Genomic and Molecular Analysis of the Exopolysaccharide Production in the Bacterium *Thauera aminoaromatica* MZ1T

A Dissertation

Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Ke Jiang

May 2011

ACKNOWLEDGEMENTS

I want take this opportunity to express my deep gratitude to my major advisor, Dr. Gary Sayler, for his valuable and close guidance and heartfelt encouragement throughout my Ph.D. education. Without his encouragement and profound insight to help me shape my project during my difficult times, I could never reach this bright point today. Although fully occupied by so many tasks in all his positions, he has generously spent a large amount of time to attend and given comments at almost all my lab meeting presentations and colloquium rehearsals and carefully read and edited all my class and publication writings. All of these trainings helped me to improve my oral presentation and writing skills which have bestowed my many successful talks at different circumstances and finally made this manuscript come together. I would like to thank my committee professors, Dr. Alison Buchan, Dr. Todd Reynolds, Dr. Erik Zinser and Dr. Jun Lin for generously taking adequate amount of time out of their busy schedule to discuss with me on my project design and progress, and give me very helpful and timing advices, suggestions and recommendations at each stage of my project. Without them kindly talking all the way with me throughout the years, I could not image how can I make it at every step. I also want to show my appreciation to Dr. Dunlap for his nice help on microelectronic graphing of MZ1T.

I would like to thank all of my lab advisors and mates at Center for Environmental Biotechnology for their help, companions, encouragement and friendship.

Dr. John Sanseverino, Dr. Ying Wang and Dr. Alice Layton have given me tremendous help and guidance on experiment skills, planning project, editing writing, and opportunity for interacting with prestigious scientists in Microbiology, and training in genome science. Dr. Steven Ripp helped me use ChemDraw software and more effectively organize my bench work. Dr. James Fleming helped me to use the LinRegPCR software for QPCR and gave me guidance for relax my study stress, and offered help to remove obscurations for my study. Dan William taught me how to do QPCR from scratch and shared his experiences for bench work which saved me plenty time. Scott Moser always helped for every tedious things and also taught me how to extract massive amount of genomic DNA from MZ1T. Jack taught me how to do immunoblotting. Dr. Jennifer DeBruyn helped me for many things and editing my writings and even after she graduated I was still receiving her help on research and writing. Matthew Schwartz helped me by spending plenty time to discus my project. Vernon McIntosh gave me encouragement. Tommy Mead is a very nice companion who I really like to work together in the lab. Other lab members, Dr. Mel Eldridge, Dr. Michael, Jun Wang and so forth are all very helpful in my stay in CEB.

Special Thanks are given to my husband, Dr. Wenchao Qu, for his continuous support of my work and my life. Without him, nothing could be done.

I would like to thank my parents, my mother-in-law, all my relatives and friends for their love, friendship, encouragement and support. I also want to thank anyone who I

may have annoyed out of my ignorance for their tolerance to me and who have made me upset for their honest and nice training to make me stronger.

Last but not the least, I would like to express my tremendous gratitude to my lama Khenpo Sodargye, for his continuous encouragement and blessing to me throughout these years that lead me to perfect my Ph.D education and to transform myself from being always helped or taken care of by others to being willing to help and take care of others. Khenpo's supreme wisdom and great compassion are what I will continuously pursue in present and future lives. May every act of my body, speech and mind benefit all sentient beings! Lama Chen!

ABSTRACT

Thauera aminoaromatica MZ1T is an exopolysaccharide (EPS)-producing Gram negative bacterium isolated from the wastewater treatment plant of a major industrial chemical manufacturer as the causal agent for poor sludge dewatering. It shares common features with other known *Thauera* spp. (i.e. *Thauera aromatica*, and *Thauera selenatis*), being capable of degrading aromatic compounds anaerobically and using acetate and succinate as carbon sources. It is unique among the *Thauera* spp. in its production of abundant EPS which results in viscous bulking and poor sludge dewaterability. In this respect, it is similar to *Azoarcus* sp. EbN1 and BH72. Thaueran is the proposed name for EPS produced by MZ1T for research purpose.

The focus of this research is to fully characterize the microorganism and identