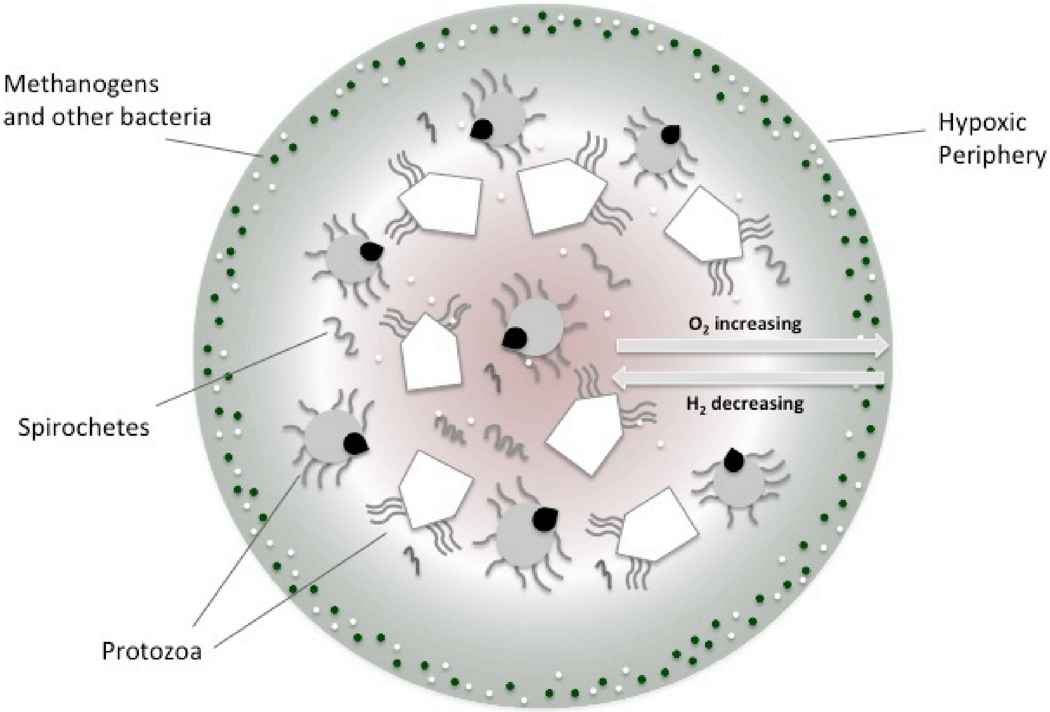


Figure 1.4 Schematic representation of the termite hindgut transect. There is a natural diffusion of O2 from the epithelium and bacteria present close to the gut wall are believed to be responsible for consuming O_2, making the center anaerobic. Adapted from J. Wertz and J. Graber

Chapter 2 Complete genome sequence of the *Opitutaceae* bacterium strain TAV5, a potential facultative methylotroph of the wood-feeding termite *Reticulitermes flavipes.*

Published: Genome Announcements 3, no. 2 (2015): e00060–15

Authors: Malini Kotak, Jantiya Isanapong, Lynne Goodwin, David Bruce, Amy Chen, Chris Detter, James Han, Shunsheng Han, Brittany Held, Marcel Huntemann, Natalia Ivanova, Miriam L. Land, Susan M. Lucas, Konstantinos Mavromatis, Matt Nolan, Amrita Pati, Len Pennacchio, Sam Pitluck, Ernest Szeto, Roxanne Tapia, Tanja Woyke and Jorge L. M. Rodrigues

Author's Contribution: JLMR, MK and JI conceived the project. JI prepared the DNA and sent for sequencing. LG, DB, AC, CD, JH, SH, BH, MH, NI, ML, SL, KM, MN, AP, LP, SP, ES, RT, TW sequenced and annotated the genome. MK analyzed the genome. JLMR, MK and JI wrote the manuscript.

2.1 Abstract

The *Opitutaceae* bacterium strain TAV5, member of the phylum *Verrucomicrobia,* was isolated from the wood-feeding termite hindgut. We report the complete genome sequence containing a chromosome and a plasmid of 7,317,842 bp and 99,831 bp, respectively. The genomic analysis reveals genes for methylotrophy, lignocellulose degradation, and ammonia and sulfate assimilation.

2.2 Introduction

The ability to grow on single carbon compounds other than CO_2 is a distinctive feature of methylotrophs. Among this group, members of the *Verrucomicrobia* have been identified as the first phylum outside of the Proteobacteria to be characterized as methanotrophs (1). Three species, isolated from geographically distant, but geothermally similar environments, had their genomes sequenced and studied (2-4). Here, we report the complete genome of the *Opitutaceae* bacterium strain TAV5, a mesophilic *Verrucomicrobium* isolated from the hindgut of the wood-feeding termite *Reticulitermes flavipes* (5), containing genes associated with methylotrophic competency.

The genomic DNA of TAV5 was isolated using a cetyltrimethylammonium bromide method and the genome was sequenced using a combination of Illumina Hiseq2000 and 454 FLEX- Titanium systems. The individual reads were assembled with the Newbler assembler (Roche). Prodigal (6) was used to identify the genes and manual curation was done using the JGI GenePRIMP pipeline (7). The translated protein-coding genes (coding sequences [CDS]) were used to search the National Center for Biotechnology Information (NCBI) nonredundant database, the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases and product description for each protein was predicted. Noncoding genes and miscellaneous features were predicted using tRNAscan-SE (8), RNAMMer (9), Rfam (10), TMHMM (6), and signalP (11).

The TAV5 genome is composed of a chromosome 7,317,842 bp long, containing 6,000 genes along with a G + C content of 63.45%. A plasmid (99,831 bp) contains 96 genes and its G + C content is 54.44%. Together, they contain 5,950 CDS, 48 tRNA genes and 1 complete rRNA operon. The number of CDS with predicted function is of 4, 086 (67.03% of the total), 1,912 have KEGG orthologs (31.36%), 2,984 have Cluster of Orthologous Group classification (48.95%). Analysis of the genome revealed presence of enzymes for formate metabolism such as formate dehydrogenase, formate transporter and enzymes of serine pathway like serine hydroxymethyltransferase, glycerate kinase and phosphoenol pyruvate carboxykinase.

The TAV5 genome contains a number of glycoside hydrolases (GH) involved in the degradation of cellulose (GH 5 and 9) and hemicellulose (GH 8, 10, 26, 28, 53), as observed for the TAV1 and TAV2 genomes (12, 13). The genome has genes for the enzymes 3-carboxymuconate cyclase (EC 5.5.1.5) and 4-carboxymuconolactone decarboxylase (EC 4.1.1.44), which are involved in the degradation of protocatechuate that is derived from lignin (14), as well as genes encoding for dioxygenases and dienelactone hydrolase (EC 3.1.1.45), known for ring cleavage of aromatic compounds. These enzymes structurally modify lignin, improving the accessibility of polysaccharides to glycoside hydrolases and increasing the efficiency of degradation (15). The TAV5 genome contains the cbb3-type cytochrome oxidase gene, implying the role of TAV5 in oxygen removal. Furthermore, genes for ammonia and sulfate assimilation, urea cycle and nitrogen fixation were observed. The availability of this genome will allow the understanding of metabolic and physiological properties carried out by members of the phylum *Verrucomicrobia* present in the termite hindgut.

2.3 Nucleotide sequence accession number

The complete genome sequence of *Opitutaceae* bacterium TAV5 was deposited in GenBank under the following accession numbers: CP007053.1 (chromosome) and CP007054.1 (plasmid).

ACKNOWLEDGMENTS

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

2.4 References

1. **Op den Camp HJ, Islam T, Stott MB, Harhangi HR, Hynes A, Schouten S, Jetten MS, Birkeland NK, Pol A, Dunfield PF.** 2009. Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. Environ Microbiol Rep **1:**293-306. http://dx.doi.org/10.1111/j.1758-2229.2009.00022.x
2. **Dunfield PF, Yuryev A, Senin P, Smirnova AV, Stott MB, Hou S, Ly B, Saw JH, Zhou Z, Ren Y, Wang J, Mountain BW, Crowe MA, Weatherby TM, Bodelier PL, Liesack W, Feng L, Wang L, Alam M.** 2007. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. Nature **450:**879-882. http://dx.doi.org/10.1038/nature06411

3. **Pol A, Heijmans K, Harhangi HR, Tedesco D, Jetten MS, Op den Camp HJ.** 2007. Methanotrophy below pH 1 by a new Verrucomicrobia species. Nature **450:**874-878. http://dx.doi.org/10.1038/nature06222

4. **Islam T, Jensen S, Reigstad LJ, Larsen O, Birkeland NK.** 2008. Methane oxidation at 55 degrees C and pH 2 by a thermoacidophilic bacterium belonging to the Verrucomicrobia phylum. Proc Natl Acad Sci U S A **105:**300-304. http://dx.doi.org/10.1073/pnas.0704162105

5. **Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, Breznak JA.** 2004. New strategies for cultivation and detection of previously uncultured microbes. Appl Environ Microbiol **70:**4748-4755. http://dx.doi.org/10.1128/AEM.70.8.4748-4755.2004

6. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119

7. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods **7:**455-457. http://dx.doi.org/10.1038/nmeth.1457

8. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955-964. http://dx.doi.org/10.1093/nar/25.5.0955

9. **Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100-3108. http://dx.doi.org/10.1093/nar/gkm160

10. **Griffiths-Jones S.** 2003. Rfam: an RNA family database. Nucleic Acids Research **31:**439-441. http://dx.doi.org/10.1093/nar/gkg006

11. **Bendtsen JD, Nielsen H, von Heijne G, Brunak S.** 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340:**783-795. http://dx.doi.org/10.1016/j.jmb.2004.05.028

12. **Isanapong J, Goodwin L, Bruce D, Chen A, Detter C, Han J, Han CS, Held B, Huntemann M, Ivanova N, Land ML, Mavromatis K, Nolan M, Pati A, Pennacchio L, Pitluck S, Szeto E, Tapia R, Woyke T, Rodrigues JL.** 2012. High-quality draft genome sequence of the Opitutaceae bacterium strain TAV1, a symbiont of the wood-feeding termite Reticulitermes flavipes. J Bacteriol **194:**2744-2745. http://dx.doi.org/10.1128/JB.00264-12

13. **Isanapong J, Sealy Hambright W, Willis AG, Boonmee A, Callister SJ, Burnum KE, Pasa-Tolic L, Nicora CD, Wertz JT, Schmidt TM, Rodrigues JL.** 2013. Development of an ecophysiological model for Diplosphaera colotermitum TAV2, a termite hindgut Verrucomicrobium. ISME J **7:**1803-1813. http://dx.doi.org/10.1038/ismej.2013.74

14. **Harwood CS, Parales RE.** 1996. The β-ketoadipate pathway and the biology of self-identity. Annual Review of Microbiology **50:**553-590. http://dx.doi.org/10.1146/annurev.micro.50.1.553

15. **Brune A.** 2014. Symbiotic digestion of lignocellulose in termite guts. Nat Rev Microbiol **12:**168-180.http://dx.doi.org/10.1038/nrmicro31

Chapter 3 Draft genome sequences of strains TAV3 and TAV4 (*Verrucomicrobial: Opitutaceae*) isolated from a wood-feeding termite and in silico analysis of their polysaccharide-degrading enzymes.

Submitted: Standards in Genomic Science

Authors: Malini Kotak, Jantiya Isanapong, Jorge L.M. Rodrigues

Author's Contribution: JLMR, MK and JI conceived the project. MK prepared samples for sequencing, submitted the genomes for annotation, submitted genome information to Genbank, and performed bioinformatics analysis. JLMR, MK and JI wrote the manuscript.

3.1 Abstract

Opitutaceae bacteria strains TAV3 and TAV4 were isolated from hindgut of wood feeding termite *Reticulitermes flavipes*. Here, we report the genome sequences, annotations and features of these Termite Associated *Verrucomicrobia* (TAV) strains. The genome sequence of strain TAV3 is 5,622,789 bp and that of TAV4 is 5,607,803 bp. TAV3 has 4,738 protein-coding genes and 54 RNA genes while TAV4 has 4,692 protein-coding genes and 54 RNA genes. Genomic characterization reveals that both strains are closely related and belong to same species. They have genes for nitrogen fixation, assimilation of inorganic nutrients, amino acids and vitamin production, and *cbb3*-type terminal cytochrome oxidase. Moreover, bioinformatic analysis carried out to identify potential lignocellulose degrading genes reveals that strains TAV3 and TAV4 have genes encoding putative cellulose and hemicellulose digesting enzymes.

Keywords: Termite Associated *Verrucomicrobia*, termite gut, lignocellulose degrading enzymes, host nutrition, CAZymes.

Abbreviations:

GH: Glycoside Hydrolases, CE: Carbohydrate Esterases, CBM: Carbohydrate-binding modules, CAZymes: Carbohydrate Active Enzymes, CAZy: Carbohydrate-Active EnZymes database.

3.2 Introduction

Termites harbor in their hindgut a wide variety of microorganisms, helping with the digestion of cellulose and contributing towards overall nutrition. A characteristic feature of termite hindgut ecosystem is the steep oxygen gradient with an anoxic core and hypoxic periphery (1, 2). Members of the termite microbial community residing in gut periphery are thought to play an important role in O_2 consumption, maintaining the core under anoxia, which is necessary for processes like methanogenesis and acetogenesis to take place (1, 2).

Previously, members of the family *Opitutaceace,* phylum *Verrucomicrobia*, were isolated from hindgut of wood-feeding termite *Reticulitermes flavipes* (3). Genomic, transcriptomic and proteomic analyses of the Termite Associated *Verrucomicrobium* (TAV) strain TAV2 revealed its microaerophilic nature and potential role in important processes such as O_2 consumption, N_2 fixation, amino acid production, and cellulose digestion in termite gut (4, 5). In addition, the strain TAV5 was reported to be a potential facultative methylotroph based on presence of genes involved in formate metabolism and serine pathway (6).

Here, we report the genome sequences, annotation, and features of strains TAV3 and TAV4. In addition, we performed a bioinformatic analysis of putative CAZymes in TAV3 and TAV4, which might be involved in digestion of cellulosic biomass. The availability of these novel genomic sequences will aid in unraveling the ecological roles of *Verrucomicrobia* in termite hindgut and provide insights into the genetic complexity and versatility of this bacterial group.

3.3 Organism information

3.3.1 Classification and features

Strains TAV3 and TAV4 were isolated from hindgut of wood feeding lower termite *R. flavipes* (Kollar) (Rhinotermitidae), which was collected near Dansville, Michigan (3). Isolates are Gram-negative, microaerophilic bacteria with optimum growth between 2 – 8% oxygen. Cells are coccoid (0.25 μm to 0.5 μm in diameter) and occur in pairs (Fig. 3.1). When grown on R2A media, they form 2 to 4 mm colonies within 7-10 days at 25-28 °C. The colonies are round, with entire margin, raised elevation and are mucoid and cream in color (Table 3-1).

To understand the relationship of strains TAV3 and TAV4 with other members of phylum *Verrucomicrobia* a whole genome phylogenetic analysis was performed. The phylogenetic tree shown in Fig. 3.2 was constructed by aligning 21 genomic sequences using REALPHY (7). *Diplosphaera colitermitum* TAV2 was used as reference and *Methylacidiphilum infernorum* V4, *Methylacidiphilum fumariolicum* SolV and *Methylacidiphilum kamchatkense* Kam1 were used as outgroups. Based on the whole genome phylogenomic analysis (Fig. 3.2) and the 16S rRNA phylogenetic analysis performed by Stevenson et al. (3), strains TAV3 and TAV4 belong to phylum *Verrucomicrobia* and are classified within the order *Opitutales* of the class *Opitutae*. The strains are most closely related to *Diplosphaera colitermitum* TAV2, which was also isolated from gut of *R. flavipes* (4) along with strains TAV1 and TAV5. In addition, a comparison of the 16S rRNA gene sequences of strains TAV3 and TAV4 indicates these strains are closely related and considered to be the same species, with 100% identity at nucleotide level. The accession numbers for 16S rRNA gene sequences of strains TAV3 and TAV4 are AY587233 and AY587234, respectively.

3.3.2 Extended features description

To gain further insights into these strains several biochemical tests were performed. Strains TAV3 and TAV4 tested negative for fermentation when grown on Triple Sugar Iron Agar and Eosin Methylene Blue Agar (8). They did not carry out nitrate reduction or urea hydrolysis when tested using nitrate and urea broths, respectively. When grown on SIM medium, which tests for sulfur reduction, indole production and motility, both TAV3 and TAV4 showed none to weak motility and neither indole production nor sulfur reduction. They showed very faint to no catalase activity when tested with hydrogen peroxide (8).

3.4 Genome sequencing information

3.4.1 Genome project history

The genomes of strains TAV3 and TAV4 were sequenced to understand the ecological roles of these strains in termite hindgut ecosystem. The sequencing and analysis was carried out in 2015. The genome projects are deposited at the Genome onLine Database (9) with Gold legacy ID Gp0007976 and Gp0007972 for TAV3 and TAV4 respectively. They are also submitted to the NCBI BioProject Database and the accession number for TAV3 genome project is PRJNA321366 and for TAV4 is PRJNA321367. Table 3-2 provides a summary of the project information and its compliance with MIGS version 2.0 (10).

3.4.2 Growth conditions and genomic DNA preparation

Strains TAV3 and TAV4 were grown in 100 ml of liquid R2B medium, commercial R2A medium lacking agar (EMD Chemicals Inc., Billerica, MA, USA), using an orbital shaker at 200 rpm at 25°C. When the cells reached an optical density (OD_{600}) of 0.3 they were harvested by centrifugation and genomic DNA was isolated using the MasterPure™ Complete DNA and RNA Purification Kit (Epicentre Biotechnologies, USA) according to the manufacturer's instructions. The purity and quantity of DNA was checked using a spectrophotometer Nanodrop2000c (ThermoFisher Scientific Corp., San Jose, CA, USA).

3.4.3 Genome sequencing and assembly

The genomes of strains TAV3 and TAV4 were sequenced at the University of Missouri DNA Core Facility using an Illumina Miseq technology with 2 x 75 paired-end reads. A total of 20,683,006 reads were obtained for TAV3 and 21,234,536 reads for TAV4 resulting in >270X sequencing coverage for both strains. Reads were analyzed for quality control using FastQC (11) and the filtered reads were *de novo* assembled using the Abyss v1.9.0 assembler (12). The final draft assembly of TAV3 contained 86 contigs and TAV4 contained 90 contigs, which were larger than 1,000 nucleotides. The total sizes of final draft assemblies were 5,622,789 bp and 5,607,803 bp for TAV3 and TAV4, respectively. A final average coverage was calculated as 270X depth. The Whole Genome Shotgun sequences and annotations for strains TAV3 and TAV4 have been deposited at DDBJ/ENA/GenBank under the accession LXWT00000000 and LXWU00000000 respectively.

3.4.4 Genome annotation

The draft genomes of strains TAV3 and TAV4 were annotated using Integrated Microbial Genomes-Expert Review (IMG ER) pipeline (13). Gene prediction was performed using Prodigal (14) and the predicted genes were used to search the NCBI nonreduntant, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases, and annotations for each gene were predicted. The tRNAscan-SE tool (15) was used to find tRNA genes, whereas ribosomal

RNAs were found by using RNAmmer (16). Transmembrane domains and signal peptides were predicted using TMHMM (17) and SignalP (18), respectively. The annotations for TAV3 and TAV4 genome described here are available through GenBank.

3.5 Genome properties

The draft genome sequence of TAV3 consists of 86 contigs containing 5,622,789 bp with a 60.8% G + C content. 4,792 genes were predicted for TAV3 genome, which included 4,738 protein-coding genes, and 54 RNA genes (47 tRNA, 3 rRNA and 4 other RNA genes). A total of 71.37% of genes were assigned a putative function and remaining genes were annotated as hypothetical proteins (Fig. 3.3, Table 3-3). A total of 46.56% of the genes could be assigned to the Clusters of Orthologous Groups (COG) of proteins (Fig. 3.3, Table 3-4).

Similarly, the TAV4 draft genome sequence is 5,607,803 bp long containing 90 contigs with a 60.8% G + C content. The genome comprises 4,746 genes out of which 4,692 are protein-coding genes, and 54 are RNA genes (47 tRNA, 3 rRNA and 4 other RNA genes). Putative functions were assigned to 73.35% genes and 46.92% genes were assigned to COG categories (Fig. 3.3, Table 3-4). A circular map of the TAV3 genome and its comparison with the TAV4 genome is shown in Fig. 3.3. Pairwise alignment of TAV3 and TAV4 genomes using blastn (*E*-value cutoff of 10^{-5}) reveals a high level of similarity between the two genomes which is evident from the continous ring of syntenous regions (Fig. 3.3, third ring from outside).

3.6 Insights from the genome sequence

The Average Nucleotide Identity (19) among strains TAV3 and TAV4 was calculated as 99.97%. A comparison of TAV3 and TAV4 genomes shows that they share 3,858 genes, with

TAV3 and TAV4 containing 444 and 428 unique genes, respectively. These unique genes mostly code for hypothetical genes and mobile elements.

The O_2 concentration in termite gut periphery, where these isolates are thought to be localized, is between 2-4% (1). Based on culture experiments, strain TAV2 is a facultative microaerophile because it grows better under 2% O_2 as compared to 20% (almost atmospheric concentration), and has all the necessary genes for expression of high-affinity *cbb3*-type terminal cytochrome oxidase (4). When the genomes of strains TAV3 and TAV4 were analyzed, it was revealed that these strains also have genes for the *cbb3*-type terminal cytochrome oxidase. They have the genes *cydA* and *B* for another high-affinity *bd*-type cytochrome oxidase. The presence of high-affinity cytochrome oxidases and an absence of other cytochrome oxidases imply that strains TAV3 and TAV4 are microaerophilic, and might be involved in O_2 consumption in termite hindgut like other TAV strains (3–5).

Strains TAV3 and TAV4 possess all the necessary genes for fixation of atmospheric N_2 into ammonia. They have *nifDHK* genes encoding nitrogenase, nif-operon regulator *nifA* gene, Mo-Fe cofactor biosynthesis and Mo^{2+} transporter genes. In addition to *nifDHK* operon, both the TAV3 and TAV4 genomes have genes for the *anfHDK* operon but lack genes for the transcriptional activator *anfA*. Both strains contain genes for glutamine synthetase, which is involved in ammonia assimilation, as well as genes for urease and ammonia and urea transporters. These results suggest that the strains are capable of fixing atmospheric N_2 as well as recycling N waste products, useful traits for survival in the termite hindgut. Genes for a nitrate transporter, nitrite reductase and nitroreductase were identified in both genomes, but no nitrate reductase genes were found.

Genes for high affinity phosphate transporters, *pstA, B, C, S*, low affinity phosphate transporter *pitA* and a regulatory gene *phoU* were also found in both genomes. Moreover, genes for sulfate assimilation were also identified in both genomes. Both the strains TAV3 and TAV4 have genes for sulfate transporter *cysA, U, P, W*, ATP sulfurylase *cysD, N*, APS kinase *cysC*, and phosphoadenosine phosphosulfate reductase cysH. The genes *cysJ, I* encoding sulfite reductase, which converts sulfite to sulfide (20), are present in TAV3, but not observed in the TAV4 genome. The cysteine synthase gene *cysK*, responsible for carrying out the final step of sulfur assimilation in cysteine (20), is present in both TAV3 and TAV4 genomes. Moreover, both TAV3 and TAV4 genomes carry genes for synthesizing at least 18 essential and non-essential amino acids and several vitamins and co-factors. Together, these results indicate that these strains have a putative nutritional role in termite hindgut.

3.7 Extended insights

Several recent studies suggest a potential role of members of phylum *Verrucomicrobia* in the degradation of polysaccharides (21, 22). The genomes of strains TAV1, TAV2 and TAV5 show the presence of a substantial amount of genes involved in lignocellulose digestion (5, 6, 23). In addition, several genes like acetyl xylan esterase and xylose isomerase that are involved in hemicellulose metabolism were identified as being expressed in TAV2 cells (5). Translated gene sequences from the TAV3 and TAV4 genomes were screened against CAZy database (24) to identify potential polysaccharide hydrolyzing enzymes, using the CAZymes Analysis Toolkit (CAT) (25). A sequence similarity based annotation was used with bit score threshold of 55 and an *E*-value cutoff of 10^{-15}. A total of 433 (8.6%) genes in TAV3 and 431 (8.7%) genes in TAV4 were identified as belonging to one of the CAZy families. The CAZy annotation identified 235 and 236 genes, for TAV3 and TAV4 respectively, as GH belonging to 52 different families (Fig. 3.4). TAV3 and TAV4 genomes also contain 83 and 85 glycosyltransferases (GT), 8 and 7 polysaccharide lyases (PL), 42 and 40 CE and 61 and 59 CBM, respectively. Digestion of extracellular cellulose requires the action of secreted enzymes (26). A total of 104 CAZymes in each of the TAV3 and TAV4 genomes contain signal peptides that suggest extracellular secretion of these proteins.

Complete hydrolysis of cellulose requires concerted activity of four different groups of enzymes called endo and exo β-1,4-glucanases, cellobiohydrolases and β-glucosidases (27, 28). The TAV3 and TAV4 genomes have β-1,4-endoglucanases belonging to families GH5 and GH17, β-glucosidases and β-1,4-exoglucanases of families GH1 and GH3 and cellobiohydrolases of family GH5 (Table 3-5). Most of the GH5 endoglucanases, 10 out of 15 in TAV3 and 11 out of 15 in TAV4 genomes have signals for secretion. CBM are believed to bring the CAZyme in close proximity of the substrate, thereby increasing enzyme efficiency (28). Both TAV3 and TAV4 have genes that code for proteins with cellulose binding CBM domains of families CBM2 and CBM6.

The genomes of strains TAV3 and TAV4 have several genes coding for potential CAZymes associated with digestion of xyloglucan, xylan and mannan backbones, and their various side chains. These are components of hemicellulose, a heterogeneous polymer (29). The genomes contain genes for β-1,4-endoglucanases, β-1,4-glucosidases and β-1,4-glucanases of families GH1, GH3 and GH17 that act on the xyloglucan backbone (Table 3-5). Both TAV3 and TAV4 genomes have genes for potential β-1,4-xylanases belonging to family GH10 and β-1,4-xylosidases of family GH39. While the GH10 xylanases have endoglycoside activity towards xylans, GH39 enzymes have exoglycoside activity. A number of putative β-mannosidases associated with families GH38, GH2, and GH130 and β-galactosidases (families GH110, GH42, GH16 and GH36) were also observed in both TAV3 and TAV4 genomes (29).

Several CE families with acetyl xylan esterase activities are known to be involved in hemicellulose digestion (27, 28). TAV3 and TAV4 have several putative acetyl xylan esterases classified as belonging to families CE1, CE4, CE6 and CE7. The CAZymes from these CE families are involved in removing the acetyl moieties that hinder the complete break down of hemicelluloses. The genomes have putative CAZymes with CBM modules of families CBM13, CBM27, CBM35, CBM4 and CBM9 involved in hemicellulose digestion.

In addition to cellulose and hemicellulose, most plant cell walls contain pectin (28). Both TAV3 and TAV4 genomes have CAZymes of families CE12, GH105, GH28, GH78, GH78|CBM6 with the potential of digesting pectin. These genomes also contain genes encoding putative sialidase (GH33), which cleaves sialic acid residues from N-linked gylcoproteins, in addition to other genes encoding enzymes like sialate O-acetylesterase and N-acetylneuraminate lyase involved in sialic acid degradation (30).

3.8 Final remarks

The present work provides information on draft genome sequences of two members of the phylum *Verrucomicrobia*, strains TAV3 and TAV4, isolated from the hindgut of *R. flavipes*. Whole genome comparisons show that 3,858 (>90%) protein-coding genes are orthologous. Analyses of these two genomes revealed that approximately 8% of open reading frames code for potential carbohydrate-active enzymes involved in the digestion of cellulose and other plant polysaccharides. The present study will serve as a framework for future studies on metabolic engineering, expression analysis, and use of cellulose digesting enzymes in the industry.

Acknowledgements

High-throughput sequencing services and genome assembly were performed at the University of Missouri DNA Core Facility and Informatics Research Core Facility, respectively. Scanning electron microscopy was done at Characterization Center for Materials and Biology at University of Texas – Arlington. The authors thank Dr. Demuth, University of Texas – Arlington, for encouraging discussions and Dr. Chrzanowski, University of Texas – Arlington, for letting us use his laboratory facility.

3.9 References

1. **Brune A**. 2014. Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. **12**:168–80.
2. **Brune A**, **Dietrich C**. 2015. The Gut Microbiota of Termites : Digesting the Diversity in the Light of Ecology and Evolution. Annu. Rev. Microbiol. **69**:145–166.
3. **Stevenson BS**, **Eichorst S a**, **Wertz JT**, **Schmidt TM**, **Breznak J a**. 2004. New strategies for cultivation and detection of previously uncultured microbes. Appl. Environ. Microbiol. **70**:4748–55.
4. **Wertz JT**, **Kim E**, **Breznak JA**, **Schmidt TM**, **Rodrigues JLM**. 2012. Genomic and physiological characterization of the Verrucomicrobia isolate Diplosphaera colitermitum gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. Appl. Environ. Microbiol. **78**:1544–1555.
5. **Isanapong J**, **Sealy Hambright W**, **Willis AG**, **Boonmee A**, **Callister SJ**, **Burnum KE**,

Paša-Tolić L, **Nicora CD**, **Wertz JT**, **Schmidt TM**, **Rodrigues JL**. 2013. Development of an ecophysiological model for Diplosphaera colotermitum TAV2, a termite hindgut Verrucomicrobium. ISME J. **7**:1803–13.

6. **Kotak M**, **Isanapong J**, **Goodwin L**, **Bruce D**, **Chen A**, **Han CS**, **Huntemann M**, **Ivanova N**, **Land ML**, **Nolan M**, **Pati A**, **Woyke T**, **Rodrigues JLM**. 2015. Complete Genome Sequence of the Opitutaceae Bacterium Strain TAV5, a Potential Facultative Methylotroph of the Wood-Feeding Termite Reticulitermes flavipes. Genome Announc. **3**:e00060-15.
7. **Bertels F**, **Silander OK**, **Pachkov M**, **Rainey PB**, **Van Nimwegen E**. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol. Biol. Evol. **31**:1077–1088.
8. **Leboffe MJ**, **Pierce BE**. 2010. Microbiology : laboratory theory and application. BOOK, Morton Pub. Co., Englewood, Colo.
9. **Reddy TBK**, **Thomas AD**, **Stamatis D**, **Bertsch J**, **Isbandi M**, **Jansson J**, **Mallajosyula J**, **Pagani I**, **Lobos EA**, **Kyrpides NC**. 2015. The Genomes OnLine Database (GOLD) v.5: A metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res. **43**:D1099–D1106.
10. **Field D**, **Garrity G**, **Gray T**, **Morrison N**, **Selengut J**, **Sterk P**. 2008. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol **26**:541–547.
11. **Cuccuru G**, **Orsini M**, **Pinna A**, **Sbardellati A**, **Soranzo N**, **Travaglione A**, **Uva P**, **Zanetti G**, **Fotia G**. 2014. Orione , a web-based framework for NGS analysis in microbiology. Bioinformatics **30**:1928–1929.

12. **Simpson JT**, **Wong K**, **Jackman SD**, **Schein JE**, **Jones SJM**. 2009. ABySS : A parallel assembler for short read sequence data. Genome Res **19**:1117–1123.

13. **Markowitz VM**, **Mavromatis K**, **Ivanova NN**, **Chen IMA**, **Chu K**, **Kyrpides NC**. 2009. IMG ER: A system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.

14. **Hyatt D**, **Chen G-L**, **LoCascio PF**, **Land ML**, **Larimer FW**, **Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119.

15. **Lowe TM**, **Eddy SR**. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. **25**:0955–0964.

16. **Lagesen K**, **Hallin P**, **Rødland EA**, **Staerfeldt H-H**, **Rognes T**, **Ussery DW**. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–8.

17. **Krogh A**, **Larsson B**, **von Heijne G**, **Sonnhammer ELL**. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. **305**:567–580.

18. **Bendtsen JD**, **Nielsen H**, **von Heijne G**, **Brunak S**. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340**:783–95.

19. **Klappenbach JA**, **Goris J**, **Vandamme P**, **Coenye T**, **Konstantinidis KT**, **Tiedje JM**. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. **57**:81–91.

20. **Valdés J**, **Veloso F**, **Jedlicki E**, **Holmes D**. 2003. Metabolic reconstruction of sulfur

assimilation in the extremophile Acidithiobacillus ferrooxidans based on genome analysis. BMC Genomics **16**:1–16.

21. **Flint HJ**, **Scott KP**, **Duncan SH**, **Louis P**, **Forano E**, **Flint HJ**, **Scott KP**, **Duncan SH**, **Louis P**, **Forano E**, **Flint HJ**, **Scott KP**, **Duncan SH**, **Louis P**, **Forano E**. 2012. Microbial degradation of complex carbohydrates in the gut Microbial degradation of complex carbohydrates in the gut. Gut Microbes **3**:289–306.

22. **Martinez-Garcia M**, **Brazel DM**, **Swan BK**, **Arnosti C**, **Chain PSG**, **Reitenga KG**, **Xie G**, **Poulton NJ**, **Gomez ML**, **Masland DED**, **Thompson B**, **Bellows WK**, **Ziervogel K**, **Lo CC**, **Ahmed S**, **Gleasner CD**, **Detter CJ**, **Stepanauskas R**. 2012. Capturing single cell genomes of active polysaccharide degraders: An unexpected contribution of verrucomicrobia. PLoS One **7**:1–11.

23. **Isanapong J**, **Goodwin L**, **Bruce D**, **Chen A**, **Detter C**, **Han J**, **Han CS**, **Held B**, **Huntemann M**, **Ivanova N**, **Land ML**, **Mavromatis K**, **Nolan M**, **Pati A**, **Pennacchio L**, **Pitluck S**, **Szeto E**, **Tapia R**, **Woyke T**, **Rodrigues JLM**. 2012. High-quality draft genome sequence of the Opitutaceae bacterium strain TAV1, a symbiont of the wood-feeding termite Reticulitermes flavipes. J. Bacteriol. **194**:2744–5.

24. **Cantarel BI**, **Coutinho PM**, **Rancurel C**, **Bernard T**, **Lombard V**, **Henrissat B**. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. **37**:233–238.

25. **Karpinets T V**. 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database **20**:1574–1584.

26. **Blifernez-Klassen O**, **Klassen V**, **Doebbe A**, **Kersting K**, **Grimm P**, **Wobbe L**, **Kruse O**. 2012. Cellulose degradation and assimilation by the unicellular phototrophic eukaryote Chlamydomonas reinhardtii. Nat. Commun. **3**:1214.

27. **Adav SS**, **Chao LT**, **Sze SK**. 2012. Quantitative Secretomic Analysis of Trichoderma reesei Strains Reveals Enzymatic Composition for Lignocellulosic Biomass Degradation. Mol. Cell. Proteomics **11**:M111.012419-M111.012419.

28. **Blackman LM**, **Cullerne DP**, **Hardham AR**. 2014. Bioinformatic characterisation of genes encoding cell wall degrading enzymes in the Phytophthora parasitica genome. BMC Genomics **15**:785.

29. **Scheller HV**, **Ulvskov P**. 2010. Hemicelluloses. Annu. Rev. plant Biol. **61**:263–289.

30. **Youssef NH**, **Farag IF**, **Rinke C**, **Hallam SJ**. 2015. In Silico Analysis of the Metabolic Potential and Niche Specialization of Candidate Phylum " Latescibacteria " (WS3). PLoS One **10**:e127499.

Table 3-1 Classification and general features of strains TAV3 and TAV4 according to the MIGS recommendations **[10]**

MIGS ID	Property	Term	Evidence code [a]
	Classification	Domain *Bacteria*	TAS [3,31]
		Phylum *Verrucomicrobia*	TAS [3,32]
		Class *Opitutae*	TAS [3,33]
		Order *Opitutales*	TAS [3,33]
		Family *Opitutaceae*	TAS [3,33]
		Genus Unclassified	TAS [3]
		Species Unclassified	TAS [3]
		Strain TAV3 and TAV4	TAS [3]
	Gram stain	Negative	TAS [3]
	Cell Shape	Coccoid	TAS [3]
	Motility	Not reported	NAS
	Sporulation	Not reported	NAS
	Temperature range	21 – 23 °C	TAS [3]
	Optimum temperature	21 – 23 °C	TAS [3]
	pH range; Optimum	5.5 – 7.5; 7.0	NAS
	Carbon source	Heterotroph	TAS [3]
MIGS-6	Habitat	Termite gut	TAS [3]
MIGS-6.3	Salinity	Not reported	NAS
MIGS-22	Oxygen requirement	Microaerophilic	TAS [3]
MIGS-15	Biotic relationship	Symbiont	TAS [3]
MIGS-14	Pathogenicity	Not reported	NAS
MIGS-4	Geographic location	Michigan, USA	TAS [3]
MIGS-5	Sample collection	2001 - 2002	TAS [3]
MIGS-4.1	Latitude	42.55598° N	NAS
MIGS-4.2	Longitude	84.3033° W	NAS
MIGS-4.4	Altitude	293 m	NAS

Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [34].

Table 3-2 Project information for genomic sequencing of strains TAV3 and TAV4**.**

MIGS ID	Property	TAV3	TAV4
MIGS 31	Finishing quality	Draft	Draft
MIGS 28	Libraries used	Illumina paired-end	Illumina paired-end
MIGS 29	Sequencing platform	Illumina Miseq	Illumina Miseq
MIGS 31.2	Fold coverage	275X	284X
MIGS 30	Assemblers	Abyss v1.9.0	Abyss v1.9.0
MIGS 32	Gene calling method	Prodigal	Prodigal
	Locus Tag	Ga01000230	Ga01000231
	Genbank ID	LXWT00000000	LXWU00000000
	Genbank Date of Release	6/1/2016	6/1/2016
	Gold ID	Gp0007976	Gp0007972
	BIOPROJECT	PRJNA321366	PRJNA321367
MIGS 13	Source Material Identifier	TAV3	TAV4
	Project relevance	Industrial	Industrial

Table 3-3 Genome statistics for strains TAV3 and TAV4**.**

	TAV3		TAV4	
Attribute	Value	% of total	Value	% of total
Genome size (bp)	5,622,789	100.00	5,607,803	100.00
DNA coding (bp)	4,910,943	87.34	4,906,267	87.49
DNA G + C (bp)	3,418,656	60.80	3,409,544	60.80
DNA scaffolds	86	100.00	90	100.00
Total genes	4,792	100.00	4,746	100.00
Protein coding genes	4,738	98.87	4,692	98.86
RNA genes	54	1.13	54	1.14
Pseudo genes	436	9.10	406	8.55
Genes in internal clusters	1,192	24.87	1,174	24.74
Genes with function prediction	3,420	71.37	3,481	73.35
Genes assigned to COGs	2,231	46.56	2,227	46.92
Genes with Pfam domains	3,365	70.22	3,356	70.71
Genes with signal peptides	595	12.42	605	12.75
Genes with transmembrane helices	1,252	26.13	1,258	26.51
CRISPR repeats	1	-	1	-

Table 3-4 Number of genes associated with general COG functional categories for the genomes sequences of strains TAV3 and TAV4**.** The total is based on the total number of protein coding genes in the genome**.**

	TAV3		TAV4		
Code	Value	%	Value	%	Description
J	155	6.39	157	6.46	Translation, ribosomal structure and biogenesis
A	0	0	0	0	RNA processing and modification
K	281	11.58	279	11.48	Transcription
L	101	4.16	104	4.28	Replication, recombination and repair
B	0	0	0	0	Chromatin structure and dynamics
D	17	0.70	18	0.74	Cell cycle control, Cell division, chromosome partitioning
V	61	2.51	62	2.55	Defense mechanisms
T	85	3.50	84	3.46	Signal transduction mechanisms
M	153	6.31	153	6.29	Cell wall/membrane/envelope biogenesis
N	111	4.58	112	4.61	Cell motility
U	104	4.29	104	4.28	Intracellular trafficking and secretion
O	101	4.16	102	4.20	Posttranslational modification, protein turnover, chaperones
C	123	5.07	122	5.02	Energy production and conversion
G	213	8.78	218	8.97	Carbohydrate transport and metabolism
E	189	7.79	190	7.82	Amino acid transport and metabolism
F	57	2.35	57	2.34	Nucleotide transport and metabolism
H	118	4.86	119	4.90	Coenzyme transport and metabolism
I	82	3.38	80	3.29	Lipid transport and metabolism
P	125	5.15	121	4.98	Inorganic ion transport and metabolism
Q	51	2.10	51	2.10	Secondary metabolites biosynthesis, transport and catabolism
R	212	8.74	211	8.67	General function prediction only
S	87	3.59	87	3.58	Function unknown
-	2125	44.34	2113	44.52	Not in COG

Table 3-5 Putative cellulolytic and hemicellulolytic genes in TAV3 and TAV4 genomes

Substrate	CAZy Family	Potential activity as	No of genes	
			TAV3	TAV4
Cellulose	GH1	β-glucosidases	6	7
	GH3	β-glucosidases	11	11
	GH5	Endoglucanase	15	15
	GH17	Endo-1,4-β-glucanase	11	11
	CBM2	Cellulose binding Domain	2	2
	CBM6	Cellulose binding Domain	3	3
Hemicellulose	GH2	β-mannosidase	14	12
	GH10	β-1,4-xylanase	4	4
	GH16	β-galactosidase	2	2
	GH36	β-galactosidase	4	4
	GH38	β-mannosidases	15	14
	GH39	β-1,4-xylosidase	19	19
	GH39|CBM9	β-1,4-xylosidase	8	8
	GH42	β-galactosidase	5	5
	GH110	β-galactosidase	3	4
	GH130	β-mannosidases	7	7
	CE1	Acetyl xylan esterase	3	3
	CE4	Acetyl xylan esterase	3	3
	CE6	Acetyl xylan esterase	3	3
	CE7	Acetyl xylan esterase	7	7
	CBM4	Xylan binding domain	9	9
	CBM9	Xylan binding domain	8	8
	CBM13	Xylan binding domain	6	5
	CBM27	Xylan binding domain	6	6
	CBM35	Xylan binding domain	5	5

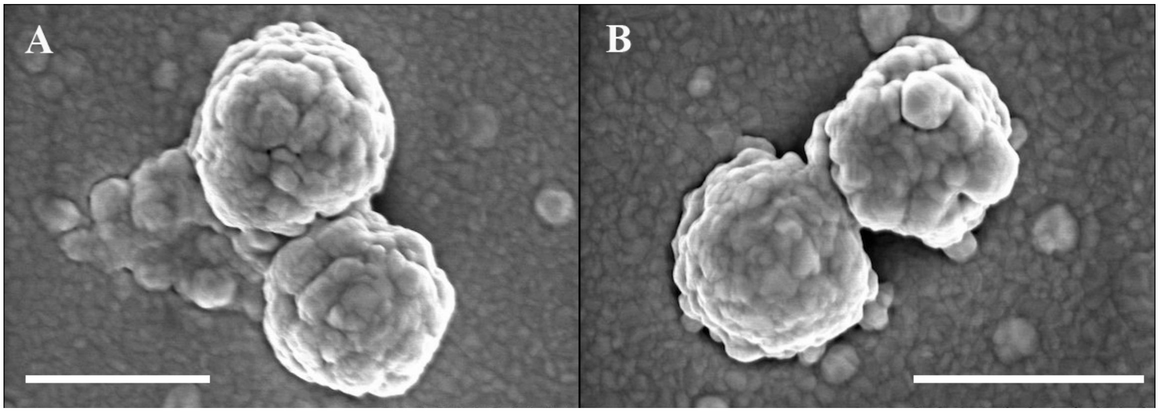


Figure 3.1 Scanning electron micrographs of strains A) TAV3 and B) TAV4 showing the diplococci morphology. Late exponential phase cells of TAV3 and TAV4 were fixed in 4% glutaraldehyde, placed on a poly-L-lysine-coated glass slide, air-dried, coated with silver and visualized by scanning electron microscopy. For both panels, scale represents 1 μm.

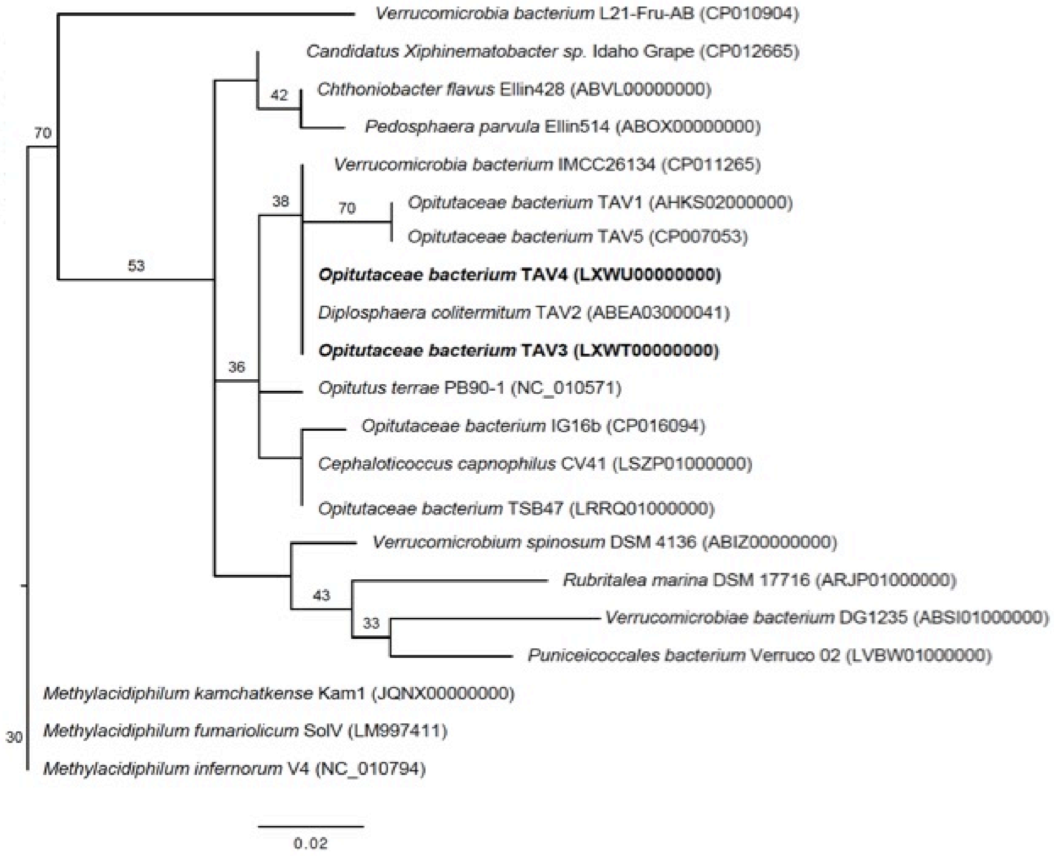


Figure 3.2 Phylogenomic tree showing position of strains TAV3 and TAV4 (shown in bold) relative to the other sequenced genomes in phylum *Verrucomicrobia*. Genomic alignments of 21 members of *Verrucomicrobia* were performed using REALPHY [7]. Bootstrap values ≥ 30% (100 iterations) are shown at branch points and GenBank accession numbers for the genomes are provided in parentheses.

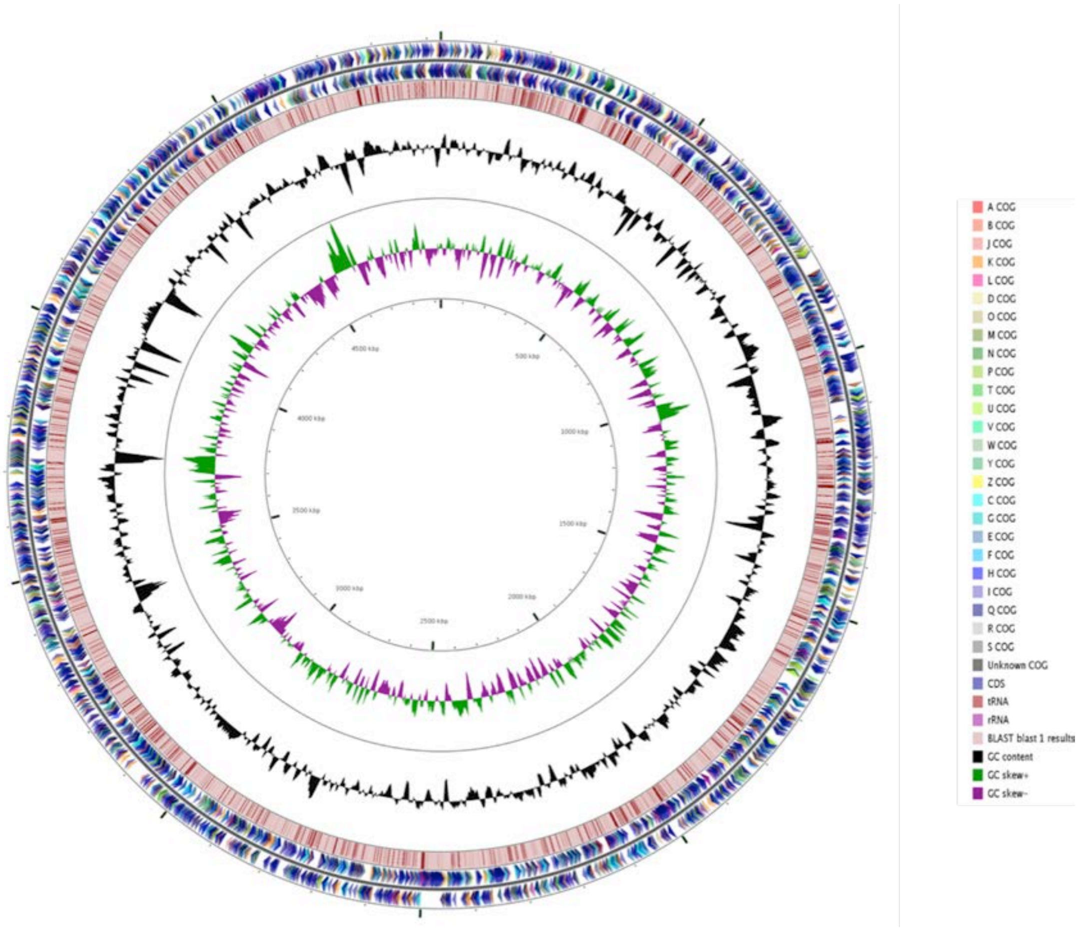


Figure 3.3 Graphical circular representation of the genome of strain TAV3 compared with TAV4 genome**.**
From outside to center: distribution of genes on forward strand (colored by COG categories), genes on reverse strand (colored by COG categories), RNA genes (tRNAs are shown in red and rRNAs are show in light purple), Blast hits (*E*-value cutoff of 10^{-5}) with strain TAV4 are shown in pink, G + C content is shown in black, GC skew (positive in green and negative in purple). The map was generated using CGView server using default parameters and blast cut off value as mentioned above [36].

3.10 Additional Files

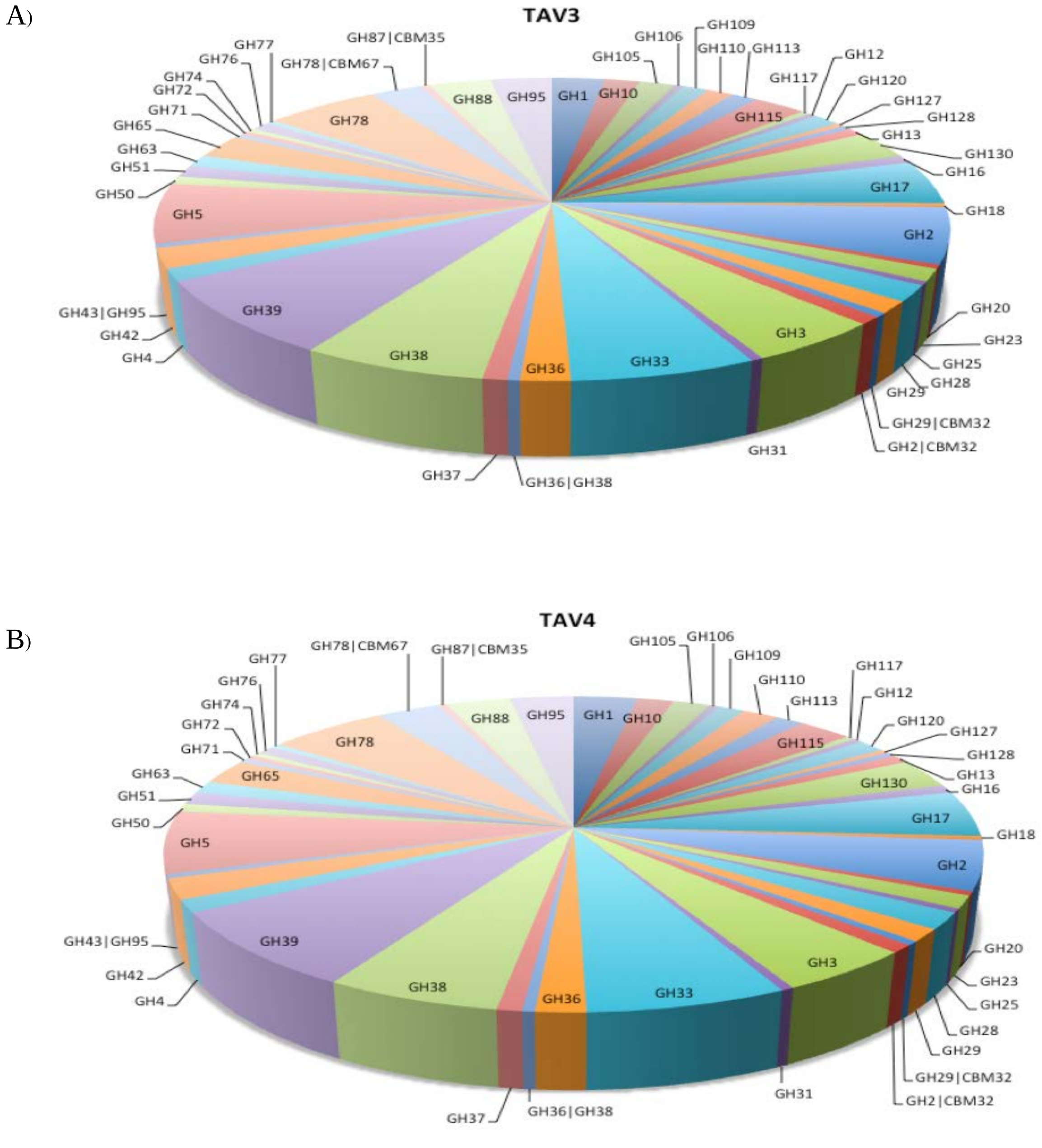


Figure 3.4 Distribution of glycoside hydrolases identified in (A) TAV3 and (B) TAV4 genomes.

Chapter 4 Comparative genomics of Termite Associated *Verrucomicrobia* (TAV) strains: insights into their microaerophilic physiology and adaptations in termite gut.

Manuscript in preparation

Authors: Malini Kotak, Jeff Demuth, Jorge L. M. Rodrigues

Author's Contribution: JLMR, MK and JD conceived the project. JD provided scirpts to prepare data for PAML and helped with data analysis. MK performed all of bioinformatics and genomic analysis. JLMR, MK and JD wrote the manuscript.

4.1 Abstract

The microorganism *Diplosphaera colitermitum* strain TAV2, a *Verrucomicrobium* isolated from the hindgut of *Reticulitermes flavipes*, has been reported as being involved in important ecophysiological processes in the termite hindgut, and revealed that it is a microaerophilic organism. These findings not only added to our knowledge of microaerophilic organisms and their importance in the termite gut, but also increased our understanding of ecological distribution and functions of the members of phylum *Verrucomicrobia*. In the present study, we compare genomes of five TAV strains including TAV2 to gain a comprehensive understanding of these strains. Our comparative genomic study shows that these strains represent two distinct species, and despite significant number of genes being conserved across the genomes; synteny of genomes between species is lost. We found that important ecological attributes like microaerophilic physiology and the ability to digest lignocellulose are conserved at the genomic level in these strains. However, high proportions of genes involved in transcription were present in species-specific genomes, suggesting that differential regulation of gene expression might be the primary feature distinguishing these otherwise highly similar genomes. Our study identified an alternative complex III in genomes of all five strains, and signatures of adaptive variation in NADH dehydrogenase and ATP synthetase enzymes, which might be linked to microaerophilic physiology. We identified several membrane proteins/enzymes involved in recombination and repair, nitrogen metabolism and lignocellulose degradation as undergoing adaptive evolution, reflecting the evolutionary mechanisms used by TAV strains to survive in the termite gut.

4.2 Introduction

The PVC superphylum comprises of bacterial phyla: *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae, Lentisphaerae* and *Omnitrophica* (1). Members of phylum *Verrucomicrobia* are found in diverse range of ecosystems that include marine, hot springs, soil and host associated environments. The members of this phylum possess some interesting features; for instance, one of the members has a homolog of eukaryotic tubulin. *Verrucomicrobium spinosum* have FtsZ divergent from other phyla of domain Bacteria, in addition, the only extreme acidophilic methanotroph know so far that belongs to this phylum (2). Although the members of *Verrucomicrobia* are ecologically important and pose interesting biological questions due to their unique characteristics, the proportion of cultured bacteria within this phylum is significantly low as compared to the phylogenetic diversity (1).

Stevenson *et. al.* isolated several symbiotic verrucomicrobia from the hindgut of wood-feeding lower termite *Reticulitermes flavipes* (3). Symbiotic microorganisms present in the gut play an important role in the termites' ability to thrive on highly recalcitrant lignocellulosic plant material and its derived residues. Termites harbor a dense community of microorganisms in their hindgut and though the individual roles of each microorganism are not yet defined, collectively the gut symbionts are known to participate in lignocellulose degradation, fermentation, host nutrition and oxygen scavenging (4). Characterization of one of the symbiotic verrucomicrobia, namely *Diplosphaera colitermitum* TAV2, revealed its autochthonous nature and important genomic and physiological features (5). Using an integrated transcriptomics and proteomics approach, Isanapong *et. al.* (6) developed an in-depth model representing the ecophysiological roles for strain TAV2 in the termite hindgut ecosystem. TAV2 was revealed to be potentially involved in nitrogen fixation, amino-acid production, hemicellulose digestion and O_2 consumption.

The lower termites have protists in addition to bacteria in their guts, and it is believed that lignocellulose digestion in lower termites is mainly done by protists. However, the presence and expression of lignocellulose digesting enzymes in strain TAV2 (6) and other bacteria isolated from lower termites (7–9) conflict this traditional view of the dominant role of protists in lignocellulose digestion. Detailed understanding of the lignocellulose degrading capacity of TAV strains will contribute towards understanding the relative roles of bacteria and protists in polysaccharide digestion in the lower termites.

A very fascinating aspect of termite gut symbiosis is its highly diverse yet structured gut environment. The microaerophilic bacteria present in the termite gut periphery use up the inward diffusing oxygen and create a steep O_2 gradient in the hindgut with microoxic periphery and an anoxic core. This structured environment is very important for maintaining the gut community structure and functions. Although the microaerophilic organisms play an important role in maintaining the gut community structure, their metabolism and ecological physiology are unclear. Identification of strain TAV2 as a microaerophile and its potential role in O_2 consumption has opened up new avenues of research on microaerophilic organisms.

So far the studies on strain TAV2 have revealed important characteristics about its genomic and metabolic capacities and its ecophysiological role in the termite gut. This information will serve as an excellent model for a detailed understanding of TAV strains and microaerophiles of termite gut in general. But to what extent the genetic and functional information obtained from this model organism can be extrapolated to the rest of the TAV strains needs to be determined. As more bacterial genomes become available, it is becoming clear that phylogenetic relatedness cannot imply functional and physiological similarities among bacteria. Microorganisms with as little as 3-10% variation in their 16S rRNA sequences have been reported to have distinct ecological distributions and functions (10).

In order to gain a comprehensive understanding of TAV strains, we compared the genomes of 5 strains with each other and with *Opitutus terrae*, which is phylogenetically related to these strains albeit a free-living organism. Our study was focused on three broad hypotheses: 1) the phylogenetic divergence is reflected in their genomes, 2) the important ecological pathways and functions are conserved in the core genome, and 3) the TAV strains are evolutionary adapted to survive in the termite gut.

4.3 Materials and Methods

4.3.1 Bacterial strains and phylogenetic analysis

The genomic sequences and related information for five TAV strains, TAV1 (taxon ID 2508501067), TAV2 (2517572100), TAV3 (2639763185), TAV4 (2639763186), TAV5 and *Opitutus terrae* (2510065049) were obtained from JGI (https://img.jgi.doe.gov) (9, 5, 8, 11).

The full-length 16S rRNA gene sequences of the 5 TAV strains were aligned using MUSCLE (12), and the phylogeny was constructed with MEGA5 (13) with the neighbor-joining method. The clustering of the sequences was tested by a bootstrap approach with 1000 repeats. Pairwise Average Nucleotide Identity (ANI) was calculated for each strain using NSimScan via the Joint Genome Institute website (14). In short, the bidirectional best hits were computed as hits of genes with 70% or more identity and at least 70% coverage for shorter genes.

4.3.2 Genome comparisons

OrthoVenn was used to carry out a comparative genomic analysis amongst 5 TAV strains (15). The inflation value was set to 1.5 and an *E* value cutoff of 10^{-15} was used to identify core, auxiliary and unique gene clusters in the TAV strains. The TAV genomes were also compared to *O. terrae* to gain more insights about these organisms. The prediction of metabolic pathways was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) (16), and BioCyc pathway schema (17).

4.3.3 Positive selection

The CODEML program from the PAML package (Version 4.7) (18) was used for detecting positive selection in the TAV strains. The sites under positive selection were identified using a site model, which allows the *dN/dS* (ω) to vary among sites. In this method a null model (M1a) which only allows sites with $\omega = 1$ *and* $\omega = 0$ is compared with a selection model (M2a) which allows sites evolving with value ω_s. Only single copy orthologous genes shared by all 5 genomes were used for the study. For each set of orthologous genes, both M1a and M2a were executed and a likelihood ratio test was used to compare M1a with M2a. If M1a was rejected in favor of M2a and if the estimate of $\omega_s > 1$ then a set of orthologous genes was considered to be under positive selection. The Benjamini-Hochberg (19) procedure was applied to control the false positive detection rate to 1%.

4.3.4 Detection of Recombination

The methods used for detecting positive selection assume that all sites share the same phylogeny, so recombination can lead to detection of false positives (20). GENECONV was used to detect evidence of recombination (21) and was run for each orthologous gene cluster that was found to be under positive selection. The software was run using "/r" (silent sites only) option and P values were calculated for 10,000 randomizations of the data. If the P values were found to be < 0.05, then the orthologous gene set was not considered as positively selected.

4.3.5 Comparative analysis of Carbohydrate Active Enzymes (CAZymes)

Translated gene sequences from all 5 TAV genomes were screened against the CAZy database (22) to identify potential polysaccharide hydrolyzing enzymes using the CAZymes Analysis Toolkit (CAT) (23). A sequence similarity based annotation was used with a bit score threshold of 55 and an E-value cutoff of 10^{-15}. Secreted proteins were identified using SignalP analyses (24).

4.4 Results

4.4.1 General Features and phylogenetic classification

TAV strains belong to phylum *Verrucomicrobia* and are classified within the order *Opitutale*s of the class *Opitutae* (3). Comparison of 16S rRNA gene sequences indicated that strains TAV3 and TAV4 are very closely related, with 100% identity at nucleotide level. They shared a 16S rRNA gene identity of 99% with strain TAV2, which indicates that they are strains of the same species (Fig 4.1). Similarly, strains TAV1 and TAV5 had a 100% 16S rRNA identity implying they are strains of the same species, but they shared 96% gene identity with strains TAV2, TAV3 and TAV4. This indicates that TAV strains belong to two distinct species. Thus these strains provided a gradient of relatedness and a unique opportunity to study interspecies and intraspecies variation.

Table 4.1 provides the general features of all 5 TAV strains being studied. The same species strain showed similarity in genome size and G + C content. The genomes of strains TAV2, TAV3 and TAV4 were approximately 5.6 Mbp while that of TAV1 and TAV5 are 7.0 Mbp. The G + C content was roughly similar with larger genomes having G + C content of 63% and small genomes having 61%. Presence of plasmids has been detected only in strain TAV5 and has 96 genes. The pairwise Average Nucleotide Identity (ANI) showed results similar to that obtained by the 16S rRNA gene analysis. It confirmed that strains TAV2, TAV3 and TAV4 belong to the same species because they share ≈97% ANI with each other while TAV1 and TAV5 form a distinct species sharing 96.93% ANI with each other and 80% with the remaining strains.

Genomic rearrangements lead to either reduction in genome size or creation of new DNA, thus they have a very huge impact on bacterial evolution (25). Comparison of genomes to establish the extent of synteny revealed that the genomes were significantly syntenic when compared within species i.e TAV1 and TAV5 or TAV2 and TAV3 or TAV4 (Fig 4.2), but the synteny broke down completely when genomes were compared between species (Fig 4.2). We estimated the number of mobile genetic elements (transposons and phage) in each TAV genome because they are known to be associated with genomic rearrangements (25). The number of mobile genetic elements (MGE) per gene was highest in TAV3 and lowest in TAV5. It was observed that TAV1 and TAV5 had less mobile genetic elements as compared to TAV2, TAV3 and TAV4. The loss of synteny might be an outcome of species diversification or a cause of divergence of TAV1 and TAV5 from the other TAV strains.

4.4.2 Core, auxiliary and unique genome

Comparative genomics of TAV strains allowed us to identify a total of 6,332 orthologous gene clusters (present in at least 2 strains) out of which 2,490 gene clusters represented the core genome of TAV strains i.e. genes present in all 5 strains (Fig 4.3). Core genome included 2,238 orthologs, which are single copy homologous genes considered to have evolved from a single ancestral gene and 252 clusters with paralogs. On average the core genome represented 50% of each TAV genome, indicating that the TAV strains have a high proportion of shared genes. Functional characterization revealed that 30% of gene clusters in the core genome could not be assigned to any COG category, while the rest of the core genome mainly had genes in functional categories associated with housekeeping processes. The core genome was significantly enriched in functional categories like amino acid transport and metabolism (6%), cell motility (3%), cell wall/cell membrane biogenesis (4%), coenzyme transport and metabolism (3.6%), energy production and conversion (4.6%), lipid transport and metabolism (2%), posttranslational modification (3.6%), replication, recombination and repair (2.7%) and translation (6%) (Fig 4.4). Using the Biocyc pathway schema, we identified 165 pathways and 52 transporters as being common to all the TAV strains.

In addition to the core genome, we also identified an auxiliary genome, which is defined as the genes present in some strains but not in all 5 strains. We focused only on species-specific auxiliary genome, i.e. genes specific to TAV5 and TAV1, but not present in either core genome or strains TAV2, TAV3 and TAV4 and vice versa. Surprisingly, only 50% of gene clusters identified in the auxiliary genome could be annotated functionally. The auxiliary genome specific to strains TAV5 and TAV1 was enriched in categories like general functions (5.6%) and inorganic metabolism (4%). It also had a high number of genes assigned to categories carbohydrate transport and metabolism (6.5%), signal transduction (2%) and transcription (7%). A detailed analysis revealed the presence of multiple copies of putative LacI family transcriptional regulators, short chain alcohol dehydrogenase and ABC transporters in the TAV5 and TAV1 specific auxiliary genome. The most abundant functional categories encoded by the TAV2, TAV3 and TAV4 specific genome were amino acid (3%) and carbohydrate transport and metabolism (3.7%), general functions (3%) and transcription (4.5%) (Fig 4.4). Though TAV2, TAV3 and TAV4 belong to the same species, 710 unique gene clusters were identified as present only in strains TAV3 and TAV4 and interestingly, only 10% of these genes could be functionally characterized.

Finally, the comparative studies also helped us to identify the unique genes present in each TAV strain. The unique genomes of all 5 strains were enriched in COG category mobilome: prophages and transposons (approx. 6-7% in each genome) (Fig 4.4), indicating the presence of unique mobile elements in each genome. Presence of high proportion of mobile elements indicates that horizontal gene transfer plays an important role in shaping the genomes of these strains. Additionally, the unique genomes of TAV1 and TAV5 contained several hypothetical proteins, multiple copies of LacI transcription factors and ABC transporters.

4.4.3 Expansion of gene families

Examination of the TAV genomes highlighted an over-representation of LacI-family transcription regulators in these genomes. Expansion of protein and gene families is known to be associated with the acquisition of novel function and regulatory structures for adaptation to a specific environment or condition (26, 27). To identify the gene families that are undergoing expansion, we compared the COG gene families present in the TAV and *O. terrae* genomes. Only those COG families, which had a minimum of 10 genes in at least one genome, were considered expanded. The analysis revealed that the most expanded family in the TAV genome was LacI-family transcription regulators (Fig 4.5). The majority of the LacI transcription factors are regulators of sugar utilization pathways (28) and the TAV genomes contain as many as 200 genes encoding these transcription factors. Another major family of transcriptional regulators found to be expanding in the TAV genomes is AraC/XylS family regulators, which regulate the degradation of sugars like arabinose, melibiose, cellobiose, raffinose, xylose, among others (29). Interestingly, these COG families were not over-represented in the *O. terrae* genome (Fig 4.5), implicating that this expansion might be linked to the ecological adaptation of the TAV strains.

Genes encoding proteins with transport and secretion function were also amplified in the TAV genomes. For example, gene families encoding type II secretion pathway protein PulG, biopolymer transport proteins and ABC-transporters were amplified. Expansion of transport related gene families was more prominent in the TAV5 and TAV1 genomes (Fig 4.5), and several ABC transporters involved in inorganic metabolism were found to be present only in the TAV1 and the TAV5 specific genomes. Lastly, genes encoding the PEP-CTERM domain containing proteins and prepilin-type N-terminal cleavage proteins were present in significant number in all the 5 TAV genomes but were absent in *O. terrae* genomes. These proteins were not assigned to any COG categories, but they are reported to be involved in exporting the proteins, mostly associated with exopolysaccharide production, to the cell exterior or the cell membrane (24, 30).

4.4.4 Comparison with *Opitutus terrae* PB90-1

The genomes of TAV strains were compared with their closest relative *O. terrae* for which the complete genome sequence is available (11). We reasoned that the genes required for survival in the termite gut should be conserved in all the 5 TAV strains, but not present in free-living *O. terrae* or they should show weak sequence similarity. A total of 1,074 gene clusters unique to the TAV strains were identified, which included genes involved in metabolism of amino acids like arginine, methionine, lysine, cysteine, tryptophan, and threonine. We also identified the TAV specific CAZymes belonging to 58 distinct CAZy families involved in degradation of cellulose, hemicellulose and host derived polysaccharides. TAV strains had molybdenum-containing and vanadium-dependent nitrogenases encoded by the *nifHDK* and *anfHDGK* operons respectively, which help with nitrogen fixation. The *O. terrae* genome did not have either *nif* or *anf* operons suggesting that the ability to fix nitrogen is an important characteristic of the TAV strains required for host-associated lifestyle.

Furthermore, genes encoding subunits A, C, D, J, K, L, M, and N of type 1 NADH dehydrogenase (NDH-1) were revealed to be non-orthologous with *O. terrae*, possibly due to the weak homology. In addition, 71% of the core gene clusters assigned to COG category transcription were found to be unique to the TAV strains. In addition, genes identified as unique to the TAV strains also included those encoding hypothetical proteins and genes not assigned to any functional category. It is possible that a few genes identified as unique to the TAV strains might have no role in host-associated lifestyle, but are absent or have weak homology with genes in *O. terrae* because of phylogenetic distance. This possibility cannot be ruled out at present due to a small number of genomes available in subdivision 4 of the *Verrucomicrobia* phylum.

However, as more genomes become available, the proposed set of unique genes will be verified and the list refined.

4.4.5 Predicted metabolism based on shared genes

Metabolic capabilities of TAV strains were inferred based on predicted shared and auxiliary genomes. Some of these metabolic capabilities are discussed below along with comparisons to *O. terrae*.

4.4.5.1 Electron transport

The TAV genomes had two different types of respiratory NADH dehydrogenases; these include the proton-pumping type I NADH dehydrogenase (NDH-1) and non-proton pumping type II NADH dehydrogenase (NDH-2). NDH-2 is a small homodimeric membrane protein that catalyzes electron transfer from NADH via FAD to quinone (31). All 5 TAV genomes had a single copy of NDH-2 and it is not present in the vicinity of NDH-1.

The TAV genomes had two sets of genes encoding NDH-1, which form gene clusters, which were conserved in all the 5 TAV strains (Fig 4.6). When compared to NDH-1 of *E. coli* (32) individual cluster was missing certain genes, but together each genome had the full set of 14 *nuo* genes required for expression of NADH:ubiquinone oxidoreductase. These clusters were either next to a t-RNA (TAV1 and TAV5 genome), or they had transposases inside the gene clusters (TAV2, TAV3 and TAV4 genome). Cluster 1 in each genome had 13 *nuo* genes (*nuoB* - *nuoN*) and was missing *nuoA*, which is involved in proton translocation. Cluster 2 was missing genes *nuoE*, *nuoF* and *nuoG*, which form NADH binding site, the FMN-binding site and a majority of Fe-S clusters necessary for the activity of NADH dehydrogenase. TAV genomes contained more than two copies of gene *nuoM*, which participate in proton translocation and ubiquinone binding (33). Comparison of the two clusters revealed that though they code for NADH:quinone oxidoreductase, the genes were are not paralogous suggesting that they did not arise from recent gene duplication. The genome of *O. terrae* also had more than one set of genes encoding for type-1 NADH dehydrogenase, thus multiple copies of NDH-1 genes is not a unique characteristic of the TAV strains. Although *O. terrae* had multiple sets of NDH-1 gene clusters, only cluster 1, was orthologous to the genes in the TAV strains while cluster 2 was unique to the TAV strains.

Unlike *E. coli* succinate dehydrogenase, which consists of 4 subunits (34), the succinate dehydrogenase of TAV strains had only 3 subunits. The structural genes for this membrane bound enzyme complex were arranged in an *sdhCAB* gene cluster, where *sdhC* encoded a cytochrome b_{558}, *sdhA* encoded a flavoprotein and *sdhB* encoded for an iron-sulfur protein. Cytochrome b_{558} is a hydrophobic protein, which is responsible for anchoring the two catalytic domains flavoprotein and iron-sulfur protein to the membrane (35).

TAV strains had genes for two different oxygen-dependent high affinity terminal cytochrome oxidases. The strains contain genes *ccoN*, *ccoO* and *ccoP*, necessary for expression of high affinity cbb_3 type terminal cytochrome oxidase. The gene *ccoN* represents the catalytic subunit of cytochrome cbb_3 oxidase, both *ccoO* and *ccoP* are believed to be responsible for the transfer of electrons from bc_1 complex to the catalytic subunit (36). None of the five TAV strains contained a homolog of gene *ccoQ*, and its function is still unclear. In addition, except TAV1, all the TAV strains also contain genes for expression of subunits CydA and B of another high affinity *bd*-type terminal oxidase. TAV1 contained the gene *cydB* but a truncated *cydA*.

However, none of the strains contained homologs of subunits CydC, CydD and CydX, which are essential for cytochrome *bd* oxidase activity (37). So it appears that the TAV strains utilize cytochrome *cbb*$_3$ as terminal cytochrome oxidase.

Previously, Wertz et. al reported the absence of genes coding for the *bc*$_1$ complex, which transfers electrons from a ubiquinone pool onto the terminal cytochrome oxidase in the TAV2 genome. Genes with significant homology to *bc*$_1$ complex were not found in any of the TAV genomes. The TAV strains utilize a high affinity *cbb*$_3$ type terminal cytochrome oxidase, which is unable to receive electrons directly from reduced quinones, so cytochrome *bc*$_1$ or another enzyme with similar function is required. An alternative complex III (ACIII) identified in *Rhodothermus marinus* and *Chloroflexus aurantiacus* has been shown to function like cytochrome *bc*$_1$ in electron transport (38). In addition to genes required for the expression of *bc*$_1$ complex, *O. terrae* also had genes for an ACIII (38). Comparison of the TAV genomes with *O. terrae* showed orthologs of the ACIII gene cluster to be present in all 5 TAV strains.

In the TAV genomes, the genes encoding ACIII formed a cluster of 7 genes (*ActABCDEFG*) (Fig 4.7). *ActA* codes for a penta-heme cytochrome *c* having five CxxCH motifs and *ActE* codes for a mono-heme cytochrome *c* with one CxxCH motif. *ActB* codes for proteins with two domains: N terminal domain is similar to molybdopterin containing proteins, and C-terminus is similar to iron-sulfur proteins. *ActC* and *ActF* encode proteins with 10 trans membrane helices each and *ActD* and *ActG* code for uncharacterized proteins (38). In *O. terrae* and TAV strains the *Act* gene cluster was followed by the *cco* gene cluster, which codes for subunits of cytochrome *cbb*$_3$ oxidase. As seen in Fig 4.7, a transposase was present in the middle of the ACIII gene cluster in strains TAV2, TAV3 and TAV4, while in strain TAV1 and *O. terrae* the gene clusters were next to a tRNA and an integrase, respectively.

All 5 strains encoded 2 copies of iron-containing superoxide dismutase, which helps in combating the oxygen stress. Surprisingly, a gene coding for enzyme catalase was identified only in the strain TAV5. The gene encoding catalase was also found to be absent in the *O. terrae* genome. These findings suggest that the TAV strains except TAV5 and *O. terrae* use only superoxide dismutase to manage reactive oxygen species.

4.4.5.2 Anaerobic Respiration and fermentation

Except TAV1, all TAV genomes had genes encoding respiratory formate-dependent nitrite reductase and nitrous reductase. The formate-dependent nitrite reductase is a membrane bound respiratory enzyme consisting of two subunits, NrfA and NrfH. NrfA is a catalytic subunit and NrfH is a membrane-bound c-type cytochrome and mediates electron transport between menaquinol and NrfA (39). Strains TAV2-TAV5 also had genes for nitrous oxide reductase that reduces nitrous oxide to dinitrogen and its accessory protein. The genes for nitrous oxide reductase and nitrite reductase were clustered together in all the 4 TAV genomes.

Moreover, the TAV genomes also had genes for assimilatory sulfate reduction. Before sulfate is reduced, enzyme ATP sulfurylase catalyzes its activation yielding adenosine-phosphosulfate (APS), which in turn is further activated by adenylylsufate kinase to form 3'-phosphoadenosine 5'-phosphosulfate (PAPS). PAPS is reduced to sulfite by PAPS reductase and eventually sulfite reductase reduces sulfite to sulfide which is assimilated (40). The above-mentioned genes for assimilation of sulfate were present in all the 5 TAV strains except TAV4, which lacks a gene for enzyme adenylylsufate kinase. These results suggest that the TAV strains have genetic potential to perform anaerobic respiration.

The genomes of all 5 strains also coded for genes involved in mixed acid fermentation. The end products of mixed acid fermentation are ethanol, acetic acid, formate, lactate and succinate (41). These strains had genes for enzymes like lactate dehydrogenase which can ferment pyruvate to lactate; fumarase and fumarate reductase which form succinate from phosphoenolpyruvate; phosphotansacetylase and acetate kinase which participate in fermenting acetyl co-A to acetate and pyruvate formate lyase which form formate from pyruvate (42).

4.4.5.3 Amino acid biosynthesis

Transcriptomic and proteomic experiments with strain TAV2 revealed expression of several genes involved in amino acid biosynthesis suggesting that it is capable of synthesizing important amino acids for the host termite. Investigation of the genomes of other TAV strains revealed that all the 5 strains had complete pathways for the synthesis of 16 essential and non-essential amino acids required for protein synthesis. The gene *hisE*, which codes for phosphoribosyl-ATP-pyrophosphohydrolase and is important for biosynthesis of histidine (43), was absent from all TAV genomes. Interestingly, the *hisE* gene was similarly found to be absent in *O. terrae*; therefore, it appears that this gene was lost in a common ancestor. It is possible that both the TAV strains and *O. terrae* do not synthesize amino acid histidine, or that these organisms utilize an alternate gene for this reaction. More experiments are required to conclude if the TAV strains can or cannot synthesize histidine. Moreover, the genomes of strains TAV1 and TAV4 are missing a key gene *aroA*, which encodes 5-enolpyruvyl-shikimate-3-phosphate (EPSP) synthase required for the biosynthesis of aromatic amino acids (44). Mutants of *aroA* gene were found to be unable to synthesize aromatic amino acids because they failed to grow on M9 minimal media (45). Therefore, it is highly unlikely that the strains TAV1 and TAV4 have the ability to synthesize aromatic amino acids like tyrosine, tryptophan and phenylalanine in the termite gut. Unlike the two strains mentioned above, all the genes required for the synthesis of aromatic amino acids were found to be present in the strains TAV2, TAV3, TAV5 and *O. terrae*.

4.4.5.4 Nitrogen metabolism

Microbial nitrogen metabolism plays a very important role in the termite gut ecosystem because termites feed on nitrogen-deficient food. Comparison of genes involved in nitrogen metabolism revealed that, except for TAV1, all TAV genomes contain two different nitrogenase systems, molybdenum (MoFe) nitrogenase and an alternative iron only (FeFe) nitrogenase. The TAV1 genome has genes only for the alternate iron (FeFe) nitrogenase. The genes encoding FeFe nitrogenase genes are organized in an *anfHDGKOR* gene cluster. All the five genomes have the gene *anfH* that encodes subunits of dinitrogenase reductase and the genes *anfD* and *anfK* that code for α- and β-subunits of dinitrogenase (46). The gene *anfG* encoding an additional δ-subunit was present in all strains except TAV1. The δ-subunit forms a $\alpha_2\beta_2\delta_2$ hexameric structure in FeFe nitrogenase, and is specific for the alternative nitrogenase (46). In addition, TAV1 did not contain the gene *anfO* and the functions of the genes *anfO* and *anfR* are unknown (46).

The proposed minimum set of genes required for Mo-dependent nitrogen fixation is *nifHDKENB*, where the *nifHDK* are catalytic genes and the *nifENB* are biosynthetic genes (47). The strains TAV2-TAV5 contained the catalytic *nifHDK* gene cluster encoding nitrogenase.

While TAV1 did not have the *nifHDK* gene cluster, it does have a *nifENB* gene cluster along with the other TAV strains. In addition, all the 5 genomes have a *nifA* gene which codes for a nif-operon regulator (48). The presence of the *nif* and *anf* genes suggests that all 5 genomes have an ability to fix atmospheric nitrogen into ammonia. This ammonia can be incorporated in glutamine and glutamate by the action of enzymes glutamine synthetase and glutamate synthase, respectively (49). The TAV genomes had genes coding for these enzymes which can assimilate ammonia into glutamine and glutamate which act as precursors of various other amino acids.

The TAV strains had the *ure* gene cluster, which encodes for subunits and accessory proteins required for expression of enzyme urease, which is capable of urea hydrolysis. TAV genomes encoded for a multimeric urease, which consists of three subunits: UreA (subunit γ), UreB (subunit β) and UreC (subunit α). The genomes also coded for the accessory proteins UreD, UreE, UreF, UreG, and UreH required for activating urease (50). Previously, upregulated expression of urease α subunit (UreC) was reported for strain TAV2 under hypoxic conditions. These results show that the TAV strains play an important role in nitrogen recycling of the termite nitrogenous waste products.

4.4.5.5 Flagellum and chemotaxis

So far the presence of a flagellum or the evidence of motility has not been observed in the TAV strains during physiological studies or under the electron microscope. However, the genomes of these strains contained all 21 genes identified as a core set of flagellar genes. The TAV strains had multiple copies of the gene *fliC* that codes for a protein, responsible for formation of flagellar filament. They had the genes *flgK* and *flgL* for hook filament junction, *flgE* for the hook, and *flgB*, *flgC*, *flgG* and *flgF* for the rod of the flagellum (51). The genes *fliF*, which code for MS ring, *fliG, fliM, fliN* coding for C ring, the motor genes *motA* and *motB* and export apparatus genes *flhA, flhB, fliI, fliP, fliR* and *fliQ* were also present (51). Finally, they had *flgD* encoding the hook-capping protein, which is required for flagellar assembly and *flgH* and *flgI* coding the L and P ring proteins situated in the outer membrane of gram-negative bacteria (51).

In addition to genes encoding the flagellum, the TAV genomes also had genes for chemotaxis, which are mediated by membrane-associated receptors that send signals to the flagellar motors via a signaling system formed by cytoplasmic proteins. The genes encoding the cytoplasmic proteins CheA, CheB, CheR, CheW and CheY and transmembrane receptors called as methyl accepting chemotaxis proteins (MCPs) are present in all the TAV genomes (52). The presence of flagellar and chemotaxis genes indicates that the TAV strains are capable of chemotactic movement of flagella in response to an external stimulus.

4.4.5.6 Regulatory functions

Approximately 7% of the core genome encoded for genes involved in regulatory functions, and a significant number of transcription related genes were also enriched in unique and auxiliary genomes of the TAV strains. The high proportion of regulatory genes in the TAV genomes implies that these strains are exposed to diverse environmental conditions in the termite gut. They had approximately 60 genes encoding two-component sensor histidine kinase-response regulatory proteins, which mediate signal transduction mechanisms. σ^{70}, σ^{28}, σ^{54}, σ^{24} and 15-20 extra cytoplasmic function (ECF) family sigma factors were present in each TAV genome. The

ECF sigma factors are part of cell-surface signaling systems (CSS) composed of three components, an alternative σ^{70} factor of ECF family, a regulator located in the cytoplasmic membrane and an outer membrane receptor (53). When the receptor perceives an extracellular signal the sigma factor is activated, which promotes the transcription of specific set of genes (53).

4.4.6 Positive selection

We investigated the core genome of the TAV strains to identify the proteins undergoing adaptation to the termite hindgut environment. Positive selection results in the fixation of advantageous mutations and leads to adaptive changes in the genes (54). A total of 2,238 clusters of single copy orthologous genes identified in the five TAV genomes were screened for genes showing evidence of positive selection. Signatures of positive selection were identified in 215 clusters, but 30 of these clusters showed evidence of recombination so they were not included in further studies. The remaining 185 (8.3%) single copy orthologous gene clusters were assigned to COG functional categories to gain further insights into the evolutionary mechanisms used by TAV strains.

Seven COG categories were significantly enriched ($P < 0.05$, binomial test) among the positively selected gene clusters when compared with the entire set of genes in the TAV2 genome (Table 4.2 and Fig. 4.8). Genes in the COG category, cell wall/membrane/envelope biogenesis were significantly enriched in positively selected genes. These genes code for membrane structures, proteins or transporters that are exposed on cell surface and might participate in interactions with host or other bacteria and phages in the gut environment. Several positively selected genes in the COG category, unknown function and the proteins that were not assigned to any COG categories were annotated as membrane proteins suggesting that surface proteins play an important role in adaptive evolution of TAV strains.

Another category showing significant enrichment was energy production and conversion. Genes encoding enzymes succinyl-CoA synthetase and isocitrate dehydrogenase, involved in the citric acid cycle were found to be under positive selection. Phosphoenolpyruvate decarboxylase, an enzyme that replenishes oxaloacetate in the citric acid cycle by converting phosphoenolpyruvate to oxaloacetate, was also identified as positively selected. In addition, the proton translocating NADH dehydrogenase (NDH-1) subunits nuoB and nuoN involved in electron transfer and proton pumping (55) showed evidence of positive selection. Two subunits of the F_0F_1 ATP synthase enzyme, that catalyzes ATP synthesis (56), were identified as undergoing positive selection. Subunit *a* of F_0 complex which is involved in proton translocation and δ subunit of F_1 complex, connecting F_0 and F_1 complex showed signs of selection. NADH dehydrogenase is the first enzyme complex in the respiratory chain and ATP synthase is the last and directly involved in ATP synthesis.

Genes like DNA polymerase II, DNA topoisomerase III and DNA helicase, assigned to functional category replication, recombination and repair were also significantly enriched. Signs of positive selection were identified in DNA processing protein DprA that participates in natural transformation (57). These results indicate that the acquisition of foreign DNA and recombination helps in the adaptive evolution of the TAV strains in addition to amino acid sequence variation. Genes assigned to COG categories that were not significantly enriched were investigated to further identify the proteins evolving under positive selection. The enzyme superoxide dismutase was found to be under positive selection suggesting that the ability to cope with reactive oxygen species is an important environmental adaptation for the microaerophilic TAV strains.

The *nifE* gene associated with the assembly of nitrogenase complex (58) and the α subunit of urease enzyme were also found to be under positive selection. Previously, we reported upregulated expression of α subunit of urease enzyme in strain TAV2 when grown with 2% oxygen (condition mimicking the termite hindgut environment) (6). Together these results strongly imply a potential role for these strains in nitrogen metabolism in the termite hindgut. To understand the evolution of lignocellulose degradation we screened the translated sequences of positively selected genes against CAZy database (22), and found 26 potential CAZymes among the 185 positively selected genes. Genes in glycosyl hydrolase (GH) families with putative cellulase, endo 1,4-xylanase, xylosidase, acetyl xylan esterase, β-glucosidase, and cellulose binding activities were found to be under positive selection. These results suggest that the termite gut environment selects for mutations that affect the lignocellulose degradation ability of the TAV strains.

4.4.7 Comparative CAZy

Carbohydrate-active enzymes (CAZymes) play an important role in biosynthesis, modification and breakdown of oligo and polysaccharides. In the termite gut these enzymes are known to be important in the breakdown of lignocellulosic biomass and also in host-microbe interactions. Previous studies on genomic characterization have revealed the presence of potential CAZymes coding genes in the genomes of TAV strains (6, 9, 5). In order to extend our knowledge and get a comprehensive understanding of the cellulolytic and hemicellulolytic capabilities of the TAV strains, an *in silico* prediction of CAZymes was performed on the core, auxiliary and unique genomes. The core genome of the TAV strains encoded approximately 241 (10.8% of the core genome) putative CAZymes (3 auxiliary activities (AA), 25 carbohydrate-binding modules (CBM), 29 carbohydrate esterases (CE), 119 glycosidehydrolases (GH), 62 glycosyltransferases (GT), 3 polysaccharide lyases (PL)), which were distributed in 90 different families. Most of these CAZymes were organized in gene clusters in the TAV genomes, with each cluster containing one or more CAZymes, associated transporters and LacI-type transcriptional regulators.

Lignocellulose, composed of cellulose, hemicellulose and lignin, requires the cooperative activity of a broad category of enzymes (59). Table 4.3 lists the core set of putative cellulolytic and hemicellulolytic genes predicted in the genomes, and its distribution in various CAZy families. Comparison of the CAZyme content of the core genome revealed that the TAV strains have genes necessary for the thorough degradation of cellulose. The core genome has seven putative endoglucanases, which hydrolyze cellulose in random positions; two cellobiohydrolases, which act on reducing ends to release disaccharide cellobiose; and nine β-glucosidases, which cleave cellobiose. Hemicelluloses are complex heteropolymers, and their breakdown requires a battery of enzymes (Fig 4.9) (60). The core genome had a higher number of putative hemicellulases than cellulases. These include endoxylanases of the family GH10 and exoxylanases represented by the family GH39, which hydrolyze xylan into short xylooligomers and xylose. In addition, putative acetyl xylan esterases of the families CE1, CE3, CE4, CE6 and CE7 that hydrolyze the acetyl substituents of xylose moieties were identified in the core genome (Fig 4.9). Putative α-l-arabinofuranosidases, galactosidases, and mannosidases, which cleave hemicellulose side chains and backbones were identified.

In addition, nine non-catalytic CBMs that are likely to bind cellulose and hemicellulose were also identified in the core genome. These CBMs increase the efficiency of catalysis by bringing the enzyme closer to the substrate. The majority of CAZymes present in the core genome have N-terminal signal peptides, indicating that they are transported to the periplasm or outer membrane during lignocellulose metabolism.

The TAV strains had enzymes like rhamnosidase, polygalacturonase, and galacturonate lyase, which are possibly involved in the hydrolysis of pectin, a polymer present in plant biomass (61). Interestingly, the core genome contains genes predicted to encode enzymes like sialidase and fucosidase, which can digest host-derived mono/oligosaccharides like mucin (62). Another interesting CAZyme identified in the core genome was the LysM domain of CBM50 family.

Analysis of the auxiliary genomes reveals that the TAV5 and TAV1 specific genome contains a total of 150 CAZymes and the TAV2, TAV3 and TAV4 specific genome codes for 83 CAZymes (Fig 4.9). The TAV2, TAV3 and TAV4 auxiliary genome mainly consisted of genes involved in hemicellulose and cellulose degradation. On the other hand, the TAV5 and TAV1 auxiliary genome had a significant number of genes involved in the breakdown of other polysaccharides like pectin, and host derived carbohydrates in addition to key CAZymes involved in lignocellulose degradation. This suggests that the strains TAV5 and TAV1 might have a highly versatile capacity in polysaccharide degradation. In contrast to the auxiliary genome, the unique genomes of TAV strains did not show any potential CAZymes.

4.5 Discussion

In an attempt to further our knowledge of the TAV strains and their evolutionary adaptation to O_2 restricted conditions, we compared their genomic and metabolic capabilities.

These strains represent two distinct species, providing an evolutionary gradient to understand inter and intra-species variations. The comparison of genomes showed that the general genomic features are similar within species but not between species, and that the synteny of genomes is completely lost between species.

We identified and functionally characterized the core genome to determine if previous findings indicating a potential role for the strain TAV2 in biological nitrogen fixation, amino-acid production, hemicellulose degradation and consumption of O_2 (6, 5), could be extended to the other TAV strains. Examination of the core genome revealed that the enzymes involved in energy production and utilization were highly conserved. A cbb_3 type terminal cytochrome oxidase was identified in all the five genomes suggesting that the microaerophilic physiology is conserved among the strains, and these strains are possibly involved in O_2 scavenging in the termite gut. Our study was able to answer the outstanding question raised by Wertz *et. al.* regarding the absence of bc_1-type cytochrome complex in electron transport chain of the strain TAV2 (5). We identified a gene cluster orthologous to the *act* genes of alternative complex (AC) III reported in *O. terrae* (38), and it is possible that instead of Complex III the TAV strains use ACIII to pass electrons to terminal cytochrome oxidase. Finally, subunits of type I NADH dehydrogenase (NDH-1) and ATP synthase showed signs of positive selection. Positively selected changes in these enzymes may be related to their roles in energy metabolism, and possibly help the TAV strains to adapt to hypoxia in the termite gut. Several studies on mammals and birds show that adaptive evolution of NADH dehydrogenase complex and ATP synthase helps survival in hypoxia related to high altitude (63).

Lignocellulose is low in nitrogen, amino acids and vitamins; therefore, termites depend heavily on their gut microorganisms for such nutrients. Our comparative metabolic study predicted that the TAV strains have the ability to recycle nitrogenous waste products, fix atmospheric nitrogen and assimilate fixed nitrogen into nutritionally valuable amino acids. Such metabolic capacity strongly implies that these strains play a potential role in host nutrition. Our genomic findings are further supported by elevated dN/dS values for important genes like urease alpha subunit and *nifE*, involved in nitrogen metabolism and *hisB*, *hisC*, *trpC* and *ilvD* involved in the biosynthesis of amino acids histidine, tryptophan, isoleucine and valine.

The comparative genomic study was also able to identify the core set of CAZymes in the TAV strains with putative roles in the digestion of cellulose, hemicellulose and other host derived or plant-based polysaccharides. Based on the core genomic content, we can predict that all the TAV strains have retained the capacity for lignocellulose digestion (Table 4.3), but the candidate gene repertoire for cellulases is not as abundant as hemicellulases. Conservation of important CAZy families involved in lignocellulose digestion reflects their common role and suggests that this characteristic is important for their survival in the termite gut. Xylan is the most abundant hemicellulose found in nature (60), and the core genome of TAV strains has an elaborate set of putative xylan-degrading enzymes containing catalytic modules (GHs and CEs) as well as carbohydrate-binding modules (CBMs) required for the complete digestion of xylan. Such abundance reflects a high capacity of these strains to digest xylans. A high proportion of the identified lignocellulose-digesting enzymes were revealed to be under positive selection (Table 4.3), implying that the termite gut environment selects TAV strains for higher CAZy content. Moreover, for degradation of plant material the CAZymes must be secreted and the presence of N-terminal signals for secretion was observed in several of these putative lignocellulose-digesting enzymes.

Additionally, the TAV strains have putative enzymes to digest other plant polysaccharides like pectin and host derived polysaccharides like mucin. CBM50 family CAZymes containing LysM domain were abundant in the core genome and were also identified to be under selection. These CBM modules are found in chitin and peptidoglycan degrading enzymes and are also found in cell wall proteins. LysM domain containing proteins are reported to be involved in host-microbe interaction processes like signaling and recognition (64) These results suggest that these organisms, which are believed to be located in close proximity of the termite gut lining, are capable of digesting the gut mucin layer and interacting with the host through signaling.

The genotypic similarities in the TAV strains were in agreement with their phylogenetic relatedness i.e. the genomes within a species were more similar than between species. Identification and functional assignment of species-specific genomes revealed that significant proportion of amplified LacI-type transcriptional regulators and AraC binding domain-containing proteins were present in species-specific genome. The number of such amplified transcription regulators was highter in the TAV1 and TAV5 specific genomes, which can be attributed to their bigger genomes. We hypothesize that an expansion of these transcription regulators is reflective of their roles in carbohydrate metabolism in the termite gut. Because the expanded gene families are represented in significant numbers in species-specific genomes, it can be inferred that the regulation of genes associate with carbohydrate metabolism might be different between species. Regulatory elements evolve faster than the rest of the genome, so a variation in gene expression is the first manifestation of species divergence and might be linked to species-specific ecological adaptation. Whether these gene families expanded due to duplication or horizontal gene transfer remains to be determined.

Our study also identified several species-specific genes; for instance, the genes encoding CE9 family N-acetylglucosamine 6-phosphate deacetylase, required for N-acetylglucosamine assimilation (65) were found only in the TAV5 and TAV1 specific genomes. Likewise, pectin methylesterases of the CE8 family involved in pectin digestion were present only in the TAV2, TAV3 and TAV4 specific genomes. The genes for glycine cleavage system were present only in the TAV1 and TAV5 specific genomes. Identification of such species-specific genes helps in understanding the specialized roles of each TAV species in the termite gut.

Although an overall pattern observed in the TAV genomes was that evolutionary related genomes shared more genes, there were several pathways and functions where such a trend was not observed. For instance, the TAV1 genome has the least number of genes in nitrogen metabolism, while the TAV5 genome has the highest number of genes in this category. Another example is that of amino acid metabolism where all the genes required for biosynthesis of aromatic amino acids were found only in the strains TAV2, TAV3 and TAV5.

4.6 References

1. **Spring S**, **Bunk B**, **Spröer C**, **Schumann P**, **Rohde M**, **Tindall BJ**, **Klenk H-P**. 2016. Characterization of the first cultured representative of *Verrucomicrobia* subdivision 5 indicates the proposal of a novel phylum. ISME J. **10**:1–16.
2. **Freitas S**, **Hatosy S**, **Fuhrman JA**, **Huse SM**, **Welch DBM**, **Sogin ML**, **Martiny AC**. 2012. Global distribution and diversity of marine *Verrucomicrobia*. ISME J. **6**:1499–505.

3. **Stevenson BS**, **Eichorst S a**, **Wertz JT**, **Schmidt TM**, **Breznak J a**. 2004. New strategies for cultivation and detection of previously uncultured microbes. Appl. Environ. Microbiol. **70**:4748–55.

4. **Brune A**. 2014. Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. **12**:168–80.

5. **Wertz JT**, **Kim E**, **Breznak JA**, **Schmidt TM**, **Rodrigues JLM**. 2012. Genomic and physiological characterization of the *Verrucomicrobi*a isolate *Diplosphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. Appl. Environ. Microbiol. **78**:1544–1555.

6. **Isanapong J**, **Sealy Hambright W**, **Willis AG**, **Boonmee A**, **Callister SJ**, **Burnum KE**, **Paša-Tolić L**, **Nicora CD**, **Wertz JT**, **Schmidt TM**, **Rodrigues JL**. 2013. Development of an ecophysiological model for *Diplosphaera colotermitum* TAV2, a termite hindgut *Verrucomicrobium*. ISME J. **7**:1803–13.

7. **Tokuda G**, **Tsuboi Y**, **Kihara K**, **Saitou S**, **Moriya S**, **Lo N**, **Kikuchi J**. 2014. Metabolomic profiling of 13C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. Proc. Biol. Sci. **281**:20140990.

8. **Kotak M**, **Isanapong J**, **Goodwin L**, **Bruce D**, **Chen A**, **Han CS**, **Huntemann M**, **Ivanova N**, **Land ML**, **Nolan M**, **Pati A**, **Woyke T**, **Rodrigues JLM**. 2015. Complete genome sequence of the *Opitutaceae* bacterium Strain TAV5, a potential facultative methylotroph of the wood-Feeding termite *Reticulitermes flavipes*. Genome Announc. **3**:e00060-15.

9. **Isanapong J**, **Goodwin L**, **Bruce D**, **Chen A**, **Detter C**, **Han J**, **Han CS**, **Held B**,

Huntemann M, **Ivanova N**, **Land ML**, **Mavromatis K**, **Nolan M**, **Pati A**, **Pennacchio L**, **Pitluck S**, **Szeto E**, **Tapia R**, **Woyke T**, **Rodrigues JLM**. 2012. High-quality draft genome sequence of the *Opitutaceae* bacterium strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes*. J. Bacteriol. **194**:2744–5.

10. **Denef VJ**, **Kalnejais LH**, **Mueller RS**, **Wilmes P**, **Baker BJ**, **Thomas BC**, **VerBerkmoes NC**, **Hettich RL**, **Banfield JF**. 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proc. Natl. Acad. Sci. U. S. A. **107**:2383–2390.
11. **van Passel MWJ**, **Kant R**, **Palva A**, **Copeland A**, **Lucas S**, **Lapidus A**, **del Rio TG**, **Pitluck S**, **Goltsman E**, **Clum A**, **Sun H**, **Schmutz J**, **Larimer FW**, **Land ML**, **Hauser L**, **Kyrpides N**, **Mikhailova N**, **Richardson PP**, **Janssen PH**, **de Vos WM**, **Smidt H**. 2011. Genome sequence of the Verrucomicrobium *Opitutus terrae* PB90-1, an abundant inhabitant of rice paddy soil ecosystems. J. Bacteriol.
12. **Edgar RC**. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797.
13. **Tamura K**, **Peterson D**, **Peterson N**, **Stecher G**, **Nei M**, **Kumar S**. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28**:2731–2739.
14. **Novichkov V**, **Kaznadzey A**, **Alexandrova N**, **Kaznadzey D**. 2016. NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy. Bioinformatics btw126.
15. **Wang Y**, **Coleman-Derr D**, **Chen G**, **Gu YQ**. 2015. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species.

Nucleic Acids Res. **43**:W78–W84.

16. **Ogata H**, **Goto S**, **Sato K**, **Fujibuchi W**, **Bono H**, **Kanehisa M**. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **27**: 29-34

17. **Caspi R**, **Billington R**, **Ferrer L**, **Foerster H**, **Fulcher CA**, **Keseler IM**, **Kothari A**, **Krummenacker M**, **Latendresse M**, **Mueller LA**, **Ong Q**, **Paley S**, **Subhraveti P**, **Weaver DS**, **Karp PD**. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **44**:D471–D480.

18. **Yang Z**. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. **24**:1586–1591.

19. **Benjamini Y**, **Hochberg Y**. 1995. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B **57**:289–300.

20. **Anisimova M**, **Nielsen R**, **Yang Z**. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics **164**:1229–1236.

21. **Sawyer S**. 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6**:526–38.

22. **Cantarel BI**, **Coutinho PM**, **Rancurel C**, **Bernard T**, **Lombard V**, **Henrissat B**. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. **37**:233–238.

23. **Karpinets T V**. 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and

analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database **20**:1574–1584.

24. **Bendtsen JD**, **Nielsen H**, **von Heijne G**, **Brunak S**. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340**:783–95.

25. **Endo A**, **Watanabe T**, **Ogata N**, **Nozawa T**, **Aikawa C**, **Arakawa S**, **Maruyama F**, **Izumi Y**, **Nakagawa I**. 2014. Comparative genome analysis and identification of competitive and cooperative interactions in a polymicrobial disease. ISME J. **9**:629–642.

26. **Treangen TJ**, **Rocha EPC**. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet. **7**:e1001284.

27. **Demuth JP**, **Hahn MW**. 2009. The life and death of gene families. BioEssays **31**:29–39.

28. **Ravcheev DA**, **Khoroshkin MS**, **Laikova ON**, **Tsoy O V.**, **Sernova N V.**, **Petrova SA**, **Rakhmaninova AB**, **Novichkov PS**, **Gelfand MS**, **Rodionov DA**. 2014. Comparative genomics and evolution of regulons of the LacI-family transcription factors. Front. Microbiol. **5**:1–1.

29. **Gallegos MT**, **Schleif R**, **Bairoch a**, **Hofmann K**, **Ramos JL**. 1997. Arac/XylS family of transcriptional regulators. Microbiol. Mol. Biol. Rev. **61**:393–410.

30. **LaPointe CF**, **Taylor RK**. 2000. The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. J. Biol. Chem. **275**:1502–10.

31. **Heikal A**, **Nakatani Y**, **Dunn E**, **Weimar MR**, **Day CL**, **Baker EN**, **Lott JS**, **Sazanov LA**, **Cook GM**. 2014. Structure of the bacterial type II NADH dehydrogenase: A monotopic membrane protein with an essential role in energy generation. Mol. Microbiol. **91**:950–964.

32. **Kao MC**, **Di Bernardo S**, **Perego M**, **Nakamaru-Ogiso E**, **Matsuno-Yagi A**, **Yagi T**. 2004. Functional roles of four conserved charged residues in the membrane domain subunit NuoA of the proton-translocating NADH-quinone oxidoreductase from *Escherichia coli*. J. Biol. Chem. **279**:32360–32366.

33. **Torres-Bacete J**, **Nakamaru-Ogiso E**, **Matsuno-Yagi A**, **Yagi T**. 2007. Characterization of the NuoM (ND4) subunit in Escherichia coli NDH-1: Conserved charged residues essential for energy-coupled activities. J. Biol. Chem. **282**:36914–36922.

34. **Cecchini G**, **Schröder I**, **Gunsalus RP**, **Maklashina E**. 2002. Succinate dehydrogenase and fumarate reductase from *Escherichia coli*. Biochim. Biophys. Acta **1553**:140–157.

35. **Schagger H**, **Cramer WA**, **Vonjagow G**. 1994. New properties of *Bacillus subtilis* succinate dehydrogenase altered at the active site The apparent active site thiol of succinate oxidoreductases is dispensable for succinate oxidation. Anal. Biochem. **217**:220–230.

36. **Pitcher RS**, **Watmough NJ**. 2004. The bacterial cytochrome cbb3 oxidases. Biochim. Biophys. Acta - Bioenerg. **1655**:388–399.

37. **Hoeser J**, **Hong S**, **Gehmann G**, **Gennis RB**, **Friedrich T**. 2014. Subunit CydX of Escherichia coli cytochrome bd ubiquinol oxidase is essential for assembly and stability of the di-heme active site. FEBS Lett. **588**:1537–1541.

38. **Refojo PN**, **Sousa FL**, **Teixeira M**, **Pereira MM**. 2010. The alternative complex III: A different architecture using known building modules. Biochim. Biophys. Acta - Bioenerg. **1797**:1869–1876.

39. **Simon J**, **Pisa R**, **Stein T**, **Eichler R**, **Klimmek O**, **Gross R**. 2001. The tetraheme

cytochrome c NrfH is required to anchor the cytochrome c nitrite reductase (NrfA) in the membrane of *Wolinella succinogenes*. Eur. J. Biochem. **268**:5776–5782.

40. **Valdés J**, **Veloso F**, **Jedlicki E**, **Holmes D**. 2003. Metabolic reconstruction of sulfur assimilation in the extremophile *Acidithiobacillus ferrooxidans* based on genome analysis. BMC Genomics **16**:1–16.

41. **Lee SJ**, **Lee D-Y**, **Kim TY**, **Kim BH**. 2005. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. Appl. Environ. Microbiol. **71**:7880–7887.

42. **San K-Y**, **Bennett GN**, **Berríos-Rivera SJ**, **Vadali R V**, **Yang Y-T**, **Horton E**, **Rudolph FB**, **Sariyar B**, **Blackwood K**. 2002. Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. Metab. Eng. **4**:182–92.

43. **Jung SI**, **Han MS**, **Kwon JH**, **Cheon CI**, **Min KH**, **Lee MS**. 1998. Cloning of the histidine biosynthetic genes of *Corynebacterium glutamicum*: organization and sequencing analysis of the hisA, impA, and hisF gene cluster. Biochem. Biophys. Res. Commun. **247**:741–745.

44. **O'Connell C**, **Pattee P a**, **Foster TJ**. 1993. Sequence and mapping of the aroA gene of *Staphylococcus aureus* 8325-4. J. Gen. Microbiol. **139**:1449–60.

45. **Jin D**, **Lu W**, **Ping S**, **Zhang W**, **Chen J**, **Dun B**, **Ma R**, **Zhao Z**, **Sha J**, **Li L**, **Yang Z**, **Chen M**, **Lin M**. 2007. Identification of a new gene encoding EPSPS with high glyphosate resistance from the metagenomic library. Curr. Microbiol. **55**:350–355.

46. **Yang J**, **Xie X**, **Wang X**, **Dixon R**, **Wang Y-P**. 2014. Reconstruction and minimal gene

requirements for the alternative iron-only nitrogenase in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 1–8.

47. **Dos Santos PC**, **Fang Z**, **Mason SW**, **Setubal JC**, **Dixon R**. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics **13**:162.

48. **Gussin GN**, **Ronson CW**, **Ausubel FM**. 1986. Regulation of Nitrogen Fixation Genes. Annu. Rev. Genet. **20**:567–591.

49. **van Heeswijk WC**, **Westerhoff H V**, **Boogerd FC**. 2013. Nitrogen Assimilation in Escherichia coli: Putting Molecular Data into a Systems Perspective. Microbiol. Mol. Biol. Rev. **77**:628–695.

50. **Konieczna I**, **Zarnowiec P**, **Kwinkowski M**, **Kolesinska B**, **Fraczyk J**, **Kaminski Z**, **Kaca W**. 2012. Bacterial urease and its role in long-lasting human diseases. Curr. Protein Pept. Sci. **13**:789–806.

51. **Liu R**, **Ochman H**. 2007. Stepwise formation of the bacterial flagellar system. Proc. Natl. Acad. Sci. U. S. A. **104**:7116–7121.

52. **Liu JD**, **Parkinson JS**. 1989. Role of CheW protein in coupling membrane receptors to the intracellular signaling system of bacterial chemotaxis. Proc. Natl. Acad. Sci. U. S. A. **86**:8703–8707.

53. **Llamas MA**, **Van Der Sar A**, **Chu BCH**, **Sparrius M**, **Vogel HJ**, **Bitter W**. 2009. A novel extracytoplasmic function (ECF) sigma factor regulates virulence in *Pseudomonas aeruginosa*. PLoS Pathog. **5**:e1000572

54. **Arbiza L**, **Dopazo J**, **Dopazo H**. 2006. Positive selection, relaxation, and acceleration in

the evolution of the human and chimp genome. PLoS Comput. Biol. **2**:288–300.

55. **Sato M**, **Torres-Bacete J**, **Sinha PK**, **Matsuno-Yagi A**, **Yagi T**. 2014. Essential regions in the membrane domain of bacterial complex I (NDH-1): The machinery for proton translocation. J. Bioenerg. Biomembr.

56. **Okuno D**, **Iino R**, **Noji H**. 2011. Rotation and structure of FoF1-ATP synthase. J. Biochem.

57. **Tadesse S**, **Graumann PL**. 2007. DprA/Smf protein localizes at the DNA uptake machinery in competent Bacillus subtilis cells. BMC Microbiol. **7**:105.

58. **Raymond J**, **Siefert JL**, **Staples CR**, **Blankenship RE**. 2004. The Natural History of Nitrogen Fixation. Mol. Biol. Evol. **21**:541–554.

59. **Ni J**, **Tokuda G**. 2013. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol. Adv. **31**:838–50.

60. **Shallom D**, **Shoham Y**. 2003. Microbial hemicellulases. Curr. Opin. Microbiol.

61. **Scheller HV**, **Ulvskov P**. 2010. Hemicelluloses. Annu. Rev. plant Biol. **61**:263–289.

62. **Bottacini F**, **O’Connell Motherway M**, **Kuczynski J**, **O’Connell KJ**, **Serafini F**, **Duranti S**, **Milani C**, **Turroni F**, **Lugli GA**, **Zomer A**, **Zhurina D**, **Riedel C**, **Ventura M**, **van Sinderen D**. 2014. Comparative genomics of the Bifidobacterium breve taxon. BMC Genomics **15**:170.

63. **Zhou T**, **Shen X**, **Irwin DM**, **Shen Y**, **Zhang Y**. 2014. Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds. Mitochondrion **18**:70–75.

64. **Cardoso AM**, **Cavalcante JJ V**, **Cantão ME**, **Thompson CE**, **Flatschart RB**, **Glogauer**

A, **Scapin SMN**, **Sade YB**, **Beltrão PJMSI**, **Gerber AL**, **Martins OB**, **Garcia ES**, **de Souza W**, **Vasconcelos ATR**. 2012. Metagenomic analysis of the microbiota from the crop of an invasive snail reveals a rich reservoir of novel genes. PLoS One **7**:e48505.

65. **Yadav V**, **Panilaitis B**, **Shi H**, **Numuta K**, **Lee K**, **Kaplan DL**. 2011. N-acetylglucosamine 6-phosphate deacetylase (nagA) is required for N-acetyl glucosamine assimilation in *Gluconacetobacter xylinus*. PLoS One **6**:e18099.

Table 4-1 General features of TAV genomes.

Feature	**TAV strains**				
	TAV1	**TAV2**	**TAV3**	**TAV4**	**TAV5**
Chromosome (bp)	7,073,483	5,693,399	5,627,312	5,612,938	7,317,842
Plasmid (bp)	ND	ND	ND	ND	99,831
Gene count	6,008	5,105	5,085	5,040	6,096
Scaffolds	2	6	109	115	1
GC (%)	63	61	61	61	63
CDS Count	5,941	5,029	5,029	4,983	6,043
RNA Count	67	76	56	57	53
rRNA Count	4	3	3	3	3
5S rRNA Count	1	1	1	1	1
16S rRNA Count	1	1	1	1	1
23S rRNA Count	2	1	1	1	1
tRNA Count	61	69	50	50	48
Pseudo Genes Count	111	133	0	0	93

Table 4-2 COG categories significantly enriched in positively selected gene clusters.

COG category	**No of genes under Positive selection**	**Binomial test *P-value***
Cell cycle control, cell division, chromosome partitioning	4	0.00068
Cell wall/membrane/envelope biogenesis	16	0.00012
Energy production and conversion	9	0.01697
Nucleotide transport and metabolism	5	0.02461
Post translational modification, protein turnover, chaperones	8	0.02407
Replication, recombination and repair	6	0.04920
Function Unknown	9	0.00246

Table 4-3 Potential cellulose and hemicellulose degrading genes in core genome of TAV strains. Table shows distribution of core genes in various CAZy families with potential role in lignocellulose degradation and also provides the number of genes identified as positively selected and having signal peptides, in each of these families.

Substrate	CAZy Family	Potential activity as	Number of genes	Positively selected	Signal Peptide
Cellulose	GH1	β-glucosidase	4	None	None
	GH3	β-glucosidase	1	None	1
	GH4	β-glucosidase	3	None	None
	GH5	Endoglucanase	4	1	2
	GH17	Endo-1,4-β-glucanase	3	1	None
	GH30	β-glucosidase	1	None	1
	GH74	Endoglucanase, cellobiohydrolase	2	None	1
	CBM2	Cellulose binding Domain	1	None	None
	CBM6	Cellulose binding Domain	1	None	None
Hemicellulose	GH2	β-mannosidase	3	1	None
	GH3	α –L-Arabinofuranosidases	1	None	1
	GH10	β-1,4-xylanase	2	1	None
	GH12	Xyloglucanases	2	1	None
	GH35	β-galactosidase	1	None	None
	GH36	β-galactosidase	3	2	None
	GH38	β-mannosidase	5	None	None
	GH39	β-1,4-xylosidase	14	1	6
	GH42	β-galactosidase	2	None	None
	GH51	α –L-Arabinofuranosidases	1	None	None

Hemicellulose	GH76	β-mannosidase	2	None	None
	GH92	β-mannosidase	5	1	None
	GH105	Rhamnogalacturonly hydrolase	5	None	None
	GH110	β-galactosidase	2	None	1
	GH113	β-mannosidase	1	None	None
	GH130	β-mannosidase	4	1	None
	CBM4	Xylan binding domain	1	None	1
	CBM9	Xylan binding domain	1	None	None
	CBM9|GH39	Xylan binding domain	3	None	None
	CBM13	Xylan binding domain	1	None	None
	CBM32	Xylan binding domain	5	1	5
	CE1	Xylan binding domain	3	None	None
	CE1|CBM22	Acetyl xylan esterase	1	None	1
	CE3	Acetyl xylan esterase	1	1	1
	CE4	Acetyl xylan esterase	3	None	2
	CE6	Acetyl xylan esterase	3	None	1
	CE7	Acetyl xylan esterase	4	1	2

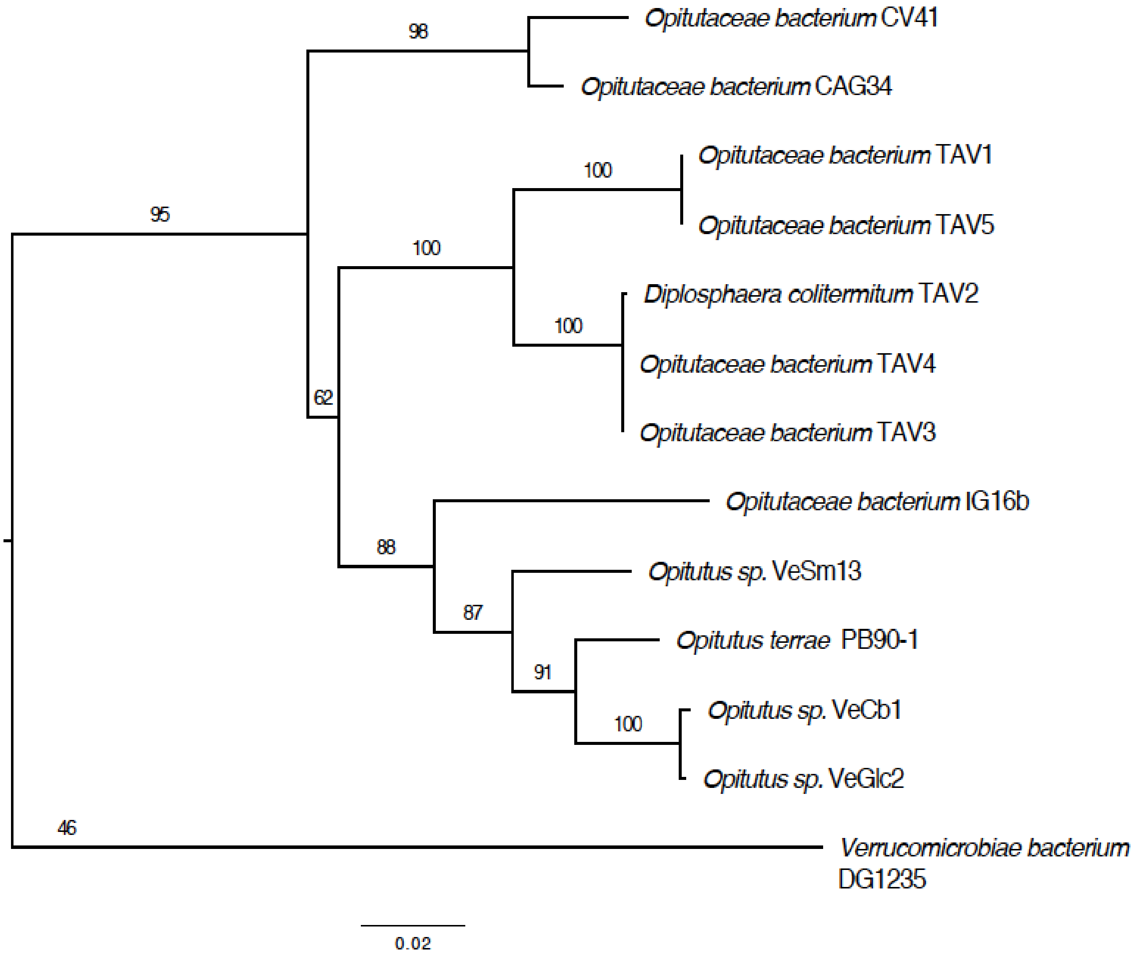


Figure 4.1 Phylogenetic tree showing position of TAV strains in phylum *Verrucomicrobia*. Full-length 16S rRNA gene sequences were aligned using MUSCLE (12) and the phylogeny was constructed with MEGA5 (13). The clustering of the sequences was tested by boostrap approach with 1000 repeats. Bootstrap values above 50 are shown. *Verrucomicrobiae bacterium* DG1235 was used as outgroup.

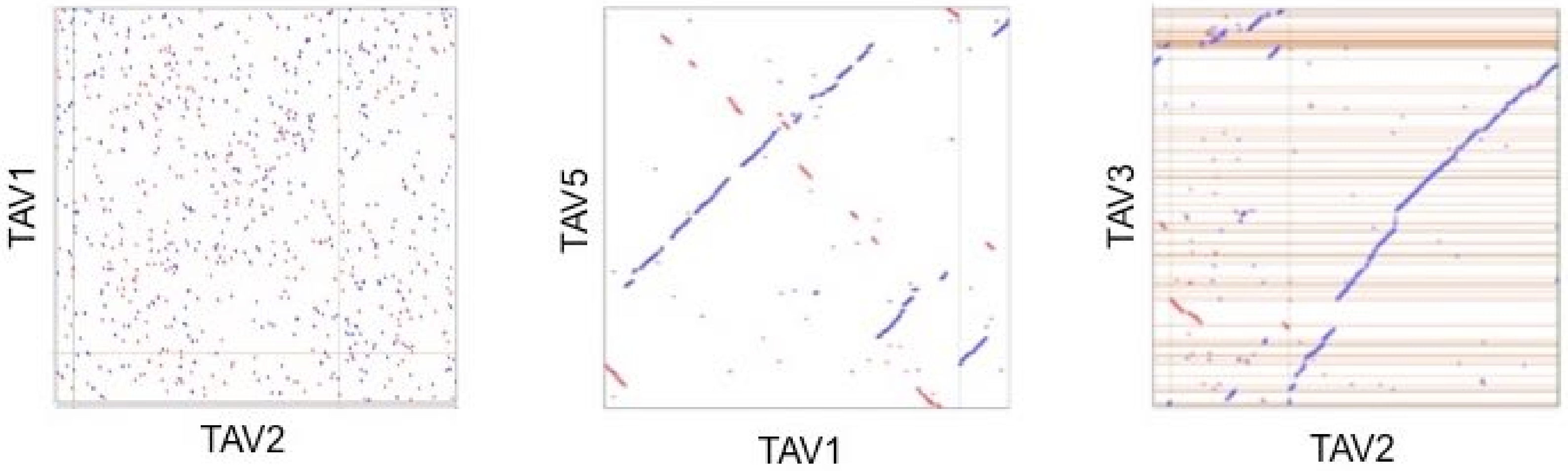


Figure 4.2 DNA-DNA dot plot analysis of TAV genomes. Within species e.g. TAV1 and TAV5 the gene order colinearity is maintained but between species e.g. TAV1 and TAV2 the colinearity is disrupted. Each dot represents a 20-bp sequence similarity exhibited between two genomes.

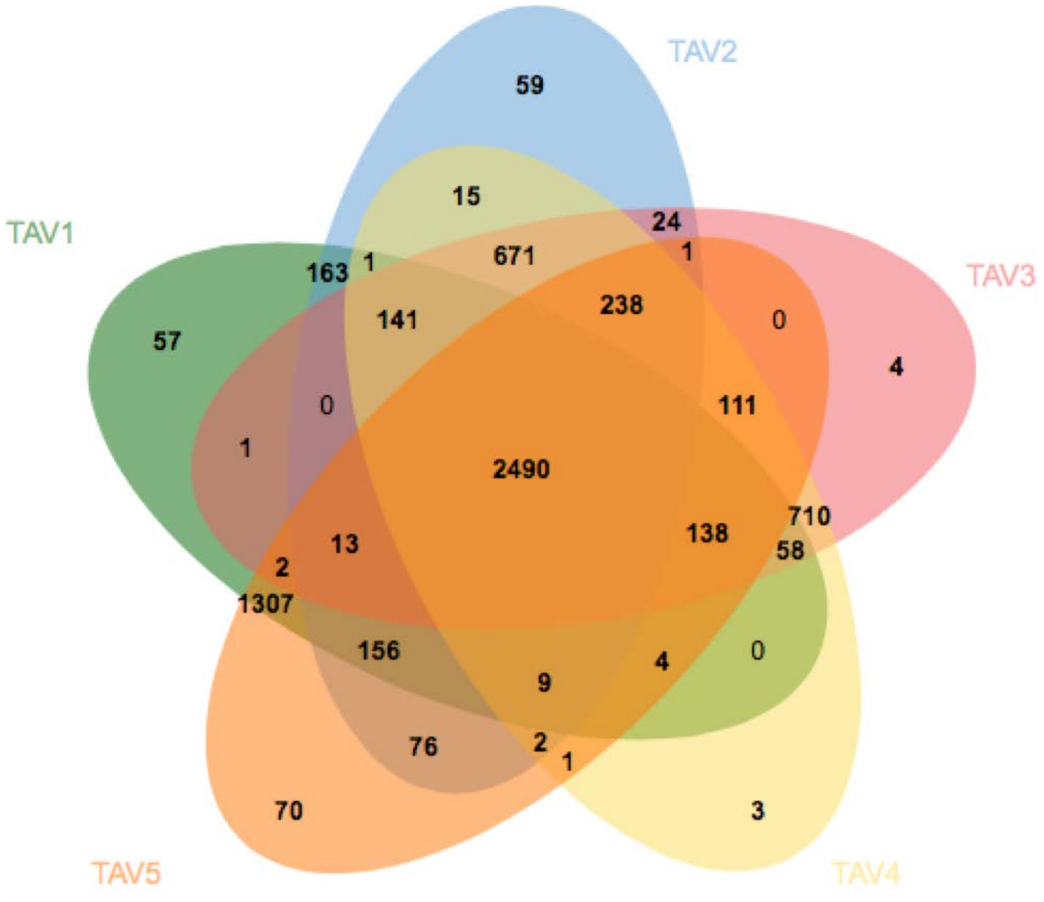


Figure 4.3 Venn diagram showing distribution of orthologous gene clusters among TAV strains. 2490 gene clusters are shared by all 5 TAV strains (core genome). 1307 gene clusters are shared between TAV1 and TAV5 and 671 gene clusters are shared between strains TAV2, TAV3 and TAV4 (species-specific/auxiliary genome).

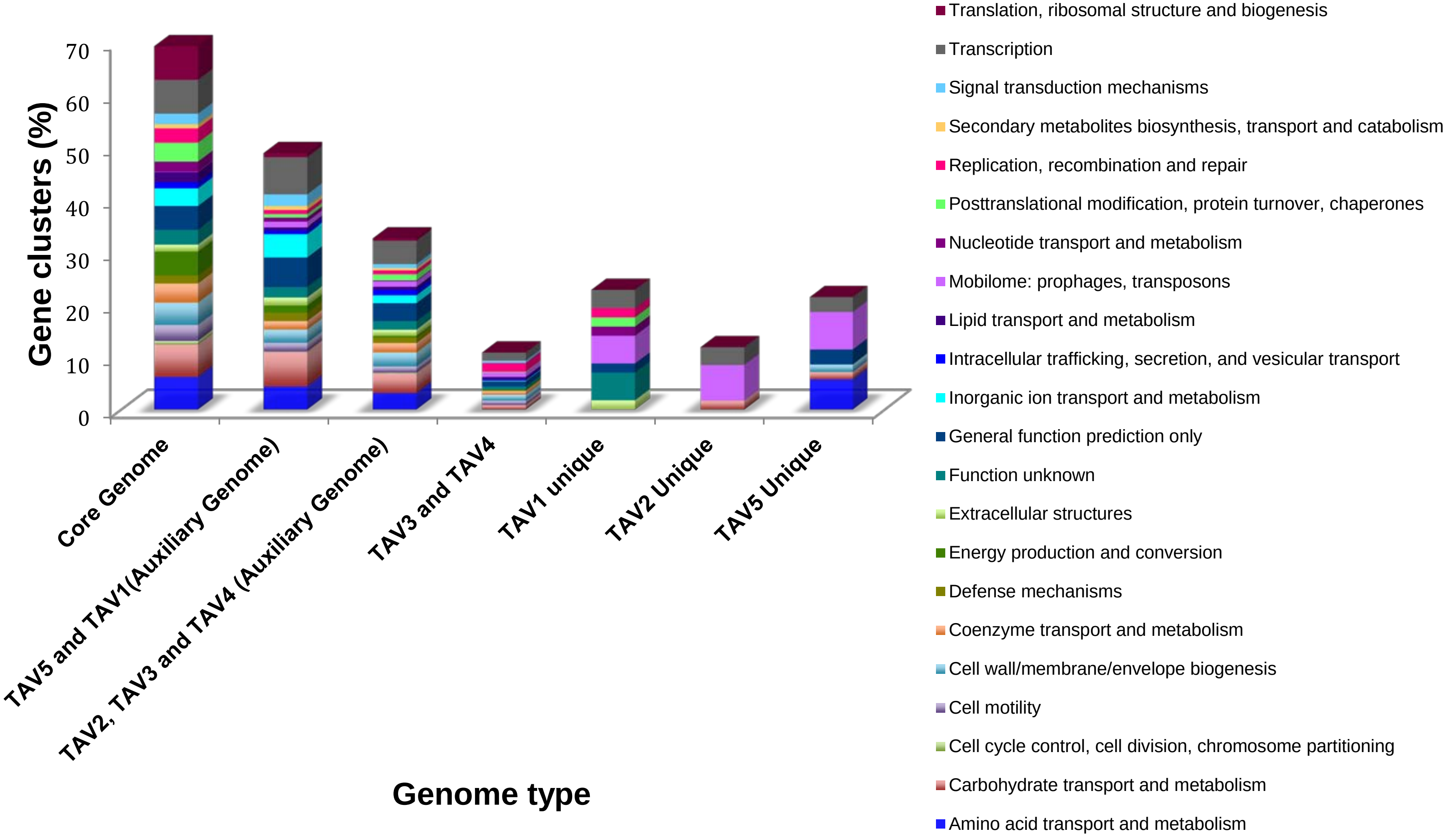
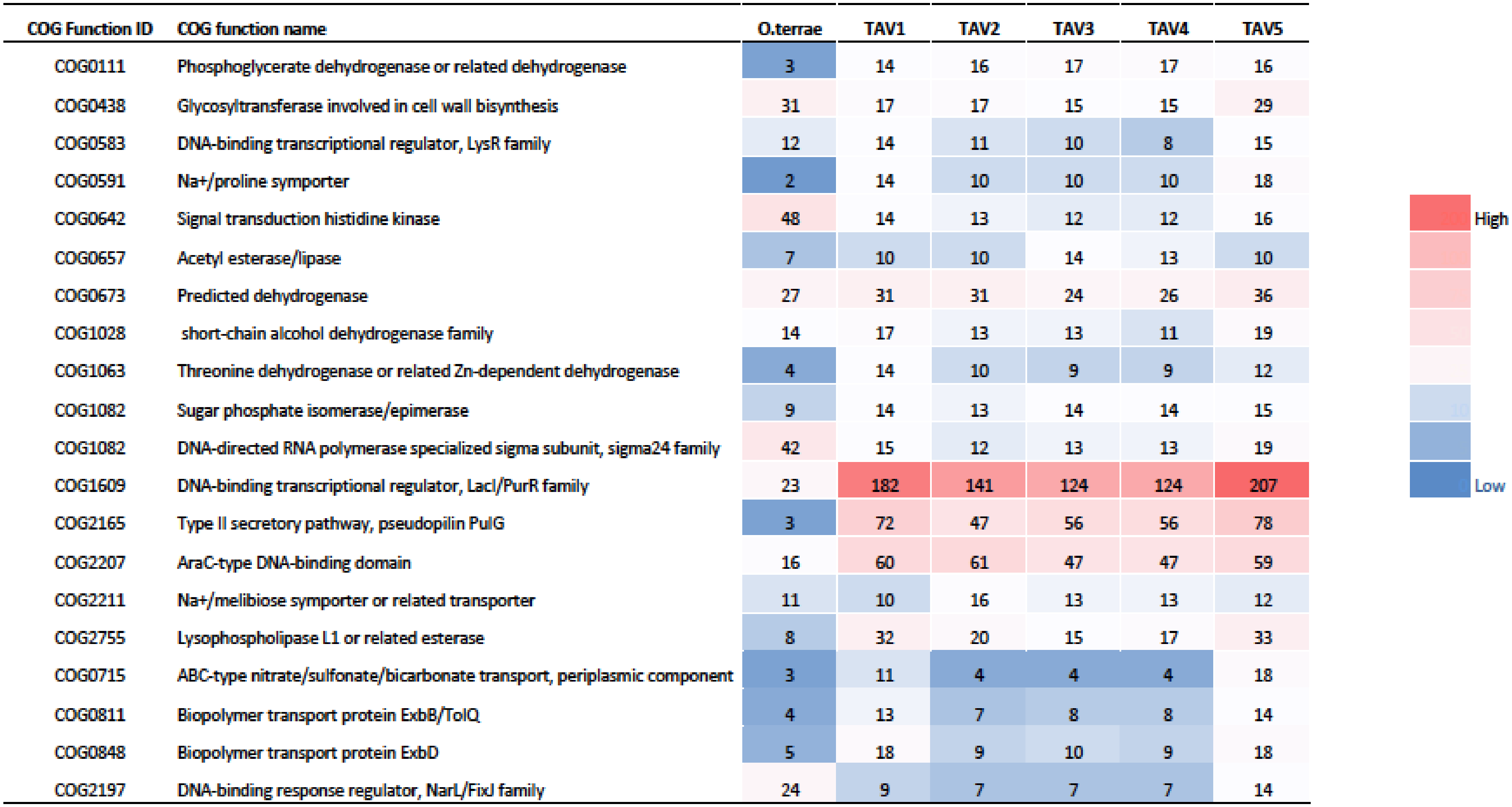


Figure 4.4 Functional annotation of core, auxiliary and unique genomes of TAV strains. Distribution of COG categories (percentage of total number of genes) in core, auxiliary and unique genomes of TAV strains.



COG Function ID	COG function name	O.terrae	TAV1	TAV2	TAV3	TAV4	TAV5
COG0111	Phosphoglycerate dehydrogenase or related dehydrogenase	3	14	16	17	17	16
COG0438	Glycosyltransferase involved in cell wall bisynthesis	31	17	17	15	15	29
COG0583	DNA-binding transcriptional regulator, LysR family	12	14	11	10	8	15
COG0591	Na+/proline symporter	2	14	10	10	10	18
COG0642	Signal transduction histidine kinase	48	14	13	12	12	16
COG0657	Acetyl esterase/lipase	7	10	10	14	13	10
COG0673	Predicted dehydrogenase	27	31	31	24	26	36
COG1028	short-chain alcohol dehydrogenase family	14	17	13	13	11	19
COG1063	Threonine dehydrogenase or related Zn-dependent dehydrogenase	4	14	10	9	9	12
COG1082	Sugar phosphate isomerase/epimerase	9	14	13	14	14	15
COG1082	DNA-directed RNA polymerase specialized sigma subunit, sigma24 family	42	15	12	13	13	19
COG1609	DNA-binding transcriptional regulator, LacI/PurR family	23	182	141	124	124	207
COG2165	Type II secretory pathway, pseudopilin PulG	3	72	47	56	56	78
COG2207	AraC-type DNA-binding domain	16	60	61	47	47	59
COG2211	Na+/melibiose symporter or related transporter	11	10	16	13	13	12
COG2755	Lysophospholipase L1 or related esterase	8	32	20	15	17	33
COG0715	ABC-type nitrate/sulfonate/bicarbonate transport, periplasmic component	3	11	4	4	4	18
COG0811	Biopolymer transport protein ExbB/TolQ	4	13	7	8	8	14
COG0848	Biopolymer transport protein ExbD	5	18	9	10	9	18
COG2197	DNA-binding response regulator, NarL/FixJ family	24	9	7	7	7	14

Figure 4.5 COG gene families undergoing expansion in TAV genomes and its comparison with *O. terrae*. The cells in the table are color-coded based on number on number of genes present in each genome for that particular COG family.

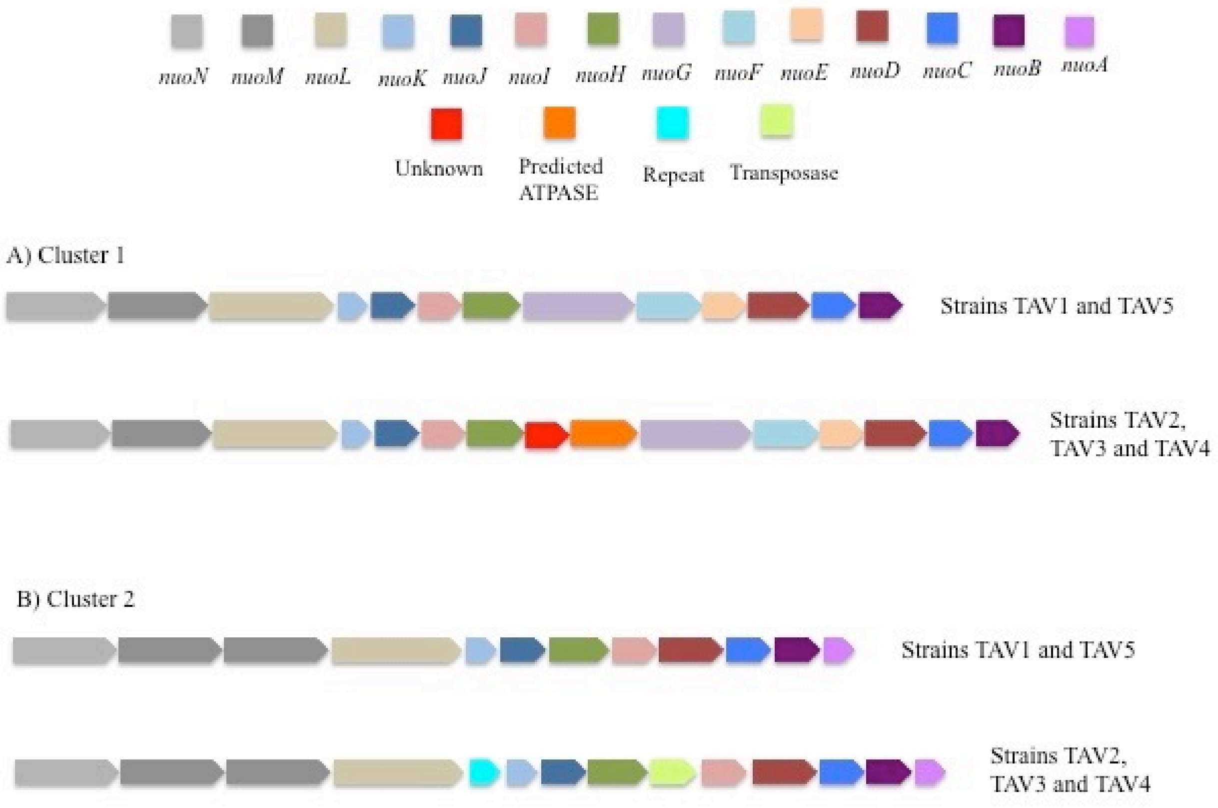


Figure 4.6 Gene clusters encoding type-1 NADH dehydrogenase A) Cluster 1 B) Cluster 2 in TAV strains.

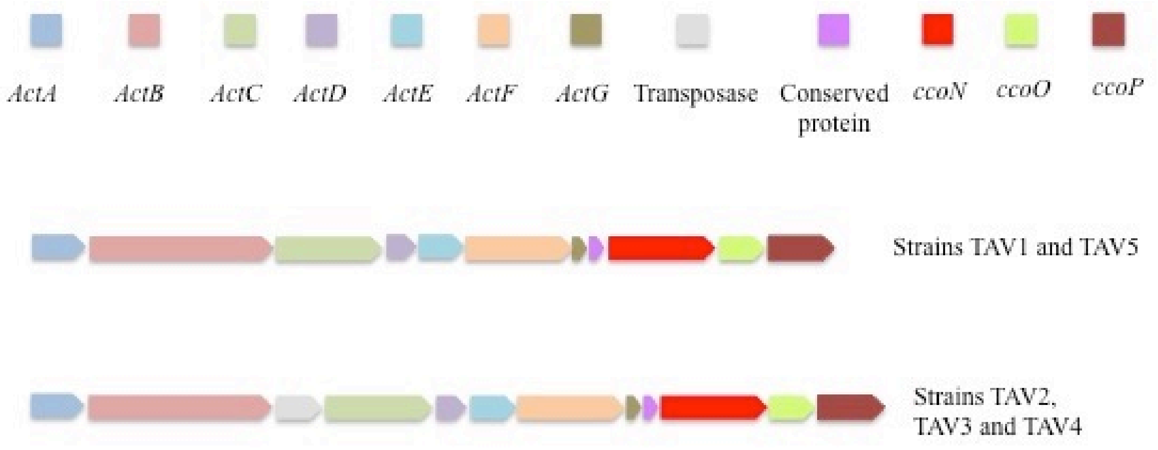


Figure 4.7 Gene clusters encoding an alternative complex III (ACIII) and *cbb3* type cytochrome oxidase.

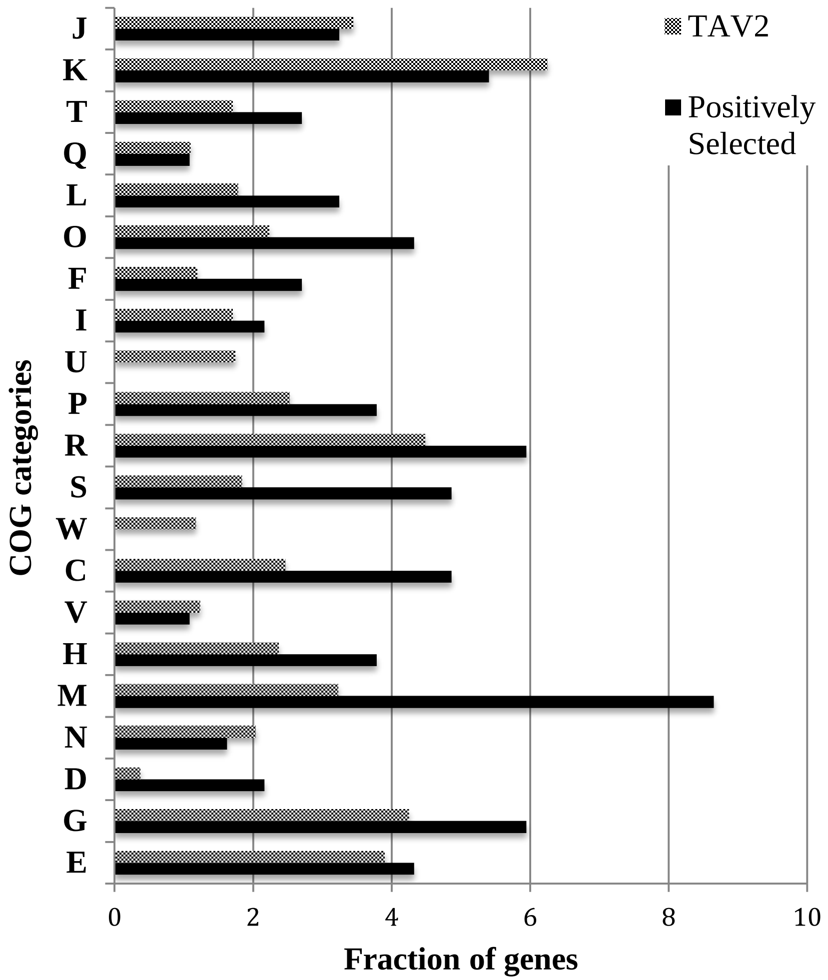


Figure 4.8 Distribution of positively selected gene clusters based on COG categories. The black bars indicate genes under positive selection in TAV strains (n=185) and the checkered bars are for all the genes annotated in strain TAV2 (n=5105). COG categories with codes D, M, C, F, O, L, and S were significantly enriched ($P<0.05$, binomial test) in positively selected genes as compared to all genes in TAV2. COG category codes are as follows: U, intracellular trafficking, secretion and vesicular transport; G, carbohydrate transport and metabolism; I, lipid transport and metabolism; R, general function prediction only; D, cell cycle control, cell division, chromosome partitioning; H, coenzyme transport and metabolism; P, inorganic ion transport and metabolism; W, extracellular structures; O, posttranslational modification, protein turnover, chaperones; J, translation, ribosomal structure and biogenesis; L, replication, recombination and repair; C, energy production and conversion; M, cell wall/membrane/envelope biogenesis; Q, secondary metabolites biosynthesis, transport and catabolism; V, defense mechanisms; E, amino acid transport and metabolism; K, transcription; N, cell motility; T, signal transduction mechanisms; F, nucleotide transport and metabolism; and S, function unknown.

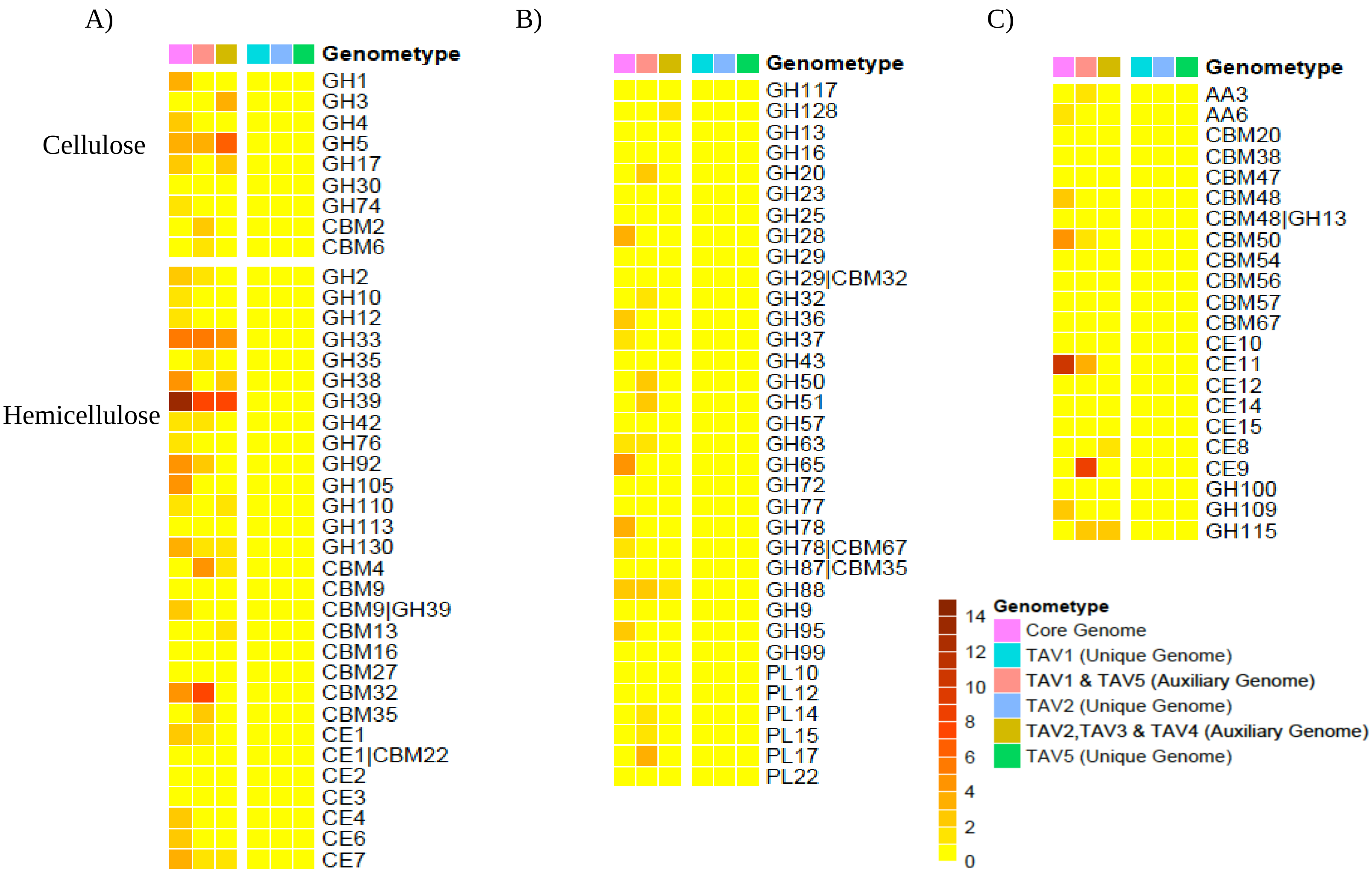


Figure 4.9 Heat map showing the number of putative CAZymes identified in core, auxiliary and unique genomes and their distribution in various CAZy families. A) CAZy families putatively involved in lignocellulose degradation. B) and C) CAZy families involved in breakdown of other plant and host derived polysaccharides.

Chapter 5 RNA-seq based analysis of the microaerophilic physiology of *Diplosphaera colitermitum* TAV2, a hindgut isolate of *R. flavipes*.

Manuscript in preparation

Authors: Malini Kotak*, Jantiya Isanapong*, Jorge L. M. Rodrigues

**co-authors*.

Author's Contribution: JLMR, MK and JI conceived the project. JI prepared the RNA samples for sequencing. MK performed all computational analysis. JLMR, MK and JI wrote the manuscript.

5.1 Abstract

Diplosphaera colitermitum TAV2, previously identified as a microaerophile, can also grow in the presence of atmospheric oxygen concentration, albeit slowly. This unique ability of TAV2 is exploited in the present study to understand its microaerophilic physiology, and further our knowledge of microaerophilic organisms in general. In the presence of 2% O_2 we observed an upregulation in expression of 261 genes while 174 genes were downregulated. As expected, the genes that encode the proteins that play a role in oxidative stress protection were downregulated in hypoxic conditions. Putative CAZymes that degrade lignocellulose and other polysaccharides like pectin represent a majority of upregulated genes. These CAZymes are arranged in a unique operon like arrangement where each operon consisted of one or more transcriptional regulators, CAZyme, putative secretion/cell surface attachment proteins. Genes associated with functional categories like motility, amino acid metabolism, and coenzyme metabolism were downregulated. Additionally, we identify the changes in electron transport chain employed by TAV2 under hypoxic and atmospheric oxygen concentrations. In the present study we attempt to determine the reason for downregulation of pyruvate dehydrogenase under microaerophilic conditions, and also identify putative oxygen sensing signal transducing proteins in the TAV2 strain. In addition to corroborating our previous findings, these results reveal the metabolic adaptations used by TAV2 to survive in hypoxic conditions and will also provide a foundation for future studies to understand the microaerophilic physiology of the TAV strains.

5.2 Introduction

Diplosphaera colitermitum TAV2 is a *Verrucomicrobia* strain isolated from the hindgut of lower termite, *Reticulitermes flavipes* (1). Termites harbor a huge variety of microorganisms in their hindgut that assist in the efficient digestion of otherwise recalcitrant lignocellulose (2). Recently, this symbiotic relationship between termites and its hindgut symbionts has garnered more interest because of its potential application in modeling efficient and sustainable industrial processes for biofuel production from plant biomass. Despite being studied for a long time the fine details of this symbiosis are still unclear due to the complexity and unculturability of the gut microorganisms (3).

An important aspect of termite gut symbiosis, which needs attention, is the understanding of the microaerophilic microorganisms present in the gut periphery**.** The termite's 1ul-volume hindgut harbors a rich diversity of microbial populations of up to 250 species and is highly structured. There is a steep O_2 gradient in the hindgut, creating a microaerophilic zone surrounding a completely anoxic core where important processes like methanogenesis and acetogenesis occur (Chapter 1, Fig 1.4) (4). However, there is a constant influx of O_2 in the hindgut that can jeopardize this highly structured environment. This highly structured environment is important for maintaining the community organization and its related functions. Therefore, the microbial community that colonizes the hindgut wall is considered to be essential for O_2 removal and maintenance of the anoxic core. Surprisingly, very limited attention has been paid to the understanding of microaerophilic microorganisms, their metabolism and ecological physiology.

Strain TAV2 has been identified as a facultative microaerophile which can survive under atmospheric oxygen (20%) but prefers 2-8% O_2 (5), mimicing the oxygen concentration in the termite gut periphery. The genomic, transcriptomic and proteomic analysis indicate a putative role for this strain in termite nutrition, hemicellulose degradation and O_2 scavenging (1, 5). Though the previous studies increased our understanding about this strain's function and metabolism in the termite gut, they do not answer important questions about its microaerophilic physiology. Also, these studies have posed some new questions regarding the strain's physiology including pyruvate metabolism (5), missing enzymes in respiratory chain (1) etc., which warrant further investigation.

The goal of this study was to understand the metabolic changes employed by the strain TAV2 to survive in the hypoxic periphery of termite guts through whole genome transcriptome analysis using high throughput RNA-seq technology. We believe that the knowledge of the TAV2 transcriptome will further our understanding about the physiology of this strain as well as the understanding of microaerophilic physiology in general. Through our previous studies we have already established that oxygen leads to significant metabolic changes in TAV2 (5), therefore through this study we aim at identifying the fine-scale changes that occur in the metabolism of TAV2 which were not captured in our previous study.

5.3. Materials and Methods

5.3.1 General considerations for experimental design

In the current study we carried out a transcriptome analysis of strain TAV2 at 2% and 20% O_2 concentration, where 2% O_2 concentration mimics the condition in the termite gut periphery. Transcriptome analysis has proven to be well suited to determine the genetic response of a bacterium to changes in environmental conditions. By characterizing a stimulon i.e. the group of genes that change expression in response to an environmental stimulus like pH change or temperature change, it is possible to understand the physiological strategy applied by that organism to cope with the corresponding environmental change (6). We hypothesized that a fine-scale change in expression will be observed which can be statistically quantified by counting Illumina reads as compared to light intensity measurements used in microarrays. We took advantage of the latest developments in high throughput sequencing to generate transcriptomic profiles (RNA-seq). This approach provides the following advantages over microarrays: 1) because no probe sequences are specified, all transcription is studied in an unbiased manner allowing discovery of novel genetic features and delineation of operons and untranslated regions; 2) mapping of sequences is more precise permitting the study of more repetitive regions of the genome (TAV genomes have high number of repetitive regions); 3) no issues of interference between genes due to non-specific hybridization of cDNA to probes; and 4) because the sequence reads do not saturate like light detection systems, RNA-seq provides a greater range in measuring variations in expression levels, and is more sensitive at low expression levels (7).

5.3.2 RNA isolation, library preparation and sequencing

TAV2 cells to be used for transcriptomic analysis were grown in R2B medium separately under 2% and 20% oxygen concentrations with three biological triplicates in each condition. The cells were harvested when the OD_{600} reached approximately 0.3 and were stored in RNA*later*^{TM}, RNA Stabilization Reagent (Qiagen) till they were ready to be used for total RNA isolation. The total RNA was extracted using the RiboPure^{TM} – Bacteria kit (Ambion) according to the manufacturer’s instructions. The total RNA isolated was subjected to ribosomal-RNA depletion using the RiboZero^{TM} rRNA removal kit (Bacteria, Epicentre) following manufacturer's instructions. The rRNA depleted RNA was stored at -20°C and transported to the Joint Genome Institute (JGI) for sequencing. Libraries were prepared using ScriptSeq mRNA-Seq Library preparation kit (Epicenter) as per manufacturer's instructions, and sequencing was performed using Illumina's HiSeq^{TM} 2000 platform.

5.3.3 Bioinformatics and Analysis

The filtered reads were aligned to the reference genome using Bowtie2 (8) with default settings. Differential expression analysis was performed using EdgeR (9) and Deseq which identifies significant differentially expressed genes from the genome-scale count data using exact tests based on a negative binomial distribution. Genes identified as differentially expressed in 2% O_2 concentration relative to 20% O_2 with a adjusted *p*-value less than 0.1 were used for further analysis. COG enrichment analysis was performed using customized scripts.

5.3.4 NADH/NAD+ analysis

TAV2 were grown in 25 mL R2A broth and incubated at 2% and 20% oxygen concentration with biological triplicates in each condition. When the OD_{600} was approximately 0.3, 1 mL culture from each replicate was harvested. The harvested cell pellets were immediately frozen in a dry ice-ethanol bath. Subsequently, 250ul of either 0.2 M HCl (for NAD^+ extraction) or 0.2 M NaOH (for NADH extraction) was added to the frozen pellets. The samples were placed in a 100° C water bath/heat block for 10 min, and then subjected to centrifugation to remove cellular debris. The dinucleotide-containing supernatants were transferred to fresh tubes and used for the assay.

Assays of the extracts containing specific dinucleotides were performed in triplicate by following Bernofsky and Swan's recycling assay (10). Coenzyme standards from 0.00005 to 0.005 mM were used to calibrate the assay.

5.4 Results

5.4.1 Differential expression of genes in response to O_2 concentration

Analysis of TAV2 transcriptome revealed that 435 genes ($P < 0.05$, FDR < 10%) were differentially regulated with 261 genes being upregulated and 174 genes downregulated when the TAV2 cells were grown in 2% and 20% O_2, respectively (Supplementary Table S2 and S3) (Table 5.1, Fig 5.2). To gain insight into the metabolic profile of the transcriptome, the expression data was examined using the pathway prediction program available on http://biocyc.org/ and assigned to COG categories. Approximately 50% of the upregulated genes were annotated as hypothetical proteins or proteins of unknown function. Carbohydrate metabolism and CAZy genes represented a large percentage of upregulated genes, while, functional categories like amino acid transport and metabolism coenzyme transport and metabolism, post-translational modification and transcription were downregulated in hypoxic condition. The exact pathways in which the various differentially expressed genes might be involved are discussed below.

5.4.2 Upregulation of extensive repertoire of CAZymes

The upregulated genes were searched against the CAZy database to identify genes with potential roles in lignocellulose degradation. The analysis revealed a high number of upregulated CAZymes when TAV2 was grown in the presence of 2% oxygen compared to 20% O_2 (Fig 5.1).

Most of the upregulated CAZymes consisted of N-terminal signal peptides suggesting that they are transported to the periplasm or outer membrane during lignocellulose digestion. The most highly induced CAZy gene was upregulated 1.66 fold and encodes a putative fucose-binding domain. Several xyloglucans have fucose containing side-chains (11) so this enzyme might be involved in hemicellulose digestion. The next most highly up-regulated CAZyme (1.61 fold) encodes a GH5 family endoglucanase with potential cellulase activity. Also found upregulated were GH74 family putative cellobiohydrolase and another GH5 family endoglucanase. Thus several potential cellulose-degrading enzymes were found to be upregulated in the TAV2 transcriptome in the presence of microoxic conditions.

Several CAZymes putatively involved in hemicellulose degradation were also upregulated. For instance, genes encoding acetyl xylan esterase that acts on 2/3-O-acetyl xylan and exo-β-1,4-xylosidase, which acts on xylooligomers and xylobiose were also upregulated. Also, upregulated expression was observed for 2 genes encoding β-mannosidase capable of digesting β-mannan based hemicelluloses. In addition to lignocellulose digesting CAZymes, pectin-digesting enzymes with potential pectin methylesterase and rhamnogalacturonly hydrolase activity were found to be upregulated.

Closer inspection of CAZy genes revealed a very interesting arrangement of these genes (Fig. 5.1). Each gene was arranged in an operon like structure with a transcription regulator, one or more CAZyme, a PEP-CTERM domain containing protein, and a prepilin like N-terminal domain containing protein. Along with CAZymes, most of these proteins present in the operons were also upregulated. PEP-CTERM domain containing proteins and prepilin like N-terminal domain containing protein have trans-membrane helices or signal peptides suggesting that these proteins might be responsible for transporting CAZymes to the outer-membrane. Interestingly, such N-terminal prepilin like domains were reported to attach the CAZyme to outer membrane (12, 13). We found such prepilin and PEP-CTERM containing domains to be present in multiple copies in all TAV genomes (Chapter 4) and a high number of genes coding for such domains, with no potential association with CAZymes were also found to be upregulated. Determining the function of these domains and if they attach/transport CAZymes to the outer membrane of TAV2 could be an interesting future study.

5.4.3 Differential regulation of carbohydrate metabolism and transport

Apart from CAZymes, the only carbohydrate metabolism enzyme found to be significantly upregulated in the presence of 2% O_2 was ribulose-5-phosphate-4-epimerase. This enzyme participates in the non-oxidative phase of pentose phosphate pathway and is also found on arabinose operon encoded by gene *araD*. Additionally, arabinose utilization genes *araA*, *araB* and *araC* were also upregulated, but the fold change was not statistically significant. However, the genes encoding subunits of glucose-6-phosphate dehydrogenase, which catalyze the oxidative branch of pentose phosphate pathway, were downregulated. Down regulation of the pentose phosphate pathway enzymes when TAV2 is grown in hypoxic condition was also observed in our previous study (5).

Our results are in agreement with earlier findings that demonstrate the downregulation of pyruvate dehydrogenase complex subunit E1 (5). This enzyme catalyzes the oxidative decarboxylation of pyruvate to form acetyl coenzyme A (acetyl-CoA). Expression of phosphoenolpyruvate carboxylase, which forms oxaloacetate from phosphoenol pyruvate, was also downregulated under microoxic conditions. Phosphenolpyruvate carboxylase plays an important role in replenishing the oxaloacetate of citric acid cycle to increase the flux through the cycle.

Additionally, the strain TAV2 induced ABC transporter genes encoding subunits MalK, MalF, MalE and porin of a maltose transporter. Also, TonB dependent transporter subunits ExbB and ExbD were found to be upregulated. Interestingly, two genes encoding arabinose efflux permease, which transport arabinose, were down regulated.

5.4.4 Energy production and related metabolism

In the TAV2 genome two different types of respiratory NADH dehydrogenases have been identified which include proton-pumping type I (NDH-1), and non-proton pumping type II (NDH-2) (Chapter 4). NDH-1 is encoded by *nuoA-N* genes, out of which *nuoL* and *nuoM* were upregulated under 2% O_2 whereas *nuoB* was downregulated. NDH-2 was also downregulated under microoxic conditions. A putative quinone oxidoreductase (QOR) and V-type H+ translocating pyrophosphatase were upregulated under 2% O_2. V-type H+-translocating pyrophosphatase is known to establish a proton gradient across the membrane, and is reported to have a role in scavenging metabolic energy (14). QOR catalyzes the transfer of electrons from NADPH to substrates, and play a role in oxidative stress resistance (15, 16). Finally, expression of a NiFe group 4 hydrogenase was also down regulated in TAV2 grown under hypoxic conditions. Hydrogenases catalyze reversible oxidation of hydrogen gas and play a role in energy metabolism, fermentation and H_2 respiration (17). Additionally, gene coding for superoxide dismutase was significantly downregulated in the presence of 2% O_2.

5.4.5 Downregulation of coenzyme metabolism

Biosynthesis of several coenzymes was downregulated when TAV2 was grown at 2% O_2 compared to 20% O_2. Thiamine (Vitamin B1) biosynthesis in bacteria is carried out by the separate formation of pyrimidine and thiazole moieties, which are subsequently coupled to form thiamine. Our results revealed that enzymes 1-deoxy-D-xylulose-5-phosphate synthase and thiazole-phosphate synthase, which participates in thiazole biosynthesis (18) and thiamine-phosphate pyrophosphorylase, which participates in pyrimidine biosynthesis, were downregulated. TAV2 also showed downregulation of enzymes lipoate synthase (LipA) and lipoate-protein ligase (LplA) responsible for the biosynthesis of lipoic acid. Lipoic acid is an important cofactor for the aerobic metabolism and functioning of pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase, and glycine cleavage system (19). Also, pyridoxine 5'-phosphate oxidase, which catalyzes the last step of vitamin B6 biosynthesis and cobalamin adenosyltransferase involved in cobalamin (B12) metabolism were downregulated. Lastly, ABC transporters involved in the transport of Fe^{3+} and cobalamin were also downregulated under 2% O_2 condition.

5.4.6 Downregulation of amino acid metabolism

Our analysis revealed that strain TAV2 downregulated the expression of 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase (*aroF*), and 3-phosphoshikimate 1-carboxyvinyltransferase (*aroA*) which participate in aromatic amino acid biosynthesis super pathway and prephanate dehydrogenase (*tyrA*) which catalyzes the reaction from phenylalanine to tyrosine. As previously reported, the cobalamin independent methionine synthase (*metE*) was also downregulated when TAV2 was grown in the presence of 2% O_2. Both subunits of 3-

isopropylmalate dehydratase (*leuD*) involved in leucine biosynthesis were downregulated. Additionally, the expression of histidinol-phosphate phosphatase (*hisK*) responsible for biosynthesis of histidine and sulfate adenylyltransferase subunit 2 (*cysD*), which eventually leads to cysteine biosynthesis, was also downregulated. Interestingly, putative threonine dehydrogenase and lysophospholipase L1 like enzymes, which are present in multiple copies in the TAV2 genome, were differentially regulated in response to O_2. Three genes encoding putative lysophospholipase L1 like enzymes were upregulated, and one such putative enzyme was downregulated. Also, a putative threonine dehydrogenase was upregulated when TAV2 was grown in hypoxic conditions. Finally, an ABC transporter involved in polar amino acid transport was downregulated.

5.4.7 Downregulation of motility chemotaxis genes

A total of 14 genes associated with motility and chemotaxis were downregulated in TAV2 in the presence of 2% O_2. The genes required for biogenesis of flagella, *motA* and *motB* genes encoding motor protein, *fliF* coding for MS ring, *fliG* coding for C ring, and export apparatus genes *flhA*, *flhB*, *flil*, *fliP*, *fliQ* and *fliR* were downregulated. Also, the expression of a CheY like chemotaxis protein was downregulated.

5.4.8 Transcription, posttranslation and signal transduction regulation

Several regulatory proteins were differentially regulated when TAV2 was grown at 2 different O_2 concentrations. An extracellular sigma factor (ECF) sigma 70 was upregulated under 2% O_2 condition. ECF sigma factors are responsible for regulating a wide variety of functions involved in sensing and reacting to conditions in the membrane, periplasm or extracellular environment (20). Also upregulated expression was observed for an AraC type transcriptional regulator, a LacI family transcription regulator and a transcription regulator, which could not be assigned to any family. A sigma 70, ECF sigma 70, 2 LysR family transcriptional regulator, AraC type transcriptional regulator, and 3 putative transcriptional regulators were downregulated.

In addition, two genes coding for proteins with chaperone activity, GrpE and DnaJ, were downregulated in 2% O_2, suggesting that oxidative stress at 20% O_2 leads to improper protein folding. Expression of SufD protein that participates in oxidative stress management was also downregulated. Four genes coding for putative two-component regulatory system proteins were significantly downregulated in TAV2 under hypoxic condition. Two of these genes were among the top 20 differentially expressed genes with >2 fold change (ObacDraft_3816 and ObacDraft_3817, Table 5.1).

5.5 Discussion

In this study, the transcriptome of *Diplosphaera colitermitum* TAV2 expressed when grown at 2% O_2 was compared to that expressed under 20% O_2. This *Verrucomicrobial* strain is possibly located in the termite gut periphery where the typical O_2 concentrations range from 2-4% (1). The aim of this study was to determine the stimulon that is differentially regulated in response to O_2, which will give us an insight into the mechanisms used by TAV2 to sense the change in O_2 concentration, and to acclimatize to hypoxic conditions in the termite gut. Under 2% O_2 concentration TAV2 induced expression of genes involved in functional categories carbohydrate metabolism and replication and recombination. The majority of the genes upregulated in the carbohydrate metabolism category were involved in polysaccharide degradation (CAZymes). This organism possesses a remarkable repertoire of CAZymes, which can potentially digest the dietary polysaccharide lignocellulose and host derived polysaccharides like mucin (Chapter 4). Our present study demonstrates that in addition to the upregulation of xylan degrading genes reported previously (5), TAV2 also induces expression of potential cellulose degrading, hemicellulose (other than xylan) degrading and pectin degrading CAZymes (Fig 5.3). Upregulation of such a high number of CAZymes suggests a very active role for TAV2 in lignocellulose digestion. Another interesting finding of this study is the arrangement of upregulated CAZymes in operon like structures, which include potential secretion or cell surface attachment domains (Fig 5.1). However, this finding needs to be confirmed by further experiments.

Previously, based on proteomics studies Isanapong *et. al* reported downregulation in expression of acetyl transferring subunit of pyruvate dehydrogenase (PDH), which links glycolysis to citric acid cycle, in TAV2 grown under 2% O_2 (5). In the present study we found downregulation of gene encoding the subunit of PDH which decarboxylates pyruvate. For several microaerophiles, a dramatic increase in $NADH/NAD^+$ ratio has been reported when they are shifted from aerobic to microaerobic conditions. Accordingly, we hypothesized that hypoxia in TAV2 results in the accumulation of NADH, a known inhibitor of PDH, and it re-oxidizes the reducing equivalents by operating the TCA cycle in the reductive direction (5). However, when $NADH/NAD^+$ concentrations were measured, we did not find NADH accumulation in TAV2 cells cultured in hypoxic conditions, but we observed a significant NADH accumulation in the presence of 20% O_2 (Fig 5.3). These results indicate that accumulation of reduced cofactors like NADH does not lead to downreglulation of PDH in TAV2. On further analysis of TAV2 transcriptome we observed that along with PDH the coenzymes required for its activity, lipoic acid and thiamine pyrophosphate (21), are also being downregulated. Several pathogens are reported to have lost the ability to de novo synthesize thiamine and lipoic acid de novo and depend completely on their host for supply of these cofactors. However, the TAV2 cells were cultured in a media containing yeast extract, which contains these cofactors making cofactor unavailability a less likely reason for PDH downregulation. Additionally, in a different study we observed an upregulation in expression of enzyme acetate kinase, when TAV2 was cultured in the presence of 2% O_2 (data not shown). *Stenoxybacter acetivorans*, an obligate microaerophile from the *R. flavipes* gut, utilizes acetate as a primary carbon source in the termite hindgut and uses acetate kinase and phosphotransacetylase for its activation to acetyl-CoA (22) thus bypassing the PDH reaction required to generate acetyl-CoA. Acetate is abundant in the termite hindgut and upregulation of acetate kinase suggests that as an adaptation to survive in termite hindgut, TAV2 could be utilizing acetate as a primary carbon source. Currently, either use of an alternate carbon source or downregulation of coenzyme biosynthesis appears to be a possible explanation of PDH downregulation, but further experiments are required to determine the exact reason.

Transcriptional profiling of TAV2 revealed differential regulation of NADH dehydrogenase (complex I) of electron transport chain. Genes encoding subunits *nuoL* and *nuoM* of NDH-1 were upregulated in 2% O_2 while *nuoB* was downregulated. Additionally, expression of non proton-pumping NDH-2 was downregulated. This is consistent with a previous study showing upregulation of NDH-1 subunits in TAV2 grown in hypoxia (5). Together, these results indicate that TAV2 prefers NDH-1 for microaerophilic physiology and NDH-2 for aerobic respiration. The choice of NDH-1 over NDH-2 seems to be unusual because NDH-1 is known to produce more reactive oxygen species as compared to NDH-2 (23), and proteins like superoxide dismutase and SufD involved in counteracting the effect of oxidative stress were downregulated with depletion in O_2 concentration. This choice can be attributed to the energy conserving and PMF generating nature of NDH-1. Additionally, upregulation of a putative NAD(P)H quinone oxidoreductase might be an adaptation of TAV2 to resist this oxidative stress. Downregulation of the *nuoB* subunit which contains 4Fe-4S cluster can be attributed to the downregulation of Fe+ transportation and metabolism; but a significant downregulation was not observed for other Fe-S containing *nuo* subunits (*nuoEFGI*). The effect of downregulation of nuoB on the activity of NDH-1 remains to be determined.

Signal transduction is a principle mechanism by which cells respond to various environmental cues. We identified two signal-transducing proteins, which can be potentially associated with sensing O_2 change in TAV2. A gene coding for PAS domain containing protein (ObacDraft_5084) was upregulated in the presence of 2% O_2. PAS domains are sensory domains and protein-protein interaction sites identified in a large variety of signal transducing proteins and they recognize several stimuli, which include light, oxygen and redox potential (24). Another likely O_2 sensing signal transduction protein can be a membrane associated histidine kinase and its corresponding DNA binding domain-containing protein (ObacDraft_3817, ObacDraft_3816), which were one of the top twenty differentially regulated proteins in this study (Table 5.1). A homology search identified these proteins as putative 2-component system phoPR, which participates in phosphate metabolism and particularly helps in scavenging phosphate. The deduced gene products of 2 coding DNA sequences (CDS) (ObacDraft_3015, ObacDraft_3014) upstream of these putative phoPR genes were alkaline phosphatase like proteins and were also downregulated in the presence of 2% O_2. Interestingly, one of the genes coding a putative phosphatase (ObacDraft_3015) was the most downregulated gene in the study. Together these results suggest that these four genes are co-regulated and form an operon possibly involved in phosphate dependent gene regulation in TAV2. Such phosphate dependent gene regulation in response to oxidative stress has been previously reported for intracellular pathogen *Mycobacterium tuberculosis* (25).

5.6 References:

1. **Wertz JT**, **Kim E**, **Breznak JA**, **Schmidt TM**, **Rodrigues JLM**. 2012. Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. Appl. Environ. Microbiol. **78**:1544–1555.
2. **Brune A**. 2014. Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. **12**:168–80.
3. **Hongoh Y**, **Sharma VK**, **Prakash T**, **Noda S**, **Taylor TD**, **Kudo T**, **Sakaki Y**, **Toyoda A**, **Hattori M**, **Ohkuma M**. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proc. Natl. Acad. Sci. U. S. A. **105**:5555–5560.
4. **Brune a.**, **Emerson D**, **Breznak J a.** 1995. The termite gut microflora as an oxygen sink: Microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol. **61**:2681–2687.

5. **Isanapong J**, **Sealy Hambright W**, **Willis AG**, **Boonmee A**, **Callister SJ**, **Burnum KE**, **Paša-Tolić L**, **Nicora CD**, **Wertz JT**, **Schmidt TM**, **Rodrigues JL**. 2013. Development of an ecophysiological model for *Diplosphaera colotermitum* TAV2, a termite hindgut *Verrucomicrobium*. ISME J. **7**:1803–13.

6. **Burkovski A**. 2008. Corynebacteria. Genomics and molecular biology comparative and functional genomics. 5-32.

7. **Croucher NJ**, **Thomson NR**. 2010. Studying bacterial transcriptomes using RNA-seq. Curr. Opin. Microbiol. **13**:619–624.

8. **Langmead B**, **Trapnell C**, **Pop M**, **Salzberg SL**. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10**:R25.

9. **Robinson MD**, **McCarthy DJ**, **Smyth GK**. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139–140.

10. **Bernofsky C**, **Swan M**. 1973. An improved cycling assay for nicotinamide adenine dinucleotide. Anal. Biochem. **53**:452–458.

11. **Scheller HV**, **Ulvskov P**. 2010. Hemicelluloses. Annu. Rev. plant Biol. **61**:263–289.

12. **Liebl W**, **Winterhalter C**, **Baumeister W**, **Armbrecht M**, **Valdez M**. 2008. Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **190**:1350–1358.

13. **Girfoglio M**, **Rossi M**, **Cannio R**. 2012. Cellulose degradation by *Sulfolobus solfataricus* requires a cell-anchored endo-β-1-4-glucanase. J. Bacteriol. **194**:5091–5100.

14. **Gifford SM**, **Sharma S**, **Booth M**, **Moran MA**. 2013. Expression patterns reveal niche

diversification in a marine microbial assemblage. ISME J. **7**:281–298.

15. **Wang G**, **Maier RJ**. 2004. An NADPH quinone reductase of *Helicobacter pylori* plays an important role in oxidative stress resistance and host colonization. Infect. Immun. **72**:1391–1396.

16. **Zheng Q**, **Song Y**, **Zhang W**, **Shaw N**, **Zhou W**, **Rao Z**. 2015. Structural views of quinone oxidoreductase from *Mycobacterium tuberculosis* reveal large conformational changes induced by the co-factor. FEBS J. **282**:2697–2707.

17. **Vignais PM**, **Colbeau A**. 2004. Molecular biology of microbial hydrogenases. Curr. Issues Mol. Biol. **6**:159–188.

18. **Dorrestein PC**, **Zhai H**, **McLafferty FW**, **Begley TP**. 2004. The biosynthesis of the thiazole phosphate moiety of thiamin: The sulfur transfer mediated by the sulfur carrier protein ThiS. Chem. Biol. **11**:1373–1381.

19. **Zhang H**, **Luo Q**, **Gao H**, **Feng Y**. 2015. A new regulatory mechanism for bacterial lipoic acid synthesis. Microbiologyopen **4**:282–300.

20. **Kazmierczak MJ**, **Wiedmann M**, **Boor KJ**. 2005. Alternative sigma factors and their roles in bacterial virulence. Microbiol. Mol. Biol. Rev. **69**:527–543.

21. **Spalding MD**, **Prigge ST**. 2010. Lipoic acid metabolism in microbial pathogens. Microbiol. Mol. Biol. Rev. **74**:200–228.

22. **Wertz JT**, **Breznak JA**. 2007. Physiological ecology of *Stenoxybacter acetivorans*, an obligate microaerophile in termite guts. Appl. Environ. Microbiol. **73**:6829–6841.

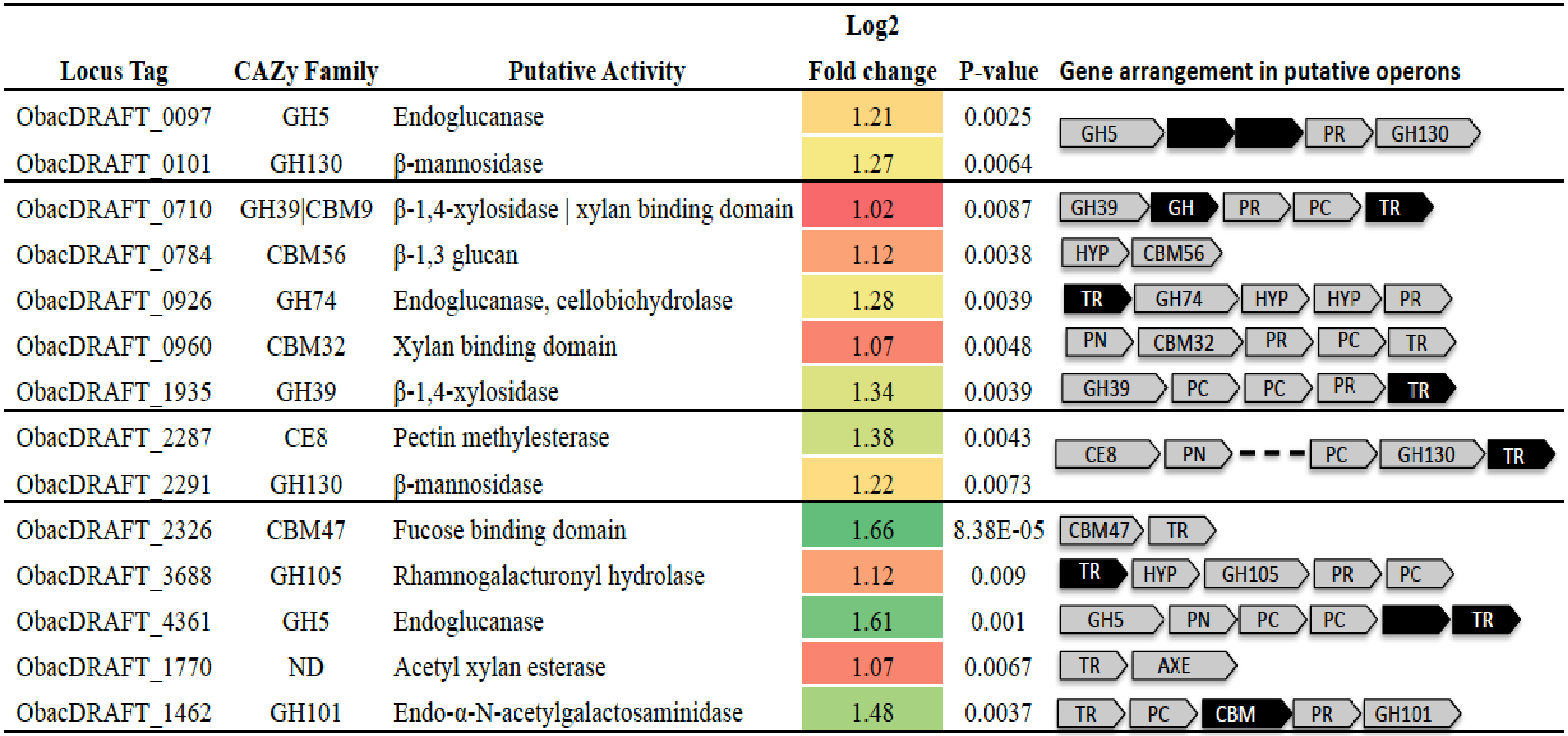
23. **Heikal A**, **Nakatani Y**, **Dunn E**, **Weimar MR**, **Day CL**, **Baker EN**, **Lott JS**, **Sazanov LA**, **Cook GM**. 2014. Structure of the bacterial type II NADH dehydrogenase: A

monotopic membrane protein with an essential role in energy generation. Mol. Microbiol. **91**:950–964.

24. **Petrova OE**, **Sauer K**. 2012. PAS domain residues and prosthetic group involved in bdladependent dispersion response by pseudomonas aeruginosa biofilms. J. Bacteriol. **194**:5817–5828.

25. **Gonzalo-Asensio J**, **Mostowy S**, **Harders-Westerveen J**, **Huygen K**, **Hernández-Pando R**, **Thole J**, **Behr M**, **Gicquel B**, **Martín C**. 2008. PhoP: A missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence. PLoS One **3**:e3496

26. **Karpinets T V**. 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database **20**:1574–1584.



Locus Tag	CAZy Family	Putative Activity	Log2 Fold change	P-value	Gene arrangement in putative operons
ObacDRAFT_0097	GH5	Endoglucanase	1.21	0.0025	GH5 ■ ■ PR GH130
ObacDRAFT_0101	GH130	β-mannosidase	1.27	0.0064	
ObacDRAFT_0710	GH39|CBM9	β-1,4-xylosidase | xylan binding domain	1.02	0.0087	GH39 GH PR PC TR
ObacDRAFT_0784	CBM56	β-1,3 glucan	1.12	0.0038	HYP CBM56
ObacDRAFT_0926	GH74	Endoglucanase, cellobiohydrolase	1.28	0.0039	TR GH74 HYP HYP PR
ObacDRAFT_0960	CBM32	Xylan binding domain	1.07	0.0048	PN CBM32 PR PC TR
ObacDRAFT_1935	GH39	β-1,4-xylosidase	1.34	0.0039	GH39 PC PC PR TR
ObacDRAFT_2287	CE8	Pectin methylesterase	1.38	0.0043	CE8 PN - - - PC GH130 TR
ObacDRAFT_2291	GH130	β-mannosidase	1.22	0.0073	
ObacDRAFT_2326	CBM47	Fucose binding domain	1.66	8.38E-05	CBM47 TR
ObacDRAFT_3688	GH105	Rhamnogalacturonyl hydrolase	1.12	0.009	TR HYP GH105 PR PC
ObacDRAFT_4361	GH5	Endoglucanase	1.61	0.001	GH5 PN PC PC ■ TR
ObacDRAFT_1770	ND	Acetyl xylan esterase	1.07	0.0067	TR AXE
ObacDRAFT_1462	GH101	Endo-α-N-acetylgalactosaminidase	1.48	0.0037	TR PC CBM PR GH101

Figure 5.1 Putative CAZyme encoding genes upregulated in TAV2 under microaerophilic condition (2% O2). Each gene was assigned to a CAZy family using CAZyme Analysis tool kit (26) and the arrangement of these CAZymes in putative operons was predicted. In predicted operon arrangement shown below, genes shown in grey were upregulated while those shown in black were not differentially regulated. PR: Prepillin like N-terminal domain containing protein, PC: PEP-CTERM domain containing protein, HYP:hypothetical protein, TR transcriptional regulator.

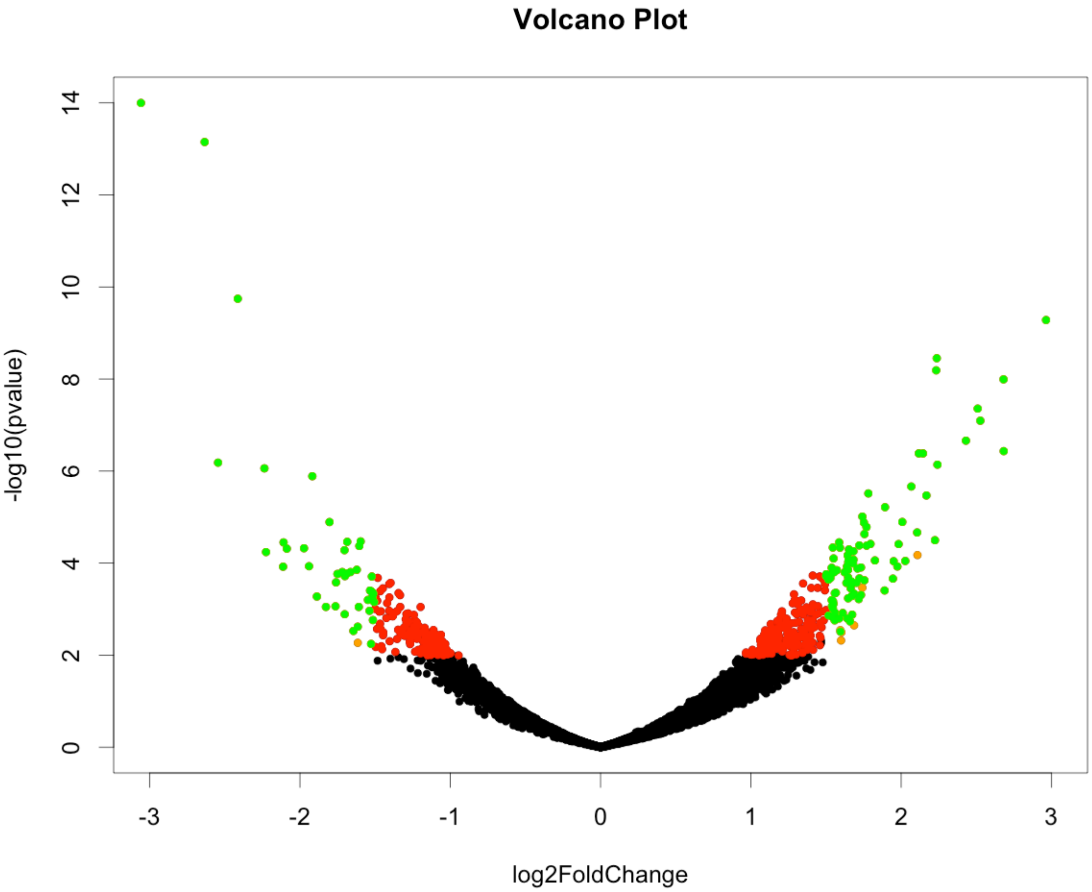


Figure 5.2 Volcano plot showing differential gene expression in TAV2 cells when cultured in presence of 2% and 20% O_2. Genes in red: adjusted *P*-value < 0.1, orange: Log_2 fold change: >1.0, green: *P*-value < 0.1 & Log_2 fold change: >1.0.

Table 5-1 Top 20 differentially regulated genes in strain TAV2 in response to oxygen.

Locus tag	**Function**	**Log2 fold change**	**P-value**
ObacDRAFT_1391	LysM domain.	2.5	8.05E-08
ObacDRAFT_1392	Uncharacterized protein conserved in bacteria	3.0	5.23E-10
ObacDRAFT_1397	Hypothetical protein	2.2	6.47E-09
ObacDRAFT_1398	Hypothetical protein	2.2	3.54E-09
ObacDRAFT_2835	Hypothetical protein	2.7	3.69E-07
ObacDRAFT_2967	Hypothetical protein	2.7	1.02E-08
ObacDRAFT_2970	SprT-like family.	2.2	7.26E-07
ObacDRAFT_2971	Hypothetical protein	2.4	2.19E-07
ObacDRAFT_3923	Hypothetical protein	2.5	4.40E-08
ObacDRAFT_4438	Hypothetical protein	2.2	3.18E-05
ObacDRAFT_3535	Hypothetical protein	-2.2	8.73E-07
ObacDRAFT_3815	PAP2 (acid phosphatase) superfamily protein	-3.1	1.00E-14
ObacDRAFT_3816	Response regulators consisting of a CheY-like receiver domain	-2.4	1.80E-10
ObacDRAFT_3817	Signal transduction histidine kinase	-2.6	7.09E-14
ObacDRAFT_4577	Flagellar biosynthesis pathway, component FliR	-2.1	3.55E-05
ObacDRAFT_4579	flagellar biosynthetic protein FliP	-2.5	6.56E-07
ObacDRAFT_4592	flagellar basal-body M-ring protein/flagellar hook-basal body protein (fliF)	-2.1	4.85E-05
ObacDRAFT_4755	3-isopropylmalate dehydratase, small subunit (EC 4.2.1.33)	-2.0	4.76E-05
ObacDRAFT_4865	Hypothetical protein	-2.2	5.75E-05
ObacDRAFT_4866	flagellar biosynthesis protein FlhA	-2.1	0.000119762

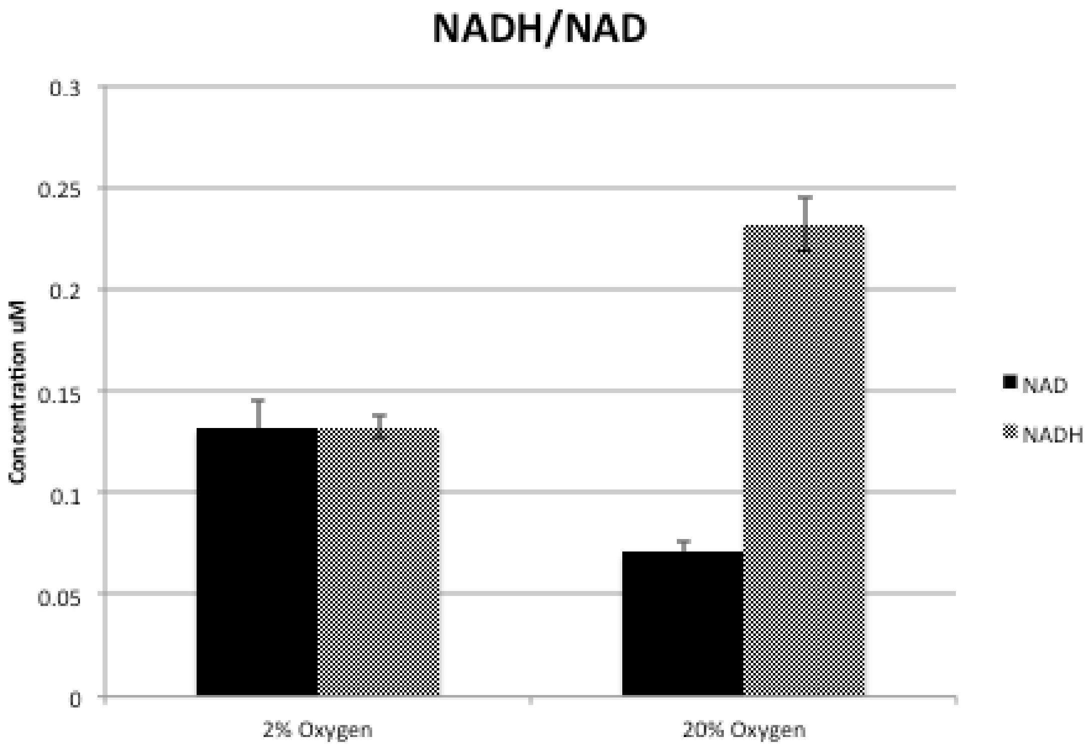


Figure 5.3 NADH and NAD$^+$ concentrations in TAV2 grown at 2% and 20% O_2. The measurements were done in triplicates as mentioned in Materials and Methods. NADH was accumulation was observed in TAV2 cells grown at 20% O_2.

Chapter 6 General Discussion

6.1 Conclusion, discussion and future prospects.

The hindguts of wood-feeding termites harbor a wide variety of microorganisms, which help termites to degrade lignocellulose. Previously, five isolates of an obscure phylum *Verrucomicrobia* were isolated from the hindgut of lower termite *R. flavipes* (1). These Termite Associated *Verrucomicrobia* (TAV strains) were confirmed to be indigenous to the termite hindgut, and *Diplosphaera colitermitum* TAV2 was subsequently used for further studies to gain insights into the ecophysiological roles of these strains. Genomic, physiological, transcriptomic and proteomic analysis carried out on the TAV2 revealed it to be a microaerophile and potentially involved in amino-acid production, nitrogen fixation, hemicellulose digestion and O_2 scavenging in the termite gut (2, 3).

In the present study we ask 1) to what extent the genetic and functional information obtained from TAV2 can be extended to other TAV strains? 2) what are the metabolic changes applied by the TAV strains to adapt to hypoxic environment in the termite gut? 3) Is the metabolic machinery of these microaerophiles evolutionarily adaptive? 4) Do these strains participate in lignocellulose digestion in the termite gut? This chapter provides an overview of important findings of this study whereas in-depth discussion, methodology and results of experiments used to address above-mentioned questions are given in previous chapters.

1) Understanding the TAV strains: core and species-specific genome

Through comparative genomic analysis we conclude that the genotypic similarities in the TAV strains are in agreement with their phylogenetic relatedness i.e. the genomes within a species are more similar than between species. The core genome of TAV strains represents 50% of TAV genes where genes related to microaerophilic respiration are highly conserved, indicating primary role for TAV strains in O_2 scavenging in the termite gut. We found a moderate level of conservation for functional attributes like amino acid production, nitrogen fixation and lignocellulose digestion. For instance, strains TAV1 and TAV4 are missing an important gene in the aromatic amino acid biosynthesis pathway (Chapter 4). These results need to be verified using phenotypic characterizations, but knowledge of such species and strain specific genomes will help us identify unique roles of these strains. Interestingly, species-specific genomes were highly enriched in transcription related genes implying that though TAV genomes are highly similar, they might be regulated differently. Gene expression analysis in TAV2 showed downregulation of a majority of core genes and upregulation of species-specific or unique genes (Supplementary Tables S2 and S3). Therefore, we hypothesize that differential regulation of orthologs and acquisition of new genes are probable mechanism leading to species diversification. Transcriptomic analysis of other TAV strains will help in verifying this hypothesis.

Additionally, through comparative genomics we found genes encoding enzymes for amino-acid production, urea recycling, and nitrogen fixation to be present in all the TAV genomes. However, TAV1 had genes only for a FeFe nitrogenase, suggesting a moderate role for this strain in nitrogen fixation as compared to other TAV strains, which also have genes for MoFe type nitrogenase. Genes for urease alpha subunit and MoFe nitrogenase biogenesis gene *nifE* were found to be undergoing selection indicating a common role of the TAV strains in host nutrition. We did not observe upregulation of genes associated with host nutrition in TAV2 growing in microaerophilic conditions. But our previous study observed upregulation of several amino-acid biosynthesis pathways and also urease expression (2).

2) Metabolic changes to survive in hypoxic environment

The fundamental adaptations of TAV strains to survive in the termite gut are to respire and generate ATP in the hypoxic periphery of its hindgut, but at the same time tolerate atmospheric O_2 concentration to be able to successfully colonize new larvae. Examination of their electron transport chains reveals presence of two types of primary dehydrogenases (NDH-1 and NDH-2) which can be used to fuel the respiratory chain, and two high affinity terminal cytochrome oxidases (cbb_3 type and *bd* type) (Fig 6.1). We also identify an alternative complex (AC) III in the core genome of TAV strains (Fig 6.1), which is possibly used to transfer electrons to quinones because traditional complex III (cytochrome bc_1) is missing in these strains. Investigation of the TAV2 transcriptome shows a preference for NDH-1 over NDH-2 in microaerophilic condition, which can be attributed to the energy conserving and PMF generating nature of NDH-1 (Chapter 5). In addition, the subunits NuoN and NuoB of NDH-1 and subunits *a* and *δ* of F_0F_1 ATPases were identified as undergoing adaptive evolution in the TAV strains. Together, these independent results suggest that the complex 1 and ATPase play important roles in the adaptation of TAV strains to the termite gut environment. We also found a putative NiFe hydrogenase, a V-type pyrophosphatase and a putative quinone:oxidoreductase being differentially regulated in response to oxygen, but their roles in the electron transport chain of the TAV strains remains to be determined. Knowledge about the roles of these proteins will advance our understanding of the microaerophilic physiology of these strains.

Several potential metabolic adjustments are involved in the successful expression of microaerophilic physiology. Current and previous studies show several key enzymes in the TCA cycle like pyruvate dehydrogenase, malate dehydrogenase and succinate dehydrogenase to be differentially regulated in the TAV2 cells exposed to hypoxia as compared to atmospheric O_2. Additionally, the TCA cycle enzymes isocitrate dehydrogenase and succinyl CoA synthetase showed evidence of positive selection. Therefore we hypothesize that the regulation of the TCA cycle is an important metabolic adaptation associated with the microaerophilic physiology of the TAV strains and further investigations will shed light on the details of this adaptation.

The aforementioned observations pose an interesting question, namely how these strains sense the change in O_2 concentration? Investigation of the expressed signal transduction proteins in TAV2 identified two putative O_2 sensing proteins. A PAS domain containing protein (ObacDRAFT_5084) was upregulated in TAV2 under 2% O_2. This protein along with a downstream CDS (ObacDRAFT_5085) was present in the core genome of TAV strains. ObacDRAFT_5085 codes for an Obg family GTPase and was also identified as being positively selected in the TAV strains. Obg is a ribosome dependent GTPase which along with a sensor kinase is predicted to be responsible for the differential regulation of translation under various stress conditions (4). Another such potential O_2 sensing protein was a dramatically downregulated phoR like sensor kinase which was also identified as being positively selected. A possible study to confirm their roles in O_2 sensing will be to create knock out mutants and analyze the microaerophilic physiology of these mutants.

Additionally, we found a large number of gene families undergoing expansion in the TAV genomes (Chapter 4). The TAV2 transcriptome showed upregulation of two such expanded gene families coding for prepilin type N-terminal domain (35 genes) and PEP-CTERM domain (14 genes). A few of these genes were associated with CAZymes, but a majority of these genes were not categorized in any functional categories. These proteins are known to be involved in exopolysaccharide (EPS) secretion and we found a downregulation of motility related genes in the TAV2 transcriptome under hypoxic condition. It is likely that when exposed to hypoxic conditions the TAV2 cells become non-motile and settle near the gut lining using an EPS like structure. However, culture experiments so far with the TAV cells have never demonstrated any sign of such behavior. The specific roles for these proteins and how these proteins are acquired remains to be determined but upregulation of gene families undergoing expansion and species-specific genes shows that probably acquiring new functions or gene family expansion and neofunctionalization plays a very important part in the adaptation of TAV strains to the termite gut.

3) Lignocellulose digestion

Our comparative genomics identified putative cellulose, hemicellulose and host derived polysaccharide-degrading CAZymes in the core genome and species-specific genomes of the TAV strains. Additionally, the genes encoding LacI-family transcription regulators and AraC family transcriptional regulators, associated with regulation of carbohydrate metabolism, were present in multiple copies in all of the genomes (Chapter 4). Furthermore, enzymes like acetyl xylan esterase (AXE), β-1,4-xylanase and β-galactosidase showed positive selection, and a significant number of CAZymes were upregulated in TAV2 under 2% O_2. Together, the gene expression and molecular evolution results suggest an active role for the TAV strains in lignocellulose digestion which needs to be confirmed by showing their cellulase and hemicellulase activity in vitro. Most of the CAZymes studied so far are induced in the presence of a carbohydrate source; therefore it is reasonable to hypothesize that we should see a significant upregulation of CAZymes when the TAV strains are cultured in the presence of complex polysaccharides.

6.2 References:

1. **Stevenson BS**, **Eichorst S a**, **Wertz JT**, **Schmidt TM**, **Breznak J a**. 2004. New strategies for cultivation and detection of previously uncultured microbes. Appl. Environ. Microbiol. **70**:4748–55.
2. **Isanapong J**, **Sealy Hambright W**, **Willis AG**, **Boonmee A**, **Callister SJ**, **Burnum KE**, **Paša-Tolić L**, **Nicora CD**, **Wertz JT**, **Schmidt TM**, **Rodrigues JL**. 2013. Development of an ecophysiological model for Diplosphaera colotermitum TAV2, a termite hindgut Verrucomicrobium. ISME J. **7**:1803–13.
3. **Wertz JT**, **Kim E**, **Breznak JA**, **Schmidt TM**, **Rodrigues JLM**. 2012. Genomic and physiological characterization of the Verrucomicrobia isolate Diplosphaera colitermitum gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. Appl. Environ. Microbiol. **78**:1544–1555.
4. **Feng B**, **Mandava CS**, **Guo Q**, **Wang J**, **Cao W**, **Li N**, **Zhang Y**, **Zhang Y**, **Wang Z**, **Wu J**, **Sanyal S**, **Lei J**, **Gao N**. 2014. Structural and Functional Insights into the Mode of Action of a Universally Conserved Obg GTPase. PLoS Biol. **12**:e1001866.

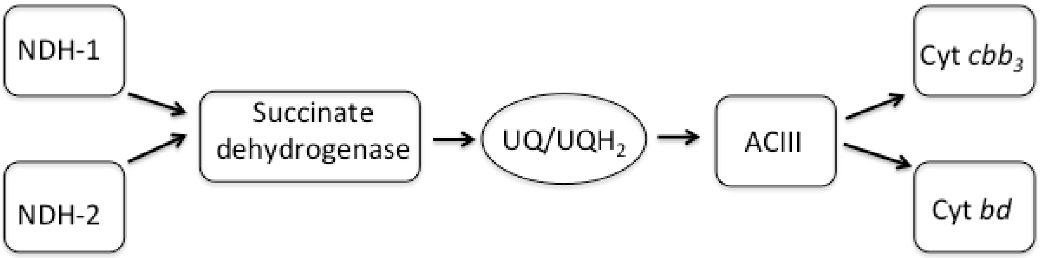


Figure 6.1 Schematic representation of the aerobic respiratory transport system of TAV strains.

Supplementary Tables

Table S 1 Single copy orthologs identified as Core Genome of TAV strains.

Gene IDs							
TAV1	**TAV2**	**TAV3**	**TAV4**	**TAV5**	**Function**	**Positively Selected**	***P*-value**
2508840353	2517754261	2642344682	2642350817	2510268146	biopolymer transport protein ExbB	-	-
2508836350	2517758457	2642343832	2642348577	2510267534	-	-	-
2508837063	2517756258	2642345692	2642350599	2510266737	glycogen synthase (ADP-glucose)	-	-
2508836434	2517753722	2642344029	2642349924	2510267437	pyruvate-ferredoxin/flavodoxin oxidoreductase	-	-
2508835203	2517755267	2642347602	2642351268	2510264480	elongation factor G	-	-
2508836381	2517755077	2642344844	2642351523	2510267482	3,4-dihydroxy 2-butanone 4-phosphate synthase	-	-
2508840140	2517758681	2642344010	2642353037	2510270014	ATP-dependent helicase HepA	-	-
2508837803	2517756645	2642346927	2642352045	2510265964	nicotinamide-nucleotide amidase	-	-
2508836421	2517754955	2642347910	2642349641	2510267450	-	Yes	0.000339711
2508840396	2517753825	2642345468	2642351885	2510268108	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508836266	2517757878	2642345177	2642349253	2510264622	Uncharacterized membrane protein YoaK, UPF0700 family	Yes	3.48625E-11
2508835982	2517758714	2642343314	2642350943	2510267920	hypothetical protein	-	-
2508838287	2517755644	2642347464	2642353106	2510266650	ribonucleoside-diphosphate reductase beta chain	-	-
2508839973	2517756063	2642343066	2642349286	2510269847	-	-	-
2508834662	2517758060	2642344480	2642349996	2510269365	dephospho-CoA kinase	-	-
2508835093	2517757529	2642346150	2642352413	2510267720	-	-	-
2508839747	2517754920	2642347953	2642349600	2510268536	ADP-ribosylglycohydrolase	-	-
2508839462	2517754548	2642343717	2642351183	2510268937	-	Yes	2.8991E-12
2508839079	2517756515	2642346793	2642352178	2510264900	-	-	-
2508837929	2517754558	2642345844	2642348741	2510265887	LacI family transcriptional regulator	-	-
2508839111	2517754940	2642347930	2642349621	2510264866	-	-	-

2508835032	2517757096	2642344578	2642350293	2510267657	membrane fusion protein, multidrug efflux system	Yes	5.90085E-24
2508835242	2517758700	2642343332	2642350960	2510264445	Sugar or nucleoside kinase, ribokinase family	-	-
2508837478	2517757742	2642343147	2642348169	2510266340	LemA protein	-	-
2508836929	2517756856	2642347378	2642352893	2510266909	-	-	-
2508835298	2517756972	2642347233	2642352746	2510264388	-	-	-
2508835730	2517757606	2642346207	2642352023	2510268202	-	-	-
2508838861	2517756868	2642347365	2642352880	2510265083	-	-	-
2508839216	2517755025	2642347840	2642349712	2510264690	MraZ protein	-	-
2508836700	2517758379	2642343450	2642351481	2510267135	putative sigma-54 modulation protein	-	-
2508836667	2517757335	2642346025	2642352540	2510267166	Lysophospholipase L1	-	-
2508836656	2517756591	2642346875	2642352098	2510265013	DNA-binding transcriptional regulator GbsR, MarR family	-	-
2508839488	2517758581	2642344960	2642352939	2510268913	-	-	-
2508835009	2517757883	2642345172	2642349247	2510269054	-	Yes	1.49784E-07
2508835400	2517757385	2642346076	2642352489	2510269470	epoxyqueuosine reductase	-	-
2508836277	2517755168	2642346564	2642351422	2510264634	DNA replication and repair protein RecF	-	-
2508835208	2517755272	2642347597	2642351263	2510264475	large subunit ribosomal protein L2	-	-
2508839402	2517756180	2642343189	2642348286	2510268984	-	Yes	3.94075E-10
2508839285	2517758266	2642347720	2642350119	2510267611	Galactose mutarotase	-	-
2508839002	2517755984	2642344787	2642350320	2510267232	-	-	-
2508837730	2517755799	2642346282	2642350521	2510266056	Predicted arabinose efflux permease, MFS family	-	-
2508837525	2517757983	2642345148	2642349223	2510266291	cation diffusion facilitator family transporter	-	-
2508836933	2517753912	2642345553	2642351801	2510266905	-	-	-
2508839211	2517755031	2642347835	2642349717	2510264695	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	-	-
2508835529	2517754122	2642345361	2642349497	2510269625	fructose-bisphosphate aldolase, class II	-	-
2508837232	2517756495	2642346772	2642352198	2510265676	protease-4	-	-
2508836283	2517754662	2642346446	2642349820	2510264640	Zn-dependent protease (includes SpoIVFB)	-	-
2508838851	2517758380	2642343451	2642351480	2510265094	-	-	-
2508838894	2517755630	2642347477	2642353119	2510265057	-	-	-

2508836291	2517754995	2642347872	2642349680	2510264649	-	-	-
2508835910	2517755924	2642345916	2642350388	2510268003	Threonine dehydrogenase	-	-
2508834746	2517757049	2642343499	2642351123	2510266048	-	-	-
2508835355	2517753899	2642345541	2642351813	2510264325	phosphate acetyltransferase	-	-
2508835360	2517756930	2642347277	2642352790	2510264309	-	-	-
2508835727	2517754443	2642345258	2642348492	2510268205	-	-	-
2508839579	2517754387	2642345319	2642352291	2510268814	carboxyl-terminal processing protease	-	-
2508839624	2517758026	2642344449	2642349964	2510268765	Protein-arginine kinase activator protein McsA	-	-
2508835206	2517755270	2642347599	2642351265	2510264477	large subunit ribosomal protein L4	-	-
2508836233	2517754623	2642346409	2642349858	2510264593	-	-	-
2508838530	2517756543	2642346818	2642352153	2510265392	Uncharacterized membrane protein YdjX	-	-
2508835495	2517754894	2642343581	2642351038	2510269563	-	-	-
2508837920	2517757196	2642347065	2642349906	2510265882	-	-	-
2508837548	2517758082	2642344502	2642350017	2510266266	LacI family transcriptional regulator	-	-
2508834905	2517756508	2642346786	2642352185	2510269165	thioredoxin reductase (NADPH)	-	-
2508835781	2517756987	2642347219	2642352732	2510268045	-	Yes	0.000139174
2508836620	2517758326	2642343349	2642351900	2510264979	Uncharacterized membrane protein YjjP, DUF1212 family	-	-
2508839202	2517755042	2642347824	2642349727	2510264704	cell division protein FtsA	-	-
2508840310	2517756061	2642343071	2642349291	2510270071	-	-	-
2508835202	2517755266	2642347603	2642351269	2510264481	small subunit ribosomal protein S7	-	-
2508838078	2517754644	2642346430	2642349836	2510265841	-	-	-
2508839150	2517755620	2642347487	2642353129	2510264827	ethanolamine utilization protein EutN	-	-
2508840087	2517756021	2642344825	2642351466	2510269962	outer membrane transport energization protein TonB	-	-
2508835752	2517754281	2642344663	2642350837	2510268177	uncharacterized zinc-type alcohol dehydrogenase-like protein	-	-
2508835554	2517756035	2642343098	2642349319	2510269655	ABC-2 type transport system permease protein	-	-
2508838335	2517754126	2642345357	2642349493	2510265471	-	-	-
2508837163	2517757136	2642346649	2642348812	2510265552	23S rRNA (cytidine1920-2'-O)	-	-
2508834939	2517755091	2642346628	2642351361	2510269137	-	-	-
2508837187	2517757134	2642346646	2642348814	2510265585	transcriptional regulator, LacI family	-	-

2508839749	2517755825	2642345964	2642350433	2510268534	-	-	-
2508837002	2517754797	2642344287	2642349426	2510266805	AraC-type DNA-binding protein	-	-
2508835228	2517754691	2642346481	2642349788	2510264455	small subunit ribosomal protein S13	-	-
2508837753	2517758776	2642344617	2642350882	2510266017	-	-	-
2508836222	2517756354	2642345790	2642350701	2510264582	-	-	-
2508837524	2517755299	2642347568	2642351232	2510266292	ribosome biogenesis GTPase	-	-
2508837353	2517758253	2642347729	2642350128	2510266447	NAD+ kinase	-	-
2508837071	2517758768	2642344610	2642350891	2510266727	-	-	-
2508839241	2517757176	2642346693	2642348767	2510267565	Outer membrane protein TolC	Yes	6.03346E-09
2508839020	2517757032	2642347176	2642352686	2510267251	-	-	-
2508838893	2517754540	2642345857	2642348754	2510265058	ABC-2 type transport system ATP-binding protein	-	-
2508840447	2517756363	2642347053	2642350277	2510264125	dCMP deaminase	-	-
2508836373	2517753915	2642345556	2642351798	2510267490	polyphosphate glucokinase	Yes	4.04784E-06
2508835048	2517758460	2642343830	2642348574	2510267675	multidrug efflux pump	-	-
2508835146	2517757083	2642344570	2642350283	2510264528	dihydroorotate oxidase A	-	-
2508837749	2517755807	2642346273	2642350530	2510266028	hypothetical protein	-	-
2508835788	2517756552	2642346827	2642352145	2510268038	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase	-	-
2508838166	2517754221	2642344726	2642350774	2510266521	formate-dependent phosphoribosylglycinamide formyltransferase	-	-
2508839287	2517757644	2642346249	2642351980	2510267613	maltoporin	-	-
2508839151	2517755621	2642347486	2642353128	2510264826	aldehyde dehydrogenase	-	-
2508836072	2517758428	2642343863	2642348601	2510267850	phosphate ABC transporter membrane protein 2, PhoT family	-	-
2508839245	2517758069	2642344490	2642350005	2510267569	myo-inositol-1(or 4)-monophosphatase	-	-
2508836187	2517755211	2642347653	2642351318	2510267739	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508837308	2517756275	2642345717	2642350625	2510266490	transcriptional regulator, LacI family	-	-
2508837035	2517756623	2642346911	2642352062	2510266766	hypothetical protein	-	-
2508836028	2517756547	2642346822	2642352149	2510267882	L-rhamnose mutarotase	-	-
2508840373	2517758353	2642343425	2642351503	2510268125	-	-	-
2508834812	2517757731	2642343136	2642348158	2510269272	DNA-binding transcriptional response regulator, NtrC family	-	-

2508834868	2517756861	2642347372	2642352887	2510269204	Uncharacterized membrane protein YckC, RDD family	-	-
2508836501	2517754594	2642345805	2642348703	2510267368	phosphoribosylaminoimidazole-succinocarboxamide synthase	-	-
2508839506	2517758306	2642343370	2642351921	2510268896	flagellar assembly protein FliH	Yes	6.83278E-15
2508837311	2517758436	2642343855	2642348593	2510266487	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+)	-	-
2508837799	2517757387	2642346078	2642352487	2510265968	hypothetical protein	-	-
2508834867	2517757563	2642346167	2642352396	2510269205	acetyl-CoA carboxylase carboxyl transferase subunit alpha	-	-
2508836238	2517754627	2642346414	2642349853	2510264597	-	Yes	2.54417E-15
2508840180	2517756114	2642343164	2642348311	2510270043	alpha-galactosidase	-	-
2508839939	2517758008	2642343274	2642349191	2510269803	tRNA-2-methylthio-N6-dimethylallyladenosine synthase	Yes	8.87634E-20
2508840571	2517754935	2642347935	2642349616	2510264227	Transglutaminase-like enzyme, putative cysteine protease	-	-
2508840559	2517754880	2642343568	2642351051	2510264215	small subunit ribosomal protein S15	-	-
2508837528	2517754451	2642345250	2642348484	2510266288	UDP-glucuronate 4-epimerase	-	-
2508835167	2517756202	2642345629	2642350540	2510264508	Glucose/arabinose dehydrogenase, beta-propeller fold	-	-
2508839238	2517758438	2642343853	2642348591	2510264668	segregation and condensation protein B	-	-
2508836925	2517755145	2642346582	2642351405	2510266913	LacI family transcriptional regulator	-	-
2508835299	2517756971	2642347234	2642352747	2510264387	Signal transduction histidine kinase	-	-
2508836399	2517755282	2642347586	2642351252	2510267468	DNA polymerase-1	-	-
2508838835	2517756775	2642345195	2642350979	2510265109	DNA repair protein RecN (Recombination protein N)	-	-
2508835804	2517757355	2642346046	2642352519	2510268020	8-amino-7-oxononanoate synthase	-	-
2508836264	2517754754	2642344240	2642349381	2510264620	-	-	-
2508837351	2517756994	2642347212	2642352725	2510266449	-	-	-
2508835595	2517755595	2642347514	2642353155	2510269696	isocitrate dehydrogenase	Yes	1.2051E-29
2508835491	2517754897	2642343584	2642351035	2510269560	transcriptional regulator, LacI family	-	-
2508835600	2517755592	2642345055	2642348920	2510269701	-	-	-
2508840468	2517758089	2642344507	2642350022	2510264148	sulfate transport system ATP-binding protein	-	-
2508839422	2517756521	2642346799	2642352171	2510268962	ATP phosphoribosyltransferase	-	-
2508838993	2517754899	2642343585	2642351034	2510267224	-	-	-
2508836374	2517756757	2642345185	2642350970	2510267489	fructose-1,6-bisphosphatase	-	-

2508837807	2517754837	2642343521	2642351101	2510265960	-	-	-
2508834922	2517758520	2642344325	2642348894	2510269149	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	-	-
2508837835	2517756164	2642343173	2642348302	2510265932	type IV pilus assembly protein PilB	-	-
2508835414	2517757162	2642346681	2642348779	2510269490	-	-	-
2508838896	2517755391	2642348028	2642351641	2510265055	-	-	-
2508839099	2517754491	2642344001	2642352282	2510264885	-	-	-
2508839904	2517757900	2642345154	2642349229	2510268332	NADH-quinone oxidoreductase subunit A	-	-
2508840144	2517756246	2642345679	2642350588	2510270018	Outer membrane receptor proteins, mostly Fe transport	-	-
2508837467	2517758559	2642344984	2642352918	2510266351	-	-	-
2508836087	2517755014	2642347852	2642349700	2510267835	amino acid/amide ABC transporter membrane protein 1, HAAT family	-	-
2508835924	2517758517	2642344328	2642348897	2510267988	UDP-2,3-diacylglucosamine pyrophosphatase LpxH	-	-
2508839584	2517754290	2642344651	2642350848	2510268801	-	-	-
2508837138	2517755076	2642344845	2642351522	2510265515	Spermidine synthase	-	-
2508834969	2517755892	2642345951	2642350420	2510269094	-	-	-
2508838887	2517754350	2642343934	2642348672	2510265064	Predicted dehydrogenase	-	-
2508839219	2517754757	2642344244	2642349384	2510264687	Predicted nucleic acid-binding protein, contains PIN domain	-	-
2508835448	2517756998	2642347208	2642352721	2510269522	-	-	-
2508835765	2517754477	2642343972	2642352253	2510268063	-	-	-
2508839214	2517755027	2642347838	2642349714	2510264692	-	-	-
2508840404	2517754482	2642343992	2642352273	2510268098	-	-	-
2508838823	2517756273	2642345707	2642350614	2510265116	-	-	-
2508835656	2517756294	2642345732	2642350642	2510268256	CRISPR-associated protein Cas2	-	-
2508837227	2517758515	2642344330	2642348899	2510265671	-	-	-
2508835136	2517754369	2642343915	2642348653	2510264538	cytochrome c oxidase accessory protein FixG	-	-
2508835724	2517755792	2642346290	2642350513	2510268208	23S rRNA (adenine2503-C2)-methyltransferase	-	-
2508835121	2517754081	2642345402	2642349541	2510267732	sulfofructose kinase	-	-
2508839152	2517755622	2642347485	2642353127	2510264825	ethanolamine utilization protein EutN	-	-
2508835362	2517756932	2642347275	2642352788	2510264307	general secretion pathway protein G	-	-
2508838418	2517756982	2642347224	2642352737	2510265452	phosphoglycerate kinase	-	-

2508835332	2517758640	2642344906	2642352994	2510264364	-	-	-
2508838074	2517758331	2642343651	2642348328	2510265845	AraC-type DNA-binding protein	Yes	4.60432E-05
2508839157	2517755627	2642347480	2642353122	2510264820	putative phosphotransacetylase	-	-
2508836911	2517754216	2642344732	2642350769	2510266927	phosphoglycerate mutase	-	-
2508839148	2517755618	2642347489	2642353131	2510264829	L-fuculose-phosphate aldolase	-	-
2508835096	2517757338	2642346028	2642352537	2510267722	-	-	-
2508839966	2517753764	2642346966	2642351569	2510269840	nitrogenase molybdenum-iron protein NifN	-	-
2508837246	2517756728	2642345229	2642351011	2510265689	Nucleoside-diphosphate-sugar epimerase	-	-
2508836169	2517757880	2642345175	2642349250	2510267757	-	Yes	7.13869E-13
2508839213	2517755029	2642347837	2642349715	2510264693	cell division protein FtsI (penicillin-binding protein 3)	Yes	9.49949E-06
2508834814	2517753972	2642345121	2642352368	2510269270	hypothetical protein	-	-
2508839911	2517757892	2642345162	2642349237	2510268325	NADH-quinone oxidoreductase subunit K	-	-
2508835556	2517756036	2642343097	2642349318	2510269657	ABC-2 type transport system permease protein	-	-
2508834877	2517756928	2642347279	2642352792	2510269195	Uncharacterized membrane protein YckC, RDD family	-	-
2508837440	2517756012	2642344814	2642351455	2510266381	-	-	-
2508837079	2517755416	2642348001	2642351613	2510266717	Xaa-Pro aminopeptidase	-	-
2508836001	2517757565	2642346169	2642352394	2510267908	histidinol-phosphate aminotransferase	-	-
2508836025	2517758471	2642343819	2642348563	2510267885	3-isopropylmalate/(R)-2-methylmalate dehydratase small subunit	-	-
2508839947	2517753871	2642345512	2642351841	2510269811	DNA-directed RNA polymerase subunit omega	-	-
2508835579	2517757562	2642346166	2642352397	2510269682	Uncharacterized conserved protein YtfP, gamma-glutamylcyclotransferase (GGCT)/AIG2-like family	-	-
2508837349	2517756517	2642346795	2642352175	2510266451	-	-	-
2508839424	2517757040	2642347168	2642352678	2510268960	undecaprenyl diphosphate synthase	-	-
2508836237	2517754626	2642346413	2642349854	2510264596	polysaccharide export outer membrane protein	-	-
2508837251	2517757301	2642345993	2642352572	2510265694	-	Yes	1.97113E-06
2508840430	2517757849	2642347538	2642351202	2510264108	-	-	-
2508839204	2517755040	2642347826	2642349725	2510264702	D-alanine-D-alanine ligase	-	-
2508835766	2517754478	2642343973	2642352254	2510268062	23S rRNA pseudouridine1911/1915/1917 synthase	-	-
2508836372	2517753916	2642345557	2642351797	2510267491	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD	-	-

2508839960	2517753858	2642345499	2642351854	2510269829	Osmotically-inducible protein OsmY, contains BON domain	-	-
2508838854	2517756379	2642347037	2642350260	2510265091	GTP-binding protein LepA	-	-
2508835664	2517756507	2642346785	2642352186	2510268248	5'-methylthioadenosine phosphorylase	-	-
2508838846	2517758668	2642344869	2642353030	2510265099	tRNA (cytidine/uridine-2'-O-)-methyltransferase	-	-
2508837277	2517753913	2642345554	2642351800	2510265718	tellurite resistance protein TerC	-	-
2508835447	2517756997	2642347209	2642352722	2510269521	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	-	-
2508839373	2517758635	2642344912	2642352988	2510269026	multicomponent Na+:H+ antiporter subunit B	-	-
2508837511	2517755687	2642347437	2642353078	2510266305	F-type H+-transporting ATPase subunit epsilon	-	-
2508836565	2517758798	2642344640	2642350860	2510267305	-	-	-
2508837609	2517755392	2642348027	2642351640	2510266215	6-phosphogluconolactonase	-	-
2508840546	2517753909	2642345550	2642351804	2510264211	putative membrane protein	-	-
2508834896	2517755127	2642346598	2642351391	2510269174	-	-	-
2508836412	2517756634	2642346922	2642352050	2510265901	alpha-galactosidase	-	-
2508836492	2517758466	2642343824	2642348568	2510267379	-	-	-
2508836221	2517756365	2642347051	2642350275	2510264581	two-component system, OmpR family	-	-
2508834879	2517757865	2642347522	2642351219	2510269193	DNA replication and repair protein RecR	-	-
2508837174	2517758143	2642347697	2642350094	2510265562	-	-	-
2508837927	2517754556	2642345846	2642348743	2510265885	-	-	-
2508836006	2517755306	2642347562	2642351226	2510267903	lipopolysaccharide export system permease protein	-	-
2508839839	2517758279	2642347706	2642350103	2510268454	-	-	-
2508839504	2517758304	2642343372	2642351923	2510268898	flagellar FliJ protein	-	-
2508836046	2517756587	2642346870	2642352103	2510267870	iron complex transport system permease protein	-	-
2508835929	2517756902	2642347325	2642352839	2510267983	Putative hemolysin	-	-
2508839487	2517758580	2642344961	2642352938	2510268914	RNA polymerase, sigma 28 subunit, SigD/FliA/WhiG	Yes	2.50326E-05
2508836022	2517757990	2642345135	2642349210	2510267888	-	-	-
2508836073	2517758429	2642343862	2642348600	2510267849	phosphate ABC transporter membrane protein 1, PhoT family	-	-
2508840060	2517754917	2642344352	2642348459	2510269924	-	-	-
2508837536	2517755510	2642347547	2642352298	2510266280	mannonate dehydratase	-	-
2508839326	2517758228	2642347816	2642350214	2510269028	4-alpha-glucanotransferase	Yes	1.86658E-29

2508840568	2517756770	2642345191	2642350976	2510264224	preprotein translocase subunit SecA	-	-
2508839262	2517754191	2642344761	2642350740	2510267592	-	-	-
2508836519	2517756535	2642346813	2642352158	2510267351	Endonuclease, Uma2 family (restriction endonuclease fold)	-	-
2508836616	2517755260	2642347609	2642351275	2510264975	Fur family transcriptional regulator, ferric uptake regulator	-	-
2508835606	2517755589	2642345052	2642348923	2510269708	-	Yes	4.75254E-14
2508837183	2517758038	2642344459	2642349974	2510265576	ribonucrease Y	-	-
2508836026	2517758472	2642343818	2642348562	2510267884	3-isopropylmalate/(R)-2-methylmalate dehydratase large subunit	-	-
2508836891	2517756436	2642346756	2642352214	2510266960	putative NAD(P)H quinone oxidoreductase, PIG3 family	-	-
2508839381	2517755481	2642343753	2642348440	2510269017	ATP-dependent DNA helicase RecQ	-	-
2508835162	2517758041	2642344462	2642349977	2510264513	RNA processing exonuclease, beta-lactamase fold, Cft2 family	-	-
2508836123	2517755746	2642346355	2642350452	2510267796	4-O-beta-D-mannosyl-D-glucose phosphorylase	-	-
2508835417	2517757166	2642346684	2642348776	2510269493	general secretion pathway protein G	-	-
2508839933	2517756848	2642347386	2642352901	2510269798	HlyD family secretion protein	-	-
2508837520	2517755697	2642347427	2642353068	2510266296	putative sigma-54 modulation protein	-	-
2508837570	2517758627	2642344920	2642352980	2510266241	3-dehydroquinate dehydratase	-	-
2508840448	2517757362	2642346055	2642352511	2510264126	-	-	-
2508837831	2517755124	2642346601	2642351388	2510265936	DUF971 family protein	-	-
2508839975	2517755104	2642346616	2642351373	2510269849	-	-	-
2508835286	2517757351	2642346042	2642352523	2510264401	transcription elongation factor GreA	-	-
2508836282	2517756947	2642347260	2642352773	2510264639	-	-	-
2508835576	2517758099	2642344517	2642350032	2510269679	-	-	-
2508835567	2517755122	2642346603	2642351386	2510269668	3-deoxy-manno-octulosonate cytidylyltransferase	-	-
2508836347	2517756525	2642346803	2642352168	2510267538	large subunit ribosomal protein L19	-	-
2508836375	2517754730	2642344213	2642349353	2510267488	integrase/recombinase XerC	-	-
2508839411	2517755645	2642347462	2642353104	2510268975	Ubiquinone/menaquinone biosynthesis C-methylase UbiE	-	-
2508839022	2517757030	2642347178	2642352688	2510267253	-	-	-
2508835209	2517755273	2642347596	2642351262	2510264474	small subunit ribosomal protein S19	-	-
2508835271	2517758615	2642344932	2642352967	2510264415	uroporphyrinogen III methyltransferase / synthase	Yes	6.50504E-07
2508834703	2517758789	2642344630	2642350869	2510269326	spermidine/putrescine transport system permease protein	Yes	3.39243E-

11

2508839416	2517757360	2642346052	2642352513	2510268968	hypoxanthine phosphoribosyltransferase	-	-
2508838634	2517755935	2642345905	2642350376	2510265359	TRAP-type mannitol/chloroaromatic compound transport system	-	-
2508835764	2517755601	2642347506	2642353148	2510268064	transcriptional regulator, LacI family	-	-
2508839298	2517758760	2642344602	2642350899	2510269436	general secretion pathway protein G	-	-
2508837029	2517756736	2642345223	2642351005	2510266770	NADH-quinone oxidoreductase subunit L	-	-
2508835510	2517754705	2642346498	2642349772	2510269577	phosphoglucosamine mutase	-	-
2508838325	2517755208	2642347659	2642351324	2510265481	NusA antitermination factor	-	-
2508839289	2517757642	2642346247	2642351982	2510267615	maltose/maltodextrin transport system permease protein	-	-
2508839036	2517758374	2642343445	2642351486	2510264947	DNA-binding transcriptional regulator, LacI/PurR family	Yes	4.8762E-21
2508836168	2517758727	2642343305	2642350935	2510267758	acyl-CoA thioester hydrolase	-	-
2508837483	2517757236	2642347112	2642352619	2510266335	-	-	-
2508836893	2517756438	2642346758	2642352212	2510266958	Predicted arabinose efflux permease, MFS family	-	-
2508835614	2517755581	2642345043	2642348931	2510269717	iron complex outermembrane recepter protein	-	-
2508840570	2517754934	2642347936	2642349615	2510264226	Uncharacterized conserved protein, Alpha-E superfamily	-	-
2508840061	2517754293	2642344648	2642350852	2510269925	-	-	-
2508836059	2517754409	2642345291	2642348523	2510267857	ABC-2 type transport system ATP-binding protein	-	-
2508838206	2517756765	2642345186	2642350971	2510266577	oligopeptide transport system substrate-binding protein	-	-
2508834617	2517758424	2642343867	2642348605	2510269440	aminotransferase	-	-
2508838209	2517756911	2642347293	2642352806	2510266580	sulfite reductase (NADPH) hemoprotein beta-component	-	-
2508839961	2517753857	2642345498	2642351855	2510269830	Kef-type potassium/proton antiporter, CPA2 family	-	-
2508839619	2517758267	2642347719	2642350118	2510268769	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	-	-
2508840475	2517758095	2642344514	2642350029	2510264155	hypothetical protein	-	-
2508836061	2517754411	2642345289	2642348521	2510267855	ABC-type uncharacterized transport system involved in gliding motility	-	-
2508840426	2517754923	2642347949	2642349603	2510268067	-	-	-
2508837751	2517755307	2642347561	2642351225	2510266026	PAS domain S-box-containing protein	-	-
2508838360	2517753754	2642346975	2642351578	2510268388	-	-	-
2508839905	2517757899	2642345155	2642349230	2510268331	NADH-quinone oxidoreductase subunit B	Yes	1.98867E-06
2508836737	2517755171	2642346560	2642351425	2510267084	uncharacterized protein	-	-

2508834687	2517754877	2642343565	2642351054	2510269348	-	-	-
2508834664	2517758056	2642344477	2642349993	2510269363	type IV pilus assembly protein PilB	-	-
2508836085	2517755016	2642347849	2642349703	2510267837	-	-	-
2508840133	2517755147	2642346578	2642351409	2510270007	starch phosphorylase	-	-
2508836572	2517754111	2642345370	2642349507	2510267268	-	-	-
2508835226	2517754693	2642346483	2642349786	2510264457	-	-	-
2508836775	2517757246	2642347121	2642352629	2510267054	beta-glucosidase	-	-
2508838758	2517755563	2642345034	2642348938	2510265176	-	-	-
2508838338	2517755230	2642347635	2642351301	2510265468	-	-	-
2508837030	2517756735	2642345224	2642351006	2510266769	NADH-quinone oxidoreductase subunit M	-	-
2508839702	2517756331	2642345767	2642350678	2510268647	2,4-dienoyl-CoA reductase	-	-
2508837887	2517757431	2642346115	2642352449	2510265855	phosphoribosylglycinamide formyltransferase-1	-	-
2508838994	2517755769	2642346330	2642350476	2510267225	-	-	-
2508837228	2517758516	2642344329	2642348898	2510265672	glucose-6-phosphate 1-dehydrogenase	-	-
2508835190	2517756877	2642347355	2642352869	2510264493	heterodimeric methylmalonyl-CoA mutase large subunit precursor	-	-
2508835450	2517757000	2642347206	2642352719	2510269524	lipopolysaccharide export system ATP-binding protein	-	-
2508834984	2517757138	2642346651	2642348810	2510269079	unsaturated rhamnogalacturonyl hydrolase	-	-
2508839790	2517754767	2642344256	2642349395	2510268495	3-oxoacyl-[acyl-carrier protein] reductase	Yes	5.53116E-16
2508836435	2517756340	2642345775	2642350686	2510267423	glycerol kinase	-	-
2508840154	2517754448	2642345253	2642348487	2510270024	3-oxoacyl-[acyl-carrier-protein] synthase II	-	-
2508837796	2517755390	2642348029	2642351642	2510265971	hypothetical protein	-	-
2508839945	2517753873	2642345514	2642351839	2510269809	large subunit ribosomal protein L35	-	-
2508839475	2517758489	2642344435	2642349951	2510268926	flagellar P-ring protein precursor FlgI	-	-
2508836419	2517756377	2642347040	2642350263	2510267452	5-methylcytosine-specific restriction endonuclease McrA	-	-
2508839731	2517757975	2642343119	2642348140	2510268542	CRP/FNR family transcriptional regulator, anaerobic regulatory protein	-	-
2508837680	2517753997	2642346728	2642349444	2510266106	-	-	-
2508838716	2517753751	2642346978	2642351581	2510265286	ArsR family transcriptional regulator	-	-
2508839471	2517758485	2642344440	2642349955	2510268930	flagellar basal-body rod protein FlgG	-	-

2508835590	2517757262	2642347135	2642352645	2510269691	phospholipid/cholesterol/gamma-HCH transport system ATP-binding protein	Yes	4.84981E-16
2508839482	2517758496	2642344428	2642349943	2510268919	-	-	-
2508834634	2517758465	2642343825	2642348569	2510269422	integration host factor subunit alpha	-	-
2508837537	2517755509	2642347546	2642352299	2510266279	transcriptional regulator, IclR family	-	-
2508839321	2517757100	2642344583	2642350297	2510269033	3-dehydroquinate synthase	-	-
2508836305	2517756400	2642347027	2642350250	2510264658	3-oxoacyl-[acyl-carrier-protein] synthase-3	-	-
2508835197	2517756885	2642347346	2642352859	2510264486	oxaloacetate decarboxylase, beta subunit	-	-
2508839270	2517757292	2642345985	2642352580	2510267600	-	-	-
2508835643	2517756280	2642345722	2642350631	2510269747	Predicted DNA-binding transcriptional regulator YafY	-	-
2508839275	2517754551	2642345850	2642348747	2510267605	dihydrolipoamide dehydrogenase	-	-
2508836289	2517754993	2642347874	2642349678	2510264647	-	-	-
2508839710	2517758595	2642344946	2642352953	2510268637	-	-	-
2508840332	2517758531	2642344308	2642348877	2510270086	putative holliday junction resolvase	-	-
2508838631	2517755932	2642345908	2642350380	2510265362	Ethanolamine ammonia-lyase light chain	-	-
2508835566	2517755123	2642346602	2642351387	2510269667	phosphocarrier protein	-	-
2508835467	2517756193	2642343200	2642348276	2510269538	-	-	-
2508836500	2517754593	2642345806	2642348704	2510267369	acetylornithine/N-succinyldiaminopimelate aminotransferase	-	-
2508836900	2517756622	2642346910	2642352063	2510266939	methylated-DNA-[protein]-cysteine S-methyltransferase	-	-
2508834665	2517758055	2642344476	2642349992	2510269362	twitching motility protein PilT	-	-
2508836180	2517758328	2642343655	2642348331	2510267746	methionyl-tRNA synthetase	-	-
2508839409	2517756406	2642347016	2642350238	2510268977	Cephalosporin-C deacetylase	-	-
2508834817	2517757259	2642347133	2642352642	2510269267	-	-	-
2508839719	2517757971	2642343122	2642348144	2510268554	-	-	-
2508835149	2517756341	2642345776	2642350687	2510264525	secondary thiamine-phosphate synthase enzyme	-	-
2508839274	2517754550	2642345851	2642348748	2510267604	2-oxoglutarate dehydrogenase E2 component	-	-
2508840450	2517758456	2642343833	2642348578	2510264128	orotate phosphoribosyltransferase	-	-
2508836662	2517758108	2642344526	2642350041	2510265018	-	-	-
2508839703	2517756420	2642347001	2642350222	2510268646	cystathione beta-lyase	-	-
2508839576	2517758518	2642344327	2642348896	2510268817	processive 1,2-diacylglycerol beta-glucosyltransferase	-	-

2508834701	2517758787	2642344628	2642350871	2510269328	3-oxoacyl-[acyl-carrier protein] reductase	-	-
2508836982	2517757336	2642346026	2642352539	2510266824	serine/threonine protein kinase	-	-
2508838886	2517754155	2642345332	2642349469	2510265065	-	Yes	1.10749E-05
2508837466	2517758558	2642347960	2642348182	2510266352	-	Yes	5.97376E-33
2508840315	2517758716	2642343312	2642350941	2510270075	glutamate-1-semialdehyde 2,1-aminomutase	-	-
2508837516	2517755693	2642347431	2642353072	2510266300	F-type H+-transporting ATPase subunit b	-	-
2508835989	2517756549	2642346823	2642352148	2510267913	1,4-dihydroxy-2-naphthoate prenyltransferase	-	-
2508836157	2517758246	2642347731	2642350130	2510267770	-	-	-
2508838964	2517755532	2642345004	2642348968	2510267193	penicillin-binding protein 2	-	-
2508840168	2517758745	2642344589	2642350914	2510270036	transcriptional regulator, BadM/Rrf2 family	-	-
2508840181	2517754392	2642345313	2642348544	2510270044	YhcH/YjgK/YiaL family protein	-	-
2508836471	2517757797	2642343607	2642350726	2510267399	Glyoxylase, beta-lactamase superfamily II	-	-
2508835331	2517758639	2642344907	2642352993	2510264365	hypothetical protein	-	-
2508834972	2517754733	2642344217	2642349357	2510269091	-	-	-
2508839084	2517755711	2642347413	2642353053	2510264895	acetolactate synthase, large subunit	-	-
2508835605	2517755590	2642345053	2642348922	2510269707	LacI family transcriptional regulator	-	-
2508836290	2517754994	2642347873	2642349679	2510264648	-	-	-
2508835069	2517754936	2642347934	2642349617	2510267695	membrane-associated protein	-	-
2508839844	2517758273	2642347712	2642350109	2510268449	prokaryotic molybdopterin-containing oxidoreductase family	-	-
2508835987	2517757195	2642347064	2642349905	2510267915	-	-	-
2508839109	2517754587	2642345812	2642348710	2510264868	glyoxylase I family protein	-	-
2508835119	2517754083	2642345400	2642349539	2510267730	-	-	-
2508840172	2517758576	2642344965	2642352934	2510270040	DNA gyrase subunit A	-	-
2508838477	2517755235	2642347630	2642351296	2510265446	hypothetical protein	-	-
2508837254	2517757298	2642345990	2642352575	2510265697	-	-	-
2508839609	2517755187	2642346544	2642351442	2510268779	Predicted dehydrogenase	-	-
2508838856	2517756863	2642347370	2642352885	2510265088	-	-	-
2508840066	2517758703	2642343329	2642350957	2510269930	-	-	-
2508835659	2517756418	2642347003	2642350224	2510268253	-	-	-

2508836605	2517755917	2642345923	2642350395	2510264964	large subunit ribosomal protein L7/L12	-	-
2508838982	2517756943	2642347264	2642352777	2510267212	-	-	-
2508839573	2517757367	2642346058	2642352508	2510268820	solute:Na+ symporter, SSS family	-	-
2508836899	2517756621	2642346909	2642352064	2510266940	AraC family transcriptional regulator, regulatory protein of adaptative response	-	-
2508839644	2517754094	2642345389	2642349527	2510268739	DNA polymerase-4	-	-
2508835047	2517758461	2642343829	2642348573	2510267674	membrane fusion protein, multidrug efflux system	-	-
2508835273	2517758622	2642344925	2642352975	2510264412	-	-	-
2508839235	2517754981	2642347884	2642349668	2510264671	-	-	-
2508838904	2517754776	2642344266	2642349405	2510265047	Predicted thiol-disulfide oxidoreductase YuxK, DCC family	-	-
2508839513	2517758313	2642343363	2642351913	2510268889	two-component system, chemotaxis family, response regulator CheB	Yes	5.32482E-14
2508840383	2517755406	2642348012	2642351624	2510268120	Uncharacterized conserved protein, DUF1800 family	-	-
2508839474	2517758488	2642344436	2642349952	2510268927	flagellar L-ring protein precursor FlgH	Yes	7.43069E-08
2508835406	2517758659	2642344888	2642353012	2510269476	Adenine deaminase	-	-
2508835963	2517758360	2642343434	2642351496	2510267947	Transglutaminase-like enzyme, putative cysteine protease	-	-
2508840293	2517754983	2642347882	2642349670	2510270056	nucleoside diphosphate kinase	-	-
2508839983	2517754051	2642345424	2642349562	2510269864	-	-	-
2508838157	2517757415	2642346105	2642352460	2510266516	-	-	-
2508839423	2517756522	2642346800	2642352170	2510268961	-	-	-
2508839000	2517757221	2642347090	2642352597	2510267230	-	-	-
2508836170	2517755679	2642347440	2642353081	2510267756	hypothetical protein	-	-
2508835743	2517753729	2642344036	2642349932	2510268186	-	-	-
2508836972	2517757052	2642345634	2642350545	2510266844	LacI family transcriptional regulator	-	-
2508839932	2517754564	2642345837	2642348735	2510268314	urease accessory protein	-	-
2508840558	2517754881	2642343569	2642351050	2510264214	polyribonucleotide nucleotidyltransferase	-	-
2508838937	2517758348	2642343420	2642351508	2510267173	NAD(P)-dependent dehydrogenase	-	-
2508835244	2517757986	2642345140	2642349215	2510264443	-	-	-
2508839185	2517754679	2642346468	2642349800	2510264712	-	-	-
2508838269	2517757186	2642346706	2642348757	2510266632	-	-	-

2508835699	2517758064	2642344484	2642350000	2510268218	Aryl-phospho-beta-D-glucosidase BglC, GH1 family	-	-
2508837018	2517756048	2642343084	2642349304	2510266781	-	-	-
2508839421	2517756520	2642346798	2642352172	2510268963	N-carbamoylputrescine amidase	-	-
2508835662	2517755798	2642346283	2642350520	2510268250	conserved hypothetical protein	Yes	3.13996E-05
2508837936	2517755822	2642345967	2642350435	2510265898	-	-	-
2508835135	2517755712	2642347412	2642353052	2510264540	tyrosyl-tRNA synthetase	-	-
2508837949	2517756627	2642346915	2642352058	2510265908	-	-	-
2508836245	2517755677	2642347442	2642353083	2510264604	-	-	-
2508835498	2517754890	2642343139	2642348161	2510269566	-	-	-
2508837485	2517754783	2642344273	2642349413	2510266329	transcriptional regulator, LacI family	-	-
2508836088	2517755013	2642347853	2642349699	2510267834	amino acid/amide ABC transporter membrane protein 2, HAAT family	-	-
2508838772	2517755642	2642347466	2642353108	2510265166	Predicted arabinose efflux permease, MFS family	-	-
2508836970	2517756723	2642345235	2642351018	2510266846	-	-	-
2508839449	2517754508	2642343706	2642351171	2510268950	flagellin	-	-
2508835826	2517754025	2642345450	2642349588	2510267287	-	-	-
2508838661	2517755901	2642345939	2642350411	2510265333	Uncharacterized conserved protein YcfL	-	-
2508836773	2517758375	2642343446	2642351485	2510267056	-	-	-
2508840358	2517758129	2642344560	2642350075	2510268140	-	-	-
2508840362	2517758052	2642344473	2642349988	2510268136	-	-	-
2508836614	2517753901	2642345543	2642351811	2510264973	-	-	-
2508840062	2517756556	2642346831	2642352141	2510269926	Phytoene dehydrogenase-related protein	-	-
2508835191	2517756878	2642347353	2642352867	2510264492	heterodimeric methylmalonyl-CoA mutase small subunit	-	-
2508836392	2517755964	2642345873	2642350342	2510267475	ABC-type transport system involved in multi-copper enzyme maturation	-	-
2508837603	2517757240	2642347115	2642352622	2510266221	glutamate synthase (NADPH/NADH) small chain	-	-
2508838838	2517756939	2642347268	2642352781	2510265106	transcriptional regulator, TraR/DksA family	Yes	3.35754E-07
2508836949	2517757653	2642346266	2642351963	2510266871	lipoprotein-releasing system permease protein	-	-
2508837166	2517756546	2642346821	2642352150	2510265555	Fucose permease	-	-
2508837829	2517755925	2642345915	2642350387	2510265938	-	Yes	2.60283E-15

2508840480	2517756059	2642343073	2642349293	2510264160	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508838421	2517755517	2642344988	2642348984	2510265449	rod shape determining protein RodA	-	-
2508835426	2517755097	2642346622	2642351367	2510269502	-	-	-
2508837405	2517756496	2642346773	2642352197	2510266397	-	-	-
2508839145	2517755974	2642345863	2642350332	2510264836	-	-	-
2508836071	2517758427	2642343864	2642348602	2510267851	phosphate transport system ATP-binding protein	-	-
2508835805	2517754642	2642346428	2642349838	2510268019	phenylacetate-CoA ligase	-	-
2508836505	2517754653	2642346438	2642349828	2510267364	AraC family transcriptional regulator, cel operon repressor	-	-
2508837119	2517756725	2642345232	2642351014	2510266676	VanZ like family protein	-	-
2508836035	2517754848	2642343532	2642351090	2510267875	peptidyl-tRNA hydrolase, PTH1 family	-	-
2508838959	2517758538	2642347981	2642348204	2510267188	general secretion pathway protein D	-	-
2508834947	2517758701	2642343331	2642350959	2510269121	ribulokinase	-	-
2508835056	2517754202	2642344749	2642350752	2510267683	-	-	-
2508837737	2517757347	2642346038	2642352528	2510266040	Acetyl esterase/lipase	-	-
2508839580	2517754388	2642345318	2642352292	2510268813	SAM-dependent methyltransferase, MidA family	-	-
2508837620	2517755203	2642347662	2642351327	2510266192	-	-	-
2508838168	2517753881	2642345522	2642351831	2510266523	Trehalose utilization protein	-	-
2508834880	2517757190	2642347059	2642349900	2510269192	adenylate kinase	-	-
2508836889	2517756433	2642346753	2642352217	2510266962	3-oxoacyl-[acyl-carrier protein] reductase	-	-
2508835674	2517753968	2642345117	2642352363	2510268238	16S rRNA (adenine1518-N6/adenine1519-N6)-dimethyltransferase	-	-
2508836366	2517753987	2642345126	2642348269	2510267522	-	-	-
2508836469	2517756731	2642345226	2642351008	2510267401	RNase HII	-	-
2508834656	2517754132	2642345351	2642349487	2510269371	succinyl-CoA synthetase alpha subunit	-	-
2508836306	2517756399	2642347028	2642350251	2510264659	phosphate:acyl-[acyl carrier protein] acyltransferase	-	-
2508837905	2517754091	2642345392	2642349531	2510265870	Regulatory P domain of the subtilisin-like proprotein convertase	-	-
2508839824	2517754137	2642345346	2642349482	2510268472	-	-	-
2508836345	2517756527	2642346805	2642352166	2510267540	small subunit ribosomal protein S16	-	-
2508837279	2517755344	2642348079	2642351690	2510265720	AraC-type DNA-binding protein	-	-
2508835640	2517754926	2642347946	2642349606	2510269737	fructose-1-phosphate kinase	-	-

2508838480	2517755148	2642346577	2642351410	2510265443	serine O-acetyltransferase	-	-
2508840355	2517755005	2642347861	2642349691	2510268143	-	-	-
2508836281	2517754363	2642343920	2642348658	2510264638	transporter, UIT6 family	-	-
2508835153	2517754859	2642343544	2642351077	2510264521	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	-	-
2508839401	2517756181	2642343190	2642348285	2510268985	-	-	-
2508839472	2517758486	2642344439	2642349954	2510268929	flagellar basal-body rod protein FlgG	-	-
2508840312	2517754970	2642347895	2642349657	2510270072	Pyridoxamine 5'-phosphate oxidase	-	-
2508840542	2517758660	2642344887	2642353013	2510264207	Fur family transcriptional regulator, ferric uptake regulator	-	-
2508834832	2517758240	2642347738	2642350136	2510269257	two-component system, NarL family, sensor kinase	-	-
2508839606	2517754255	2642344690	2642350808	2510268782	-	-	-
2508835611	2517755584	2642345047	2642348928	2510269713	-	-	-
2508835728	2517757608	2642346209	2642352021	2510268204	sulfonate transport system substrate-binding protein	-	-
2508838361	2517753753	2642346976	2642351579	2510268387	-	-	-
2508839410	2517757172	2642346689	2642348771	2510268976	-	Yes	9.6783E-14
2508837747	2517758801	2642344643	2642350857	2510266030	Sugar phosphate isomerase/epimerase	-	-
2508839286	2517757645	2642346250	2642351979	2510267612	AraC-type DNA-binding protein	-	-
2508838683	2517754721	2642346512	2642349758	2510265316	arginine decarboxylase	-	-
2508836192	2517754063	2642345411	2642349550	2510267734	AraC-type DNA-binding protein	-	-
2508835229	2517754690	2642346480	2642349789	2510264454	small subunit ribosomal protein S11	-	-
2508837249	2517757303	2642345995	2642352570	2510265692	prepilin-type N-terminal cleavage/methylation domain-containing protein	-	-
2508838708	2517754910	2642344360	2642348452	2510265294	superoxide dismutase, Fe-Mn family	-	-
2508837533	2517754964	2642347901	2642349651	2510266283	nicotinate-nucleotide adenylyltransferase	-	-
2508840333	2517758532	2642344307	2642348876	2510270087	mannose-6-phosphate isomerase, type 2	-	-
2508839856	2517757173	2642346690	2642348770	2510268337	Hemolysin, contains CBS domains	-	-
2508835795	2517755364	2642348055	2642351666	2510268031	solute:Na+ symporter, SSS family	-	-
2508838901	2517758359	2642343433	2642351497	2510265050	-	-	-
2508839049	2517756262	2642345696	2642350603	2510264934	Sugar phosphate isomerase/epimerase	-	-
2508839447	2517758340	2642343416	2642351512	2510268952	HD-like signal output (HDOD) domain, no enzymatic activity	-	-
2508836934	2517756529	2642346807	2642352164	2510266904	spermidine/putrescine transport system permease protein	-	-

2508835214	2517755278	2642347591	2642351257	2510264469	large subunit ribosomal protein L29	-	-
2508837200	2517756260	2642345694	2642350601	2510265594	-	-	-
2508838926	2517756056	2642343076	2642349296	2510265025	ATP-binding cassette, subfamily B, MsbA	-	-
2508838690	2517758684	2642344864	2642348844	2510265309	monovalent cation:H+ antiporter-2, CPA2 family	-	-
2508835701	2517758066	2642344487	2642350002	2510268216	-	-	-
2508835434	2517753954	2642345108	2642352353	2510269510	DNA polymerase elongation subunit (family B)	Yes	6.25094E-13
2508835480	2517757873	2642344380	2642348866	2510269549	cyclohexadieny/prephenate dehydrogenase	-	-
2508838476	2517757526	2642346147	2642352416	2510265447	-	-	-
2508838707	2517758750	2642344593	2642350910	2510265295	alcohol dehydrogenase	-	-
2508840369	2517758350	2642343422	2642351506	2510268128	transcriptional regulator, LacI family	-	-
2508835750	2517757592	2642346192	2642352037	2510268179	-	Yes	1.07282E-05
2508836711	2517755023	2642347842	2642349710	2510267121	Predicted dehydrogenase	-	-
2508840470	2517758091	2642344509	2642350024	2510264150	sulfate transport system permease protein	-	-
2508837193	2517757189	2642347058	2642349899	2510265591	-	-	-
2508838958	2517758537	2642347982	2642348205	2510267187	general secretion pathway protein E	-	-
2508838714	2517753748	2642346980	2642351583	2510265288	arsenate reductase	-	-
2508836655	2517756592	2642346876	2642352097	2510265012	UDP-glucuronate decarboxylase	-	-
2508839210	2517755032	2642347833	2642349718	2510264696	Phospho-N-acetylmuramoyl-pentapeptide-transferase	-	-
2508839440	2517754359	2642343924	2642348662	2510268954	PAS domain S-box-containing protein	-	-
2508839625	2517755167	2642346565	2642351421	2510268764	branched-chain amino acid aminotransferase	-	-
2508834965	2517756183	2642343192	2642348283	2510269098	-	-	-
2508838849	2517756889	2642347341	2642352855	2510265096	adenylosuccinate lyase	-	-
2508835735	2517757601	2642346202	2642352028	2510268197	Benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit	-	-
2508835588	2517757264	2642347137	2642352647	2510269689	phospholipid transport system substrate-binding protein	-	-
2508837892	2517754621	2642346407	2642349860	2510265860	acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	-	-
2508835700	2517758065	2642344486	2642350001	2510268217	Lysophospholipase L1	-	-
2508836003	2517755358	2642348061	2642351672	2510267906	pyruvate formate lyase activating enzyme	-	-
2508837510	2517758597	2642344943	2642352956	2510266306	Uncharacterized conserved protein, contains tandem ACT domains	-	-

2508840371	2517758352	2642343424	2642351504	2510268126	-	Yes	4.0776E-10
2508837212	2517756565	2642346846	2642352126	2510265603	-	-	-
2508836964	2517756721	2642345237	2642351020	2510266855	LacI family transcriptional regulator	-	-
2508836689	2517757991	2642345134	2642349209	2510267149	-	-	-
2508837371	2517756620	2642346908	2642352065	2510266430	SSU ribosomal protein S12P methylthiotransferase	-	-
2508836054	2517754950	2642347916	2642349635	2510267862	cell division protease FtsH	-	-
2508834759	2517755146	2642346581	2642351406	2510269283	-	-	-
2508837800	2517757386	2642346077	2642352488	2510265967	Predicted dehydrogenase	-	-
2508839451	2517754510	2642343708	2642351173	2510268948	-	-	-
2508840382	2517755405	2642348013	2642351625	2510268121	Uncharacterized conserved protein, DUF1501 family	-	-
2508835469	2517756186	2642343195	2642348280	2510269540	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508838814	2517755886	2642345957	2642350426	2510265124	-	-	-
2508838950	2517757125	2642343683	2642348263	2510267179	Nicotinamidase-related amidase	-	-
2508835212	2517755276	2642347593	2642351259	2510264471	small subunit ribosomal protein S3	-	-
2508836626	2517758319	2642343354	2642351905	2510264985	vitamin B12 transporter	-	-
2508837472	2517758565	2642344978	2642352924	2510266346	mannobiose 2-epimerase	-	-
2508835783	2517754755	2642344241	2642349382	2510268043	anthranilate synthase component 1	Yes	1.36689E-13
2508840054	2517754534	2642343229	2642349009	2510269918	-	-	-
2508839969	2517753761	2642346969	2642351572	2510269843	(2Fe-2S) ferredoxin	-	-
2508836185	2517757210	2642347076	2642349918	2510267741	-	-	-
2508839154	2517755624	2642347483	2642353125	2510264823	acetate kinase	-	-
2508835218	2517754701	2642346491	2642349778	2510264465	large subunit ribosomal protein L5	-	-
2508835294	2517754669	2642346454	2642349812	2510264392	Predicted nucleic acid-binding protein, contains PIN domain	-	-
2508836798	2517754804	2642343467	2642351155	2510267040	-	-	-
2508838826	2517754962	2642347903	2642349648	2510265113	hypothetical protein	Yes	1.4486E-25
2508840336	2517754344	2642345976	2642348217	2510270090	2-aminoadipate transaminase	-	-
2508836262	2517757283	2642345979	2642348214	2510264618	Glycosyltransferase involved in cell wall bisynthesis	-	-
2508836204	2517754958	2642347907	2642349644	2510264564	transcription-repair coupling factor (superfamily II helicase)	-	-
2508836603	2517755919	2642345921	2642350393	2510264962	DNA-directed RNA polymerase subunit beta'	-	-

2508837112	2517754612	2642346400	2642349867	2510266683	-	-	-
2508839964	2517756312	2642345750	2642350660	2510269833	Glycosyltransferase involved in cell wall bisynthesis	-	-
2508835698	2517758063	2642344483	2642349999	2510268219	-	-	-
2508840561	2517753835	2642345478	2642351875	2510264217	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family	-	-
2508834830	2517757252	2642347127	2642352636	2510269259	DNA-binding transcriptional regulator GbsR, MarR family	-	-
2508838302	2517755221	2642347644	2642351310	2510265502	tRNA pseudouridine55 synthase	-	-
2508835333	2517758641	2642344905	2642352995	2510264363	nitrogenase molybdenum-iron protein beta chain	-	-
2508836716	2517755131	2642346592	2642351396	2510267116	-	Yes	4.4291E-20
2508836198	2517756849	2642347385	2642352900	2510264558	ATP-dependent RNA helicase RhlE	-	-
2508840402	2517754480	2642343990	2642352271	2510268102	-	-	-
2508836557	2517758554	2642347970	2642348192	2510267314	Phosphoglycerate dehydrogenase	-	-
2508839841	2517758277	2642347708	2642350105	2510268452	Cytochrome c, mono- and diheme variants	-	-
2508839709	2517758339	2642343415	2642351513	2510268640	-	-	-
2508836394	2517755966	2642345871	2642350340	2510267473	acylphosphatase	-	-
2508837830	2517755057	2642344177	2642348378	2510265937	Holliday junction DNA helicase subunit RuvB	-	-
2508838100	2517755713	2642347411	2642353051	2510265822	dihydroxy-acid dehydratase	-	-
2508840425	2517754922	2642347950	2642349602	2510268068	hypothetical protein	-	-
2508838657	2517755903	2642345937	2642350409	2510265337	-	-	-
2508836244	2517754723	2642346514	2642349756	2510264603	pyrophosphate--fructose-6-phosphate 1-phosphotransferase	-	-
2508839515	2517754166	2642344768	2642350735	2510268887	HD-like signal output (HDOD) domain, no enzymatic activity	-	-
2508835733	2517757603	2642346204	2642352026	2510268199	putative ABC transport system ATP-binding protein	-	-
2508838984	2517756941	2642347266	2642352779	2510267214	-	-	-
2508840557	2517754541	2642345856	2642348753	2510264213	prepilin-type N-terminal cleavage	-	-
2508839226	2517753727	2642344035	2642349931	2510264680	DNA mismatch repair protein MutS	-	-
2508835578	2517757248	2642347123	2642352631	2510269681	pyruvate kinase	-	-
2508835642	2517757394	2642346085	2642352480	2510269746	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase	-	-
2508835415	2517757163	2642346682	2642348778	2510269491	-	-	-
2508834720	2517756314	2642345752	2642350662	2510269321	transcriptional regulator, LacI family	-	-
2508834990	2517755494	2642343740	2642349121	2510269073	Fur family transcriptional regulator, peroxide stress response regulator	-	-

2508840150	2517756245	2642345678	2642350587	2510270019	ATP-dependent RNA helicase RhlE	-	-
2508837699	2517757200	2642347068	2642349909	2510266084	acetyl-CoA carboxylase, biotin carboxylase subunit	-	-
2508837908	2517756428	2642346748	2642352222	2510265873	homoserine O-acetyltransferase	-	-
2508839078	2517756514	2642346792	2642352179	2510264901	23S rRNA (cytosine1962-C5)-methyltransferase	-	-
2508834813	2517753969	2642345118	2642352364	2510269271	-	-	-
2508837430	2517754356	2642343927	2642348665	2510264133	DNA-binding transcriptional regulator, LysR family	-	-
2508838075	2517755540	2642345014	2642348958	2510265844	2-dehydro-3-deoxyphosphooctonate aldolase (KDO 8-P synthase)	-	-
2508835978	2517754959	2642347906	2642349645	2510267924	Transcription antitermination factor NusG	-	-
2508834869	2517756860	2642347373	2642352888	2510269203	Uncharacterized membrane protein SpoIIM, required for sporulation	-	-
2508839509	2517758309	2642343367	2642351918	2510268893	flagellar hook-basal body complex protein FliE	-	-
2508840350	2517754264	2642344679	2642350820	2510268149	outer membrane transport energization protein TonB	Yes	1.41725E-07
2508834667	2517757547	2642346160	2642352403	2510269360	type IV pilus assembly protein PilC	-	-
2508837607	2517757237	2642347113	2642352620	2510266217	monothiol glutaredoxin	-	-
2508836342	2517758589	2642344952	2642352947	2510267543	serine protease Do	-	-
2508836844	2517756328	2642345764	2642350674	2510267001	-	-	-
2508837354	2517758254	2642347728	2642350127	2510266446	tRNA-specific 2-thiouridylase	-	-
2508836663	2517754706	2642346499	2642349771	2510267170	-	Yes	8.29814E-16
2508840182	2517754393	2642345311	2642348543	2510270045	23S rRNA (adenine2030-N6)-methyltransferase	-	-
2508836812	2517758469	2642343821	2642348565	2510267030	Predicted dehydrogenase	-	-
2508839829	2517754685	2642346474	2642349794	2510268467	carboxyl-terminal processing protease	Yes	6.19658E-35
2508836091	2517755010	2642347856	2642349696	2510267831	urease subunit gamma	-	-
2508840189	2517758230	2642347814	2642350212	2510270052	KUP system potassium uptake protein	-	-
2508837198	2517756352	2642345788	2642350699	2510265592	-	-	-
2508837917	2517755809	2642346271	2642350532	2510265880	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	-	-
2508835464	2517757424	2642346108	2642352457	2510269535	Di- and tricarboxylate transporter	-	-
2508836821	2517757107	2642343663	2642348245	2510267020	-	-	-
2508840076	2517754110	2642345371	2642349508	2510269944	ribose-phosphate pyrophosphokinase	-	-
2508840433	2517755081	2642344839	2642351528	2510264111	L-ascorbate metabolism protein UlaG, beta-lactamase superfamily	-	-

2508835539	2517758507	2642344339	2642348907	2510269637	L-glyceraldehyde 3-phosphate reductase	-	-
2508837775	2517755084	2642346636	2642351353	2510265991	chloramphenicol-sensitive protein RarD	Yes	2.3706E-06
2508835582	2517754215	2642344743	2642350758	2510269683	-	-	-
2508837832	2517755125	2642346600	2642351389	2510265935	phosphoribosyl-AMP cyclohydrolase	-	-
2508839728	2517757973	2642343121	2642348142	2510268545	-	-	-
2508837535	2517755511	2642347548	2642352297	2510266281	NAD(P)-dependent dehydrogenase	-	-
2508840493	2517753951	2642345104	2642352349	2510264172	-	-	-
2508838483	2517756077	2642343054	2642349273	2510265440	-	-	-
2508837419	2517754823	2642343493	2642351129	2510266393	23S rRNA (adenine2503-C2)-methyltransferase	-	-
2508839096	2517754494	2642344004	2642352285	2510264888	-	-	-
2508837682	2517756067	2642343063	2642349283	2510266104	-	-	-
2508838285	2517755288	2642347578	2642351243	2510266648	octaprenyl-diphosphate synthase	Yes	5.63522E-07
2508838502	2517756011	2642344813	2642351454	2510265420	transcriptional regulator, LacI family	-	-
2508836009	2517757207	2642347074	2642349916	2510267900	DNA-binding transcriptional regulator, FrmR family	-	-
2508836276	2517754108	2642345373	2642349511	2510264633	glutamine synthetase	-	-
2508836854	2517754826	2642343508	2642351114	2510266991	3-oxoacyl-[acyl-carrier-protein] synthase-1	Yes	0.000146547
2508838888	2517755130	2642346594	2642351395	2510265063	protein-tyrosine phosphatase	-	-
2508838286	2517755643	2642347465	2642353107	2510266649	ribonucleoside-diphosphate reductase alpha chain	Yes	2.15754E-22
2508834986	2517755498	2642343736	2642349125	2510269077	probable FeS assembly SUF system protein SufT	-	-
2508836942	2517754765	2642344253	2642349392	2510266886	DNA-binding transcriptional regulator, LacI/PurR family	Yes	1.02381E-13
2508836851	2517754834	2642343517	2642351105	2510266994	oligopeptide transport system substrate-binding protein	Yes	1.71723E-21
2508836295	2517757086	2642344573	2642350286	2510264653	pyruvate dehydrogenase E1 component alpha subunit	-	-
2508835586	2517754724	2642346515	2642349755	2510269687	-	-	-
2508839419	2517758695	2642344853	2642348856	2510268965	uracil phosphoribosyltransferase	-	-
2508836253	2517754729	2642344212	2642349352	2510264609	hypothetical protein	-	-
2508839243	2517755055	2642344175	2642348380	2510267567	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	-	-
2508836497	2517753879	2642345520	2642351833	2510267374	8-oxo-dGTP pyrophosphatase MutT, NUDIX family	-	-
2508837159	2517757984	2642345147	2642349222	2510265548	Uncharacterized membrane protein	-	-

2508836847	2517757988	2642345138	2642349213	2510266998	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II	-	-
2508836872	2517755173	2642346557	2642351428	2510266974	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508838296	2517755165	2642346567	2642351419	2510266668	DNA helicase-2 / ATP-dependent DNA helicase PcrA	-	-
2508835975	2517758530	2642344309	2642348878	2510267927	tRNA pseudouridine38-40 synthase	-	-
2508835950	2517757308	2642345999	2642352566	2510267960	16S rRNA (cytosine967-C5)-methyltransferase	-	-
2508835451	2517757001	2642347205	2642352718	2510269525	Hpr(Ser) kinase/phosphatase	-	-
2508836031	2517756985	2642347221	2642352734	2510267879	beta-glucosidase	-	-
2508839178	2517754866	2642343552	2642351067	2510264721	dihydrolipoamide dehydrogenase	-	-
2508836397	2517754106	2642345375	2642349513	2510267470	protein phosphatase	Yes	1.07699E-17
2508835583	2517754214	2642344742	2642350759	2510269684	ATP-binding protein involved in chromosome partitioning	-	-
2508838863	2517756382	2642347035	2642350258	2510265081	signal peptidase I	-	-
2508838885	2517754113	2642345368	2642349505	2510265066	ketopantoate hydroxymethyltransferase	-	-
2508839100	2517755668	2642347451	2642353093	2510264884	hypothetical protein	-	-
2508840446	2517757069	2642347149	2642352659	2510264124	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508835413	2517757161	2642346680	2642348780	2510269489	AraC-type DNA-binding protein	-	-
2508837933	2517754560	2642345842	2642348739	2510265889	-	-	-
2508836783	2517755161	2642346571	2642351416	2510267052	sulfate adenylyltransferase subunit 2	-	-
2508839948	2517753870	2642345511	2642351842	2510269812	two component transcriptional regulator, LuxR family	-	-
2508836384	2517755708	2642347416	2642353056	2510267479	tryptophanyl-tRNA synthetase	-	-
2508838539	2517757025	2642347183	2642352693	2510265383	-	-	-
2508839244	2517754608	2642346395	2642349873	2510267568	TatD DNase family protein	-	-
2508839582	2517754288	2642344653	2642350846	2510268805	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508839322	2517757099	2642344582	2642350296	2510269032	-	-	-
2508836217	2517758670	2642344867	2642353032	2510264577	Carbonic anhydrase or acetyltransferase, isoleucine patch superfamily	-	-
2508840481	2517758586	2642344955	2642352944	2510264161	glutaminyl-tRNA synthetase	-	-
2508835786	2517757538	2642346157	2642352406	2510268040	DNA processing protein	Yes	1.22E-10
2508839757	2517756963	2642347246	2642352758	2510268527	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	-	-
2508839516	2517754165	2642344769	2642350734	2510268886	two-component system, chemotaxis family, sensor kinase CheA	-	-
2508840365	2517755198	2642347667	2642351333	2510268132	Uncharacterized membrane protein	-	-

2508834920	2517756079	2642343051	2642349270	2510269151	acyl-[acyl-carrier-protein]-phospholipid O-acyltransferase	-	-
2508836888	2517756432	2642346752	2642352218	2510266964	Predicted dehydrogenase	-	-
2508837645	2517755633	2642347475	2642353117	2510269175	-	-	-
2508837569	2517758626	2642344921	2642352979	2510266242	Predicted dehydrogenase	-	-
2508837928	2517754557	2642345845	2642348742	2510265886	-	-	-
2508835999	2517755166	2642346566	2642351420	2510267910	O-acetylhomoserine (thiol)-lyase	-	-
2508835220	2517754699	2642346489	2642349780	2510264463	small subunit ribosomal protein S8	-	-
2508836344	2517753999	2642346726	2642349446	2510267541	elongation factor P	-	-
2508837052	2517754606	2642346393	2642349875	2510266748	-	-	-
2508835125	2517755098	2642346621	2642351368	2510264552	transcriptional regulator, LacI family	-	-
2508837406	2517755128	2642346596	2642351393	2510266396	-	Yes	5.21513E-20
2508837342	2517754085	2642345398	2642349537	2510266458	Lysophospholipase L1	-	-
2508835697	2517758062	2642344482	2642349998	2510268220	-	-	-
2508839922	2517758545	2642347974	2642348196	2510268319	hypothetical protein	-	-
2508837441	2517756081	2642343048	2642349267	2510266380	16S rRNA (uracil1498-N3)-methyltransferase	-	-
2508837469	2517758562	2642344981	2642352921	2510266349	-	-	-
2508838353	2517754578	2642345823	2642348721	2510265459	lipid kinase, YegS/Rv2252/BmrU family	-	-
2508835928	2517756903	2642347324	2642352838	2510267984	1-acyl-sn-glycerol-3-phosphate acyltransferase	Yes	7.21664E-06
2508838739	2517756719	2642345239	2642351022	2510265192	16S rRNA (guanine527-N7)-methyltransferase	-	-
2508839903	2517757901	2642345153	2642349228	2510268333	porphobilinogen synthase	-	-
2508837544	2517757245	2642347120	2642352628	2510266272	-	-	-
2508839616	2517756375	2642347042	2642350265	2510268772	methionyl-tRNA formyltransferase	-	-
2508835990	2517756550	2642346825	2642352147	2510267912	-	-	-
2508835702	2517758067	2642344488	2642350003	2510268215	-	-	-
2508837507	2517756302	2642345740	2642350650	2510266309	molecular chaperone DnaJ	-	-
2508836223	2517755980	2642344783	2642350324	2510264583	alpha-mannosidase	-	-
2508839234	2517754979	2642347886	2642349666	2510264672	LSU ribosomal protein L21P	-	-
2508835559	2517755311	2642348117	2642351728	2510269660	phosphate transport system substrate-binding protein	Yes	5.50148E-08

2508838479	2517755666	2642347453	2642353095	2510265444	adenosylhomocysteinase	-	-
2508840432	2517754379	2642345323	2642352288	2510264110	ribose 5-phosphate isomerase B	-	-
2508839070	2517758347	2642343419	2642351509	2510264908	malate dehydrogenase	-	-
2508835211	2517755275	2642347594	2642351260	2510264472	large subunit ribosomal protein L22	-	-
2508834875	2517756926	2642347282	2642352795	2510269197	-	-	-
2508837186	2517756310	2642345748	2642350658	2510265578	-	-	-
2508836617	2517755261	2642347608	2642351274	2510264976	-	-	-
2508838973	2517756082	2642343047	2642349266	2510267203	acyl-[acyl-carrier-protein]-phospholipid O-acyltransferase	-	-
2508835774	2517754000	2642346725	2642349447	2510268051	-	-	-
2508835138	2517754371	2642343913	2642348651	2510264536	-	-	-
2508835142	2517754734	2642344218	2642349358	2510264532	-	-	-
2508839935	2517756846	2642347388	2642352903	2510269800	putative ABC transport system permease protein	-	-
2508839643	2517756242	2642345675	2642350584	2510268740	MFS transporter, FSR family, fosmidomycin resistance protein	-	-
2508837339	2517754087	2642345396	2642349535	2510266460	-	-	-
2508836256	2517754475	2642343966	2642352248	2510264612	-	-	-
2508837519	2517755696	2642347428	2642353069	2510266297	-	-	-
2508836229	2517755828	2642345961	2642350430	2510264589	two-component system, OmpR family, sensor histidine kinase CreC	-	-
2508839427	2517757043	2642347165	2642352675	2510268957	regulator of sigma E protease	Yes	2.28269E-05
2508836475	2517758034	2642344455	2642349969	2510267395	Predicted arabinose efflux permease, MFS family	-	-
2508838652	2517755905	2642345935	2642350407	2510265342	-	Yes	7.1834E-17
2508839910	2517757893	2642345161	2642349236	2510268326	NADH dehydrogenase subunit J	-	-
2508839271	2517754545	2642345854	2642348751	2510267601	translation initiation factor 1	-	-
2508840077	2517755685	2642347439	2642353080	2510269945	Predicted oxidoreductase	-	-
2508839316	2517754971	2642347894	2642349658	2510269038	-	-	-
2508835696	2517758061	2642344481	2642349997	2510268221	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508840419	2517755597	2642347510	2642353152	2510268073	-	-	-
2508836493	2517754882	2642343570	2642351049	2510267378	-	-	-
2508837303	2517755396	2642348023	2642351635	2510266495	NAD(P)H dehydrogenase (quinone)	-	-
2508835224	2517754695	2642346485	2642349784	2510264459	LSU ribosomal protein L15P	-	-

2508837161	2517754879	2642343567	2642351052	2510265550	[acyl-carrier-protein] S-malonyltransferase	-	-
2508835201	2517755265	2642347604	2642351270	2510264482	small subunit ribosomal protein S12	-	-
2508840364	2517755196	2642347669	2642351335	2510268133	hypothetical protein	-	-
2508838635	2517755936	2642345904	2642350375	2510265358	TRAP-type mannitol/chloroaromatic compound transport system	-	-
2508836008	2517757852	2642347534	2642351206	2510267901	5,10-methylenetetrahydrofolate reductase (NAD(P))	-	-
2508837886	2517755767	2642346332	2642350474	2510265854	GTP-binding protein HflX	-	-
2508835967	2517758046	2642344467	2642349982	2510267942	molecular chaperone DnaK	-	-
2508838741	2517758127	2642344558	2642350073	2510265188	3-hydroxyisobutyrate dehydrogenase	-	-
2508835571	2517754058	2642345416	2642349555	2510269674	ABC-type uncharacterized transport system, permease component	-	-
2508836183	2517757398	2642346088	2642352477	2510267743	cysteine synthase A	-	-
2508838633	2517755934	2642345906	2642350377	2510265360	TRAP transporter, DctM subunit	Yes	6.57049E-08
2508834982	2517757140	2642346655	2642348806	2510269081	Regulator of RNase E activity RraA	-	-
2508838682	2517754720	2642346511	2642349759	2510265317	preprotein translocase subunit YajC	-	-
2508835613	2517755582	2642345045	2642348930	2510269716	-	-	-
2508834828	2517757253	2642347128	2642352637	2510269260	dTDP-glucose 4,6-dehydratase	-	-
2508839385	2517758106	2642344523	2642350038	2510269013	-	-	-
2508837173	2517758142	2642347696	2642350093	2510265561	Outer membrane protein assembly factor BamD, BamD/ComL family	-	-
2508836948	2517757654	2642346267	2642351962	2510266872	lysyl-tRNA synthetase, class II	-	-
2508838131	2517753743	2642346985	2642351588	2510265773	FecR family protein	-	-
2508838154	2517755784	2642346312	2642350492	2510266513	ABC-type uncharacterized transport system involved in gliding motility	-	-
2508835965	2517758362	2642343437	2642351494	2510267945	Transglutaminase-like enzyme, putative cysteine protease	-	-
2508838715	2517753749	2642346979	2642351582	2510265287	arsenite transporter, ACR3 family	-	-
2508839209	2517755033	2642347832	2642349719	2510264697	UDP-N-acetylmuramoylalanine--D-glutamate ligase	-	-
2508834635	2517758271	2642347716	2642350113	2510269421	tryptophan synthase, alpha chain	-	-
2508838853	2517753902	2642345544	2642351810	2510265092	Fur family transcriptional regulator, ferric uptake regulator	-	-
2508836747	2517755500	2642343732	2642349127	2510267072	-	-	-
2508835985	2517756073	2642343056	2642349276	2510267917	DNA polymerase (family 10)	-	-
2508837028	2517756737	2642345222	2642351004	2510266771	NADH dehydrogenase subunit K	-	-

2508840418	2517756991	2642347215	2642352728	2510268074	septum formation protein	-	-
2508839707	2517754813	2642343480	2642351141	2510268642	LacI family transcriptional regulator	-	-
2508837168	2517758794	2642344635	2642350864	2510265557	-	-	-
2508837350	2517756516	2642346794	2642352177	2510266450	-	-	-
2508837889	2517757433	2642346117	2642352447	2510265857	aspartate carbamoyltransferase	-	-
2508839317	2517755019	2642347846	2642349706	2510269037	pyrimidine-nucleoside phosphorylase	-	-
2508834929	2517754830	2642343513	2642351109	2510269143	-	-	-
2508836474	2517755249	2642347619	2642351285	2510267396	16S rRNA (cytosine967-C5)-methyltransferase	-	-
2508838946	2517754916	2642344353	2642348458	2510267176	-	-	-
2508839514	2517758314	2642343362	2642351912	2510268888	chemotaxis protein CheD	-	-
2508835522	2517755083	2642346637	2642351352	2510269618	-	-	-
2508838995	2517754821	2642343491	2642351131	2510267226	hemolysin III	-	-
2508840363	2517755195	2642347670	2642351336	2510268134	high-affinity iron transporter	Yes	2.61245E-10
2508839908	2517757896	2642345158	2642349233	2510268328	NADH-quinone oxidoreductase subunit I	-	-
2508837513	2517755690	2642347434	2642353075	2510266303	F-type H+-transporting ATPase subunit gamma	-	-
2508839391	2517757517	2642346133	2642352429	2510269005	-	-	-
2508838531	2517754351	2642343933	2642348671	2510265391	N-acetylglucosaminyl deacetylase, LmbE family	-	-
2508835658	2517756419	2642347002	2642350223	2510268254	23S rRNA pseudouridine2605 synthase	-	-
2508834755	2517754229	2642344716	2642350783	2510269287	-	-	-
2508835040	2517756028	2642343104	2642349325	2510267665	phosphoglycolate phosphatase	-	-
2508834668	2517758592	2642344949	2642352950	2510269359	DNA polymerase-3 subunit epsilon	-	-
2508838963	2517755533	2642345006	2642348966	2510267192	-	-	-
2508836205	2517754957	2642347908	2642349643	2510264565	-	-	-
2508835323	2517753782	2642346950	2642351553	2510264373	molybdate transport system ATP-binding protein	-	-
2508836967	2517755339	2642348084	2642351695	2510266849	-	-	-
2508838847	2517758132	2642344563	2642350078	2510265098	inorganic phosphate transporter, PiT family	-	-
2508834918	2517758521	2642344324	2642348893	2510269153	-	-	-
2508837959	2517754779	2642344269	2642349408	2510265927	cysteine desulfurase / selenocysteine lyase	-	-
2508835637	2517757381	2642346072	2642352493	2510269734	tRNA-Thr(GGU) m(6)t(6)A37 methyltransferase TsaA	-	-

2508834706	2517758791	2642344632	2642350867	2510269324	spermidine/putrescine transport system substrate-binding protein	-	-
2508836653	2517756593	2642346877	2642352096	2510265010	UDPglucose 6-dehydrogenase	-	-
2508840161	2517754391	2642345314	2642348545	2510270029	zinc protease	Yes	4.82864E-15
2508836045	2517756588	2642346872	2642352101	2510267871	iron complex transport system ATP-binding protein	-	-
2508837679	2517753995	2642346730	2642349442	2510266107	chromosome partitioning protein, ParB family	-	-
2508837572	2517758629	2642344918	2642352982	2510266239	transcriptional regulator, LacI family	-	-
2508835557	2517756037	2642343096	2642349317	2510269658	ribonuclease-3	-	-
2508836292	2517754996	2642347871	2642349681	2510264650	type IV pilus assembly protein PilM	-	-
2508837269	2517756353	2642345789	2642350700	2510265713	-	-	-
2508837099	2517754222	2642344725	2642350775	2510266696	Predicted dehydrogenase	-	-
2508834685	2517754361	2642343922	2642348660	2510269350	phosphatidylserine decarboxylase	-	-
2508836376	2517754731	2642344215	2642349355	2510267487	-	-	-
2508839637	2517756595	2642346879	2642352094	2510268747	-	-	-
2508836965	2517756720	2642345238	2642351021	2510266854	-	-	-
2508835195	2517756883	2642347349	2642352862	2510264488	Biotin carboxyl carrier protein	-	-
2508835721	2517756339	2642345774	2642350685	2510268211	glycerol-3-phosphate dehydrogenase	-	-
2508840482	2517757987	2642345139	2642349214	2510264162	putative hydrolases of HD superfamily	-	-
2508835592	2517755183	2642346548	2642351438	2510269693	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	-	-
2508836721	2517757065	2642347153	2642352663	2510267111	-	-	-
2508837025	2517756741	2642345218	2642351001	2510266774	NADH-quinone oxidoreductase subunit H	-	-
2508838979	2517754725	2642346516	2642349754	2510267209	anthranilate synthase component 2	-	-
2508834876	2517756927	2642347281	2642352794	2510269196	-	-	-
2508836938	2517758330	2642343652	2642348329	2510266900	hypothetical protein	Yes	9.08968E-09
2508837106	2517756254	2642345687	2642350595	2510266689	-	-	-
2508838855	2517756862	2642347371	2642352886	2510265090	Endo-1,4-beta-xylanase, GH35 family	-	-
2508838345	2517755258	2642347611	2642351277	2510265461	homospermidine synthase (spermidine-specific)	-	-
2508839510	2517758310	2642343366	2642351917	2510268892	flagellar basal-body rod protein FlgC	-	-
2508836155	2517758244	2642347734	2642350132	2510267772	Lysophospholipase L1	-	-
2508836754	2517755649	2642347458	2642353100	2510266762	-	-	-

2508839593	2517754528	2642343221	2642351193	2510268795	-	-	-
2508839201	2517755044	2642347823	2642349728	2510264705	cell division protein FtsZ	-	-
2508837083	2517755380	2642348037	2642351649	2510266713	Lipid-binding SYLF domain-containing protein	-	-
2508837123	2517754400	2642345301	2642348533	2510265506	RNA methyltransferase, TrmH family	-	-
2508835745	2517757595	2642346195	2642352034	2510268184	-	-	-
2508837157	2517754404	2642345297	2642348529	2510265535	lipid-A-disaccharide kinase	-	-
2508835489	2517756648	2642345244	2642351027	2510269558	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508836261	2517754456	2642343951	2642352233	2510264617	Uncharacterized membrane protein	-	-
2508837447	2517757429	2642346112	2642352452	2510266376	lipoyl(octanoyl) transferase	-	-
2508839698	2517757219	2642347088	2642352595	2510268652	Phosphoenolpyruvate carboxylase, type 1	Yes	4.66213E-32
2508838947	2517757123	2642343680	2642348261	2510267178	-	-	-
2508838526	2517754939	2642347931	2642349620	2510265396	carboxypeptidase Taq	-	-
2508835560	2517755310	2642348118	2642351729	2510269661	A/G-specific DNA-adenine glycosylase	-	-
2508837026	2517756739	2642345219	2642351002	2510266773	NADH dehydrogenase subunit I	-	-
2508840085	2517756018	2642344822	2642351463	2510269960	biopolymer transport protein ExbD	-	-
2508838969	2517755294	2642347574	2642351239	2510267199	LSU ribosomal protein L13P	-	-
2508835194	2517756882	2642347350	2642352863	2510264489	-	Yes	4.23432E-07
2508838997	2517757225	2642347094	2642352600	2510267227	Cu(I)/Ag(I) efflux system membrane protein CusA/SilA	-	-
2508837559	2517756303	2642345741	2642350651	2510266255	-	-	-
2508836047	2517756586	2642346869	2642352104	2510267869	iron complex transport system substrate-binding protein	-	-
2508834938	2517755090	2642346629	2642351360	2510269138	protein SCO1/2	Yes	9.71009E-18
2508835562	2517755308	2642347560	2642351224	2510269663	-	-	-
2508835801	2517754434	2642345267	2642348501	2510268023	-	-	-
2508835988	2517757194	2642347063	2642349904	2510267914	Phytoene dehydrogenase-related protein	-	-
2508835411	2517753738	2642346992	2642351594	2510269483	-	-	-
2508837448	2517754376	2642343908	2642348645	2510266375	D-lactate dehydrogenase	-	-
2508838909	2517757016	2642347191	2642352702	2510265041	arabinose-5-phosphate isomerase	-	-
2508836013	2517754424	2642345278	2642348511	2510267896	threonine synthase	-	-

2508835477	2517757333	2642346023	2642352542	2510269546	-	-	-
2508837278	2517753914	2642345555	2642351799	2510265719	Uncharacterized conserved protein	-	-
2508835036	2517756033	2642343100	2642349321	2510267661	hypothetical protein	-	-
2508836340	2517754205	2642344746	2642350755	2510267550	-	-	-
2508837074	2517758769	2642344611	2642350890	2510266722	-	Yes	1.16437E-18
2508836307	2517756398	2642347029	2642350252	2510264660	large subunit ribosomal protein L32	-	-
2508835808	2517754639	2642346425	2642349842	2510268016	succinate dehydrogenase / fumarate reductase iron-sulfur subunit	-	-
2508836081	2517754210	2642344735	2642350766	2510267841	outer membrane protein, multidrug efflux system	-	-
2508837036	2517757233	2642347110	2642352616	2510266764	adenylate cyclase	-	-
2508837960	2517754780	2642344270	2642349409	2510265928	nitrogen fixation protein NifU	Yes	2.75545E-05
2508836232	2517754231	2642344714	2642350785	2510264592	Thymidylate synthase ThyX	-	-
2508838167	2517753880	2642345521	2642351832	2510266522	Predicted dehydrogenase	-	-
2508840295	2517755377	2642348042	2642351653	2510270058	-	-	-
2508835729	2517757607	2642346208	2642352022	2510268203	sulfonate transport system permease protein	-	-
2508838283	2517756119	2642343169	2642348306	2510266645	-	-	-
2508840473	2517758093	2642344512	2642350027	2510264153	sulfate transport system substrate-binding protein	-	-
2508838978	2517754726	2642346517	2642349753	2510267208	transcriptional repressor NrdR	-	-
2508836848	2517754437	2642345264	2642348498	2510266997	glucose-6-phosphate isomerase	-	-
2508839928	2517757532	2642346153	2642352410	2510267345	DNA-binding response regulator, OmpR family	-	-
2508835504	2517756649	2642345243	2642351026	2510269572	L-ribulokinase	-	-
2508836813	2517758470	2642343820	2642348564	2510267029	LacI family transcriptional regulator	-	-
2508834997	2517754002	2642346722	2642349450	2510269066	membrane protein DedA, SNARE-associated domain	-	-
2508836247	2517755982	2642344785	2642350322	2510264606	-	-	-
2508838091	2517754651	2642346436	2642349830	2510265827	L-aspartate oxidase	-	-
2508840163	2517754986	2642347880	2642349672	2510270031	phosphoribosylformylglycinamidine synthase	Yes	4.33386E-25
2508838891	2517755547	2642345018	2642348954	2510265060	-	-	-
2508835156	2517754857	2642343541	2642351081	2510264519	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C	-	-
2508836787	2517754536	2642343231	2642349007	2510267048	peptide deformylase	-	-

2508836018	2517758533	2642344306	2642348875	2510267891	polyphosphate kinase	-	-
2508835418	2517756212	2642345639	2642350550	2510269494	-	-	-
2508839107	2517758755	2642344598	2642350905	2510264877	dTDP-4-amino-4,6-dideoxygalactose transaminase	-	-
2508840349	2517754266	2642344678	2642350821	2510268150	-	-	-
2508839478	2517758492	2642344432	2642349948	2510268923	-	-	-
2508835946	2517757312	2642346003	2642352562	2510267964	hypothetical protein	Yes	4.0595E-08
2508837571	2517758628	2642344919	2642352981	2510266240	2-desacetyl-2-hydroxyethyl bacteriochlorophyllide A dehydrogenase	-	-
2508839485	2517758578	2642344963	2642352936	2510268916	chemotaxis protein MotB	-	-
2508840068	2517758705	2642343327	2642350955	2510269290	Uncharacterized conserved protein, DUF433 family	-	-
2508838837	2517756938	2642347269	2642352782	2510265107	Isoleucyl-tRNA synthetase	-	-
2508837678	2517753994	2642346731	2642349441	2510266108	sulfate permease, SulP family	-	-
2508834835	2517758243	2642347735	2642350133	2510269253	two-component system, NarL family, sensor kinase	-	-
2508840089	2517756023	2642344827	2642351468	2510269964	-	-	-
2508837743	2517756625	2642346913	2642352060	2510266034	Na+-dependent transporter, SNF family	-	-
2508839229	2517758623	2642344924	2642352976	2510264677	3,4-dihydroxy 2-butanone 4-phosphate synthase	-	-
2508839808	2517755756	2642346345	2642350462	2510268481	alpha-mannosidase	-	-
2508840059	2517755021	2642347844	2642349708	2510269923	enoyl-[acyl-carrier protein] reductase	-	-
2508835551	2517754118	2642345364	2642349500	2510269650	ATP-dependent DNA helicase DinG	-	-
2508837031	2517756733	2642345225	2642351007	2510266768	NADH-quinone oxidoreductase subunit N	Yes	2.05564E-22
2508836624	2517758322	2642343352	2642351903	2510264983	iron complex transport system permease protein	-	-
2508835612	2517755583	2642345046	2642348929	2510269714	Uncharacterized conserved protein, DUF362 family	-	-
2508835255	2517757294	2642345987	2642352578	2510264431	GDPmannose 4,6-dehydratase	-	-
2508840086	2517756019	2642344823	2642351464	2510269961	biopolymer transport protein ExbD	-	-
2508835216	2517755280	2642347589	2642351255	2510264467	large subunit ribosomal protein L14	-	-
2508836032	2517754845	2642343529	2642351093	2510267878	large subunit ribosomal protein L9	-	-
2508836504	2517755298	2642347569	2642351233	2510267365	DTW domain-containing protein YfiP	-	-
2508840352	2517754262	2642344681	2642350818	2510268147	outer membrane transport energization protein ExbD	-	-
2508837918	2517757610	2642346210	2642352020	2510265881	-	-	-
2508839457	2517754516	2642343713	2642351178	2510268942	-	-	-

2508835773	2517754001	2642346724	2642349448	2510268052	diaminopimelate decarboxylase	-	-
2508835570	2517755297	2642347570	2642351235	2510269673	tRNA nucleotidyltransferase (CCA-adding enzyme)	-	-
2508839378	2517754102	2642345381	2642349519	2510269020	Negative regulator of sigma E activity	-	-
2508837117	2517754619	2642346405	2642349862	2510266678	-	-	-
2508836119	2517756575	2642346857	2642352116	2510267801	hydrogenase-4 component F	-	-
2508836610	2517755912	2642345928	2642350400	2510264969	preprotein translocase subunit SecE	-	-
2508834904	2517758816	2642346832	2642352140	2510269166	type IV pilus assembly protein PilB	-	-
2508840567	2517754844	2642343528	2642351094	2510264223	Lysophospholipase L1	-	-
2508834970	2517757147	2642346671	2642348789	2510269093	2-desacetyl-2-hydroxyethyl bacteriochlorophyllide A dehydrogenase	-	-
2508836007	2517754901	2642343587	2642351032	2510267902	Metal-dependent hydrolase, endonuclease	-	-
2508836219	2517754915	2642344354	2642348457	2510264579	pyruvate formate lyase activating enzyme	-	-
2508834897	2517755126	2642346599	2642351390	2510269173	alpha-N-arabinofuranosidase	-	-
2508838609	2517757358	2642346049	2642352516	2510265374	ATP-dependent Clp protease adaptor protein ClpS	-	-
2508839376	2517758631	2642344916	2642352984	2510269022	multisubunit sodium/proton antiporter, MrpF subunit	-	-
2508834702	2517758788	2642344629	2642350870	2510269327	spermidine/putrescine transport system permease protein	-	-
2508836637	2517756613	2642346902	2642352071	2510264993	Nucleoside-diphosphate-sugar epimerase	-	-
2508837304	2517756279	2642345721	2642350630	2510266494	-	-	-
2508835784	2517757802	2642343602	2642350721	2510268042	glucosamine--fructose-6-phosphate aminotransferase (isomerizing)	-	-
2508834987	2517755497	2642343737	2642349124	2510269076	Fe-S cluster assembly protein SufD	-	-
2508838417	2517756983	2642347223	2642352736	2510265453	triosephosphate isomerase	-	-
2508840439	2517755240	2642347625	2642351291	2510264116	-	-	-
2508836294	2517754998	2642347869	2642349683	2510264652	8-oxo-dGTP diphosphatase	-	-
2508835289	2517755120	2642346605	2642351384	2510264397	two-component system, OmpR family	-	-
2508839963	2517756313	2642345751	2642350661	2510269832	Predicted glycosyl hydrolase, GH43/DUF377 family	Yes	1.74057E-05
2508836845	2517754403	2642345298	2642348530	2510267000	-	-	-
2508834616	2517758425	2642343866	2642348604	2510269441	DNA-binding transcriptional regulator, Lrp family	-	-
2508836303	2517756402	2642347025	2642350248	2510264656	-	-	-
2508835088	2517758598	2642344942	2642352957	2510267715	peptide-methionine (S)-S-oxide reductase	Yes	1.01758E-05

2508836194	2517755056	2642344176	2642348379	2510264554	galactokinase	-	-
2508836308	2517754703	2642346493	2642349776	2510264661	Pyridoxamine 5'-phosphate oxidase	-	-
2508839179	2517754777	2642344267	2642349406	2510264720	dihydroxy-acid dehydratase	Yes	2.37288E-16
2508839788	2517754769	2642344258	2642349398	2510268497	-	Yes	3.67826E-27
2508839176	2517754773	2642344261	2642349401	2510264723	-	Yes	2.01608E-11
2508837214	2517755748	2642346353	2642350454	2510265658	ribulose 1,5-bisphosphate synthetase/thiazole synthase	-	-
2508839110	2517754941	2642347929	2642349622	2510264867	serine/threonine-protein kinase HipA	-	-
2508837423	2517758728	2642343304	2642350934	2510266389	argininosuccinate synthase	-	-
2508839113	2517758767	2642344607	2642350894	2510264864	Predicted Zn-dependent protease	Yes	3.89272E-06
2508839640	2517755287	2642347582	2642351247	2510268744	Glycosyltransferase involved in cell wall bisynthesis	-	-
2508839403	2517756179	2642343188	2642348287	2510268983	-	-	-
2508840167	2517758746	2642344590	2642350913	2510270035	hypothetical protein	-	-
2508839396	2517757512	2642346129	2642352434	2510269000	general secretion pathway protein G	-	-
2508838906	2517754128	2642345355	2642349491	2510265045	fumarase, class I, homodimeric	-	-
2508836136	2517755455	2642343769	2642348426	2510267784	GntR family transcriptional regulator, arabinose operon transcriptional repressor	-	-
2508835041	2517756027	2642343105	2642349326	2510267666	exodeoxyribonuclease V alpha subunit	-	-
2508839043	2517757018	2642347189	2642352700	2510264940	LacI family transcriptional regulator	-	-
2508838667	2517755330	2642348095	2642351706	2510265327	nitrogen regulatory protein P-II family	-	-
2508836890	2517756435	2642346755	2642352215	2510266961	Predicted dehydrogenase	-	-
2508837540	2517756038	2642343095	2642349316	2510266276	transcriptional regulator, IclR family	-	-
2508837160	2517754383	2642345322	2642352289	2510265549	RNA polymerase primary sigma factor	-	-
2508834631	2517757856	2642347530	2642351210	2510269425	condensin subunit ScpA	-	-
2508839007	2517757391	2642346082	2642352483	2510267237	probable rRNA maturation factor	Yes	1.40127E-10
2508840443	2517755237	2642347628	2642351294	2510264120	beta-phosphoglucomutase	-	-
2508839613	2517758048	2642344469	2642349984	2510268775	Threonine/homoserine efflux transporter RhtA	-	-
2508838077	2517754643	2642346429	2642349837	2510265842	transcriptional regulator, LacI family	-	-
2508840090	2517756024	2642344828	2642351469	2510269965	-	-	-

2508835345	2517758687	2642344861	2642348847	2510264341	-	-	-
2508836401	2517754230	2642344715	2642350784	2510267466	small subunit ribosomal protein S20	-	-
2508835593	2517757002	2642347204	2642352717	2510269694	nitrogen regulatory protein P-II family	-	-
2508837915	2517755615	2642347492	2642353134	2510265879	-	-	-
2508839153	2517755623	2642347484	2642353126	2510264824	-	-	-
2508836520	2517754061	2642345413	2642349552	2510267350	16S rRNA (cytosine967-C5)-methyltransferase	-	-
2508840537	2517758372	2642343443	2642351488	2510264202	-	Yes	1.79436E-10
2508840151	2517756244	2642345677	2642350586	2510270020	GTP cyclohydrolase I	-	-
2508838983	2517756942	2642347265	2642352778	2510267213	DNA-binding regulatory protein, YebC/PmpR family	-	-
2508837203	2517755959	2642345877	2642350346	2510265597	Uncharacterized protein YqfA, UPF0365 family	Yes	1.52507E-11
2508835429	2517758013	2642343804	2642348548	2510269505	iron-sulfur cluster assembly protein	-	-
2508840417	2517757354	2642346045	2642352520	2510268075	-	-	-
2508837471	2517758564	2642344979	2642352923	2510266347	-	-	-
2508836543	2517757371	2642346062	2642352503	2510267330	heptosyltransferase I	-	-
2508838155	2517755785	2642346311	2642350493	2510266514	ABC-2 type transport system permease protein	-	-
2508838692	2517758419	2642343873	2642348611	2510265307	-	-	-
2508839061	2517757392	2642346083	2642352482	2510264917	-	-	-
2508836391	2517755963	2642345874	2642350343	2510267476	-	-	-
2508839491	2517758583	2642344958	2642352941	2510268910	flagellar biosynthetic protein FlhB	-	-
2508839825	2517754136	2642345347	2642349483	2510268471	-	-	-
2508835536	2517757408	2642346097	2642352468	2510269634	beta-galactosidase	-	-
2508839008	2517755256	2642347613	2642351279	2510267238	indole-3-glycerol phosphate synthase	Yes	1.15084E-14
2508835657	2517755531	2642345002	2642348970	2510268255	-	-	-
2508840064	2517756537	2642346815	2642352156	2510269928	3-oxoacyl-[acyl-carrier-protein] synthase II	-	-
2508835133	2517755102	2642346617	2642351372	2510264541	recombination protein RecA	-	-
2508837114	2517754616	2642346402	2642349865	2510266681	transcriptional regulator, LacI family	-	-
2508839232	2517754977	2642347888	2642349664	2510264674	-	-	-
2508839305	2517757840	2642344371	2642349260	2510267645	UTP--GlnB (protein PII) uridylyltransferase, GlnD	-	-

2508838087	2517757806	2642343598	2642350717	2510265831	-	-	-
2508836422	2517755415	2642348002	2642351614	2510267449	-	-	-
2508838639	2517754838	2642343522	2642351100	2510265347	-	-	-
2508836075	2517758431	2642343860	2642348598	2510267847	Phosphate-selective porin	-	-
2508839502	2517758302	2642343374	2642351925	2510268900	-	-	-
2508837888	2517757432	2642346116	2642352448	2510265856	dihydroorotase	-	-
2508839767	2517754831	2642343514	2642351108	2510268517	-	-	-
2508839456	2517754515	2642343712	2642351177	2510268943	chemotaxis protein CheX	-	-
2508838951	2517757126	2642343684	2642348264	2510267180	3'-5' exoribonuclease	-	-
2508835465	2517754130	2642345353	2642349489	2510269536	K(+)-stimulated pyrophosphate-energized sodium pump	-	-
2508834978	2517757143	2642346658	2642348802	2510269085	-	-	-
2508835158	2517754855	2642343539	2642351083	2510264518	phenylalanyl-tRNA synthetase beta subunit	-	-
2508839018	2517757034	2642347174	2642352684	2510267249	-	-	-
2508838945	2517754228	2642344717	2642350782	2510267175	O-antigen ligase	-	-
2508835682	2517758336	2642343412	2642351516	2510268230	3-isopropylmalate dehydrogenase	-	-
2508834966	2517756185	2642343193	2642348282	2510269097	AraC family transcriptional regulator	-	-
2508834900	2517756558	2642346837	2642352136	2510269170	-	-	-
2508838524	2517757734	2642343088	2642349309	2510265398	LacI family transcriptional regulator	-	-
2508836979	2517755134	2642346589	2642351399	2510266827	lipid-A-disaccharide synthase	-	-
2508835787	2517758793	2642344634	2642350865	2510268039	-	-	-
2508835530	2517757402	2642346092	2642352473	2510269629	-	-	-
2508839450	2517754509	2642343707	2642351172	2510268949	flagellar hook-associated protein 2	-	-
2508838250	2517758587	2642344954	2642352945	2510266609	-	-	-
2508835073	2517757390	2642346081	2642352484	2510267699	NADH dehydrogenase subunit A	-	-
2508837956	2517758753	2642344596	2642350907	2510265924	ornithine carbamoyltransferase	-	-
2508838871	2517754952	2642347913	2642349638	2510265078	glycine hydroxymethyltransferase	-	-
2508840169	2517758574	2642344968	2642352932	2510270037	DNA gyrase subunit B	-	-
2508837891	2517754622	2642346408	2642349859	2510265859	3-hydroxyacyl-[acyl-carrier-protein] dehydratase	-	-
2508836550	2517755066	2642344183	2642348372	2510267323	-	-	-

2508835039	2517756029	2642343103	2642349324	2510267664	phospholipid/cholesterol/gamma-HCH transport system permease protein	-	-
2508838295	2517755602	2642347505	2642353147	2510266667	glyoxylase I family protein	-	-
2508835970	2517758135	2642347688	2642350086	2510267939	Nif-specific regulatory protein	Yes	3.59003E-14
2508840063	2517756538	2642346816	2642352155	2510269927	acyl carrier protein	-	-
2508838660	2517755902	2642345938	2642350410	2510265334	hypothetical protein	-	-
2508837752	2517758775	2642344616	2642350883	2510266018	GntR family transcriptional regulator, arabinose operon transcriptional repressor	-	-
2508836390	2517757111	2642343668	2642348249	2510267477	Na+/melibiose symporter	-	-
2508840499	2517758363	2642343438	2642351493	2510264178	-	-	-
2508835771	2517755355	2642348068	2642351679	2510268056	oligopeptide transport system permease protein	-	-
2508839912	2517757890	2642345164	2642349239	2510268324	NADH-quinone oxidoreductase subunit L	-	-
2508838490	2517755829	2642345960	2642350429	2510265430	Predicted dioxygenase of extradiol dioxygenase family	-	-
2508837020	2517756748	2642345211	2642350994	2510266779	NADH-quinone oxidoreductase subunit C	-	-
2508835264	2517754814	2642343483	2642351138	2510264423	-	-	-
2508839503	2517758303	2642343373	2642351924	2510268899	Flagellar motility protein MotE, a chaperone for MotC folding	-	-
2508836033	2517754846	2642343530	2642351092	2510267877	single-strand binding protein	-	-
2508835980	2517754961	2642347904	2642349647	2510267922	dinuclear metal center protein, YbgI/SA1388 family	-	-
2508836994	2517757106	2642348093	2642351704	2510266813	-	-	-
2508835222	2517754697	2642346487	2642349782	2510264461	large subunit ribosomal protein L18	-	-
2508839050	2517756261	2642345695	2642350602	2510264933	Predicted dehydrogenase	-	-
2508834885	2517755050	2642344171	2642348384	2510269185	coproporphyrinogen oxidase	-	-
2508838208	2517753736	2642346995	2642351596	2510266579	-	-	-
2508839098	2517754492	2642344002	2642352283	2510264886	oxygen-independent coproporphyrinogen-3 oxidase	Yes	1.21511E-19
2508834754	2517755246	2642347620	2642351286	2510269288	D-3-phosphoglycerate dehydrogenase	-	-
2508835297	2517756973	2642347232	2642352745	2510264389	-	-	-
2508839252	2517755747	2642346354	2642350453	2510267579	putative endonuclease	-	-
2508835506	2517756651	2642345241	2642351024	2510269574	L-ribulose-5-phosphate 4-epimerase	-	-
2508838712	2517755603	2642347504	2642353146	2510265290	6-phosphofructokinase 1	-	-
2508836668	2517755631	2642347476	2642353118	2510267165	-	-	-

2508836257	2517754473	2642343964	2642352245	2510264613	geranylgeranyl diphosphate synthase, type II	-	-
2508838836	2517755172	2642346558	2642351427	2510265108	-	-	-
2508838935	2517757869	2642344383	2642348869	2510267171	lysine 2,3-aminomutase	-	-
2508839159	2517755973	2642345864	2642350333	2510264818	diaminopimelate epimerase	-	-
2508837464	2517754192	2642344760	2642350741	2510266354	glycolate oxidase	-	-
2508840088	2517756022	2642344826	2642351467	2510269963	-	Yes	3.03381E-15
2508839412	2517755060	2642344179	2642348376	2510268972	iron complex transport system ATP-binding protein	-	-
2508840560	2517756332	2642345768	2642350679	2510264216	-	-	-
2508836121	2517756573	2642346855	2642352118	2510267798	Ni,Fe-hydrogenase III small subunit	-	-
2508837417	2517756646	2642346929	2642352044	2510266395	-	-	-
2508837253	2517757299	2642345991	2642352574	2510265696	transcriptional regulator, IclR family	-	-
2508838852	2517755336	2642348088	2642351699	2510265093	putative transport protein	-	-
2508838709	2517754911	2642344359	2642348453	2510265293	tRNA(adenine34) deaminase	-	-
2508840348	2517754267	2642344677	2642350822	2510268151	-	-	-
2508839382	2517754116	2642345365	2642349501	2510269016	protein phosphatase	-	-
2508836741	2517755508	2642343724	2642349135	2510267080	transcriptional regulator, LacI family	-	-
2508836116	2517756578	2642346861	2642352112	2510267804	hydrogenase-4 component B	-	-
2508836561	2517755185	2642346546	2642351440	2510267310	Peroxiredoxin	-	-
2508836311	2517754056	2642345418	2642349557	2510264663	lipoate-protein ligase A	-	-
2508839765	2517758036	2642344457	2642349972	2510268519	-	-	-
2508839448	2517754507	2642343705	2642351170	2510268951	flagellin	-	-
2508836947	2517756777	2642345196	2642350980	2510266873	LacI family transcriptional regulator	-	-
2508839751	2517755827	2642345962	2642350431	2510268532	tRNA threonylcarbamoyladenosine biosynthesis protein TsaE	-	-
2508836757	2517755006	2642347860	2642349692	2510266758	general secretion pathway protein G	-	-
2508836017	2517754417	2642345286	2642348519	2510267892	D-3-phosphoglycerate dehydrogenase	-	-
2508836280	2517754749	2642344235	2642349376	2510264637	cephalosporin-C deacetylase	-	-
2508835277	2517754682	2642346470	2642349798	2510264408	transcriptional regulator, LacI family	-	-
2508840477	2517754154	2642345333	2642349470	2510264157	-	-	-
2508836558	2517758693	2642344855	2642348854	2510267313	NADH-FMN oxidoreductase RutF, flavin reductase (DIM6/NTAB) family	-	-

2508838908	2517757017	2642347190	2642352701	2510265042	Predicted dehydrogenase	-	-
2508840351	2517754263	2642344680	2642350819	2510268148	biopolymer transport protein ExbD	-	-
2508835792	2517754442	2642345259	2642348493	2510268034	cardiolipin synthase	-	-
2508836936	2517756403	2642347024	2642350247	2510266902	carbamoyl-phosphate synthase small subunit	-	-
2508837726	2517755794	2642346288	2642350515	2510266060	tRNA modification GTPase	Yes	3.71022E-20
2508836736	2517756060	2642343072	2642349292	2510267085	shikimate dehydrogenase	-	-
2508840469	2517758090	2642344508	2642350023	2510264149	sulfate transport system permease protein	-	-
2508837552	2517755706	2642347418	2642353058	2510266262	-	-	-
2508835793	2517755362	2642348058	2642351669	2510268033	LacI family transcriptional regulator	-	-
2508836230	2517754871	2642343557	2642351062	2510264590	Uncharacterized conserved protein YafD	-	-
2508839826	2517754135	2642345348	2642349484	2510268470	transcriptional regulator, DeoR family	-	-
2508835526	2517753724	2642344031	2642349926	2510269622	cytochrome o ubiquinol oxidase subunit 3	Yes	7.23602E-09
2508839292	2517757639	2642346244	2642351985	2510267618	-	-	-
2508835296	2517756974	2642347231	2642352744	2510264390	Mg2+-importing ATPase	-	-
2508836945	2517758475	2642343815	2642348559	2510266881	NADH dehydrogenase	-	-
2508839405	2517756176	2642343185	2642348290	2510268981	AraC family transcriptional regulator	-	-
2508837564	2517756309	2642345747	2642350657	2510266250	-	-	-
2508834957	2517757009	2642347197	2642352710	2510269111	2-deoxy-D-gluconate 3-dehydrogenase	-	-
2508837446	2517754390	2642345316	2642352294	2510266379	lipoic acid synthetase	-	-
2508836819	2517757109	2642343665	2642348247	2510267022	-	-	-
2508838889	2517755549	2642345020	2642348952	2510265062	two-component system, OmpR family, phosphate regulon sensor histidine kinase PhoR	-	-
2508839946	2517753872	2642345513	2642351840	2510269810	chaperonin GroEL	-	-
2508836816	2517758511	2642344335	2642348903	2510267026	Enoyl-[acyl-carrier-protein] reductase [NADH]	-	-
2508835497	2517754891	2642343140	2642348162	2510269565	-	-	-
2508837340	2517754086	2642345397	2642349536	2510266459	-	-	-
2508835505	2517756650	2642345242	2642351025	2510269573	L-arabinose isomerase	-	-
2508836251	2517757743	2642343148	2642348170	2510264607	tRNA-dihydrouridine synthase	-	-
2508835543	2517758233	2642347811	2642350209	2510269642	ABC-2 type transport system ATP-binding protein	-	-

2508838934	2517755629	2642347478	2642353120	2510265019	2-dehydro-3-deoxyphosphogluconate aldolase lase	-	-
2508837930	2517754559	2642345843	2642348740	2510265888	-	-	-
2508835157	2517754854	2642343538	2642351084	2510264517	-	-	-
2508835908	2517755922	2642345918	2642350390	2510268005	-	-	-
2508839930	2517754566	2642345835	2642348733	2510268316	urease accessory protein	-	-
2508840573	2517755194	2642347671	2642351337	2510264229	-	-	-
2508839641	2517756064	2642343065	2642349285	2510268743	starvation-inducible DNA-binding protein	-	-
2508835527	2517753723	2642344030	2642349925	2510269623	-	-	-
2508836983	2517757337	2642346027	2642352538	2510266823	RNA polymerase, sigma 54 subunit, RpoN/SigL	-	-
2508839705	2517756422	2642346999	2642350220	2510268644	Molecular chaperone IbpA, HSP20 family	-	-
2508837482	2517757435	2642346119	2642352445	2510266336	Putative SOS response-associated peptidase YedK	-	-
2508836670	2517756109	2642347966	2642348188	2510267163	-	-	-
2508838088	2517757807	2642343597	2642350716	2510265830	-	-	-
2508834996	2517757746	2642343151	2642348173	2510269067	Heat shock protein HslJ	-	-
2508836625	2517758321	2642343353	2642351904	2510264984	iron complex transport system substrate-binding protein	-	-
2508839375	2517758632	2642344915	2642352985	2510269023	multicomponent Na+:H+ antiporter subunit E	-	-
2508835501	2517755144	2642346583	2642351404	2510269569	-	-	-
2508839158	2517755669	2642347450	2642353092	2510264819	-	-	-
2508835213	2517755277	2642347592	2642351258	2510264470	large subunit ribosomal protein L16	-	-
2508836758	2517754954	2642347911	2642349640	2510267067	-	-	-
2508837381	2517754007	2642346717	2642349455	2510267583	-	-	-
2508837470	2517758563	2642344980	2642352922	2510266348	-	-	-
2508835947	2517757311	2642346002	2642352563	2510267963	iron(III) transport system ATP-binding protein	-	-
2508837065	2517756259	2642345693	2642350600	2510266734	-	-	-
2508836389	2517755129	2642346595	2642351394	2510267478	DNA-3-methyladenine glycosylase	Yes	1.83246E-09
2508835528	2517756946	2642347261	2642352774	2510269624	putative membrane protein	-	-
2508836259	2517757350	2642346041	2642352524	2510264615	Rad3-related DNA helicase	-	-
2508838298	2517756255	2642345688	2642350596	2510266673	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508839230	2517756767	2642345188	2642350973	2510264676	putative MFS transporter, AGZA family, xanthine/uracil permease	-	-

2508839395	2517757513	2642346130	2642352433	2510269001	-	-	-
2508835789	2517756551	2642346826	2642352146	2510268037	dihydroneopterin aldolase	-	-
2508836200	2517755441	2642343785	2642348411	2510264562	dihydrodipicolinate reductase	-	-
2508836767	2517757243	2642347118	2642352625	2510267060	-	-	-
2508835545	2517757441	2642346125	2642352438	2510269644	haloacid dehalogenase superfamily	-	-
2508840080	2517757881	2642345174	2642349249	2510269955	16S rRNA (guanine966-N2)-methyltransferase	-	-
2508836069	2517754967	2642347898	2642349654	2510267853	23S rRNA (cytosine1962-C5)-methyltransferase	-	-
2508835198	2517756886	2642347345	2642352858	2510264485	-	-	-
2508835207	2517755271	2642347598	2642351264	2510264476	large subunit ribosomal protein L23	-	-
2508836698	2517758377	2642343448	2642351483	2510267137	23S rRNA pseudouridine1911/1915/1917 synthase	-	-
2508838334	2517754479	2642343974	2642352255	2510265472	iron (metal) dependent repressor, DtxR family	-	-
2508836858	2517754829	2642343512	2642351110	2510266988	prepilin-type N-terminal cleavage/methylation domain-containing protein	-	-
2508839071	2517756054	2642343078	2642349298	2510264907	-	-	-
2508839183	2517754677	2642346466	2642349802	2510264714	-	-	-
2508836571	2517754112	2642345369	2642349506	2510267269	DNA repair protein RadC	-	-
2508838537	2517758741	2642343288	2642350917	2510265385	argininosuccinate lyase	-	-
2508835335	2517758645	2642344902	2642352998	2510264360	nitrogenase iron protein NifH	-	-
2508836461	2517758782	2642344623	2642350876	2510267409	Threonine dehydrogenase	-	-
2508834992	2517753998	2642346727	2642349445	2510269071	4-hydroxy-tetrahydrodipicolinate synthase	-	-
2508835251	2517753964	2642345113	2642352358	2510264436	transcriptional regulator, LacI family	-	-
2508837276	2517756359	2642345794	2642350706	2510265717	rhamnulokinase	Yes	3.44297E-19
2508838300	2517757198	2642347066	2642349907	2510265504	Thioredoxin-related protein	Yes	1.75755E-07
2508836560	2517755184	2642346547	2642351439	2510267311	aconitate hydratase	-	-
2508836703	2517755085	2642346635	2642351354	2510267132	flotillin	-	-
2508838892	2517753961	2642345109	2642352354	2510265059	aspartyl-tRNA synthetase	-	-
2508837062	2517754599	2642346384	2642349884	2510266738	transcriptional regulator, LacI family	-	-
2508835603	2517755591	2642345054	2642348921	2510269704	DNA helicase-2 / ATP-dependent DNA helicase PcrA	-	-
2508836499	2517754592	2642345807	2642348705	2510267370	N-acetylglutamate kinase	-	-

2508836400	2517756564	2642346845	2642352127	2510267467	Uncharacterized OsmC-related protein	-	-
2508839389	2517754539	2642343239	2642349000	2510269007	-	-	-
2508835326	2517753784	2642346948	2642351551	2510264370	molybdate transport system substrate-binding protein	-	-
2508835405	2517758255	2642347727	2642350126	2510269475	methionine synthase (B12-independent)	-	-
2508835574	2517758102	2642344519	2642350034	2510269677	lipoprotein-releasing system ATP-binding protein	-	-
2508836491	2517756553	2642346828	2642352144	2510267380	-	-	-
2508834873	2517756924	2642347284	2642352797	2510269199	-	-	-
2508840556	2517754542	2642345855	2642348752	2510264212	-	-	-
2508835154	2517754097	2642345386	2642349524	2510267799	-	-	-
2508839823	2517754138	2642345345	2642349481	2510268473	hypothetical protein	-	-
2508839813	2517757345	2642346036	2642352530	2510268480	Endo-1,4-beta-xylanase, GH35 family	Yes	1.26231E-12
2508835340	2517753785	2642346947	2642351550	2510264355	molybdenum-pterin binding domain-containing protein	-	-
2508840407	2517754134	2642345349	2642349485	2510268090	-	-	-
2508834968	2517757148	2642346672	2642348788	2510269095	Phosphoglycerate dehydrogenase	-	-
2508834756	2517756638	2642346926	2642352046	2510269286	xylose isomerase	-	-
2508838342	2517755225	2642347639	2642351305	2510265464	membrane fusion protein, multidrug efflux system	-	-
2508835466	2517754129	2642345354	2642349490	2510269537	-	-	-
2508835392	2517757436	2642346120	2642352444	2510269443	Enamine deaminase RidA, house cleaning of reactive enamine intermediates	-	-
2508839713	2517755159	2642346573	2642351414	2510268634	type IV pilus assembly protein PilC	-	-
2508839854	2517755233	2642347632	2642351298	2510268339	Glyoxylase, beta-lactamase superfamily II	-	-
2508835452	2517756047	2642343085	2642349305	2510269526	-	-	-
2508836335	2517756980	2642347226	2642352739	2510267555	multidrug resistance protein, MATE family	-	-
2508839822	2517754139	2642345344	2642349480	2510268474	Sugar or nucleoside kinase, ribokinase family	-	-
2508838083	2517754648	2642346433	2642349833	2510265836	-	-	-
2508834993	2517757861	2642347526	2642351214	2510269070	two-component system, NarL family, sensor kinase	-	-
2508836661	2517758136	2642347689	2642350087	2510265017	Pyruvate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component	-	-
2508837806	2517754836	2642343520	2642351102	2510265961	-	-	-
2508839227	2517754425	2642345277	2642348510	2510264679	Pimeloyl-ACP methyl ester carboxylesterase	-	-

2508840359	2517754440	2642345261	2642348495	2510268139	hypothetical protein	-	-
2508836701	2517758553	2642347971	2642348193	2510267134	Uncharacterized membrane protein YccC	-	-
2508838923	2517755524	2642344995	2642348977	2510265028	-	-	-
2508835285	2517756034	2642343099	2642349320	2510264402	chromosome partitioning protein	-	-
2508839170	2517757736	2642343142	2642348164	2510264809	glutamate 5-kinase	-	-
2508836717	2517755361	2642348059	2642351670	2510267115	L-alanine dehydrogenase	-	-
2508836058	2517758718	2642343310	2642350939	2510267858	seryl-tRNA synthetase	-	-
2508839184	2517754678	2642346467	2642349801	2510264713	-	Yes	7.01643E-43
2508839012	2517755897	2642345945	2642350415	2510267241	hypothetical protein	Yes	1.39421E-15
2508839704	2517756421	2642347000	2642350221	2510268645	HSP20 family protein	-	-
2508837192	2517757128	2642346640	2642348821	2510265590	Acetyl esterase/lipase	-	-
2508836489	2517754453	2642345248	2642348482	2510267382	Holliday junction endonuclease RuvC	-	-
2508838688	2517754407	2642345293	2642348525	2510265311	-	-	-
2508838845	2517756243	2642345676	2642350585	2510265100	glycyl-tRNA synthetase	-	-
2508836615	2517756392	2642347034	2642350257	2510264974	L-rhamnose isomerase	-	-
2508835159	2517754853	2642343537	2642351085	2510264516	phenylalanyl-tRNA synthetase, alpha subunit	-	-
2508837527	2517755433	2642343793	2642348403	2510266289	-	-	-
2508837142	2517754469	2642347017	2642350239	2510269531	Formylglycine-generating enzyme, required for sulfatase activity	-	-
2508839162	2517757519	2642346135	2642352427	2510264815	-	-	-
2508837728	2517755516	2642347554	2642348390	2510266058	Glycosyltransferase, catalytic subunit of cellulose synthase	-	-
2508836312	2517754470	2642343962	2642352243	2510264664	Lipoprotein-anchoring transpeptidase ErfK/SrfK	Yes	1.99292E-09
2508835231	2517754688	2642346478	2642349791	2510264452	DNA-directed RNA polymerase subunit alpha	-	-
2508836563	2517757871	2642344382	2642348868	2510267307	Predicted RNA-binding protein, contains PUA-like domain	-	-
2508835292	2517758191	2642343384	2642351935	2510264394	Ser-tRNA(Thr) hydrolase /threonyl-tRNA synthetase	-	-
2508838966	2517755743	2642346358	2642350449	2510267195	-	-	-
2508836981	2517758504	2642344422	2642349936	2510266825	ATP-dependent Clp protease ATP-binding subunit ClpX	Yes	2.13004E-11
2508840337	2517758617	2642344930	2642352969	2510270092	hydroxymethylbilane synthase	-	-
2508837210	2517756567	2642346849	2642352124	2510265601	-	-	-

2508836460	2517758783	2642344624	2642350875	2510267410	AraC-type DNA-binding protein	-	-
2508835287	2517758192	2642343385	2642351936	2510264400	cysteinyl-tRNA synthetase	-	-
2508835949	2517757309	2642346000	2642352565	2510267961	iron(III) transport system permease protein	-	-
2508837218	2517755445	2642346441	2642349825	2510265662	-	Yes	8.3769E-12
2508835648	2517756292	2642345730	2642350640	2510268263	-	-	-
2508836425	2517755788	2642346307	2642350497	2510267446	-	Yes	1.83217E-12
2508839500	2517758300	2642343376	2642351927	2510268902	flagellar hook protein FlgE	-	-
2508834825	2517754257	2642344687	2642350811	2510269607	lipopolysaccharide transport system ATP-binding protein	-	-
2508839942	2517756569	2642346851	2642352122	2510269806	S-formylglutathione hydrolase FrmB	-	-
2508838981	2517758696	2642344852	2642348857	2510267211	enolase	-	-
2508840545	2517753908	2642345549	2642351805	2510264210	transcriptional regulator, DeoR family	-	-
2508839208	2517755034	2642347831	2642349720	2510264698	LysM repeat-containing protein	-	-
2508836549	2517755065	2642344182	2642348373	2510267324	-	-	-
2508836379	2517754732	2642344216	2642349356	2510267484	PhoH-like ATPase	-	-
2508839418	2517756975	2642347230	2642352743	2510268966	holo-[acyl-carrier protein] synthase	-	-
2508834924	2517754873	2642343559	2642351060	2510269148	pyruvate phosphate dikinase	-	-
2508840055	2517754535	2642343230	2642349008	2510269919	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	Yes	1.85219E-11
2508835092	2517757528	2642346149	2642352414	2510267719	biopolymer transport protein ExbD	-	-
2508839617	2517756374	2642347043	2642350266	2510268771	-	-	-
2508838883	2517754435	2642345266	2642348500	2510265068	-	-	-
2508840465	2517758085	2642344505	2642350020	2510264145	sulfate transport system permease protein	Yes	1.70308E-16
2508834816	2517755399	2642348021	2642351633	2510269268	dicarboxylate/amino acid:cation (Na+ or H+) symporter, DAACS family	-	-
2508838693	2517758408	2642343883	2642348621	2510265304	N-acetylneuraminate lyase	-	-
2508840569	2517754933	2642347937	2642349614	2510264225	Uncharacterized conserved protein, circularly permuted ATPgrasp superfamily	Yes	3.0647E-24
2508839291	2517757640	2642346245	2642351984	2510267617	multiple sugar transport system ATP-binding protein	-	-
2508834963	2517757151	2642346675	2642348785	2510269100	Lysophospholipase L1	-	-
2508835778	2517756907	2642347297	2642352810	2510268047	-	Yes	1.52117E-20

2508838161	2517754103	2642345380	2642349518	2510266520	hydroxypyruvate isomerase	-	-
2508836902	2517757178	2642346696	2642348765	2510266937	phosphotransferase system, enzyme I, PtsI	-	-
2508835217	2517755281	2642347588	2642351254	2510264466	large subunit ribosomal protein L24	-	-
2508835427	2517756015	2642344819	2642351460	2510269503	N-acetylmuramoyl-L-alanine amidase	Yes	1.46281E-07
2508837337	2517754089	2642345394	2642349533	2510266462	transcriptional regulator, LacI family	-	-
2508837491	2517755426	2642347992	2642351605	2510266323	beta-1,4-mannooligosaccharide	-	-
2508839714	2517755322	2642348104	2642351715	2510268633	condensin subunit Smc	-	-
2508835676	2517755292	2642347576	2642351241	2510268236	superoxide dismutase, Fe-Mn family	Yes	7.44589E-17
2508839596	2517754100	2642345383	2642349521	2510268792	cysteine synthase A	-	-
2508834878	2517757866	2642347521	2642351220	2510269194	hypothetical protein	-	-
2508839406	2517756175	2642343184	2642348291	2510268980	-	Yes	1.33957E-13
2508838640	2517754362	2642343921	2642348659	2510265346	signal recognition particle subunit FFH/SRP54 (srp54)	-	-
2508837531	2517755658	2642347457	2642353099	2510266285	hypothetical protein	Yes	2.12655E-09
2508837158	2517754840	2642343524	2642351098	2510265547	phosphoribosylformylglycinamidine cyclo-ligase	-	-
2508839959	2517753859	2642345500	2642351853	2510269828	-	-	-
2508839913	2517757889	2642345165	2642349240	2510268323	NADH-quinone oxidoreductase subunit M	-	-
2508839072	2517757247	2642347122	2642352630	2510264906	putative membrane protein	-	-
2508836202	2517755439	2642343787	2642348409	2510264560	twitching motility protein PilT	-	-
2508834942	2517757267	2642347144	2642352652	2510269134	miniconductance mechanosensitive channel	-	-
2508839660	2517755820	2642345969	2642350437	2510268724	-	-	-
2508836645	2517757257	2642347132	2642352641	2510269596	Glycosyltransferase involved in cell wall bisynthesis	-	-
2508840143	2517757060	2642347158	2642352668	2510270017	DNA-binding transcriptional regulator, LysR family	-	-
2508838655	2517755908	2642345932	2642350404	2510265339	-	-	-
2508835678	2517754447	2642345254	2642348488	2510268234	fused signal recognition particle receptor	-	-
2508836473	2517758125	2642344543	2642350058	2510267397	1-deoxy-D-xylulose-5-phosphate synthase	-	-
2508837718	2517755761	2642346338	2642350468	2510266067	DNA polymerase-3 subunit beta	-	-
2508836120	2517756574	2642346856	2642352117	2510267800	Ni,Fe-hydrogenase III large subunit	-	-
2508840479	2517756058	2642343074	2642349294	2510264159	-	-	-

2508836494	2517754883	2642343571	2642351048	2510267377	peptide chain release factor 1	-	-
2508836002	2517757564	2642346168	2642352395	2510267907	histidinol dehydrogenase	-	-
2508835749	2517757593	2642346193	2642352036	2510268180	7,8-dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase	-	-
2508837378	2517758655	2642344892	2642353008	2510266422	-	-	-
2508834636	2517758005	2642343271	2642349194	2510269420	xylulokinase	-	-
2508835393	2517757437	2642346121	2642352443	2510269444	-	-	-
2508835591	2517757261	2642347134	2642352644	2510269692	phospholipid/cholesterol/gamma-HCH transport system permease protein	-	-
2508836687	2517757993	2642345131	2642349206	2510267151	MoxR-like ATPase	-	-
2508836683	2517757425	2642346109	2642352456	2510267156	-	-	-
2508836450	2517755001	2642347866	2642349686	2510267420	-	-	-
2508837309	2517757026	2642347182	2642352692	2510266489	-	-	-
2508835449	2517756999	2642347207	2642352720	2510269523	-	Yes	9.78246E-05
2508836609	2517755913	2642345927	2642350399	2510264968	transcription antitermination protein nusG	-	-
2508837421	2517755143	2642346584	2642351403	2510266391	-	-	-
2508835071	2517754938	2642347932	2642349619	2510267697	divalent cation tolerance protein	-	-
2508839846	2517755133	2642346590	2642351398	2510268447	biopolymer transport protein ExbD	-	-
2508840431	2517754378	2642345324	2642352287	2510264109	-	-	-
2508839934	2517756847	2642347387	2642352902	2510269799	putative ABC transport system ATP-binding protein	Yes	4.55713E-10
2508837403	2517757085	2642344572	2642350285	2510266399	acetolactate synthase, large subunit	-	-
2508835215	2517755279	2642347590	2642351256	2510264468	small subunit ribosomal protein S17	-	-
2508838288	2517756913	2642347291	2642352804	2510266651	LacI family transcriptional regulator	-	-
2508839494	2517758294	2642343382	2642351933	2510268908	flagellar biosynthetic protein FliQ	-	-
2508837167	2517758795	2642344636	2642350863	2510265556	Molecular chaperone IbpA, HSP20 family	-	-
2508837043	2517753840	2642345484	2642351869	2510266757	alpha-L-fucosidase	-	-
2508836804	2517754808	2642343473	2642351149	2510267036	-	-	-
2508836706	2517753910	2642345551	2642351803	2510267125	hypothetical protein	-	-
2508837080	2517755797	2642346284	2642350519	2510266716	formamidopyrimidine-DNA glycosylase	-	-
2508836016	2517754418	2642345285	2642348518	2510267893	glycerol-3-phosphate acyltransferase PlsY	-	-

2508835807	2517754640	2642346426	2642349841	2510268017	succinate dehydrogenase subunit A	-	-
2508836467	2517758525	2642344315	2642348883	2510267403	-	-	-
2508836191	2517755219	2642347646	2642351312	2510267735	-	-	-
2508839835	2517758283	2642347702	2642350099	2510268458	-	-	-
2508835204	2517755268	2642347601	2642351267	2510264479	small subunit ribosomal protein S10	-	-
2508835254	2517757293	2642345986	2642352579	2510264432	GDP-L-fucose synthase	-	-
2508835521	2517756908	2642347296	2642352809	2510269582	WD-40 repeat-containing protein	-	-
2508837487	2517754785	2642344275	2642349415	2510266327	-	-	-
2508836382	2517754405	2642345296	2642348528	2510267481	aspartate aminotransferase	-	-
2508834871	2517756858	2642347376	2642352891	2510269201	MoxR-like ATPase	-	-
2508835732	2517757604	2642346205	2642352025	2510268200	putative ABC transport system permease protein	-	-
2508836396	2517754107	2642345374	2642349512	2510267471	-	-	-
2508835509	2517754224	2642344722	2642350778	2510269576	-	-	-
2508840053	2517754533	2642343228	2642349010	2510269917	-	-	-
2508838858	2517756865	2642347368	2642352883	2510265086	Phosphoglycerate dehydrogenase	-	-
2508836512	2517754658	2642346442	2642349824	2510267358	alpha-mannosidase	-	-
2508838865	2517755000	2642347867	2642349685	2510265079	-	-	-
2508840153	2517754194	2642344757	2642350744	2510270022	-	-	-
2508834994	2517757860	2642347527	2642351213	2510269069	two component transcriptional regulator, LuxR family	-	-
2508836410	2517757116	2642343673	2642348254	2510267459	ATP-binding cassette protein, ChvD family	-	-
2508839792	2517754003	2642346721	2642349451	2510268493	purine nucleosidase	-	-
2508838336	2517754127	2642345356	2642349492	2510265470	small subunit ribosomal protein S2	-	-
2508836184	2517757208	2642347075	2642349917	2510267742	DtxR family transcriptional regulator, Mn-dependent transcriptional regulator	-	-
2508840134	2517757203	2642347071	2642349912	2510270008	iron-sulfur cluster insertion protein	-	-
2508837632	2517756413	2642347008	2642350229	2510266157	-	-	-
2508837136	2517755438	2642343788	2642348408	2510265512	Predicted Zn-dependent protease, minimal metalloprotease (MMP)-like domain	-	-
2508838478	2517755667	2642347452	2642353094	2510265445	methionine adenosyltransferase	-	-
2508838535	2517755427	2642347991	2642351604	2510265387	-	-	-
2508837137	2517753735	2642346996	2642351597	2510265514	-	-	-

2508837306	2517756277	2642345719	2642350627	2510266492	prepilin-type N-terminal cleavage	-	-
2508834633	2517758464	2642343826	2642348570	2510269423	endopeptidase Clp ATP-binding regulatory subunit (clpX)	-	-
2508839610	2517758047	2642344468	2642349983	2510268778	Acetyl esterase/lipase	-	-
2508834927	2517758138	2642347691	2642350089	2510269145	-	-	-
2508838413	2517755149	2642346576	2642351411	2510265457	Predicted oxidoreductase	-	-
2508839766	2517758585	2642344956	2642352943	2510268518	-	Yes	2.20837E-07
2508836849	2517754438	2642345263	2642348497	2510266996	-	-	-
2508836521	2517756944	2642347263	2642352776	2510267349	5'-nucleotidase /3'-nucleotidase /exopolyphosphatase	-	-
2508836284	2517754663	2642346447	2642349819	2510264641	[LSU ribosomal protein L3P]-glutamine N5-methyltransferase	-	-
2508837484	2517754782	2642344272	2642349412	2510266330	-	-	-
2508840357	2517758808	2642346668	2642348793	2510265570	-	-	-
2508836785	2517758818	2642343235	2642349004	2510267050	-	-	-
2508835430	2517758009	2642343275	2642349190	2510269506	4-hydroxythreonine-4-phosphate dehydrogenase	-	-
2508837778	2517755608	2642347499	2642353141	2510265988	-	-	-
2508839119	2517754408	2642345292	2642348524	2510265312	raffinose/stachyose/melibiose transport system substrate-binding protein	-	-
2508836664	2517754707	2642346500	2642349770	2510267169	-	-	-
2508837231	2517756492	2642346770	2642352200	2510265675	-	-	-
2508840451	2517758455	2642343834	2642348579	2510264129	topoisomerase-4 subunit B	-	-
2508836786	2517754537	2642343233	2642349005	2510267049	-	-	-
2508839013	2517755898	2642345943	2642350414	2510267242	phosphate starvation-inducible protein PhoH	-	-
2508837245	2517757324	2642346015	2642352550	2510265688	LacI family transcriptional regulator,	-	-
2508836036	2517754849	2642343533	2642351089	2510267874	large subunit ribosomal protein L25	-	-
2508834967	2517757150	2642346674	2642348786	2510269096	Acetyl esterase/lipase	-	-
2508834743	2517755236	2642347629	2642351295	2510269306	2-dehydro-3-deoxygluconokinase	-	-
2508838706	2517756936	2642347272	2642352785	2510265296	cytochrome d ubiquinol oxidase subunit II	-	-
2508840105	2517754661	2642346445	2642349821	2510269980	GTP-binding protein	-	-
2508835573	2517758103	2642344520	2642350035	2510269676	lipoprotein-releasing system permease protein	-	-
2508836293	2517754997	2642347870	2642349682	2510264651	-	-	-
2508839240	2517754123	2642345360	2642349496	2510264666	peptide chain release factor 2	-	-

2508840136	2517754472	2642343963	2642352244	2510270010	-	-	-
2508836004	2517755359	2642348060	2642351671	2510267905	formate C-acetyltransferase	-	-
2508837248	2517757304	2642345996	2642352569	2510265691	-	Yes	3.44832E-34
2508839439	2517757058	2642347160	2642352670	2510268955	hypothetical protein	Yes	6.59574E-06
2508836235	2517754625	2642346412	2642349855	2510264595	capsular exopolysaccharide family	-	-
2508835329	2517758637	2642344909	2642352991	2510264367	Nif-specific regulatory protein	-	-
2508836272	2517756531	2642346809	2642352162	2510264629	ribonuclease PH	-	-
2508834901	2517756557	2642346836	2642352137	2510269169	-	-	-
2508836636	2517756614	2642346903	2642352070	2510264992	putative colanic acid biosysnthesis UDP-glucose lipid carrier transferase	-	-
2508834989	2517755495	2642343739	2642349122	2510269074	Fe-S cluster assembly ATP-binding protein	-	-
2508837797	2517758556	2642347968	2642348190	2510265970	-	-	-
2508836079	2517754209	2642344734	2642350767	2510267843	membrane fusion protein, multidrug efflux system	Yes	1.30958E-21
2508837556	2517756427	2642346747	2642352223	2510266258	3-methyladenine DNA glycosylase AlkC	-	-
2508836547	2517755069	2642344180	2642348375	2510267326	MoxR-like ATPase	-	-
2508839636	2517756596	2642346880	2642352093	2510268749	CDP-glucose 4,6-dehydratase	Yes	2.42898E-11
2508839480	2517758494	2642344430	2642349945	2510268921	flagellar assembly factor FliW	-	-
2508838169	2517753882	2642345524	2642351829	2510266524	-	-	-
2508836346	2517756526	2642346804	2642352167	2510267539	tRNA (guanine37-N1)-methyltransferase	-	-
2508839207	2517755036	2642347830	2642349721	2510264699	cell division protein FtsW	Yes	4.31246E-07
2508835575	2517758100	2642344518	2642350033	2510269678	-	-	-
2508835936	2517757317	2642346008	2642352557	2510267977	prepilin-type N-terminal cleavage	-	-
2508835308	2517758441	2642343843	2642348587	2510264377	oligopeptidase A	-	-
2508838797	2517756993	2642347213	2642352726	2510265141	molecular chaperone HtpG	-	-
2508834980	2517757142	2642346657	2642348804	2510269083	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	-	-
2508839108	2517756299	2642345737	2642350647	2510264876	-	-	-
2508839374	2517758633	2642344914	2642352986	2510269024	multicomponent Na+:H+ antiporter subunit D	-	-
2508839004	2517755628	2642347479	2642353121	2510267234	Outer membrane phospholipase A	Yes	5.70565E-16

2508838249	2517756490	2642346767	2642352203	2510266608	phosphonoacetate hydrolase	-	-
2508839763	2517756053	2642343079	2642349299	2510268521	ATP-dependent Clp protease, protease subunit	-	-
2508834940	2517756201	2642345628	2642350539	2510269136	-	-	-
2508836115	2517756579	2642346862	2642352111	2510267805	PTS system, nitrogen regulatory IIA component	-	-
2508837708	2517754676	2642346465	2642349803	2510266075	LacI family transcriptional regulator, fructose operon transcriptional repressor	-	-
2508840296	2517755378	2642348041	2642351652	2510270059	membrane glycosyltransferase	-	-
2508839479	2517758493	2642344431	2642349947	2510268922	flagellar hook-associated protein 3 FlgL	-	-
2508836523	2517755053	2642344173	2642348382	2510267347	oxygen-dependent protoporphyrinogen oxidase	Yes	4.83292E-19
2508839483	2517758497	2642344427	2642349942	2510268918	two-component system, NtrC family, response regulator HydG	-	-
2508836239	2517754628	2642346415	2642349852	2510264598	-	-	-
2508835161	2517754850	2642343534	2642351088	2510264514	large subunit ribosomal protein L35	-	-
2508836696	2517757998	2642343263	2642349201	2510267139	Beta-barrel assembly machine subunit BamA	-	-
2508838505	2517753971	2642345120	2642352366	2510265417	Signal transduction histidine kinase	-	-
2508838960	2517755535	2642345008	2642348964	2510267189	rod shape-determining protein MreB	-	-
2508839845	2517758272	2642347714	2642350111	2510268448	-	-	-
2508835155	2517754858	2642343542	2642351080	2510264520	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit A	-	-
2508839490	2517758582	2642344959	2642352940	2510268911	flagellar biosynthesis protein FlhA	-	-
2508835015	2517754638	2642346424	2642349843	2510269048	hypothetical protein	-	-
2508838905	2517758349	2642343421	2642351507	2510265046	tRNA-U20-dihydrouridine synthase	-	-
2508838414	2517755150	2642346575	2642351412	2510265456	-	-	-
2508839481	2517758495	2642344429	2642349944	2510268920	carbon storage regulator, CsrA	-	-
2508837714	2517755759	2642346340	2642350466	2510266071	mRNA interferase RelE/StbE	-	-
2508840440	2517755239	2642347626	2642351292	2510264117	-	-	-
2508835240	2517758698	2642343334	2642350962	2510264447	-	-	-
2508840103	2517757382	2642346073	2642352492	2510269978	arsenate reductase	-	-
2508836824	2517757230	2642347107	2642352613	2510267016	DNA polymerase-3 subunit epsilon	-	-
2508836811	2517758468	2642343822	2642348566	2510267031	Sugar phosphate isomerase/epimerase	-	-
2508839828	2517754686	2642346476	2642349793	2510268468	Membrane associated serine protease, rhomboid family	-	-
2508839990	2517754867	2642343553	2642351066	2510269871	-	-	-

2508839848	2517756571	2642346853	2642352120	2510268445	-	-	-
2508836920	2517758502	2642344424	2642349938	2510266918	-	-	-
2508836117	2517756577	2642346859	2642352114	2510267803	Formate hydrogenlyase subunit 4	-	-
2508836818	2517757110	2642343666	2642348248	2510267024	DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family	Yes	9.04487E-22
2508837230	2517756049	2642343083	2642349303	2510265674	probable FeS assembly SUF system protein SufT	-	-
2508838848	2517758133	2642344564	2642350079	2510265097	hypothetical protein	-	-
2508837765	2517755812	2642347405	2642353045	2510266001	Predicted nuclease of restriction endonuclease-like (RecB) superfamily	-	-
2508836179	2517758329	2642343653	2642348330	2510267747	-	-	-
2508834988	2517755496	2642343738	2642349123	2510269075	Fe-S cluster assembly protein SufB	-	-
2508837518	2517755695	2642347429	2642353070	2510266298	F-type H+-transporting ATPase subunit a	Yes	1.90182E-10
2508835139	2517754372	2642343912	2642348650	2510264535	cytochrome c oxidase cbb3-type subunit I/II	-	-
2508839967	2517753763	2642346967	2642351570	2510269841	nitrogen fixation protein NifB	-	-
2508835796	2517755369	2642348051	2642351662	2510268030	-	-	-
2508838082	2517754647	2642346432	2642349834	2510265837	-	-	-
2508836005	2517755305	2642347564	2642351228	2510267904	Signal transduction histidine kinase	-	-
2508838710	2517754912	2642344358	2642348454	2510265292	translation initiation factor IF-1	-	-
2508835661	2517756414	2642347007	2642350228	2510268251	pyrroline-5-carboxylate reductase	-	-
2508835594	2517758134	2642347687	2642350085	2510269695	L-fuculose-phosphate aldolase	-	-
2508836665	2517754574	2642345828	2642348726	2510267168	acyl carrier protein	-	-
2508837336	2517757112	2642343669	2642348250	2510266463	AraC family transcriptional regulator, transcriptional activator of pobA	-	-
2508838076	2517755538	2642345012	2642348960	2510265843	CTP synthase	-	-
2508836604	2517755918	2642345922	2642350394	2510264963	DNA-directed RNA polymerase subunit beta	-	-
2508837715	2517755760	2642346339	2642350467	2510266070	CDP-diacylglycerol---serine O-phosphatidyltransferase	-	-
2508835948	2517757310	2642346001	2642352564	2510267962	iron(III) transport system substrate-binding protein	-	-
2508837825	2517754052	2642345423	2642349561	2510265942	Sugar phosphate isomerase/epimerase	-	-
2508837178	2517758147	2642347701	2642350098	2510265566	-	-	-
2508839511	2517758311	2642343365	2642351916	2510268891	flagellar basal-body rod protein FlgB	-	-
2508835541	2517758231	2642347813	2642350211	2510269640	ABC-2 type transport system permease protein	-	-

2508836357	2517753985	2642345124	2642352371	2510267527	two component transcriptional regulator, LuxR family	-	-
2508839177	2517754495	2642343693	2642351158	2510264722	-	-	-
2508837834	2517756165	2642343174	2642348301	2510265933	-	-	-
2508837021	2517756747	2642345212	2642350995	2510266778	NADH-quinone oxidoreductase subunit D	-	-
2508840572	2517758805	2642346659	2642348801	2510264228	serine protease Do	-	-
2508835546	2517756869	2642347364	2642352879	2510269645	histidinol-phosphate aminotransferase	Yes	1.03032E-07
2508835725	2517754878	2642343566	2642351053	2510268207	-	-	-
2508836705	2517756203	2642345630	2642350541	2510267126	dGTPase	-	-
2508839605	2517757864	2642347523	2642351218	2510268783	peptide chain release factor 3	-	-
2508837111	2517754610	2642346398	2642349869	2510266684	-	-	-
2508836573	2517758229	2642347815	2642350213	2510267267	1,4-Dihydroxy-2-naphthoyl-CoA synthase	-	-
2508835596	2517753874	2642345515	2642351838	2510269697	Radical SAM superfamily enzyme YgiQ, UPF0313 family	Yes	1.58901E-13
2508835791	2517757902	2642345152	2642349227	2510268035	4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis	-	-
2508838815	2517754357	2642343926	2642348664	2510265123	BirA family transcriptional regulator, biotin operon repressor	-	-
2508839608	2517755186	2642346545	2642351441	2510268780	AraC family transcriptional regulator, L-rhamnose operon transcriptional activator RhaR	-	-
2508837078	2517755417	2642348000	2642351612	2510266718	hypothetical protein	-	-
2508838245	2517758751	2642344594	2642350909	2510266604	-	-	-
2508838624	2517755749	2642346352	2642350455	2510265369	-	Yes	1.03126E-07
2508837623	2517755435	2642343791	2642348405	2510266189	-	-	-
2508835568	2517754620	2642346406	2642349861	2510269669	topoisomerase-4 subunit A	-	-
2508836853	2517754832	2642343515	2642351107	2510266992	3-hydroxyacyl-[acyl-carrier-protein] dehydratase	-	-
2508837686	2517757349	2642346040	2642352526	2510266100	phosphoadenosine phosphosulfate reductase	-	-
2508839838	2517758280	2642347705	2642350102	2510268455	Uncharacterized protein, possibly involved in nitrogen fixation	-	-
2508836809	2517754809	2642343476	2642351145	2510267033	-	-	-
2508839621	2517757546	2642346159	2642352404	2510268768	putative peptidoglycan lipid II flippase	-	-
2508836188	2517755214	2642347650	2642351315	2510267738	-	-	-
2508835494	2517754895	2642343582	2642351037	2510269562	-	-	-
2508836886	2517756430	2642346750	2642352220	2510266966	LacI family transcriptional regulator	-	-

2508837108	2517758561	2642344982	2642352920	2510266687	-	-	-
2508836158	2517758247	2642347730	2642350129	2510267769	L-arabinose isomerase	-	-
2508838246	2517758752	2642344595	2642350908	2510266605	TIGR00288 family protein	-	-
2508836840	2517754348	2642345972	2642348222	2510267004	-	-	-
2508834700	2517758786	2642344627	2642350872	2510269329	Nicotinamidase-related amidase	-	-
2508839236	2517757322	2642346014	2642352551	2510264670	ribonuclease D	-	-
2508838420	2517755518	2642344989	2642348983	2510265450	ribonuclease G	-	-
2508839577	2517758519	2642344326	2642348895	2510268816	Glycosyltransferase involved in cell wall bisynthesis	-	-
2508835085	2517758523	2642344317	2642348885	2510267712	trigger factor	-	-
2508840464	2517758084	2642344504	2642350019	2510264144	sulfate transport system permease protein	-	-
2508838067	2517754872	2642343558	2642351061	2510265852	lactoylglutathione lyase	-	-
2508838251	2517755805	2642346275	2642350528	2510266610	-	-	-
2508837610	2517755659	2642347456	2642353098	2510266214	Sugar phosphate isomerase/epimerase	Yes	1.65064E-06
2508837180	2517755920	2642345920	2642350392	2510265573	-	-	-
2508839415	2517754389	2642345317	2642352293	2510268969	glycerol-3-phosphate dehydrogenase (NAD(P)+)	-	-
2508839428	2517757044	2642347164	2642352674	2510268956	-	-	-
2508839248	2517754226	2642344720	2642350780	2510267573	Excinuclease ABC subunit B	-	-
2508834632	2517757332	2642346022	2642352543	2510269424	demethylmenaquinone methyltransferase	-	-
2508840428	2517755314	2642348114	2642351725	2510264102	-	-	-
2508839941	2517758006	2642343272	2642349193	2510269805	guanylate kinase	-	-
2508838933	2517758555	2642347969	2642348191	2510265020	Competence protein ComGC	-	-
2508835663	2517757070	2642347148	2642352658	2510268249	adenosine deaminase	-	-
2508838343	2517755224	2642347640	2642351306	2510265463	DNA-binding transcriptional regulator, MarR family	-	-
2508839635	2517756597	2642346881	2642352092	2510268750	CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase	-	-
2508837801	2517758758	2642344600	2642350902	2510265966	-	-	-
2508836203	2517756378	2642347039	2642350262	2510264559	two-component system, NtrC family, nitrogen regulation response regulator GlnG	Yes	8.13272E-08
2508836843	2517754401	2642345300	2642348532	2510267002	-	-	-
2508838500	2517756009	2642344811	2642351452	2510265422	Predicted glycosyl hydrolase, GH43/DUF377 family	-	-
2508838816	2517754358	2642343925	2642348663	2510265122	nicotinate-nucleotide pyrophosphorylase [carboxylating]	-	-

2508839014	2517755899	2642345942	2642350413	2510267243	Diadenosine tetraphosphate (Ap4A) hydrolase	-	-
2508837165	2517756774	2642345194	2642350978	2510265554	-	-	-
2508835479	2517754133	2642345350	2642349486	2510269548	Phosphopantetheine adenylyltransferase	-	-
2508838072	2517758333	2642343649	2642348326	2510265847	-	-	-
2508836426	2517755787	2642346308	2642350496	2510267445	-	-	-
2508834935	2517755087	2642346633	2642351356	2510269141	putative membrane protein	-	-
2508838522	2517757733	2642347328	2642352842	2510265400	-	-	-
2508838086	2517757805	2642343599	2642350718	2510265832	-	-	-
2508840472	2517758088	2642344511	2642350026	2510264152	-	-	-
2508840342	2517757525	2642346146	2642352417	2510264098	-	-	-
2508838665	2517755328	2642348097	2642351708	2510265329	-	-	-
2508835584	2517754060	2642345414	2642349553	2510269685	Metal-sulfur cluster biosynthetic enzyme	-	-
2508834949	2517756411	2642347011	2642350232	2510269119	glycolate oxidase	-	-
2508834883	2517753903	2642345545	2642351809	2510269188	-	-	-
2508839980	2517756892	2642347338	2642352852	2510269855	ABC-type glycerol-3-phosphate transport system, substrate-binding protein	-	-
2508840058	2517755022	2642347843	2642349709	2510269922	D-tyrosyl-tRNA(Tyr) deacylase	-	-
2508835049	2517758458	2642343831	2642348576	2510267676	Outer membrane protein TolC	-	-
2508838684	2517756756	2642347398	2642352912	2510265315	-	Yes	2.30225E-05
2508834831	2517758239	2642347739	2642350137	2510269258	two component transcriptional regulator, LuxR family	-	-
2508839273	2517754547	2642345852	2642348749	2510267603	2-oxoglutarate dehydrogenase E1 component	-	-
2508839203	2517755041	2642347825	2642349726	2510264703	cell division protein FtsQ	-	-
2508835052	2517754196	2642344755	2642350746	2510267679	DNA-binding transcriptional regulator, MarR family	-	-
2508838301	2517755222	2642347643	2642351309	2510265503	riboflavin kinase / FMN adenylyltransferase	-	-
2508837526	2517757904	2642345150	2642349225	2510266290	hypothetical protein	-	-
2508836697	2517757997	2642343262	2642349202	2510267138	replicative DNA helicase	Yes	2.95705E-15
2508837382	2517758654	2642344893	2642353007	2510266418	LacI family transcriptional regulator	-	-
2508836513	2517758689	2642344859	2642348850	2510267357	Uncharacterized conserved protein YloU, alkaline shock protein (Asp23) family	-	-
2508834870	2517756859	2642347375	2642352890	2510269202	Uncharacterized conserved protein, DUF58 family, contains vWF domain	-	-

2508840339	2517758619	2642344928	2642352972	2510270095	acetolactate synthase, small subunit	-	-
2508836196	2517754781	2642344271	2642349410	2510264556	-	-	-
2508835794	2517755363	2642348056	2642351667	2510268032	-	-	-
2508835615	2517755578	2642345040	2642348934	2510269718	uncharacterized zinc-type alcohol dehydrogenase-like protein	-	-
2508835193	2517756880	2642347351	2642352865	2510264490	propionyl-CoA carboxylase beta chain	-	-
2508834995	2517757858	2642347529	2642351211	2510269068	glutamate dehydrogenase (NADP+)	-	-
2508839843	2517758275	2642347710	2642350107	2510268450	prokaryotic molybdopterin-containing oxidoreductase family, membrane subunit	-	-
2508835087	2517754683	2642346472	2642349796	2510267714	-	-	-
2508836108	2517754077	2642345405	2642349544	2510267814	-	-	-
2508838953	2517758105	2642344522	2642350037	2510267182	Outer membrane protein TolC	-	-
2508835470	2517758376	2642343447	2642351484	2510269541	-	-	-
2508837700	2517756253	2642345686	2642350594	2510266083	phosphoserine phosphatase	-	-
2508836015	2517754419	2642345284	2642348517	2510267894	homoserine dehydrogenase	-	-
2508838977	2517754727	2642346518	2642349752	2510267207	-	-	-
2508839497	2517758297	2642343379	2642351930	2510268905	-	-	-
2508837422	2517753737	2642346994	2642351595	2510266390	ribosomal protein L11 methyltransferase	-	-
2508839940	2517758007	2642343273	2642349192	2510269804	-	-	-
2508836024	2517757409	2642346098	2642352467	2510267886	citrate synthase	-	-
2508835140	2517754373	2642343911	2642348649	2510264534	-	-	-
2508839206	2517755037	2642347829	2642349722	2510264700	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	-	-
2508839156	2517755626	2642347481	2642353123	2510264821	-	-	-
2508836086	2517755015	2642347850	2642349702	2510267836	amino acid/amide ABC transporter substrate-binding protein, HAAT family	-	-
2508836568	2517758754	2642344597	2642350906	2510267302	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	-	-
2508838256	2517755193	2642347672	2642351339	2510266614	Phosphoglycerol transferase MdoB	-	-
2508839837	2517758281	2642347704	2642350101	2510268456	cytochrome c oxidase cbb3-type subunit 1	-	-
2508835394	2517757438	2642346122	2642352441	2510269445	-	-	-
2508838968	2517755295	2642347573	2642351238	2510267198	-	-	-
2508838204	2517755264	2642347605	2642351271	2510266575	2,5-diketo-D-gluconate reductase A	-	-

2508839929	2517754567	2642345834	2642348732	2510268317	urease accessory protein	-	-
2508836704	2517758070	2642344491	2642350006	2510267131	inner membrane protein	-	-
2508835803	2517755408	2642348010	2642351622	2510268021	glutamine amidotransferase	-	-
2508839623	2517758027	2642344450	2642349965	2510268766	hypothetical protein	-	-
2508839604	2517755201	2642347664	2642351329	2510268784	transcriptional regulator, LacI family	-	-
2508839269	2517757281	2642345977	2642348216	2510267599	magnesium transporter	-	-
2508834945	2517755080	2642344841	2642351526	2510269131	phosphoglucomutase	-	-
2508836714	2517757095	2642344577	2642350292	2510267118	branched-chain amino acid aminotransferase	-	-
2508840302	2517756850	2642347384	2642352899	2510270063	3-oxoacyl-[acyl-carrier protein] reductase	-	-
2508835349	2517754153	2642345334	2642349471	2510264337	-	-	-
2508836092	2517755009	2642347857	2642349695	2510267830	urease subunit beta	-	-
2508839761	2517754464	2642343957	2642352239	2510268523	riboflavin synthase alpha chain	-	-
2508839729	2517757974	2642343120	2642348141	2510268544	Uncharacterized conserved protein YeaO, DUF488 family	-	-
2508838827	2517758383	2642343457	2642351475	2510265112	putative ATPase	-	-
2508835336	2517753917	2642345558	2642351796	2510264359	-	-	-
2508837317	2517754966	2642347899	2642349653	2510266482	cysteine desulfuration protein SufE	-	-
2508840309	2517758264	2642347721	2642350120	2510270070	alpha-galactosidase	-	-
2508840474	2517758094	2642344513	2642350028	2510264154	phosphate-selective porin OprO and OprP	-	-
2508835210	2517755274	2642347595	2642351261	2510264473	-	-	-
2508837307	2517756276	2642345718	2642350626	2510266491	-	Yes	8.79775E-05
2508840344	2517756505	2642346783	2642352188	2510264100	-	-	-
2508839323	2517754865	2642343551	2642351068	2510269031	-	-	-
2508838238	2517755781	2642346316	2642350488	2510266597	haloacetate dehalogenase	-	-
2508835516	2517755410	2642348008	2642351620	2510267792	-	-	-
2508837318	2517754090	2642345393	2642349532	2510266481	alpha-L-fucosidase	-	-
2508839155	2517755625	2642347482	2642353124	2510264822	-	-	-
2508838809	2517754711	2642346504	2642349766	2510265129	HlyD family secretion protein	-	-
2508835339	2517758647	2642344900	2642353000	2510264356	homocitrate synthase NifV	-	-
2508836220	2517756366	2642347050	2642350274	2510264580	phosphoenolpyruvate carboxykinase (GTP)	-	-

2508837022	2517756746	2642345213	2642350996	2510266777	NADH dehydrogenase subunit E	-	-
2508836566	2517758797	2642344638	2642350861	2510267304	undecaprenyl-diphosphatase	-	-
2508835253	2517753967	2642345116	2642352362	2510264433	-	-	-
2508835984	2517754467	2642343961	2642352242	2510267918	-	-	-
2508839760	2517754825	2642343507	2642351115	2510268524	-	-	-
2508838324	2517755206	2642347657	2642351323	2510265482	translation initiation factor IF-2	-	-
2508837181	2517755921	2642345919	2642350391	2510265574	Isopentenyldiphosphate isomerase	-	-
2508840414	2517757352	2642346043	2642352522	2510268078	carbonic anhydrase	-	-
2508838774	2517754125	2642345358	2642349494	2510265161	-	-	-
2508837621	2517755887	2642345956	2642350425	2510266191	phosphate transport system substrate-binding protein	Yes	2.0766E-09
2508840466	2517758086	2642344506	2642350021	2510264146	sulfate transport system substrate-binding protein	-	-
2508834964	2517756182	2642343191	2642348284	2510269099	-	-	-
2508840370	2517758351	2642343423	2642351505	2510268127	-	Yes	2.52879E-12
2508837201	2517755962	2642345875	2642350344	2510265595	membrane-bound serine protease (ClpP class)	-	-
2508836522	2517755054	2642344174	2642348381	2510267348	histidine triad (HIT) family protein	-	-
2508834985	2517757137	2642346650	2642348811	2510269078	transcriptional regulator, IclR family	-	-
2508837122	2517756330	2642345766	2642350677	2510265505	glutamate-5-semialdehyde dehydrogenase	-	-
2508835636	2517754968	2642347897	2642349655	2510269733	IMP dehydrogenase	-	-
2508840078	2517754232	2642344712	2642350787	2510269946	serine protease Do	-	-
2508837226	2517758514	2642344331	2642348900	2510265670	6-phosphogluconolactonase	-	-
2508840492	2517753950	2642345103	2642352347	2510264171	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508839073	2517756512	2642346790	2642352181	2510264905	Uncharacterized protein, contains SIS (Sugar ISomerase) phosphosugar binding domain	-	-
2508837037	2517757235	2642347111	2642352618	2510266763	Glyoxylase, beta-lactamase superfamily II	-	-
2508838260	2517755970	2642345867	2642350336	2510266624	multiple antibiotic resistance protein	-	-
2508836974	2517758451	2642343840	2642348584	2510266842	-	-	-
2508838160	2517754104	2642345379	2642349517	2510266519	-	-	-
2508839581	2517756969	2642347237	2642352750	2510268812	-	-	-
2508837694	2517757218	2642347087	2642352594	2510269299	transcriptional regulator, LacI family	-	-
2508840141	2517755170	2642346562	2642351424	2510270015	STE24 endopeptidase	-	-

2508839268	2517757282	2642345978	2642348215	2510267598	magnesium transporter	-	-
2508839426	2517757042	2642347166	2642352676	2510268958	1-deoxy-D-xylulose 5-phosphate reductoisomerase	-	-
2508835523	2517753725	2642344034	2642349930	2510269619	azurin	-	-
2508835968	2517758045	2642344466	2642349981	2510267941	chaperonin GroES	-	-
2508836531	2517757650	2642346256	2642351973	2510267339	-	-	-
2508839805	2517757077	2642347139	2642352649	2510268484	oligoendopeptidase F	-	-
2508837534	2517755512	2642347549	2642352296	2510266282	Regulator of RNase E activity RraA	-	-
2508837681	2517754660	2642346444	2642349822	2510266105	-	Yes	5.59027E-05
2508835395	2517757440	2642346124	2642352439	2510269446	-	-	-
2508836398	2517758692	2642344856	2642348853	2510267469	phosphatidylglycerol:prolipoprotein diacylglycerol transferase	-	-
2508837252	2517757300	2642345992	2642352573	2510265695	-	-	-
2508839379	2517754101	2642345382	2642349520	2510269019	RNA polymerase sigma-70 factor, ECF subfamily	Yes	1.06898E-10
2508836707	2517753911	2642345552	2642351802	2510267124	Uncharacterized membrane protein	-	-
2508839304	2517757632	2642346236	2642351993	2510267638	beta-galactosidase	-	-
2508840494	2517753952	2642345105	2642352350	2510264173	-	-	-
2508835051	2517754195	2642344756	2642350745	2510267678	transcriptional regulator, LysR family	-	-
2508840316	2517757075	2642347141	2642352651	2510270076	NAD(P)H-hydrate epimerase	-	-
2508838680	2517754718	2642346509	2642349761	2510265319	single-stranded-DNA-specific exonuclease	-	-
2508840347	2517754268	2642344676	2642350823	2510268152	-	-	-
2508836567	2517758796	2642344637	2642350862	2510267303	-	-	-
2508835538	2517758508	2642344338	2642348906	2510269636	RNA methyltransferase, RsmE family	-	-
2508839477	2517758491	2642344433	2642349949	2510268924	-	-	-
2508838316	2517756223	2642345654	2642350563	2510265490	glycoside/pentoside/hexuronide:cation symporter, GPH family	Yes	1.05636E-27
2508834999	2517754597	2642346379	2642349890	2510269064	hypothetical protein	-	-
2508837481	2517756570	2642346852	2642352121	2510266337	putative hydrolase of the HAD superfamily	Yes	2.27956E-12
2508839324	2517757845	2642347543	2642351198	2510269030	-	-	-
2508839083	2517755432	2642343794	2642348402	2510264896	-	-	-
2508837701	2517757003	2642347203	2642352716	2510266082	-	-	-

2508840438	2517755241	2642347624	2642351290	2510264115	-	-	-
2508837746	2517756166	2642343175	2642348300	2510266031	spermidine/putrescine transport system permease protein	-	-
2508836171	2517754976	2642347889	2642349663	2510267755	apolipoprotein N-acyltransferase	Yes	7.22708E-28
2508835608	2517755587	2642345050	2642348925	2510269710	-	-	-
2508836260	2517754457	2642343952	2642352234	2510264616	-	-	-
2508837532	2517754963	2642347902	2642349649	2510266284	ribosome-associated protein	-	-
2508836691	2517758031	2642344452	2642349967	2510267147	Nucleotide-binding universal stress protein, UspA family	-	-
2508838897	2517755189	2642347673	2642351340	2510265054	-	-	-
2508835017	2517754636	2642346422	2642349845	2510269046	putative ABC transport system permease protein	-	-
2508838873	2517756493	2642346771	2642352199	2510265075	-	-	-
2508835359	2517756929	2642347278	2642352791	2510264310	unsaturated rhamnogalacturonyl hydrolase	-	-
2508839441	2517754360	2642343923	2642348661	2510268953	HD-like signal output (HDOD) domain, no enzymatic activity	-	-
2508835420	2517757012	2642347195	2642352707	2510269496	-	Yes	0.00012712
2508836195	2517758599	2642344941	2642352958	2510264555	Predicted ATPase	-	-
2508839212	2517755030	2642347836	2642349716	2510264694	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	-	-
2508835935	2517757318	2642346009	2642352556	2510267978	-	Yes	1.34747E-21
2508835983	2517754120	2642345363	2642349499	2510267919	large subunit ribosomal protein L28	-	-
2508839925	2517757533	2642346154	2642352409	2510267344	Signal transduction histidine kinase	Yes	7.16423E-19
2508836713	2517757094	2642344576	2642350291	2510267119	-	-	-
2508838419	2517756981	2642347225	2642352738	2510265451	glyceraldehyde 3-phosphate dehydrogenase	-	-
2508840452	2517758584	2642344957	2642352942	2510264130	2-dehydropantoate 2-reductase	-	-
2508836062	2517754412	2642345288	2642348520	2510267854	-	-	-
2508838195	2517754394	2642345310	2642348542	2510266566	-	-	-
2508837271	2517756356	2642345792	2642350704	2510265715	L-lactate dehydrogenase complex protein LldF	-	-
2508835145	2517757202	2642347070	2642349911	2510264529	aromatic amino acid transport protein AroP	-	-
2508840366	2517755199	2642347666	2642351331	2510268131	putative ABC transport system permease protein	Yes	3.55411E-16
2508834637	2517756052	2642343080	2642349300	2510269419	tryptophan synthase beta chain	-	-
2508838272	2517757183	2642346702	2642348760	2510266635	-	-	-

2508836715	2517754885	2642343573	2642351046	2510267117	-	-	-
2508839147	2517755617	2642347490	2642353132	2510264830	L-lactate dehydrogenase	-	-
2508835038	2517756030	2642343102	2642349323	2510267663	phospholipid/cholesterol/gamma-HCH transport system ATP-binding protein	-	-
2508836056	2517756978	2642347228	2642352741	2510267860	large subunit ribosomal protein L31	-	-
2508839906	2517757898	2642345156	2642349231	2510268330	NADH-quinone oxidoreductase subunit C	-	-
2508836973	2517757051	2642345635	2642350546	2510266843	-	Yes	8.0336E-12
2508838811	2517754713	2642346506	2642349764	2510265127	putative ABC transport system permease protein	-	-
2508835325	2517753783	2642346949	2642351552	2510264371	molybdate transport system permease protein	-	-
2508838952	2517758104	2642344521	2642350036	2510267181	RND family efflux transporter, MFP subunit	-	-
2508839082	2517755431	2642343795	2642348401	2510264897	glutamate-1-semialdehyde 2,1-aminomutase	-	-
2508839507	2517758307	2642343369	2642351920	2510268895	flagellar motor switch protein FliG	-	-
2508838866	2517757119	2642343676	2642348257	2510267455	prepilin-type N-terminal cleavage/methylation domain-containing protein	-	-
2508835609	2517755586	2642345049	2642348926	2510269711	-	-	-
2508836978	2517755135	2642346588	2642351400	2510266828	Predicted dehydrogenase	Yes	6.03399E-17
2508839288	2517757643	2642346248	2642351981	2510267614	maltose/maltodextrin transport system substrate-binding protein	-	-
2508837204	2517755810	2642346270	2642350533	2510265599	Membrane associated serine protease, rhomboid family	-	-
2508835205	2517755269	2642347600	2642351266	2510264478	large subunit ribosomal protein L3	-	-
2508835785	2517757803	2642343601	2642350720	2510268041	glucose-1-phosphate adenylyltransferase	-	-
2508836629	2517756530	2642346808	2642352163	2510264987	6-phosphogluconate dehydrogenase	-	-
2508839249	2517755205	2642347660	2642351325	2510267574	hypothetical protein	-	-
2508836718	2517755821	2642345968	2642350436	2510267114	Nucleotide-binding universal stress protein, UspA family	Yes	7.28888E-05
2508840084	2517756017	2642344821	2642351462	2510269959	outer membrane transport energization protein ExbB	-	-
2508838319	2517754142	2642345341	2642349478	2510265488	-	-	-
2508837719	2517755762	2642346337	2642350469	2510266066	-	-	-
2508836859	2517755182	2642346549	2642351437	2510266987	-	-	-
2508835120	2517754082	2642345401	2642349540	2510267731	-	-	-
2508839597	2517758748	2642344591	2642350912	2510268791	phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase	-	-
2508836420	2517754956	2642347909	2642349642	2510267451	N-acetylglutamate synthase	-	-

2508835163	2517756729	2642345228	2642351010	2510264512	ApaG protein	-	-
2508840165	2517755661	2642347454	2642353096	2510270033	L-aspartate 1-decarboxylase	-	-
2508836044	2517756589	2642346873	2642352100	2510267872	oxygen-independent coproporphyrinogen-3 oxidase	-	-
2508837502	2517755699	2642347425	2642353066	2510266314	cystathionine gamma-synthase	-	-
2508840073	2517758710	2642343318	2642350947	2510269936	xylose isomerase	-	-
2508836296	2517757087	2642344574	2642350287	2510264654	pyruvate dehydrogenase E1 component beta subunit	-	-
2508836156	2517758245	2642347733	2642350131	2510267771	transcriptional regulator, IclR family	-	-
2508838799	2517756992	2642347214	2642352727	2510265139	ATP-binding cassette, subfamily F, uup	-	-
2508838636	2517758109	2642344527	2642350042	2510265356	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 5/6)	Yes	5.53138E-16
2508836559	2517758694	2642344854	2642348855	2510267312	PTS system IIA component, Fru family	-	-
2508835973	2517758528	2642344311	2642348880	2510267929	acetyl-CoA carboxylase carboxyl transferase subunit beta	-	-
2508835338	2517753760	2642346970	2642351573	2510264357	-	-	-
2508835095	2517757348	2642346039	2642352527	2510267721	beta-glucuronidase	Yes	3.11828E-14
2508840294	2517755376	2642348043	2642351654	2510270057	glucans biosynthesis protein	Yes	2.22525E-17
2508839003	2517755983	2642344786	2642350321	2510267233	Sugar or nucleoside kinase, ribokinase family	-	-
2508836574	2517756857	2642347377	2642352892	2510267266	GMP synthase (glutamine-hydrolysing)	-	-
2508837293	2517757414	2642346104	2642352461	2510265732	5-(carboxyamino)imidazole ribonucleotide mutase	Yes	3.11055E-05
2508837703	2517754674	2642346463	2642349805	2510266080	-	-	-
2508836712	2517755024	2642347841	2642349711	2510267120	NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family	-	-
2508838705	2517756935	2642347273	2642352786	2510265297	cytochrome bd-I ubiquinol oxidase subunit 1 apoprotein	-	-
2508835300	2517756970	2642347235	2642352748	2510264386	two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	-	-
2508839944	2517757250	2642347125	2642352634	2510269808	N-acetyl-gamma-glutamyl-phosphate reductase	-	-
2508836961	2517758450	2642343841	2642348585	2510266858	Predicted deacylase	-	-
2508838276	2517757181	2642346700	2642348762	2510266638	transcriptional regulator, LacI family	-	-
2508840308	2517757400	2642346090	2642352475	2510270069	-	-	-
2508836695	2517757999	2642343264	2642349200	2510267140	periplasmic chaperone for outer membrane proteins Skp	-	-
2508834723	2517756321	2642345757	2642350667	2510269318	-	-	-

2508836814	2517758509	2642344337	2642348905	2510267028	haloalkane dehalogenase	-	-
2508835734	2517757602	2642346203	2642352027	2510268198	-	-	-
2508840374	2517758355	2642343427	2642351502	2510268124	-	-	-
2508836935	2517756528	2642346806	2642352165	2510266903	spermidine/putrescine transport system ATP-binding protein	-	-
2508835192	2517756879	2642347352	2642352866	2510264491	-	-	-
2508839161	2517754193	2642344758	2642350743	2510264816	acetyl-CoA hydrolase	-	-
2508837179	2517756405	2642347022	2642350244	2510265567	ribonuclease R	-	-
2508836892	2517756437	2642346757	2642352213	2510266959	hypothetical protein	-	-
2508836799	2517754805	2642343468	2642351154	2510267039	transcriptional regulator, LacI family	-	-
2508838999	2517757222	2642347091	2642352598	2510267229	Outer membrane protein TolC	-	-
2508837175	2517758144	2642347698	2642350095	2510265563	biopolymer transport protein ExbB	-	-
2508836514	2517758688	2642344860	2642348849	2510267356	-	-	-
2508837053	2517754605	2642346391	2642349876	2510266747	LacI family transcriptional regulator	-	-
2508838653	2517755906	2642345934	2642350406	2510265341	biopolymer transport protein ExbB	-	-
2508835151	2517755397	2642348022	2642351634	2510264523	peptidyl-prolyl cis-trans isomerase B (cyclophilin B)	-	-
2508835681	2517756979	2642347227	2642352740	2510268231	Nitroreductase	-	-
2508835512	2517756217	2642345647	2642350556	2510269579	RNA polymerase, sigma 70 subunit, RpoD	-	-
2508838632	2517755933	2642345907	2642350378	2510265361	Ethanolamine ammonia-lyase heavy chain	-	-
2508837748	2517755757	2642346342	2642350465	2510266029	3-oxoacyl-(acyl-carrier-protein) synthase III	-	-
2508837282	2517757410	2642346099	2642352466	2510265723	-	-	-
2508838339	2517755229	2642347636	2642351302	2510265467	-	-	-
2508839708	2517758338	2642343414	2642351514	2510268641	hypothetical protein	-	-
2508834936	2517755088	2642346632	2642351357	2510269140	protoheme IX farnesyltransferase	-	-
2508840338	2517758618	2642344929	2642352971	2510270093	ketol-acid reductoisomerase	-	-
2508839750	2517755826	2642345963	2642350432	2510268533	thiamine-monophosphate kinase	-	-
2508839772	2517754735	2642344219	2642349359	2510268513	glutamyl-tRNA synthetase	-	-
2508838748	2517754794	2642344284	2642349423	2510267623	-	-	-
2508835091	2517757527	2642346148	2642352415	2510267718	Cell division and transport-associated protein TolQ	-	-
2508836846	2517756948	2642347259	2642352772	2510266999	23S rRNA pseudouridine1911/1915/1917 synthase	-	-

2508836524	2517755051	2642344172	2642348383	2510267346	uroporphyrinogen decarboxylase	-	-
2508839572	2517758546	2642347973	2642348195	2510268821	Cu+-exporting ATPase	-	-
2508840142	2517755169	2642346563	2642351423	2510270016	Metal-dependent hydrolase, endonuclease/exonuclease/phosphatase family	-	-
2508838810	2517754712	2642346505	2642349765	2510265128	putative ABC transport system ATP-binding protein	-	-
2508836218	2517754914	2642344355	2642348456	2510264578	ribonucleoside-triphosphate reductase class III catalytic subunit	-	-
2508838792	2517758258	2642347722	2642350121	2510265158	Lysophospholipase L1	-	-
2508839594	2517754527	2642343220	2642351192	2510268794	transcription termination factor Rho	Yes	6.87867E-12
2508836946	2517756778	2642345197	2642350981	2510266874	-	-	-
2508838357	2517753787	2642346945	2642351548	2510268391	-	-	-
2508839473	2517758487	2642344437	2642349953	2510268928	-	-	-
2508835977	2517758434	2642343857	2642348595	2510267925	phosphatidylglycerophosphatase A	-	-
2508835074	2517757389	2642346080	2642352485	2510267700	Diadenylate cyclase (c-di-AMP synthetase), DisA_N domain	-	-
2508836030	2517758800	2642344642	2642350858	2510267880	-	-	-
2508837541	2517756039	2642343094	2642349315	2510266275	glucarate dehydratase	-	-
2508840132	2517756200	2642345627	2642350538	2510270006	myo-inositol-1(or 4)-monophosphatase	-	-
2508835346	2517758686	2642344862	2642348846	2510264340	preprotein translocase subunit SecG	-	-
2508834649	2517757307	2642345998	2642352567	2510269413	-	-	-
2508840307	2517757401	2642346091	2642352474	2510270068	aspartate aminotransferase	-	-
2508836297	2517757093	2642344575	2642350290	2510264655	pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase)	-	-
2508839851	2517755164	2642346568	2642351418	2510268341	SsrA-binding protein	-	-
2508839748	2517754921	2642347951	2642349601	2510268535	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508838859	2517756866	2642347367	2642352882	2510265085	-	-	-
2508836107	2517754079	2642345404	2642349543	2510267815	-	-	-
2508836455	2517755527	2642346440	2642349826	2510267415	-	-	-
2508835660	2517756416	2642347005	2642350226	2510268252	methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase	-	-
2508838263	2517755889	2642345954	2642350423	2510266200	-	-	-
2508837938	2517756633	2642346921	2642352051	2510265902	-	-	-
2508836093	2517755008	2642347858	2642349694	2510267829	urease subunit alpha	Yes	7.97175E-30

2508835969	2517758044	2642344465	2642349980	2510267940	chaperonin GroEL	-	-
2508837512	2517755688	2642347436	2642353077	2510266304	ATP synthase F1 subcomplex beta subunit	-	-
2508838528	2517756545	2642346820	2642352151	2510265394	-	-	-
2508836118	2517756576	2642346858	2642352115	2510267802	hydrogenase-4 component E	-	-
2508836608	2517755914	2642345926	2642350398	2510264967	large subunit ribosomal protein L11	-	-
2508839693	2517755257	2642347612	2642351278	2510268653	-	Yes	8.61176E-13
2508836710	2517758345	2642343417	2642351511	2510267122	ubiquinone biosynthesis protein	-	-
2508835160	2517754851	2642343535	2642351087	2510264515	large subunit ribosomal protein L20	-	-
2508836089	2517755012	2642347854	2642349698	2510267833	urea transport system ATP-binding protein	-	-
2508835007	2517757882	2642345173	2642349248	2510269056	-	-	-
2508836124	2517755745	2642346356	2642350451	2510267795	LacI family transcriptional regulator, fructose operon transcriptional repressor	-	-
2508835018	2517754635	2642346421	2642349846	2510269045	putative ABC transport system ATP-binding protein	-	-
2508835361	2517756931	2642347276	2642352789	2510264308	-	-	-
2508835473	2517756768	2642345189	2642350974	2510269544	-	Yes	5.60546E-13
2508834866	2517758107	2642344524	2642350039	2510269206	threonine dehydratase	-	-
2508835806	2517754641	2642346427	2642349840	2510268018	-	-	-
2508834679	2517757076	2642347140	2642352650	2510269356	haloacid dehalogenase superfamily	-	-
2508836263	2517757291	2642345984	2642352581	2510264619	-	-	-
2508839215	2517755026	2642347839	2642349713	2510264691	16S rRNA (cytosine1402-N4)-methyltransferase	-	-
2508835972	2517758527	2642344312	2642348881	2510267930	dihydrofolate synthase / folylpolyglutamate synthase	-	-
2508839400	2517754285	2642344656	2642350843	2510268986	alpha-mannosidase	-	-
2508840399	2517753815	2642345458	2642351896	2510268105	-	Yes	1.46251E-05
2508839840	2517758278	2642347707	2642350104	2510268453	-	-	-
2508836150	2517757616	2642346219	2642352011	2510267931	haloacid dehalogenase superfamily, subfamily IA,	-	-
2508840156	2517758024	2642344447	2642349962	2510270026	-	-	-
2508836129	2517758653	2642344894	2642353006	2510267790	transcriptional regulator, IclR family	-	-
2508839272	2517754546	2642345853	2642348750	2510267602	-	-	-
2508837813	2517753773	2642346957	2642351560	2510265954	LacI family transcriptional regulator	-	-

2508839016	2517755259	2642347610	2642351276	2510267245	-	-	-
2508838948	2517757124	2642343681	2642348262	2510267177	-	-	-
2508839265	2517754188	2642344764	2642350738	2510267595	YidC/Oxa1 family membrane protein insertase	Yes	9.4694E-10
2508835737	2517757600	2642346201	2642352029	2510268195	-	-	-
2508837567	2517758624	2642344923	2642352977	2510266244	putative NAD(P)H quinone oxidoreductase, PIG3 family	-	-
2508837107	2517758560	2642344983	2642352919	2510266688	mRNA interferase RelE/StbE	-	-
2508837044	2517754609	2642346396	2642349871	2510266756	LacI family transcriptional regulator	Yes	1.11943E-10
2508838352	2517754577	2642345825	2642348723	2510265460	ATP-binding cassette, subfamily F, member 3	-	-
2508835639	2517757378	2642346069	2642352496	2510269736	multiple sugar transport system permease protein	-	-
2508834618	2517758423	2642343868	2642348606	2510269439	-	-	-
2508835695	2517757081	2642344026	2642349922	2510268222	alanyl-tRNA synthetase	-	-
2508836490	2517756555	2642346829	2642352143	2510267381	porin, OprB family	Yes	7.7504E-06
2508834944	2517755079	2642344842	2642351525	2510269132	hypothetical protein	-	-
2508835301	2517757801	2642343603	2642350722	2510264385	-	-	-
2508836252	2517757744	2642343149	2642348171	2510264608	hypothetical protein	-	-
2508839499	2517758299	2642343377	2642351928	2510268903	-	-	-
2508835607	2517755588	2642345051	2642348924	2510269709	-	-	-
2508838092	2517754652	2642346437	2642349829	2510265826	pantothenate synthetase	-	-
2508839393	2517757515	2642346131	2642352431	2510269003	-	-	-
2508837683	2517756068	2642343062	2642349282	2510266103	mannose-6-phosphate isomerase, type 1	-	-
2508837745	2517756167	2642343176	2642348299	2510266032	spermidine/putrescine transport system substrate-binding protein	-	-
2508835514	2517755968	2642345869	2642350338	2510269581	starch synthase	-	-
2508834628	2517757853	2642347533	2642351207	2510269428	-	-	-
2508838850	2517758432	2642343859	2642348597	2510265095	sec-independent protein translocase protein TatA	-	-
2508835703	2517758068	2642344489	2642350004	2510268214	-	-	-
2508837523	2517755300	2642347567	2642351231	2510266293	RNA-binding protein	-	-
2508835744	2517753730	2642344037	2642349933	2510268185	iron complex outermembrane recepter protein	-	-
2508834943	2517755078	2642344843	2642351524	2510269133	FKBP-type peptidyl-prolyl cis-trans isomerase	-	-
2508840137	2517757399	2642346089	2642352476	2510270011	-	-	-

2508837479	2517753877	2642345518	2642351835	2510266339	hypothetical protein	Yes	2.02688E-12
2508839001	2517757220	2642347089	2642352596	2510267231	Sugar phosphate isomerase/epimerase	-	-
2508839794	2517754005	2642346719	2642349453	2510268491	transcriptional regulator, LacI family	-	-
2508839914	2517757888	2642345166	2642349241	2510268322	NADH-quinone oxidoreductase subunit M	-	-
2508839392	2517757516	2642346132	2642352430	2510269004	-	-	-
2508838967	2517755296	2642347572	2642351237	2510267196	Uncharacterized membrane protein YgdD, TMEM256/DUF423 family	-	-
2508838689	2517754869	2642343555	2642351064	2510265310	phosphoribosyl 1,2-cyclic phosphate phosphodiesterase	-	-
2508837946	2517756631	2642346919	2642352053	2510265907	-	-	-
2508838323	2517755207	2642347656	2642351322	2510265483	ribosome-binding factor A	-	-
2508836699	2517758378	2642343449	2642351482	2510267136	hypothetical protein	-	-
2508840354	2517755007	2642347859	2642349693	2510268145	transcriptional regulator, LacI family	-	-
2508837866	2517756134	2642344077	2642349025	2510268389	-	-	-
2508835610	2517755585	2642345048	2642348927	2510269712	-	-	-
2508836815	2517758510	2642344336	2642348904	2510267027	3-oxoacyl-[acyl-carrier-protein] synthase-3	-	-
2508837826	2517754053	2642345422	2642349560	2510265941	Predicted dehydrogenase	-	-
2508839231	2517756988	2642347218	2642352731	2510264675	-	-	-
2508838529	2517756544	2642346819	2642352152	2510265393	Exodeoxyribonuclease VII large subunit	-	-
2508840368	2517757193	2642347061	2642349903	2510268129	putative ABC transport system ATP-binding protein	Yes	9.94872E-07
2508835507	2517754048	2642345427	2642349565	2510269575	AraC-type DNA-binding protein	-	-
2508839171	2517757737	2642343143	2642348165	2510264808	ribosome recycling factor	-	-
2508838200	2517758389	2642343904	2642348641	2510266571	peroxiredoxin Q/BCP	-	-
2508839242	2517757177	2642346695	2642348766	2510267566	-	-	-
2508840401	2517756890	2642347340	2642352854	2510268103	Outer membrane receptor proteins, mostly Fe transport	-	-
2508834834	2517758242	2642347736	2642350134	2510269254	transcriptional regulator, TetR family	-	-
2508835291	2517755118	2642346607	2642351382	2510264395	oxaloacetate decarboxylase, beta subunit	-	-
2508836709	2517758346	2642343418	2642351510	2510267123	Polyhydroxyalkanoate synthesis regulator phasin	-	-
2508838259	2517758593	2642344948	2642352951	2510266623	Cephalosporin-C deacetylase	Yes	5.83829E-17
2508835070	2517754937	2642347933	2642349618	2510267696	-	-	-

2508840375	2517754671	2642346458	2642349809	2510268123	-	-	-
2508835706	2517755409	2642348009	2642351621	2510264890	LacI family transcriptional regulator	-	-
2508835275	2517758051	2642344472	2642349987	2510264410	-	-	-
2508838911	2517756239	2642345672	2642350581	2510265039	LacI family transcriptional regulator	-	-
2508838954	2517757799	2642343605	2642350724	2510267183	2-isopropylmalate synthase	-	-
2508834903	2517758814	2642346834	2642352138	2510269167	-	-	-
2508839711	2517757863	2642347524	2642351217	2510268636	-	-	-
2508839320	2517757122	2642343679	2642348260	2510269034	-	-	-
2508835524	2517757078	2642344033	2642349928	2510269620	cytochrome c oxidase subunit 2	-	-
2508839172	2517757738	2642343144	2642348166	2510264807	uridylate kinase	-	-
2508835016	2517754637	2642346423	2642349844	2510269047	-	-	-
2508836612	2517755910	2642345930	2642350402	2510264971	elongation factor Tu	-	-
2508835726	2517754444	2642345257	2642348491	2510268206	Predicted PurR-regulated permease PerM	Yes	7.6364E-09
2508836496	2517753878	2642345519	2642351834	2510267375	-	-	-
2508839764	2517758037	2642344458	2642349973	2510268520	peptide-methionine (R)-S-oxide reductase	-	-
2508840361	2517754225	2642344721	2642350779	2510268137	-	-	-
2508836823	2517757232	2642347109	2642352615	2510267017	Predicted DNA-binding transcriptional regulator YafY, contains an HTH and WYL domains	-	-
2508835565	2517758732	2642343301	2642350930	2510269666	ATP-binding cassette, subfamily B	-	-
2508835555	2517756962	2642347247	2642352759	2510269656	-	-	-
2508840478	2517756057	2642343075	2642349295	2510264158	-	-	-
2508839762	2517754465	2642343959	2642352240	2510268522	DNA polymerase	Yes	2.26557E-05
2508839501	2517758301	2642343375	2642351926	2510268901	-	-	-
2508836186	2517755210	2642347654	2642351319	2510267740	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508839397	2517757511	2642346128	2642352435	2510268999	-	-	-
2508839692	2517754824	2642343494	2642351128	2510268661	1,4-alpha-glucan branching enzyme	Yes	9.32088E-41
2508837082	2517755551	2642345022	2642348950	2510266714	N-acetylneuraminate lyase	-	-
2508837209	2517756568	2642346850	2642352123	2510265600	LacI family transcriptional regulator	-	-
2508835419	2517758421	2642343871	2642348609	2510269495	glycoside/pentoside/hexuronide:cation symporter, GPH family	-	-

2508838773	2517755641	2642347467	2642353109	2510265165	16S rRNA pseudouridine516 synthase	-	-
2508839909	2517757894	2642345160	2642349235	2510268327	NADH-quinone oxidoreductase subunit H	-	-
2508836423	2517755790	2642346305	2642350499	2510267448	transcriptional regulator, LacI family	-	-
2508835731	2517757605	2642346206	2642352024	2510268201	sulfonate transport system ATP-binding protein	-	-
2508840314	2517755795	2642346287	2642350516	2510270074	-	-	-
2508839097	2517754493	2642344003	2642352284	2510264887	-	-	-
2508838681	2517754719	2642346510	2642349760	2510265318	SecD/SecF fusion protein	-	-
2508837102	2517758690	2642344858	2642348851	2510266693	-	-	-
2508835971	2517755967	2642345870	2642350339	2510267938	-	-	-
2508839943	2517757249	2642347124	2642352632	2510269807	-	-	-
2508838069	2517758334	2642343648	2642348325	2510265850	-	-	-
2508839715	2517755324	2642348101	2642351712	2510268632	-	-	-
2508837275	2517756358	2642345793	2642350705	2510265716	-	-	-
2508835525	2517757079	2642344032	2642349927	2510269621	cytochrome c oxidase subunit 1	-	-
2508839319	2517756518	2642346796	2642352174	2510269035	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	-	-
2508840135	2517755302	2642347565	2642351229	2510270009	-	-	-
2508835230	2517754689	2642346479	2642349790	2510264453	small subunit ribosomal protein S4	-	-
2508840395	2517753826	2642345469	2642351884	2510268109	-	-	-
2508836810	2517758467	2642343823	2642348567	2510267032	D-glucuronate isomerase	-	-
2508836027	2517758473	2642343817	2642348561	2510267883	DNA-binding transcriptional regulator, LysR family	-	-
2508839149	2517755619	2642347488	2642353130	2510264828	Carboxysome shell and ethanolamine utilization microcompartment protein CcmK/EutM	-	-
2508839486	2517758579	2642344962	2642352937	2510268915	chemotaxis protein MotA	-	-
2508838903	2517758356	2642343431	2642351499	2510265048	ATP-binding cassette, subfamily B	-	-
2508837611	2517755660	2642347455	2642353097	2510266213	Predicted dehydrogenase	-	-
2508836000	2517757566	2642346170	2642352393	2510267909	imidazoleglycerol-phosphate dehydratase	Yes	3.60242E-10
2508839968	2517753762	2642346968	2642351571	2510269842	ferritin	-	-
2508837702	2517755454	2642343770	2642348425	2510266081	Sugar phosphate isomerase/epimerase	-	-
2508835053	2517754199	2642344752	2642350749	2510267680	membrane fusion protein, multidrug efflux system	-	-

2508836852	2517754833	2642343516	2642351106	2510266993	-	-	-
2508838662	2517755900	2642345940	2642350412	2510265332	hypothetical protein	-	-
2508837292	2517758123	2642344541	2642350056	2510265731	Predicted pyrophosphatase or phosphodiesterase, AlkP superfamily	-	-
2508837338	2517754088	2642345395	2642349534	2510266461	-	-	-
2508836090	2517755011	2642347855	2642349697	2510267832	amino acid/amide ABC transporter ATP-binding protein 2, HAAT family	-	-
2508837220	2517755156	2642345274	2642348507	2510265664	-	-	-
2508837509	2517756408	2642347014	2642350236	2510266307	-	-	-
2508835544	2517758234	2642347810	2642350208	2510269643	HlyD family secretion protein	-	-
2508835939	2517755562	2642345033	2642348939	2510267974	hypothetical protein	Yes	1.39807E-16
2508835589	2517757263	2642347136	2642352646	2510269690	-	-	-
2508836201	2517755440	2642343786	2642348410	2510264561	4-hydroxy-tetrahydrodipicolinate synthase	-	-
2508836622	2517758325	2642343350	2642351901	2510264981	cob(I)alamin adenosyltransferase	-	-
2508835046	2517758462	2642343828	2642348572	2510267673	DNA-binding transcriptional regulator, MarR family	-	-
2508835200	2517756888	2642347342	2642352856	2510264483	oxaloacetate decarboxylase, alpha subunit	-	-
2508840139	2517756297	2642345735	2642350645	2510270013	secondary thiamine-phosphate synthase enzyme	-	-
2508836014	2517754420	2642345283	2642348516	2510267895	aspartate kinase	-	-
2508835122	2517754080	2642345403	2642349542	2510267733	L-fucose isomerase	-	-
2508837477	2517757740	2642343146	2642348168	2510266341	Zn-dependent protease with chaperone function	-	-
2508834753	2517755245	2642347621	2642351287	2510269289	L-2-hydroxycarboxylate dehydrogenase (NAD+)	-	-
2508838178	2517755985	2642344788	2642350319	2510266539	-	-	-
2508835974	2517758529	2642344310	2642348879	2510267928	comF family protein	Yes	6.52351E-07
2508836702	2517755086	2642346634	2642351355	2510267133	-	-	-
2508837027	2517756738	2642345221	2642351003	2510266772	NADH-quinone oxidoreductase subunit J	-	-
2508839380	2517755480	2642343754	2642348439	2510269018	endoglucanase	-	-
2508837517	2517755694	2642347430	2642353071	2510266299	-	-	-
2508840436	2517755244	2642347622	2642351288	2510264113	AraC-type DNA-binding protein	-	-
2508836803	2517754807	2642343472	2642351150	2510267037	general secretion pathway protein G	-	-
2508840367	2517757191	2642347060	2642349901	2510268130	putative ABC transport system permease protein	-	-

2508837017	2517757113	2642343670	2642348251	2510266782	-	-	-
2508837684	2517756070	2642343060	2642349280	2510266102	DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family	-	-
2508836788	2517754799	2642344291	2642349430	2510267041	-	-	-
2508838654	2517755907	2642345933	2642350405	2510265340	-	-	-
2508839622	2517758028	2642344451	2642349966	2510268767	ATP-dependent Clp protease ATP-binding subunit ClpC	-	-
2508836240	2517754629	2642346416	2642349851	2510264599	-	Yes	8.14527E-34
2508836402	2517754775	2642344265	2642349404	2510267465	small subunit ribosomal protein S1	-	-
2508836937	2517756590	2642346874	2642352099	2510266901	-	-	-
2508839408	2517758235	2642347809	2642350207	2510268978	oxygen-independent coproporphyrinogen-3 oxidase	Yes	1.67965E-13
2508839988	2517757061	2642347157	2642352667	2510269869	ATP-dependent helicase HrpB	-	-
2508836271	2517754055	2642345420	2642349558	2510264628	Lysophospholipase L1	-	-
2508838504	2517756532	2642346810	2642352161	2510265418	type III pantothenate kinase	-	-
2508835350	2517757059	2642347159	2642352669	2510264331	-	-	-
2508837169	2517757985	2642345141	2642349216	2510265558	-	-	-
2508838278	2517756116	2642343165	2642348310	2510266640	-	-	-
2508838970	2517755293	2642347575	2642351240	2510267200	small subunit ribosomal protein S9	-	-
2508837802	2517758756	2642344599	2642350904	2510265965	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	-	-
2508839836	2517758282	2642347703	2642350100	2510268457	-	-	-
2508834688	2517755407	2642348011	2642351623	2510269347	alanine racemase	-	-
2508835311	2517758439	2642343852	2642348589	2510264374	Beta-barrel assembly machine subunit BamA	-	-
2508840405	2517754884	2642343572	2642351047	2510268092	two component transcriptional regulator, LuxR family	-	-
2508839246	2517756615	2642346904	2642352069	2510267570	diaminohydroxyphosphoribosylaminopyrimidine deaminase	-	-
2508839957	2517753861	2642345502	2642351851	2510269826	Membrane-anchored ribosome-binding protein, inhibits growth in stationary phase	-	-
2508838484	2517758542	2642347977	2642348200	2510265439	-	-	-
2508838925	2517755523	2642344994	2642348978	2510265027	raffinose/stachyose/melibiose transport system substrate-binding protein	-	-
2508839017	2517757035	2642347173	2642352683	2510267248	-	Yes	2.33438E-07
2508836545	2517753876	2642345517	2642351836	2510267328	amidophosphoribosyltransferase	-	-
2508836498	2517754927	2642347945	2642349607	2510267371	urease accessory protein	Yes	5.02522E-

							13
2508838610	2517757359	2642346050	2642352515	2510265373	-	-	-
2508837177	2517758146	2642347700	2642350097	2510265565	biopolymer transport protein ExbD	-	-
2508836258	2517754752	2642344238	2642349379	2510264614	-	-	-
2508835769	2517755349	2642348067	2642351678	2510268057	oligopeptide transport system permease protein	-	-
2508836285	2517754989	2642347879	2642349673	2510264642	-	-	-
2508836606	2517755916	2642345924	2642350396	2510264965	LSU ribosomal protein L10P	-	-
2508839413	2517756252	2642345685	2642350593	2510268971	CrcB protein	-	-
2508837677	2517754062	2642345412	2642349551	2510266109	hypothetical protein	-	-
2508839574	2517757366	2642346057	2642352509	2510268819	-	-	-
2508835549	2517758715	2642343313	2642350942	2510269648	2-isopropylmalate synthase	-	-
2508836424	2517755789	2642346306	2642350498	2510267447	-	-	-
2508835225	2517754694	2642346484	2642349785	2510264458	protein translocase subunit secY/sec61 alpha	-	-
2508838608	2517757357	2642346048	2642352517	2510265375	-	-	-
2508839237	2517757989	2642345136	2642349211	2510264669	-	-	-
2508835295	2517754668	2642346453	2642349813	2510264391	-	-	-
2508837281	2517754233	2642344711	2642350788	2510265722	endonuclease G	Yes	6.39104E-14
2508837561	2517756305	2642345742	2642350652	2510266253	-	Yes	3.81131E-10
2508835150	2517754449	2642345252	2642348486	2510264524	methylmalonyl-CoA mutase metallochaperone MeaB	-	-
2508835348	2517754841	2642343525	2642351096	2510264338	phosphoribosylamine--glycine ligase	-	-
2508834666	2517758054	2642344475	2642349991	2510269361	type IV pilus assembly protein PilB	-	-
2508839163	2517754292	2642344649	2642350851	2510264814	LacI family transcriptional regulator	Yes	1.81961E-22
2508836468	2517755742	2642346359	2642350448	2510267402	-	-	-
2508839495	2517758295	2642343381	2642351932	2510268907	flagellar biosynthetic protein FliP	-	-
2508839015	2517758039	2642344460	2642349975	2510267244	peptidyl-prolyl cis-trans isomerase B (cyclophilin B)	-	-
2508837162	2517758124	2642344542	2642350057	2510265551	-	-	-
2508835432	2517757175	2642346692	2642348768	2510269508	tRNA A37 threonylcarbamoyladenosine dehydratase	-	-
2508836393	2517755965	2642345872	2642350341	2510267474	ABC-2 type transport system ATP-binding protein	-	-

2508835561	2517755309	2642347559	2642351223	2510269662	Sugar phosphate isomerase/epimerase	-	-
2508839101	2517756396	2642347030	2642350253	2510264883	phosphoribosylanthranilate isomerase	-	-
2508840343	2517756506	2642346784	2642352187	2510264099	branched-chain amino acid aminotransferase	-	-
2508836029	2517758799	2642344641	2642350859	2510267881	GTP-binding protein	Yes	1.28531E-14
2508839388	2517758820	2642343238	2642349001	2510269008	-	-	-
2508835354	2517753898	2642345540	2642351814	2510264326	hypothetical protein	-	-
2508839377	2517758630	2642344917	2642352983	2510269021	multicomponent Na+:H+ antiporter subunit G	-	-
2508836310	2517754702	2642346492	2642349777	2510264662	dTDP-4-dehydrorhamnose reductase	-	-
2508837563	2517756308	2642345746	2642350656	2510266251	AraC-type DNA-binding protein	-	-
2508837113	2517754613	2642346401	2642349866	2510266682	-	-	-
2508839508	2517758308	2642343368	2642351919	2510268894	flagellar M-ring protein FliF	-	-
2508836855	2517754827	2642343509	2642351113	2510266990	Short-chain dehydrogenase	-	-
2508840437	2517755243	2642347623	2642351289	2510264114	-	-	-
2508835638	2517757379	2642346070	2642352495	2510269735	carbohydrate ABC transporter membrane protein 1, CUT1 family	-	-
2508838415	2517755151	2642346574	2642351413	2510265455	-	-	-
2508835172	2517754709	2642346502	2642349768	2510264503	ATPase components of ABC transporters with duplicated ATPase domains	-	-
2508839453	2517754512	2642343710	2642351175	2510268946	flagellar protein FliS	-	-
2508836231	2517754870	2642343556	2642351063	2510264591	-	-	-
2508835747	2517757594	2642346194	2642352035	2510268182	sulfate transport system substrate-binding protein	-	-
2508839965	2517753765	2642346965	2642351568	2510269839	nitrogenase molybdenum-cofactor synthesis protein NifE	Yes	2.08725E-10
2508839035	2517757567	2642346171	2642352392	2510264948	adenylosuccinate synthase	-	-
2508839280	2517755350	2642348066	2642351677	2510267606	ATP-binding cassette, subfamily B, MsbA	-	-
2508836304	2517756401	2642347026	2642350249	2510264657	-	-	-
2508834934	2517756050	2642343082	2642349302	2510269142	TIGR00255 family protein	-	-
2508834833	2517758241	2642347737	2642350135	2510269255	-	-	-
2508834919	2517756080	2642343049	2642349268	2510269152	MFS transporter, LPLT family, lysophospholipid transporter	-	-
2508836199	2517755443	2642343784	2642348412	2510264563	Holliday junction DNA helicase subunit RuvA	-	-
2508836856	2517754828	2642343510	2642351112	2510266989	Glycosyltransferase involved in cell wall bisynthesis	-	-

2508839372	2517758636	2642344910	2642352990	2510269027	multicomponent Na+:H+ antiporter subunit A	-	-
2508838158	2517757416	2642346106	2642352459	2510266517	hypothetical protein	-	-
2508836074	2517758430	2642343861	2642348599	2510267848	phosphate ABC transporter substrate-binding protein, PhoT family	-	-
2508836619	2517758327	2642343656	2642348332	2510264978	Uncharacterized membrane protein YjjB, DUF3815 family	-	-
2508838207	2517756766	2642345187	2642350972	2510266578	-	-	-
2508839827	2517754524	2642343217	2642351189	2510268469	-	-	-
2508837947	2517756629	2642346917	2642352056	2510265906	-	-	-
2508835422	2517755353	2642348072	2642351684	2510269498	serine/threonine protein kinase	-	-
2508836149	2517756269	2642345703	2642350610	2510267932	Phosphoglycerate dehydrogenase	-	-
2508837330	2517755813	2642347410	2642353050	2510266474	unsaturated rhamnogalacturonyl hydrolase	-	-
2508835232	2517754687	2642346477	2642349792	2510264451	large subunit ribosomal protein L17	-	-
2508839958	2517753860	2642345501	2642351852	2510269827	-	-	-
2508838961	2517755534	2642345007	2642348965	2510267190	rod shape-determining protein MreC	-	-
2508838079	2517754645	2642346431	2642349835	2510265840	-	-	-
2508835679	2517754446	2642345255	2642348489	2510268233	NusB antitermination factor	-	-
2508837024	2517756744	2642345215	2642350998	2510266775	-	-	-
2508839006	2517756199	2642345626	2642350537	2510267236	histidinol-phosphatase (PHP family)	-	-
2508836782	2517755160	2642346572	2642351415	2510267053	sulfate adenylyltransferase subunit 1	-	-
2508835511	2517754704	2642346497	2642349773	2510269578	anthranilate phosphoribosyltransferase	-	-
2508839956	2517753862	2642345503	2642351850	2510269825	Uncharacterized membrane protein YqjE	-	-
2508834639	2517755990	2642344793	2642350314	2510269417	-	-	-
2508840476	2517758096	2642344515	2642350030	2510264156	transcriptional regulator, BadM/Rrf2 family	-	-
2508835445	2517756966	2642347241	2642352754	2510269519	-	-	-
2508835241	2517758699	2642343333	2642350961	2510264446	dTMP kinase	-	-
2508840406	2517758454	2642343836	2642348580	2510268091	UDP-N-acetylmuramate: L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	-	-
2508839498	2517758298	2642343378	2642351929	2510268904	flagellar motor switch protein FliM	-	-
2508839907	2517757897	2642345157	2642349232	2510268329	NADH dehydrogenase subunit D	-	-
2508837156	2517755082	2642344838	2642351529	2510265534	O-succinylbenzoate synthase	-	-
2508836763	2517757241	2642347116	2642352623	2510267062	transcriptional regulator, LacI family	-	-

2508839390	2517757518	2642346134	2642352428	2510269006	-	-	-
2508834921	2517754525	2642343218	2642351190	2510269150	orotidine-5'-phosphate decarboxylase	-	-
2508837522	2517755301	2642347566	2642351230	2510266294	Uncharacterized membrane protein YeiH	-	-
2508839716	2517755325	2642348100	2642351711	2510268631	Leucyl aminopeptidase (aminopeptidase T)	-	-
2508835585	2517754059	2642345415	2642349554	2510269686	-	-	-
2508835563	2517755709	2642347415	2642353055	2510269664	-	-	-
2508838839	2517756940	2642347267	2642352780	2510265105	signal peptidase II	-	-
2508837515	2517755692	2642347432	2642353073	2510266301	-	Yes	0.000124346
2508835976	2517758433	2642343858	2642348596	2510267926	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	-	-
2508835444	2517756965	2642347242	2642352755	2510269518	LacI family transcriptional regulator	-	-
2508838857	2517756864	2642347369	2642352884	2510265087	transcriptional regulator, LacI family	-	-
2508837698	2517757199	2642347067	2642349908	2510266085	acetyl-CoA carboxylase biotin carboxyl carrier protein	-	-
2508839175	2517754774	2642344263	2642349402	2510264724	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	-	-
2508835986	2517756072	2642343057	2642349277	2510267916	1,4-dihydroxy-2-naphthoyl-CoA hydrolase	-	-
2508836694	2517758691	2642344857	2642348852	2510267141	-	-	-
2508839425	2517757041	2642347167	2642352677	2510268959	-	-	-
2508837721	2517755830	2642345958	2642350427	2510266064	MFS transporter, YNFM family, putative membrane transport protein	-	-
2508839842	2517758276	2642347709	2642350106	2510268451	-	-	-
2508839791	2517754004	2642346720	2642349452	2510268494	ribokinase	-	-
2508840540	2517757874	2642344378	2642348864	2510264205	replication restart DNA helicase PriA	-	-
2508835474	2517756769	2642345190	2642350975	2510269545	GMP synthase (glutamine-hydrolysing)	-	-
2508837674	2517757804	2642343600	2642350719	2510266112	autotransporter-associated beta strand repeat-containing protein	-	-
2508837084	2517755381	2642348036	2642351648	2510266712	N-acetylglucosamine transport system permease protein	-	-
2508838971	2517756423	2642346998	2642350219	2510267201	AMP nucleosidase	-	-
2508835037	2517756032	2642343101	2642349322	2510267662	-	-	-
2508840543	2517758661	2642344886	2642353014	2510264208	-	-	-
2508840442	2517755238	2642347627	2642351293	2510264119	kojibiose phosphorylase	-	-
2508836607	2517755915	2642345925	2642350397	2510264966	large subunit ribosomal protein L1	-	-

2508839578	2517754386	2642345320	2642352290	2510268815	N6-L-threonylcarbamoyladenine synthase	-	-
2508839758	2517756964	2642347245	2642352757	2510268526	thiamine-phosphate pyrophosphorylase	-	-
2508839915	2517757887	2642345167	2642349242	2510268321	NADH-quinone oxidoreductase subunit N	Yes	7.13273E-17
2508835057	2517754203	2642344748	2642350753	2510267684	-	-	-
2508836971	2517757053	2642345633	2642350544	2510266845	-	Yes	5.39685E-25
2508835909	2517755923	2642345917	2642350389	2510268004	3-hydroxyethyl bacteriochlorophyllide a dehydrogenase	-	-
2508835290	2517755119	2642346606	2642351383	2510264396	-	-	-
2508839712	2517758713	2642343315	2642350944	2510268635	AraC family transcriptional regulator, L-rhamnose operon transcriptional activator RhaR	-	-
2508835433	2517754459	2642343954	2642352236	2510269509	formyltetrahydrofolate deformylase	-	-
2508834630	2517757855	2642347531	2642351209	2510269426	-	-	-
2508835227	2517754692	2642346482	2642349787	2510264456	methionyl aminopeptidase	-	-
2508837508	2517756301	2642345739	2642350649	2510266308	molecular chaperone GrpE	-	-
2508840152	2517754460	2642343955	2642352237	2510270021	-	-	-
2508837229	2517756511	2642346789	2642352182	2510265673	-	-	-
2508836048	2517756580	2642346863	2642352110	2510267868	-	-	-
2508836969	2517755341	2642348082	2642351693	2510266847	-	-	-
2508837860	2517756157	2642344124	2642349071	2510265760	FecR family protein	-	-
2508840544	2517758662	2642344885	2642353015	2510264209	quinolinate synthetase	-	-
2508839614	2517758049	2642344470	2642349985	2510268774	iron(III) transport system permease protein	-	-
2508834619	2517758422	2642343870	2642348608	2510269438	GTP-binding protein	-	-
2508835424	2517755354	2642348071	2642351683	2510269500	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508838073	2517758332	2642343650	2642348327	2510265846	prepilin-type N-terminal cleavage/methylation domain-containing protein	-	-
2508837310	2517758435	2642343856	2642348594	2510266488	-	-	-
2508839239	2517758437	2642343854	2642348592	2510264667	chorismate mutase	-	-
2508837727	2517755515	2642347553	2642348391	2510266059	DNA-binding transcriptional regulator, LysR family	-	-
2508834704	2517758790	2642344631	2642350868	2510269325	spermidine/putrescine transport system ATP-binding protein	-	-
2508839591	2517754532	2642343227	2642349011	2510268797	cyclase	-	-
2508837957	2517755640	2642347468	2642353110	2510265925	tryptophan synthase beta chain	-	-

2508836927	2517756840	2642347395	2642352909	2510266911	-	-	-
2508836943	2517754766	2642344254	2642349393	2510266885	-	-	-
2508839095	2517758543	2642347976	2642348199	2510264889	polyphosphate:nucleotide phosphotransferase, PPK2 family	Yes	1.67271E-07
2508837366	2517755928	2642345912	2642350384	2510266435	Protein N-acetyltransferase, RimJ/RimL family	-	-
2508835587	2517757265	2642347138	2642352648	2510269688	phospholipid-binding lipoprotein MlaA	-	-
2508837023	2517756745	2642345214	2642350997	2510266776	NADH-quinone oxidoreductase subunit F	-	-
2508840056	2517757850	2642347537	2642351203	2510269920	manganese-dependent inorganic pyrophosphatase	-	-
2508835399	2517757384	2642346075	2642352490	2510269469	shikimate kinase	-	-
2508836273	2517755804	2642346276	2642350527	2510264630	putative ABC transport system permease protein	-	-
2508836688	2517757992	2642345132	2642349207	2510267150	Uncharacterized conserved protein, DUF58 family, contains vWF domain	-	-
2508835542	2517758232	2642347812	2642350210	2510269641	ABC-2 type transport system permease protein	-	-
2508839789	2517754768	2642344257	2642349397	2510268496	galactonate dehydratase	-	-
2508836912	2517758534	2642344305	2642348874	2510266926	-	-	-
2508835276	2517758050	2642344471	2642349986	2510264409	ATP-dependent Clp protease, protease subunit	-	-
2508836472	2517758126	2642344544	2642350059	2510267398	Exodeoxyribonuclease VII small subunit	-	-
2508836034	2517754847	2642343531	2642351091	2510267876	small subunit ribosomal protein S6	-	-
2508835243	2517756541	2642346817	2642352154	2510264444	ferredoxin	-	-
2508836216	2517758669	2642344868	2642353031	2510264576	large subunit ribosomal protein L33	-	-
2508835723	2517755793	2642346289	2642350514	2510268209	lipopolysaccharide export system permease protein	-	-
2508839852	2517755163	2642346569	2642351417	2510268340	-	-	-
2508836685	2517756989	2642347217	2642352730	2510267153	fumarase, class II	-	-
2508835564	2517758731	2642343302	2642350932	2510269665	ATP-binding cassette, subfamily B	Yes	1.78287E-20
2508839281	2517755766	2642346333	2642350473	2510267607	cytidyltransferase-like domain-containing protein	-	-
2508836409	2517757114	2642343671	2642348252	2510267460	alpha-galactosidase	Yes	1.01106E-37
2508837001	2517754812	2642343479	2642351142	2510266806	Predicted dehydrogenase	-	-
2508838606	2517758500	2642344426	2642349940	2510265377	-	-	-
2508839982	2517758130	2642344561	2642350076	2510269863	Enamine deaminase RidA, house cleaning of reactive enamine intermediates	-	-
2508837697	2517755105	2642346615	2642351374	2510266086	chorismate synthase	-	-

2508835934	2517755802	2642346279	2642350524	2510267979	-	-	-
2508835680	2517754445	2642345256	2642348490	2510268232	6,7-dimethyl-8-ribityllumazine synthase	-	-
2508838860	2517756867	2642347366	2642352881	2510265084	-	-	-
2508838089	2517757808	2642343596	2642350715	2510265829	transcriptional regulator, LacI family	-	-
2508837499	2517755700	2642347424	2642353065	2510266317	exodeoxyribonuclease-3	-	-
2508839689	2517755710	2642347414	2642353054	2510268664	-	-	-
2508836630	2517757320	2642346011	2642352554	2510264988	-	-	-
2508836383	2517754406	2642345294	2642348526	2510267480	DNA topoisomerase-3	Yes	5.08677E-32
2508835274	2517758621	2642344926	2642352974	2510264411	aspartate-semialdehyde dehydrogenase	-	-
2508838656	2517755904	2642345936	2642350408	2510265338	-	-	-
2508834699	2517758785	2642344626	2642350873	2510269330	Acetylornithine/succinyldiaminopimelate/putrescine aminotransferase	-	-
2508838664	2517755326	2642348099	2642351710	2510265330	-	-	-
2508836666	2517754575	2642345827	2642348725	2510267167	3-oxoacyl-[acyl-carrier-protein] reductase	-	-
2508838813	2517754820	2642343490	2642351132	2510265125	dihydropteroate synthase	-	-
2508835481	2517757872	2642344381	2642348867	2510269550	-	-	-
2508840471	2517758092	2642344510	2642350025	2510264151	-	-	-
2508834698	2517758784	2642344625	2642350874	2510269331	Ser/Thr protein kinase RdoA involved in Cpx stress response, MazF antagonist	-	-
2508835223	2517754696	2642346486	2642349783	2510264460	small subunit ribosomal protein S5	-	-
2508840074	2517758042	2642344463	2642349978	2510269940	putative protease	-	-
2508837299	2517756618	2642346906	2642352067	2510265739	-	-	-
2508834946	2517758702	2642343330	2642350958	2510269122	transcriptional regulator, LacI family	-	-
2508840463	2517758083	2642344503	2642350018	2510264143	Uncharacterized NAD(P)/FAD-binding protein YdhS	-	-
2508835143	2517756062	2642343067	2642349287	2510264531	Heat shock protein HslJ	-	-
2508838881	2517754347	2642345973	2642348221	2510265070	Uncharacterized conserved protein, DUF1015 family	-	-
2508840498	2517756904	2642347323	2642352837	2510264177	hypothetical protein	-	-
2508837098	2517754223	2642344724	2642350776	2510266697	ATP-dependent RNA helicase DeaD	-	-
2508834815	2517753973	2642345122	2642352369	2510269269	ATP-dependent Clp protease ATP-binding subunit ClpB	-	-
2508839931	2517754565	2642345836	2642348734	2510268315	urease accessory protein	-	-

2508839290	2517757641	2642346246	2642351983	2510267616	maltose/maltodextrin transport system permease protein	-	-
2508837729	2517756215	2642345644	2642350554	2510266057	histidyl-tRNA synthetase	-	-
2508834758	2517754931	2642347939	2642349612	2510269284	magnesium chelatase family protein	-	-
2508836395	2517755972	2642345865	2642350334	2510267472	-	-	-
2508837135	2517754396	2642345309	2642348541	2510265511	tRNA (uracil-5-)-methyltransferase	-	-
2508836931	2517756842	2642347393	2642352907	2510266907	-	-	-
2508835086	2517757798	2642343606	2642350725	2510267713	Uncharacterized conserved protein YbbC, DUF1343 family	-	-
2508836070	2517758426	2642343865	2642348603	2510267852	phosphate transport system protein	-	-
2508836255	2517754476	2642343967	2642352249	2510264611	-	-	-
2508837250	2517757302	2642345994	2642352571	2510265693	-	-	-
2508836910	2517757179	2642346697	2642348764	2510266928	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	Yes	2.24724E-12
2508835288	2517755121	2642346604	2642351385	2510264398	-	-	-
2508839318	2517756364	2642347052	2642350276	2510269036	putative transcriptional regulator	-	-
2508838214	2517754607	2642346394	2642349874	2510264193	-	Yes	3.41543E-16
2508837935	2517754562	2642345840	2642348737	2510265897	-	-	-
2508836462	2517758781	2642344622	2642350877	2510267408	L-fucose mutarotase	-	-
2508834886	2517755639	2642347469	2642353111	2510269184	transcriptional regulator, LacI family	-	-
2508839984	2517754050	2642345425	2642349563	2510269865	-	-	-
2508836246	2517755981	2642344784	2642350323	2510264605	-	-	-
2508839233	2517754978	2642347887	2642349665	2510264673	large subunit ribosomal protein L27	-	-
2508838759	2517754156	2642345331	2642349468	2510265175	queuine tRNA-ribosyltransferase	-	-
2508839417	2517756976	2642347229	2642352742	2510268967	pyridoxine 5-phosphate synthase	-	-
2508838320	2517754143	2642345340	2642349477	2510265487	-	-	-
2508837449	2517754375	2642343909	2642348646	2510266374	AraC-type DNA-binding protein	-	-
2508839645	2517757331	2642346021	2642352544	2510268738	Predicted oxidoreductase	-	-
2508837514	2517755691	2642347433	2642353074	2510266302	F-type H+-transporting ATPase subunit alpha	-	-
2508839847	2517755132	2642346591	2642351397	2510268446	biopolymer transport protein ExbB	-	-
2508836928	2517756841	2642347394	2642352908	2510266910	-	-	-
2508835045	2517755231	2642347634	2642351300	2510267672	cephalosporin-C deacetylase	-	-

2508836358	2517753984	2642345123	2642352370	2510267526	two-component system, NarL family, sensor kinase	-	-
2508835199	2517756887	2642347344	2642352857	2510264484	BirA family transcriptional regulator, biotin operon repressor / biotin-[acetyl-CoA-carboxylase] ligase	Yes	1.10275E-06
2508839895	2517756831	2642345581	2642351773	2510264746	-	-	-
2508838890	2517755548	2642345019	2642348953	2510265061	two-component system, OmpR family, phosphate regulon response regulator PhoB	-	-
2508839493	2517758293	2642343383	2642351934	2510268909	flagellar biosynthetic protein FliR	-	-
2508835351	2517755537	2642345011	2642348961	2510264327	arginyl-tRNA synthetase	Yes	3.70282E-06
2508836887	2517756431	2642346751	2642352219	2510266965	Threonine dehydrogenase	-	-
2508836470	2517756730	2642345227	2642351009	2510267400	ribonuclease HIII	-	-
2508839518	2517758671	2642344866	2642353033	2510268884	purine-binding chemotaxis protein CheW	-	-
2508835431	2517757867	2642347520	2642351221	2510269507	acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	-	-
2508835137	2517754370	2642343914	2642348652	2510264537	-	-	-
2508835397	2517757321	2642346013	2642352552	2510269467	-	-	-
2508839420	2517756519	2642346797	2642352173	2510268964	agmatine deiminase	-	-
2508835398	2517757383	2642346074	2642352491	2510269468	FKBP-type peptidyl-prolyl cis-trans isomerase FkpA	-	-
2508836060	2517754410	2642345290	2642348522	2510267856	ABC-2 type transport system permease protein	-	-
2508837489	2517755707	2642347417	2642353057	2510266325	Phosphoglycerate dehydrogenase	-	-
2508834971	2517758513	2642344332	2642348901	2510269092	hypothetical protein	-	-
2508836759	2517754953	2642347912	2642349639	2510267066	16S rRNA (cytidine1402-2'-O)-methyltransferase	-	-
2508837720	2517755744	2642346357	2642350450	2510266065	uracil permease	Yes	1.6288E-08
2508837439	2517756013	2642344815	2642351456	2510266382	RNA polymerase sigma-70 factor, ECF subfamily	Yes	1.06019E-07
2508837075	2517758770	2642344612	2642350889	2510266721	-	-	-
2508840075	2517754109	2642345372	2642349509	2510269943	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	-	-
2508840160	2517758596	2642344945	2642352954	2510270028	3-deoxy-D-manno-octulosonic-acid transferase	-	-
2508838326	2517755436	2642343790	2642348406	2510265480	-	-	-
2508837394	2517756510	2642346788	2642352183	2510266407	beta-galactosidase	-	-
2508838864	2517754999	2642347868	2642349684	2510265080	DNA polymerase-3 subunit gamma/tau	-	-
2508836820	2517757108	2642343664	2642348246	2510267021	-	-	-
2508837890	2517757434	2642346118	2642352446	2510265858	pyrimidine operon attenuation protein / uracil	-	-

					phosphoribosyltransferase		
2508839505	2517758305	2642343371	2642351922	2510268897	flagellum-specific ATP synthase	-	-
2508839307	2517754433	2642345268	2642348502	2510267647	putative oxidoreductase	-	-
2508835503	2517754886	2642343574	2642351045	2510269571	-	Yes	2.04055E-05
2508836963	2517756722	2642345236	2642351019	2510266856	-	-	-
2508835655	2517756293	2642345731	2642350641	2510268257	CRISP-associated protein Cas1	-	-
2508837019	2517756749	2642345210	2642350993	2510266780	NADH dehydrogenase subunit B	-	-
2508835221	2517754698	2642346488	2642349781	2510264462	large subunit ribosomal protein L6	-	-
2508836850	2517754439	2642345262	2642348496	2510266995	-	-	-
2508834872	2517756923	2642347285	2642352798	2510269200	-	-	-
2508837176	2517758145	2642347699	2642350096	2510265564	biopolymer transport protein ExbD	-	-
2508837937	2517756636	2642346924	2642352048	2510265899	transcriptional regulator, IclR family	-	-
2508834663	2517758057	2642344478	2642349994	2510269364	transcription termination factor Rho	Yes	7.65848E-24
2508839972	2517753758	2642346972	2642351575	2510269846	-	-	-
2508838128	2517753865	2642345506	2642351847	2510265776	two-component system, chemotaxis family, CheB/CheR fusion protein	-	-
2508837120	2517756724	2642345234	2642351017	2510266675	DNA ligase (NAD+)	-	-
2508834657	2517754131	2642345352	2642349488	2510269370	succinyl-CoA synthetase beta subunit	Yes	4.10676E-25
2508836278	2517755232	2642347633	2642351299	2510264635	hypothetical protein	-	-
2508838862	2517758318	2642343356	2642351907	2510265082	-	-	-
2508839981	2517758131	2642344562	2642350077	2510269862	hypothetical protein	-	-
2508835780	2517756986	2642347220	2642352733	2510268046	Rhodanese-related sulfurtransferase	Yes	2.20128E-08
2508839955	2517753863	2642345504	2642351849	2510269824	-	-	-
2508840427	2517755313	2642348115	2642351726	2510264101	AraC-type DNA-binding protein	-	-
2508839266	2517754187	2642344765	2642350737	2510267596	DNA-binding regulatory protein, YebC/PmpR family	-	-
2508838686	2517757885	2642345170	2642349245	2510265313	hypothetical protein	-	-
2508835219	2517754700	2642346490	2642349779	2510264464	small subunit ribosomal protein S14	-	-
2508836463	2517758780	2642344621	2642350878	2510267407	Aldo/keto reductase	-	-
2508840341	2517758672	2642344012	2642353040	2510264097	chromosomal replication initiator protein	-	-

2508836274	2517758594	2642344947	2642352952	2510264631	-	-	-
2508838284	2517756120	2642343170	2642348305	2510266646	alpha,alpha-trehalase	Yes	1.04585E-07
2508838294	2517755464	2642347287	2642352800	2510264717	-	-	-
2508839414	2517755382	2642348040	2642351651	2510268970	-	-	-
2508835478	2517757334	2642346024	2642352541	2510269547	ribulose-phosphate 3-epimerase	-	-
2508835979	2517754960	2642347905	2642349646	2510267923	3-dehydroquinate dehydratase	-	-
2508840155	2517758025	2642344448	2642349963	2510270025	-	-	-
2508835641	2517757397	2642346087	2642352478	2510269738	-	-	-
2508837077	2517758773	2642344614	2642350885	2510266719	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508837529	2517758588	2642344953	2642352946	2510266287	-	-	-
2508839399	2517758337	2642343413	2642351515	2510268997	queuine tRNA-ribosyltransferase	-	-
2508839263	2517754190	2642344762	2642350739	2510267593	-	-	-
2508837692	2517755100	2642346619	2642351370	2510264550	-	-	-
2508835334	2517758643	2642344903	2642352997	2510264361	Fe-only nitrogenase FeFe protein subunit AnfD precursor	-	-
2508839021	2517757031	2642347177	2642352687	2510267252	-	-	-
2508836998	2517757102	2642343209	2642348237	2510269105	DNA-binding transcriptional regulator, LacI/PurR family	-	-
2508835014	2517757884	2642345171	2642349246	2510269049	-	-	-
2508839019	2517757033	2642347175	2642352685	2510267250	-	Yes	1.4746E-08
2508836885	2517755750	2642346351	2642350456	2510268835	LacI family transcriptional regulator	-	-
2508836950	2517757652	2642346264	2642351965	2510266870	lipoprotein-releasing system ATP-binding protein	-	-
2508839168	2517755352	2642345145	2642349220	2510264811	-	-	-
2508836552	2517755072	2642344185	2642348369	2510267322	-	-	-
2508839188	2517756214	2642345642	2642350553	2510266256	ATP-dependent DNA helicase RecG	-	-
2508839626	2517755284	2642347585	2642351250	2510268763	prepilin-type N-terminal cleavage/methylation domain-containing protein	-	-
2508835278	2517756240	2642345673	2642350582	2510264407	-	-	-
2508840072	2517758709	2642343319	2642350948	2510269935	-	-	-
2508836316	2517757427	2642346111	2642352454	2510267564	-	-	-
2508838607	2517757356	2642346047	2642352518	2510265376	-	-	-
2508839602	2517754944	2642347926	2642349625	2510268786	-	-	-

2508836842	2517754864	2642343550	2642351070	2510267003	-	-	-
2508834733	2517756326	2642345762	2642350672	2510268806	-	-	-
2508838536	2517758742	2642343287	2642350916	2510265386	DNA mismatch repair protein MutL	-	-
2508837722	2517755414	2642348003	2642351615	2510266063	-	-	-
2508837859	2517756156	2642344125	2642349072	2510265759	RNA polymerase sigma-70 factor, ECF subfamily	-	-
2508836189	2517756098	2642347649	2642351314	2510267737	-	-	-
2508840563	2517753834	2642345477	2642351876	2510264219	-	-	-
2508836623	2517758324	2642343351	2642351902	2510264982	iron complex transport system ATP-binding protein	-	-
2508839454	2517754513	2642343711	2642351176	2510268945	-	-	-
2508836658	2517756534	2642346812	2642352159	2510265015	creatinine amidohydrolase	-	-
2508840190	2517757534	2642348063	2642351674	2510270053	Nitrate/nitrite transporter NarK	-	-
2508836618	2517755262	2642347607	2642351273	2510264977	zinc protease	-	-
2508838786	2517755971	2642345866	2642350335	2510265159	hypothetical protein	-	-
2508838812	2517754715	2642346508	2642349762	2510265126	-	-	-
2508839463	2517758316	2642343360	2642351911	2510268936	Signal transduction histidine kinase	-	-
2508834981	2517754786	2642344276	2642349416	2510269082	glycoside/pentoside/hexuronide:cation symporter, GPH family	-	-
2508838098	2517757599	2642346200	2642352030	2510265824	RNA methyltransferase, TrmH family	-	-
2508834629	2517757854	2642347532	2642351208	2510269427	-	-	-
2508836343	2517758590	2642344951	2642352948	2510267542	leucyl-tRNA synthetase	-	-
2508838270	2517757185	2642346704	2642348758	2510266633	-	-	-
2508838156	2517755786	2642346309	2642350495	2510266515	ABC-2 type transport system ATP-binding protein	-	-
2508836234	2517754624	2642346410	2642349857	2510264594	L-threonylcarbamoyladenylate synthase	Yes	2.91972E-17
2508834991	2517757535	2642348064	2642351675	2510269072	Lysophospholipase L1	-	-
2508835472	2517754908	2642344362	2642348450	2510269543	DNA polymerase-3 subunit alpha	-	-
2508835572	2517754057	2642345417	2642349556	2510269675	glutamyl-tRNA reductase	-	-
2508838938	2517754227	2642344719	2642350781	2510267174	-	-	-
2508840449	2517757361	2642346053	2642352512	2510264127	iron(III) transport system ATP-binding protein	-	-

Table S 2 Upregulated genes

Locus_Tag	Function	Fold Change	Padj	COG	CAZy	Positive Selection	Core Genome	Species specific genome	Unique genome
ObacDRAFT_0034	Type IV secretory pathway	1.25	0.063	-	-	-	-	-	-
ObacDRAFT_0050	hypothetical protein	1.08	0.079	-	-	-	-	Yes	-
ObacDRAFT_0080	PEP-CTERM protein sorting domain	2.07	0.001	-	-	-	-	-	-
ObacDRAFT_0082	hypothetical protein	1.24	0.063	-	-	-	-	-	-
ObacDRAFT_0097	Endoglucanase	1.20	0.050	-	GH5	Yes	Yes	-	-
ObacDRAFT_0100	prepilin-type N-terminal cleavage	1.48	0.031	-	-	-	-	Yes	-
ObacDRAFT_0101	Predicted glycosylase	1.26	0.077	carbohydrate	GH130	-	-	Yes	-
ObacDRAFT_0113	prepilin-type N-terminal cleavage	1.30	0.097	-	-	-	-	Yes	-
ObacDRAFT_0124	hypothetical protein	1.65	0.007	-	-	-	-	Yes	-
ObacDRAFT_0205	hypothetical protein	1.37	0.085	-	-	-	-	-	-
ObacDRAFT_0227	hypothetical protein	1.34	0.091	-	-	-	-	-	-
ObacDRAFT_0324	hypothetical protein	1.22	0.075	-	-	-	-	-	-
ObacDRAFT_0325	hypothetical protein	1.45	0.027	-	-	-	-	-	-
ObacDRAFT_0326	hypothetical protein	1.30	0.068	-	-	-	-	-	-

ObacDRAFT_0328	Site-specific recombinase XerD	1.15	0.063	Replication	-	-	-	Yes	-
ObacDRAFT_0350	prepilin-type N-terminal cleavage	1.40	0.031	-	-	-	-	-	-
ObacDRAFT_0412	vacuolar-type H(+)-translocating pyrophosphatase	1.67	0.033	Energy	-	-	Yes	-	-
ObacDRAFT_0635	Concanavalin A-like lectin/glucanases superfamily.	1.46	0.010	-	-	-	-	-	-
ObacDRAFT_0636	prepilin-type N-terminal cleavage	1.14	0.080	-	-	-	-	Yes	-
ObacDRAFT_0710	hypothetical protein	1.01	0.092	-	CBM9|GH39	-	-	Yes	-
ObacDRAFT_0712	prepilin-type N-terminal cleavage	1.68	0.005	cell motility	-	-	-	-	-
ObacDRAFT_0783	hypothetical protein	1.75	0.002	-	-	-	-	Yes	-
ObacDRAFT_0784	hypothetical protein	1.11	0.060	-	CBM56	-	-	Yes	-
ObacDRAFT_0823	prepilin-type N-terminal cleavage	1.06	0.063	cell motility	-	-	Yes	-	-
ObacDRAFT_0926	hypothetical protein	1.27	0.060	-	GH74	-	Yes	-	-
ObacDRAFT_0927	hypothetical protein	1.76	0.011	-	-	-	Yes	-	-
ObacDRAFT_0928	hypothetical protein	1.39	0.042	-	-	-	-	-	-
ObacDRAFT_0929	prepilin-type N-terminal cleavage	1.42	0.049	-	-	-	Yes	-	-
ObacDRAFT_0959	PEP-CTERM protein sorting domain	1.08	0.085	-	-	-	Yes	-	-
ObacDRAFT_0960	Pectate lyase superfamily protein	1.06	0.066	-	CBM32	Yes	Yes	-	-
ObacDRAFT_0961	prepilin-type N-terminal cleavage	1.59	0.005	-	-	-	Yes	-	-
ObacDRAFT_0962	Concanavalin A-like lectin/glucanases superfamily.	1.30	0.033	-	-	-	-	Yes	-
ObacDRAFT_1099	Concanavalin A-like lectin/glucanases superfamily.	1.34	0.022	-	-	-	-	-	-
ObacDRAFT_1153	Metal-dependent hydrolase	1.00	0.097	General	-	-	Yes	-	-
ObacDRAFT_1154	Lactoylglutathione lyase and related lyases	1.25	0.049	Secondary metabolites	-	-	Yes	-	-
ObacDRAFT_1177	prepilin-type N-terminal cleavage	1.33	0.051	-	-	-	Yes	-	-
ObacDRAFT_1178	hypothetical protein	1.41	0.047	-	-	-	-	-	-
ObacDRAFT_1225	hypothetical protein	1.21	0.031	-	-	-	-	Yes	-
ObacDRAFT_1273	ORF6N domain.	1.35	0.031	-	-	-	-	Yes	-
ObacDRAFT_1327	hypothetical protein	1.72	0.011	-	-	-	-	Yes	-
ObacDRAFT_1377	PEP-CTERM protein sorting domain	1.27	0.043	-	-	-	-	-	-
ObacDRAFT_1391	LysM domain.	2.53	0.000	-	-	-	-	-	-
ObacDRAFT_1392	Uncharacterized protein conserved in bacteria	2.96	0.000	-	-	-	-	-	-

ObacDRAFT_1393	hypothetical protein	1.73	0.009	-	-	-	-	-	-
ObacDRAFT_1394	hypothetical protein	1.67	0.012	-	-	-	-	-	-
ObacDRAFT_1395	hypothetical protein	1.83	0.007	-	-	-	-	-	-
ObacDRAFT_1396	Orthopoxvirus protein of unknown function (DUF830).	1.64	0.008	-	-	-	-	-	-
ObacDRAFT_1397	hypothetical protein	2.23	0.000	-	-	-	-	-	-
ObacDRAFT_1398	hypothetical protein	2.24	0.000	-	-	-	-	Yes	-
ObacDRAFT_1458	Transcriptional regulators	1.38	0.070	-	-	-	-	Yes	-
ObacDRAFT_1459	PEP-CTERM protein sorting domain	1.53	0.011	-	-	-	-	-	-
ObacDRAFT_1461	prepilin-type N-terminal cleavage	1.65	0.021	cell motility	-	-	-	-	-
ObacDRAFT_1462	hypothetical protein	1.46	0.059	-	-	-	Yes	-	-
ObacDRAFT_1484	hypothetical protein	1.08	0.098	-	-	-	-	-	-
ObacDRAFT_1502	Zn-finger domain associated with topoisomerase type I	1.19	0.039	-	-	-	-	Yes	-
ObacDRAFT_1523	prepilin-type N-terminal cleavage	1.29	0.031	-	-	-	Yes	-	-
ObacDRAFT_1598	hypothetical protein	1.28	0.081	-	-	-	-	-	-
ObacDRAFT_1599	PEP-CTERM motif.	1.54	0.009	-	-	-	-	-	-
ObacDRAFT_1668	hypothetical protein	1.55	0.023	-	-	-	-	-	-
ObacDRAFT_1669	prepilin-type N-terminal cleavage	1.17	0.083	cell motility	-	-	-	-	-
ObacDRAFT_1696	Bacterial pre-peptidase C-terminal domain.	1.40	0.051	-	-	-	Yes	-	-
ObacDRAFT_1747	PEP-CTERM protein sorting domain	1.56	0.032	-	-	-	-	-	-
ObacDRAFT_1769	AraC-type DNA-binding domain-containing proteins	1.17	0.041	Transcription	-	-	-	-	-
ObacDRAFT_1770	Alpha/beta hydrolase family.	1.07	0.080	-	-	-	-	Yes	-
ObacDRAFT_1802	prepilin-type N-terminal cleavage	1.32	0.077	cell motility	-	-	-	-	-
ObacDRAFT_1809	prepilin-type N-terminal cleavage	1.12	0.085	-	-	-	Yes	-	-
ObacDRAFT_1824	prepilin-type N-terminal cleavage	1.98	0.008	-	-	-	-	-	-
ObacDRAFT_1825	hypothetical protein	1.95	0.007	-	-	-	-	-	-
ObacDRAFT_1851	prepilin-type N-terminal cleavage	1.26	0.098	-	-	-	-	-	-
ObacDRAFT_1908	Carbon dioxide concentrating mechanism	1.05	0.099	-	-	-	Yes	-	-
ObacDRAFT_1935	hypothetical protein	1.33	0.060	-	GH39	-	-	-	-
ObacDRAFT_1936	hypothetical protein	1.17	0.070	-	-	-	-	-	-

ObacDRAFT_1937	hypothetical protein	1.56	0.010	-	-	-	-	-	-
ObacDRAFT_1938	prepilin-type N-terminal cleavage	1.76	0.004	-	-	-	-	-	-
ObacDRAFT_1996	Site-specific recombinase XerC	1.18	0.070	-	-	-	-	-	-
ObacDRAFT_1997	hypothetical protein	1.71	0.009	-	-	-	-	-	-
ObacDRAFT_2002	hypothetical protein	1.27	0.060	-	-	-	-	-	-
ObacDRAFT_2003	hypothetical protein	1.72	0.005	-	-	-	-	-	-
ObacDRAFT_2004	hypothetical protein	1.55	0.007	-	-	-	-	-	-
ObacDRAFT_2006	hypothetical protein	1.32	0.041	-	-	-	-	-	-
ObacDRAFT_2007	hypothetical protein	1.74	0.002	-	-	-	-	-	-
ObacDRAFT_2008	hypothetical protein	1.62	0.010	-	-	-	-	-	-
ObacDRAFT_2009	hypothetical protein	1.29	0.066	-	-	-	-	-	-
ObacDRAFT_2011	hypothetical protein	1.17	0.083	-	-	-	-	-	-
ObacDRAFT_2012	Dehydrogenases with different specificities	1.31	0.025	General	-	-	-	-	-
ObacDRAFT_2021	Acetyltransferase (isoleucine patch superfamily)	1.38	0.039	General	GT8	-	-	-	-
ObacDRAFT_2036	autotransporter-associated beta strand repeat	1.18	0.059	-	-	-	-	Yes	-
ObacDRAFT_2053	PEP-CTERM protein sorting domain	1.44	0.015	-	-	-	-	-	-
ObacDRAFT_2116	hypothetical protein	1.41	0.085	-	-	-	-	-	Yes
ObacDRAFT_2120	hypothetical protein	2.11	0.003	-	-	-	-	-	-
ObacDRAFT_2171	prepilin-type N-terminal cleavage	1.28	0.024	-	-	-	Yes	-	-
ObacDRAFT_2176	PEP-CTERM protein sorting domain	1.11	0.072	-	-	-	-	-	-
ObacDRAFT_2257	Virulence protein	1.10	0.065	Function unknown	-	-	-	-	-
ObacDRAFT_2287	Esterase/lipase	1.37	0.063	Lipid	CE8	-	-	Yes	-
ObacDRAFT_2290	hypothetical protein	1.39	0.048	-	-	-	-	Yes	-
ObacDRAFT_2291	Predicted glycosylase	1.21	0.084	carbohydrate	GH130	-	Yes	-	-
ObacDRAFT_2322	hypothetical protein	1.36	0.075	-	-	-	-	-	-
ObacDRAFT_2326	Methyltransferase domain.	1.65	0.007	-	CBM47	-	-	-	-
ObacDRAFT_2327	Transposase and inactivated derivatives	1.39	0.042	-	-	-	-	-	-
ObacDRAFT_2380	hypothetical protein	1.15	0.058	-	-	-	Yes	-	-
ObacDRAFT_2387	hypothetical protein	1.24	0.074	-	-	-	-	-	-

ObacDRAFT_2389	PEP-CTERM protein sorting domain	1.77	0.005	-	-	-	-	-	-
ObacDRAFT_2390	Lysophospholipase L1 and related esterases	1.35	0.030	Amino Acid	-	-	-	-	-
ObacDRAFT_2391	prepilin-type N-terminal cleavage	1.12	0.079	cell motility	-	-	Yes	-	-
ObacDRAFT_2397	prepilin-type N-terminal cleavage	1.29	0.018	-	-	-	-	-	-
ObacDRAFT_2403	hypothetical protein	1.98	0.005	-	-	-	-	-	-
ObacDRAFT_2412	His-Xaa-Ser system protein HsxD	1.23	0.048	-	-	-	-	-	-
ObacDRAFT_2414	TonB family C-terminal domain	2.17	0.001	-	-	-	-	-	-
ObacDRAFT_2415	hypothetical protein	1.65	0.009	-	-	-	-	-	-
ObacDRAFT_2440	hypothetical protein	1.17	0.050	-	-	-	-	-	-
ObacDRAFT_2454	prepilin-type N-terminal cleavage	1.49	0.013	-	-	-	-	-	-
ObacDRAFT_2479	Predicted transcriptional regulators	1.07	0.075	-	-	-	-	-	-
ObacDRAFT_2488	prepilin-type N-terminal cleavage	1.35	0.054	-	-	-	-	-	-
ObacDRAFT_2529	hypothetical protein	1.34	0.061	-	-	-	-	Yes	-
ObacDRAFT_2537	RNA polymerase sigma factor, sigma-70 family	1.69	0.018	Transcription	-	-	Yes	-	-
ObacDRAFT_2550	Right handed beta helix region.	1.57	0.018	-	-	-	-	-	-
ObacDRAFT_2569	CRISPR type III-B/RAMP module RAMP protein Cmr6	1.10	0.092	-	-	-	-	-	-
ObacDRAFT_2570	putative CRISPR-associated protein, APE2256 family	1.19	0.060	defense	-	-	-	-	-
ObacDRAFT_2601	KilA-N domain.	1.13	0.052	-	-	-	-	Yes	-
ObacDRAFT_2604	prepilin-type N-terminal cleavage	1.26	0.099	-	-	-	-	-	-
ObacDRAFT_2630	prepilin-type N-terminal cleavage	1.30	0.060	cell motility	-	-	-	-	-
ObacDRAFT_2631	Autotransporter-associated beta strand repeat.	1.35	0.012	-	-	-	-	-	-
ObacDRAFT_2666	TIGR02436 family protein	1.18	0.054	-	-	-	-	-	-
ObacDRAFT_2668	hypothetical protein	1.23	0.041	-	-	-	-	-	-
ObacDRAFT_2689	hypothetical protein	1.65	0.005	-	-	-	-	Yes	-
ObacDRAFT_2692	S23 ribosomal protein.	1.01	0.085	-	-	-	-	Yes	-
ObacDRAFT_2713	Zn-dependent dehydrogenases	1.12	0.060	Amino Acid	-	-	Yes	-	-
ObacDRAFT_2714	Predicted dehydrogenases and related proteins	1.01	0.092	General	-	-	Yes	-	-
ObacDRAFT_2718	putative NAD(P)H quinone oxidoreductase, PIG3 family	1.28	0.055	Energy	-	-	Yes	-	-

ObacDRAFT_2719	Predicted metal-dependent hydrolase	1.38	0.060	General	-	-	Yes	-	-
ObacDRAFT_2742	hypothetical protein	1.11	0.083	-	-	-	-	-	Yes
ObacDRAFT_2781	prepilin-type N-terminal cleavage	1.21	0.090	cell motility	-	-	-	-	-
ObacDRAFT_2822	Subtilase family.	1.13	0.053	-	-	-	-	-	-
ObacDRAFT_2835	hypothetical protein	2.68	0.000	-	-	-	Yes	-	-
ObacDRAFT_2847	prepilin-type N-terminal cleavage	1.25	0.059	cell motility	-	-	Yes	-	-
ObacDRAFT_2929	hypothetical protein	1.49	0.016	-	-	-	-	-	-
ObacDRAFT_2933	L-ribulose 5-phosphate 4-epimerase (EC 5.1.3.4)	1.29	0.059	carbohydrate	-	-	Yes	-	-
ObacDRAFT_2934	hypothetical protein	1.39	0.049	-	-	-	-	-	-
ObacDRAFT_2938	hypothetical protein	1.39	0.076	-	-	-	-	-	-
ObacDRAFT_2943	Putative phage tail protein.	1.20	0.045	-	-	-	-	-	-
ObacDRAFT_2951	hypothetical protein	1.29	0.069	-	-	-	-	-	-
ObacDRAFT_2952	hypothetical protein	1.45	0.032	-	-	-	-	-	-
ObacDRAFT_2959	Bacteriophage tail assembly protein	1.23	0.050	Mobile	-	-	-	-	-
ObacDRAFT_2965	Phage integrase family.	1.10	0.053	-	-	-	-	-	-
ObacDRAFT_2966	ASCH domain.	1.21	0.033	-	-	-	-	-	-
ObacDRAFT_2967	hypothetical protein	2.68	0.000	-	-	-	-	-	-
ObacDRAFT_2968	ATPase associated with various cellular activities (AAA).	2.12	0.000	-	-	-	-	-	-
ObacDRAFT_2969	hypothetical protein	1.89	0.001	-	-	-	-	-	-
ObacDRAFT_2970	SprT-like family.	2.24	0.000	-	-	-	-	-	-
ObacDRAFT_2971	hypothetical protein	2.43	0.000	-	-	-	-	-	-
ObacDRAFT_2972	hypothetical protein	1.15	0.077	-	-	-	-	-	-
ObacDRAFT_2977	hypothetical protein	1.36	0.058	-	-	-	-	-	-
ObacDRAFT_2983	hypothetical protein	1.80	0.005	-	-	-	-	-	Yes
ObacDRAFT_2984	hypothetical protein	1.64	0.006	-	-	-	-	-	-
ObacDRAFT_2985	hypothetical protein	1.68	0.008	-	-	-	-	-	-
ObacDRAFT_2986	VRR-NUC domain.	1.25	0.053	-	-	-	-	-	-
ObacDRAFT_2988	PD-(D/E)XK nuclease superfamily.	1.59	0.005	-	-	-	-	-	-
ObacDRAFT_2989	hypothetical protein	2.03	0.007	-	-	-	-	-	-

ObacDRAFT_2990	hypothetical protein	1.05	0.087	-	-	-	-	-	-
ObacDRAFT_2991	hypothetical protein	1.40	0.015	-	-	-	-	-	-
ObacDRAFT_2992	hypothetical protein	1.54	0.010	-	-	-	-	-	-
ObacDRAFT_2994	hypothetical protein	2.15	0.000	-	-	-	-	-	-
ObacDRAFT_2997	hypothetical protein	1.00	0.098	-	-	-	-	-	-
ObacDRAFT_3022	hypothetical protein	1.37	0.072	-	-	-	-	-	-
ObacDRAFT_3024	Domain of unknown function (DUF4276).	1.39	0.031	-	-	-	-	Yes	-
ObacDRAFT_3175	hypothetical protein	1.57	0.009	-	-	-	-	-	-
ObacDRAFT_3176	prepilin-type N-terminal cleavage	1.56	0.018	cell motility	-	-	-	-	-
ObacDRAFT_3200	hypothetical protein	1.09	0.059	-	-	-	-	-	-
ObacDRAFT_3312	hypothetical protein	1.11	0.053	-	-	-	Yes	-	-
ObacDRAFT_3313	hypothetical protein	1.22	0.046	-	-	-	Yes	-	-
ObacDRAFT_3314	hypothetical protein	1.65	0.011	-	-	-	Yes	-	-
ObacDRAFT_3316	prepilin-type N-terminal cleavage	1.67	0.019	cell motility	-	-	Yes	-	-
ObacDRAFT_3333	prepilin-type N-terminal cleavage	1.11	0.080	cell motility	-	Yes	Yes	-	-
ObacDRAFT_3414	hypothetical protein	1.78	0.001	-	-	-	-	-	-
ObacDRAFT_3415	hypothetical protein	1.41	0.010	-	-	-	-	-	-
ObacDRAFT_3497	hypothetical protein	1.47	0.049	-	-	-	-	Yes	-
ObacDRAFT_3500	hypothetical protein	1.37	0.045	-	-	-	-	-	-
ObacDRAFT_3514	hypothetical protein	1.34	0.094	-	-	-	-	Yes	-
ObacDRAFT_3526	PEP-CTERM protein sorting domain	1.51	0.010	-	-	-	Yes	-	-
ObacDRAFT_3554	hypothetical protein	1.38	0.076	-	-	-	-	-	-
ObacDRAFT_3559	hypothetical protein	1.67	0.012	-	-	-	-	-	-
ObacDRAFT_3626	hypothetical protein	1.41	0.039	-	-	-	-	-	-
ObacDRAFT_3627	prepilin-type N-terminal cleavage	1.56	0.039	cell motility	-	-	-	Yes	-
ObacDRAFT_3650	hypothetical protein	1.10	0.088	-	-	-	Yes	-	-
ObacDRAFT_3680	Predicted transcriptional regulator	1.09	0.080	Transcription	-	-	-	-	-
ObacDRAFT_3687	hypothetical protein	1.14	0.095	-	-	-	-	-	-
ObacDRAFT_3688	hypothetical protein	1.11	0.095	-	GH105	-	-	-	-

ObacDRAFT_3689	prepilin-type N-terminal cleavage	1.27	0.095	-	-	-	-	-	-
ObacDRAFT_3690	hypothetical protein	1.13	0.060	-	-	-	-	-	-
ObacDRAFT_3728	hypothetical protein	1.31	0.074	-	-	-	-	-	-
ObacDRAFT_3741	hypothetical protein	1.13	0.044	-	-	-	-	-	-
ObacDRAFT_3746	Protein of unknown function (DUF4043).	1.70	0.012	-	-	-	-	-	-
ObacDRAFT_3747	hypothetical protein	1.46	0.042	-	-	-	-	-	-
ObacDRAFT_3748	hypothetical protein	1.49	0.039	-	-	-	-	-	-
ObacDRAFT_3751	hypothetical protein	1.46	0.061	-	-	-	-	-	-
ObacDRAFT_3752	hypothetical protein	1.66	0.005	-	-	-	-	-	-
ObacDRAFT_3753	hypothetical protein	1.30	0.073	-	-	-	-	-	-
ObacDRAFT_3761	Replicative DNA helicase	1.20	0.077	-	-	-	-	-	-
ObacDRAFT_3790	hypothetical protein	1.41	0.070	-	-	-	-	-	-
ObacDRAFT_3791	Predicted transcriptional regulators	1.54	0.036	Transcription	-	-	-	-	-
ObacDRAFT_3792	GTPases - translation elongation factors	1.54	0.022	-	-	-	-	-	-
ObacDRAFT_3793	hypothetical protein	1.77	0.003	-	-	-	-	-	-
ObacDRAFT_3794	hypothetical protein	1.64	0.012	-	-	-	-	-	-
ObacDRAFT_3804	hypothetical protein	1.27	0.099	-	-	-	-	-	-
ObacDRAFT_3855	Predicted cobalamin binding protein	1.41	0.025	Energy	-	-	-	Yes	-
ObacDRAFT_3920	PEP-CTERM protein sorting domain	1.37	0.029	-	-	-	-	-	-
ObacDRAFT_3923	hypothetical protein	2.51	0.000	-	-	-	Yes	-	-
ObacDRAFT_3924	carbohydrate ABC transporter ATP-binding protein	2.01	0.002	carbohydrate	-	-	Yes	-	-
ObacDRAFT_3925	maltooligosaccharide ABC transporter membrane protein	1.61	0.032	carbohydrate	-	-	Yes	-	-
ObacDRAFT_3927	maltooligosaccharide-binding protein	1.41	0.027	carbohydrate	-	-	Yes	-	-
ObacDRAFT_3928	Maltoporin (phage lambda and maltose receptor)	1.55	0.025	carbohydrate	-	-	Yes	-	-
ObacDRAFT_3941	hypothetical protein	1.55	0.029	-	-	-	-	-	-
ObacDRAFT_3958	hypothetical protein	1.73	0.018	-	-	-	-	-	Yes
ObacDRAFT_3959	hypothetical protein	1.36	0.054	-	-	-	-	-	Yes
ObacDRAFT_3979	hypothetical protein	0.97	0.092	-	-	-	-	-	-
ObacDRAFT_4008	hypothetical protein	1.40	0.059	-	-	-	-	-	-

ObacDRAFT_4032	hypothetical protein	1.20	0.033	-	-	-	-	-	Yes
ObacDRAFT_4103	prepilin-type N-terminal cleavage	1.29	0.082	-	-	-	-	-	-
ObacDRAFT_4113	hypothetical protein	1.64	0.037	-	-	-	-	-	-
ObacDRAFT_4116	Transposase and inactivated derivatives, IS1 family	1.23	0.065	Mobile	-	-	-	-	-
ObacDRAFT_4117	hypothetical protein	1.68	0.007	-	-	-	-	-	Yes
ObacDRAFT_4118	hypothetical protein	1.72	0.021	-	-	-	-	-	-
ObacDRAFT_4119	hypothetical protein	1.43	0.033	-	-	-	-	-	-
ObacDRAFT_4168	hypothetical protein	1.11	0.077	-	-	-	Yes	-	-
ObacDRAFT_4173	NADH dehydrogenase subunit M (EC 1.6.5.3)	1.24	0.038	Energy	-	-	Yes	-	-
ObacDRAFT_4174	NADH dehydrogenase subunit L (EC 1.6.5.3)	1.18	0.066	Energy	-	-	Yes	-	-
ObacDRAFT_4175	PD-(D/E)XK nuclease superfamily.	1.50	0.030	Replication	-	-	-	Yes	-
ObacDRAFT_4266	hypothetical protein	1.54	0.027	-	-	-	-	-	-
ObacDRAFT_4273	hypothetical protein	1.06	0.085	-	-	-	Yes	-	-
ObacDRAFT_4286	prepilin-type N-terminal cleavage	1.14	0.065	-	-	-	-	-	-
ObacDRAFT_4327	Lysophospholipase L1 and related esterases	1.29	0.022	Amino Acid	-	-	-	Yes	-
ObacDRAFT_4347	hypothetical protein	1.50	0.010	-	-	-	Yes	-	-
ObacDRAFT_4361	Endoglucanase	1.59	0.037	carbohydrate	GH5	-	-	Yes	-
ObacDRAFT_4362	prepilin-type N-terminal cleavage	1.66	0.015	-	-	-	-	-	-
ObacDRAFT_4363	hypothetical protein	1.07	0.087	-	-	-	-	Yes	-
ObacDRAFT_4364	PEP-CTERM protein sorting domain	1.33	0.029	-	-	-	-	Yes	-
ObacDRAFT_4428	Biopolymer transport proteins	1.28	0.068	Intracellular	-	-	Yes	-	-
ObacDRAFT_4430	Biopolymer transport protein	1.24	0.058	Intracellular	-	-	Yes	-	-
ObacDRAFT_4431	hypothetical protein	1.48	0.011	-	-	-	Yes	-	-
ObacDRAFT_4438	hypothetical protein	2.22	0.005	-	-	-	-	-	Yes
ObacDRAFT_4439	hypothetical protein	1.60	0.054	-	-	-	-	-	-
ObacDRAFT_4440	hypothetical protein	1.46	0.075	-	-	-	-	-	-
ObacDRAFT_4447	hypothetical protein	1.38	0.073	-	-	-	-	-	-
ObacDRAFT_4462	hypothetical protein	1.02	0.097	-	-	-	-	-	-
ObacDRAFT_4467	hypothetical protein	1.89	0.016	-	-	-	-	-	-

ObacDRAFT_4468	Mu-like prophage major head subunit gpT.	1.66	0.041	-	-	-	-	-	Yes
ObacDRAFT_4471	hypothetical protein	1.95	0.011	-	-	-	-	-	-
ObacDRAFT_4480	hypothetical protein	1.37	0.027	-	-	-	-	-	-
ObacDRAFT_4542	Lysophospholipase L1 and related esterases	1.40	0.031	Amino Acid	-	-	Yes	-	-
ObacDRAFT_4543	PEP-CTERM protein sorting domain	1.50	0.010	-	-	-	-	-	-
ObacDRAFT_4547	hypothetical protein	1.17	0.060	-	-	-	-	-	-
ObacDRAFT_4554	hypothetical protein	1.35	0.033	-	-	-	-	-	-
ObacDRAFT_4617	Concanavalin A-like lectin/glucanases superfamily.	1.11	0.074	-	-	-	Yes	-	-
ObacDRAFT_4637	hypothetical protein	1.52	0.033	-	-	-	Yes	-	-
ObacDRAFT_4685	autotransporter-associated beta strand repeat	1.07	0.095	-	-	-	-	-	-
ObacDRAFT_4691	prepilin-type N-terminal cleavage	1.33	0.059	-	-	-	-	-	-
ObacDRAFT_4942	hypothetical protein	1.52	0.011	-	-	-	-	-	-
ObacDRAFT_5048	PEP-CTERM protein sorting domain	1.30	0.042	-	-	-	-	-	-
ObacDRAFT_5061	hypothetical protein	1.55	0.005	-	-	-	-	Yes	-
ObacDRAFT_5062	hypothetical protein	0.97	0.095	-	-	-	-	-	-
ObacDRAFT_5084	PAS domain S-box	1.12	0.063	-	-	-	Yes	-	-

Table S 3 Downregulated genes

Locus_Tag	Function	Fold Change	Padj	COG	CAZy	Positive Selection	Core Genome	Species specific genome	Unique genome
ObacDRAFT_0028	Predicted ATPase (AAA+ superfamily)	-1.19	0.081	General	-	-	-	Yes	-
ObacDRAFT_0195	integral membrane protein, TerC family	-1.22	0.061	Inorganic Transport	-	-	Yes	-	-
ObacDRAFT_0232	RNA polymerase sigma factor, sigma-70 family	-1.02	0.097	Transcription	-	-	Yes	-	-
ObacDRAFT_0234	hypothetical protein	-1.49	0.051	-	-	-	Yes	-	-
ObacDRAFT_0280	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	-1.76	0.012	Amino Acid	GH33	-	Yes	-	-
ObacDRAFT_0284	Uncharacterized membrane-associated protein	-1.05	0.068	Function unknown	-	-	Yes	-	-
ObacDRAFT_0339	glutamyl-tRNA reductase (EC 1.2.1.70)	-1.41	0.020	-	-	-	Yes	-	-
ObacDRAFT_0340	ABC-type uncharacterized transport system	-1.69	0.005	General	-	-	Yes	-	-
ObacDRAFT_0384	hypothetical protein	-1.94	0.008	-	-	-	Yes	-	-
ObacDRAFT_0405	peptide chain release factor 2	-1.40	0.012	Translation	-	-	Yes	-	-
ObacDRAFT_0448	Predicted signal transduction protein	-1.51	0.018	Signal transduction	-	-	Yes	-	-
ObacDRAFT_0514	periplasmic serine protease, Do/DeqQ family	-1.40	0.068	Posttranslational	-	-	Yes	-	-
ObacDRAFT_0542	Lauroyl/myristoyl acyltransferase	-1.04	0.084	Lipid	-	-	-	-	-
ObacDRAFT_0655	hypothetical protein	-1.09	0.068	-	-	-	Yes	-	-
ObacDRAFT_0656	hypothetical protein	-1.50	0.029	-	-	-	-	-	-
ObacDRAFT_0665	RNA polymerase sigma factor, sigma-70 family	-1.22	0.054	Transcription	-	-	Yes	-	-
ObacDRAFT_0672	lipoate synthase	-1.17	0.060	coenzyme	-	-	Yes	-	-
ObacDRAFT_0676	Uncharacterized conserved protein	-1.41	0.013	coenzyme	-	-	Yes	-	-
ObacDRAFT_0731	methylmalonyl-CoA mutase metallochaperone MeaB	-1.03	0.085	Posttranslational	-	-	Yes	-	-
ObacDRAFT_0739	hypothetical protein	-1.33	0.019	-	-	-	Yes	-	-
ObacDRAFT_0779	hypothetical protein	-1.16	0.095	-	-	-	-	-	-
ObacDRAFT_0860	Sphingosine kinase and diacylglycerol kinase	-1.12	0.089	General	-	-	Yes	-	-
ObacDRAFT_0918	ABC-type transport system	-1.70	0.005	cell wall	-	-	Yes	-	-
ObacDRAFT_0919	hypothetical protein	-1.27	0.041	-	-	-	Yes	-	-

ObacDRAFT_0978	ribosomal protein S5, bacterial/organelle type	-1.45	0.059	Translation	-	-	Yes	-	-
ObacDRAFT_1009	hypothetical protein	-1.06	0.075	-	-	-	Yes	-	-
ObacDRAFT_1039	Predicted nucleic acid-binding protein, contains PIN domain	-1.08	0.080	General	-	-	Yes	-	-
ObacDRAFT_1056	S-adenosylmethionine--tRNA ribosyltransferase-isomerase	-1.06	0.072	Translation	-	-	Yes	-	-
ObacDRAFT_1103	channel protein, hemolysin III family	-1.66	0.010	Intracellular	-	-	Yes	-	-
ObacDRAFT_1104	RNA-binding proteins (RRM domain)	-1.61	0.027	-	-	-	-	Yes	-
ObacDRAFT_1160	hypothetical protein	-1.13	0.077	-	-	-	Yes	-	-
ObacDRAFT_1181	Plasmid stabilization system protein	-1.17	0.049	Mobile	-	-	Yes	-	-
ObacDRAFT_1207	hypothetical protein	-1.37	0.032	-	-	-	-	Yes	-
ObacDRAFT_1214	hypothetical protein	-1.61	0.005	-	-	-	-	Yes	-
ObacDRAFT_1218	Uncharacterized membrane-associated protein	-1.00	0.095	Function unknown	-	-	Yes	-	-
ObacDRAFT_1235	probable S-adenosylmethionine-dependent methyltransferase	-1.23	0.090	Translation	-	-	Yes	-	-
ObacDRAFT_1236	Exopolyphosphatase	-1.23	0.041	-	-	-	Yes	-	-
ObacDRAFT_1252	Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5)	-1.47	0.031	coenzyme	-	-	Yes	-	-
ObacDRAFT_1283	Dioxygenases related to 2-nitropropane dioxygenase	-1.52	0.017	-	-	-	Yes	-	-
ObacDRAFT_1331	Subtilisin-like serine proteases	-1.38	0.034	Posttranslational	CBM5	-	-	-	-
ObacDRAFT_1337	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase (EC 2.5.1.54)	-1.35	0.052	Amino Acid	-	-	Yes	-	-
ObacDRAFT_1371	Uncharacterized protein required for cytochrome oxidase assembly	-1.24	0.033	coenzyme	-	-	-	-	-
ObacDRAFT_1373	hypothetical protein	-1.11	0.052	-	-	-	Yes	-	-
ObacDRAFT_1414	Biopolymer transport proteins	-1.04	0.095	Intracellular	-	-	Yes	-	-
ObacDRAFT_1425	integral membrane protein	-1.32	0.059	-	-	-	Yes	-	-
ObacDRAFT_1434	Transcriptional regulators	-1.70	0.010	Transcription	-	-	-	-	-
ObacDRAFT_1443	sulfate adenylyltransferase subunit 2 (EC 2.7.7.4)	-1.26	0.054	Amino Acid	-	-	Yes	-	-
ObacDRAFT_1475	Phosphoglycerol transferase and related proteins	-1.41	0.031	cell wall	-	-	Yes	-	-
ObacDRAFT_1491	Translation initiation factor IF-2, N-terminal region.	-1.35	0.031	-	-	-	-	-	-
ObacDRAFT_1505	transcriptional regulator, LysR family	-1.55	0.022	Transcription	-	-	-	-	-
ObacDRAFT_1570	Geranylgeranyl pyrophosphate synthase	-1.19	0.055	coenzyme	-	Yes	Yes	-	-
ObacDRAFT_1776	Fe2+/Zn2+ uptake regulation proteins	-1.51	0.023	Inorganic Transport	-	-	Yes	-	-

ObacDRAFT_1779	Iron-regulated ABC transporter permease protein SufD	-1.13	0.099	Posttranslational	-	-	Yes	-	-
ObacDRAFT_1781	Transglutaminase-like enzymes, putative cysteine proteases	-1.08	0.059	Posttranslational	-	-	-	-	-
ObacDRAFT_1816	rod shape-determining protein MreC	-1.29	0.043	cell cycle	-	Yes	Yes	-	-
ObacDRAFT_1830	Response regulators consisting of a CheY-like receiver domain	-1.03	0.073	Signal transduction	GT2	-	Yes	-	-
ObacDRAFT_1923	pseudouridine synthase	-1.10	0.085	Translation	-	-	Yes	-	-
ObacDRAFT_1924	Arabinose efflux permease	-1.09	0.074	carbohydrate	-	-	Yes	-	-
ObacDRAFT_1961	Predicted Zn-dependent protease	-1.15	0.095	Posttranslational	-	-	Yes	-	-
ObacDRAFT_2046	amino acid ABC transporter membrane protein	-1.08	0.098	Amino Acid	-	-	-	Yes	-
ObacDRAFT_2057	AsmA-like C-terminal region.	-1.07	0.060	-	-	-	-	-	-
ObacDRAFT_2089	Predicted permeases	-1.48	0.011	General	-	-	Yes	-	-
ObacDRAFT_2345	Protein of unknown function	-1.32	0.056	-	-	-	Yes	-	-
ObacDRAFT_2347	hypothetical protein	-1.52	0.019	-	-	-	-	-	-
ObacDRAFT_2348	hypothetical protein	-1.20	0.027	-	-	-	-	-	-
ObacDRAFT_2367	Predicted membrane protein (DUF2339).	-1.51	0.039	-	-	-	-	-	-
ObacDRAFT_2481	histidinol-phosphate phosphatase	-1.22	0.059	Amino Acid	-	-	Yes	-	-
ObacDRAFT_2526	GTP cyclohydrolase I	-1.06	0.070	coenzyme	-	-	Yes	-	-
ObacDRAFT_2583	Molecular chaperone GrpE (heat shock protein)	-1.16	0.058	Posttranslational	-	-	Yes	-	-
ObacDRAFT_2584	chaperone protein DnaJ	-1.27	0.068	Posttranslational	-	-	Yes	-	-
ObacDRAFT_2815	Uncharacterized protein, putative amidase	-1.13	0.059	coenzyme	-	-	-	Yes	-
ObacDRAFT_2816	Creatinine amidohydrolase.	-1.15	0.089	-	-	-	Yes	-	-
ObacDRAFT_2825	Uncharacterized conserved protein	-1.28	0.032	Function unknown	-	-	Yes	-	-
ObacDRAFT_2842	thiazole-phosphate synthase	-1.43	0.037	coenzyme	-	-	-	-	-
ObacDRAFT_2843	thiamine-phosphate pyrophosphorylase	-1.92	0.000	coenzyme	-	-	-	Yes	-
ObacDRAFT_2856	Ni,Fe-hydrogenase III large subunit	-1.14	0.082	Energy	-	-	Yes	-	-
ObacDRAFT_2857	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	-1.20	0.095	Energy	-	-	Yes	-	-
ObacDRAFT_2858	Hydrogenase 4 membrane component (E)	-1.25	0.049	Energy	-	-	Yes	-	-
ObacDRAFT_2864	prepilin-type N-terminal cleavage/methylation domain	-1.54	0.031	-	-	-	-	-	-
ObacDRAFT_2872	hypothetical protein	-1.33	0.027	-	-	-	Yes	-	-

ObacDRAFT_2873	Predicted transcriptional regulators	-1.35	0.063	Transcription	-	-	Yes	-	-
ObacDRAFT_2902	SSU ribosomal protein S12P methylthiotransferase	-1.13	0.060	Translation	-	-	Yes	-	-
ObacDRAFT_2905	Uncharacterized protein conserved in bacteria	-1.30	0.038	Function unknown	GT30	-	Yes	-	-
ObacDRAFT_2927	competence/damage-inducible protein cinA	-1.29	0.033	General	-	-	Yes	-	-
ObacDRAFT_3131	Superfamily II DNA and RNA helicases	-1.27	0.053	Replication	CE11	-	Yes	-	-
ObacDRAFT_3132	Dehydrogenases with different specificities	-1.07	0.095	General	-	-	Yes	-	-
ObacDRAFT_3219	uncharacterized radical SAM protein YgiQ	-1.80	0.002	-	-	-	-	Yes	-
ObacDRAFT_3224	Uncharacterized conserved protein	-1.11	0.084	Transcription	-	-	Yes	-	-
ObacDRAFT_3225	exosortase/archaeosortase family protein	-1.27	0.073	-	-	-	Yes	-	-
ObacDRAFT_3268	Rhodanese-related sulfurtransferase	-1.30	0.048	Inorganic Transport	-	Yes	Yes	-	-
ObacDRAFT_3272	Lysophospholipase L1 and related esterases	-1.27	0.062	Amino Acid	-	-	-	Yes	-
ObacDRAFT_3321	PEP-CTERM motif.	-1.15	0.099	-	-	-	-	Yes	-
ObacDRAFT_3352	adenosine deaminase (EC 3.5.4.4)	-1.15	0.084	nucleotide transport	-	-	Yes	-	-
ObacDRAFT_3358	haloacid dehalogenase superfamily, subfamily IA	-0.94	0.099	carbohydrate	GH65	-	Yes	-	-
ObacDRAFT_3370	Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component	-1.01	0.095	Energy	-	-	Yes	-	-
ObacDRAFT_3379	RND family efflux transporter, MFP subunit	-1.11	0.055	cell wall	-	Yes	Yes	-	-
ObacDRAFT_3380	Cation/multidrug efflux pump	-1.16	0.063	-	-	-	-	-	-
ObacDRAFT_3394	Major Facilitator Superfamily.	-1.00	0.095	carbohydrate	-	-	Yes	-	-
ObacDRAFT_3455	hypothetical protein	-1.15	0.064	-	-	Yes	Yes	-	-
ObacDRAFT_3456	Hemolysins and related proteins containing CBS domains	-1.34	0.018	General	-	-	Yes	-	-
ObacDRAFT_3464	Transcriptional regulators	-1.08	0.075	Transcription	-	-	Yes	-	-
ObacDRAFT_3502	Phosphoenolpyruvate carboxylase, type 1	-1.17	0.080	Energy	-	Yes	Yes	-	-
ObacDRAFT_3505	Outer membrane protein	-1.75	0.010	cell wall	-	-	Yes	-	-
ObacDRAFT_3508	heavy metal efflux pump (cobalt-zinc-cadmium)	-1.10	0.095	Inorganic Transport	-	-	Yes	-	-
ObacDRAFT_3535	hypothetical protein	-2.24	0.000	-	-	-	Yes	-	-
ObacDRAFT_3538	ABC-type polysaccharide/polyol phosphate transport system	-1.69	0.010	carbohydrate	CE1	-	-	-	-
ObacDRAFT_3575	TPR repeat.	-1.14	0.074	-	-	-	Yes	-	-
ObacDRAFT_3576	hypothetical protein	-1.53	0.016	-	-	-	Yes	-	-

ObacDRAFT_3700	Uncharacterized protein conserved in bacteria	-1.17	0.059	Function unknown	-	-	Yes	-	-
ObacDRAFT_3713	lipoate-protein ligase B	-1.11	0.060	coenzyme	-	-	Yes	-	-
ObacDRAFT_3814	Phosphoglycerol transferase and related proteins	-1.72	0.010	cell wall	-	-	-	-	-
ObacDRAFT_3815	PAP2 (acid phosphatase) superfamily protein	-3.06	0.000	General	-	-	-	-	-
ObacDRAFT_3816	Response regulators consisting of a CheY-like receiver domain	-2.41	0.000	Signal transduction	-	-	Yes	-	-
ObacDRAFT_3817	Signal transduction histidine kinase	-2.63	0.000	Signal transduction	-	Yes	Yes	-	-
ObacDRAFT_3821	exosortase/archaeosortase family protein	-1.03	0.085	-	-	-	-	Yes	-
ObacDRAFT_3938	lysyl-tRNA synthetase, class II (EC 6.1.1.6)	-1.05	0.087	Translation	-	-	Yes	-	-
ObacDRAFT_4029	ADP-ribose pyrophosphatase	-1.15	0.059	nucleotide transport	GT87	-	-	-	-
ObacDRAFT_4132	hypothetical protein	-1.47	0.043	-	-	-	-	Yes	-
ObacDRAFT_4157	Prephenate dehydrogenase	-1.33	0.054	Amino Acid	-	-	Yes	-	-
ObacDRAFT_4162	Predicted membrane protein	-1.03	0.085	Function unknown	-	Yes	Yes	-	-
ObacDRAFT_4163	ferrochelatase (EC 4.99.1.1)	-1.08	0.060	coenzyme	-	-	-	-	-
ObacDRAFT_4164	Superfamily II DNA/RNA helicases, SNF2 family	-1.14	0.063	-	-	Yes	Yes	-	-
ObacDRAFT_4183	NADH-quinone oxidoreductase, B subunit	-1.06	0.095	Energy	-	Yes	Yes	-	-
ObacDRAFT_4298	transcriptional regulator, LysR family	-1.14	0.095	Transcription	-	-	-	Yes	-
ObacDRAFT_4318	Arabinose efflux permease	-1.45	0.015	carbohydrate	-	-	Yes	-	-
ObacDRAFT_4325	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	-1.06	0.080	Translation	-	-	Yes	-	-
ObacDRAFT_4332	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	-1.11	0.075	-	-	-	Yes	-	-
ObacDRAFT_4333	ABC-type Fe3+ transport system, permease component	-1.07	0.083	Inorganic Transport	-	-	Yes	-	-
ObacDRAFT_4409	1-deoxy-D-xylulose-5-phosphate synthase (EC 2.2.1.7)	-1.05	0.099	coenzyme	-	-	Yes	-	-
ObacDRAFT_4411	3-hydroxyisobutyrate dehydrogenase and related dehydrogenases	-1.30	0.059	Lipid	GT1	-	Yes	-	-
ObacDRAFT_4517	ABC-type multidrug transport system, ATPase component	-1.18	0.068	defense	-	-	Yes	-	-
ObacDRAFT_4539	methionine synthase (B12-independent) (EC 2.1.1.14)	-1.49	0.022	Amino Acid	-	-	Yes	-	-
ObacDRAFT_4577	Flagellar biosynthesis pathway, component FliR	-2.11	0.005	cell motility	-	-	Yes	-	-
ObacDRAFT_4578	Flagellar biosynthesis pathway, component FliQ	-1.76	0.027	cell motility	-	-	Yes	-	-
ObacDRAFT_4579	flagellar biosynthetic protein FliP	-2.55	0.000	cell motility	-	-	Yes	-	-

ObacDRAFT_4580	Flagellar biogenesis protein	-1.70	0.033	cell motility	-	-	-	-	-
ObacDRAFT_4581	flagellar motor switch protein FliN	-1.50	0.079	-	-	-	Yes	-	-
ObacDRAFT_4582	Flagellar motor switch protein	-1.64	0.053	cell motility	-	-	Yes	-	-
ObacDRAFT_4585	Flagellar hook capping protein	-1.45	0.084	-	-	-	Yes	-	-
ObacDRAFT_4588	Flagellar FliJ protein.	-1.83	0.027	-	-	-	Yes	-	-
ObacDRAFT_4589	type III secretion system ATPase, FliI/YscN (EC 3.6.3.15)	-1.53	0.073	cell motility	-	Yes	Yes	-	-
ObacDRAFT_4590	Flagellar biosynthesis/type III secretory pathway protein	-1.62	0.048	cell motility	-	Yes	Yes	-	-
ObacDRAFT_4591	flagellar motor switch protein FliG	-1.46	0.077	cell motility	-	-	Yes	-	-
ObacDRAFT_4592	flagellar basal-body M-ring protein	-2.09	0.005	cell motility	-	-	Yes	-	-
ObacDRAFT_4597	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain	-1.33	0.056	cell motility	-	Yes	Yes	-	-
ObacDRAFT_4605	ABC-type Fe3+-hydroxamate transport system	-1.16	0.053	Inorganic Transport	-	-	Yes	-	-
ObacDRAFT_4606	ABC-type Fe3+-siderophore transport system, permease component	-1.21	0.046	Inorganic Transport	-	-	Yes	-	-
ObacDRAFT_4608	ABC-type cobalamin/Fe3+-siderophores transport system	-1.38	0.034	coenzyme	-	-	Yes	-	-
ObacDRAFT_4609	cob(I)alamin adenosyltransferase	-1.22	0.063	coenzyme	-	-	Yes	-	-
ObacDRAFT_4662	Short-chain dehydrogenases of various substrate specificities	-1.22	0.042	General	-	-	Yes	-	-
ObacDRAFT_4725	oligopeptidase A (EC:3.4.24.70). Metallo peptidase. MEROPS family M03A	-1.10	0.085	Amino Acid	-	-	Yes	-	-
ObacDRAFT_4750	hypothetical protein	-1.27	0.068	-	-	-	Yes	-	-
ObacDRAFT_4755	3-isopropylmalate dehydratase, small subunit (EC 4.2.1.33)	-1.97	0.005	Amino Acid	-	-	Yes	-	-
ObacDRAFT_4756	3-isopropylmalate dehydratase, large subunit (EC 4.2.1.33)	-1.46	0.031	Amino Acid	-	-	Yes	-	-
ObacDRAFT_4760	prepilin-type N-terminal cleavage/methylation domain	-1.42	0.024	-	-	-	-	Yes	-
ObacDRAFT_4785	Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain	-1.31	0.050	-	-	-	-	-	-
ObacDRAFT_4799	Glucose-6-phosphate dehydrogenase subunit.	-1.23	0.059	-	-	-	Yes	-	-
ObacDRAFT_4800	glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49)	-1.52	0.010	carbohydrate	-	-	Yes	-	-
ObacDRAFT_4805	hypothetical protein	-1.50	0.020	-	-	-	Yes	-	-
ObacDRAFT_4806	ABC-type transport system involved in cytochrome c biogenesis, permease component	-1.06	0.059	Posttranslational	-	-	-	-	-
ObacDRAFT_4811	folylpolyglutamate synthase/dihydrofolate synthase	-1.29	0.059	coenzyme	-	Yes	Yes	-	-
ObacDRAFT_4839	Competence protein ComGC	-1.09	0.059	Mobile	-	-	Yes	-	-

ObacDRAFT_4850	pseudouridine synthase, RluA family	-1.21	0.085	Translation	-	-	-	-	-
ObacDRAFT_4851	hypothetical protein	-1.46	0.051	-	-	-	-	-	-
ObacDRAFT_4862	Flagellar motor protein	-1.45	0.055	cell motility	-	-	Yes	-	-
ObacDRAFT_4863	Flagellar motor component	-1.37	0.091	cell motility	-	-	Yes	-	-
ObacDRAFT_4865	hypothetical protein	-2.23	0.005	-	-	-	Yes	-	-
ObacDRAFT_4866	flagellar biosynthesis protein FlhA	-2.11	0.008	cell motility	-	-	Yes	-	-
ObacDRAFT_4867	Flagellar biosynthesis pathway, component FlhB	-1.89	0.019	cell motility	-	-	Yes	-	-
ObacDRAFT_4873	Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain	-1.11	0.072	Posttranslational	-	-	Yes	-	-
ObacDRAFT_4878	hypothetical protein	-1.62	0.009	-	-	-	Yes	-	-
ObacDRAFT_4879	Dicarboxylate transport.	-1.47	0.017	-	-	-	Yes	-	-
ObacDRAFT_4987	AraC-type DNA-binding domain-containing proteins	-1.60	0.005	Transcription	-	-	Yes	-	-
ObacDRAFT_5005	3-phosphoshikimate 1-carboxyvinyltransferase	-1.09	0.073	Amino Acid	-	-	-	-	-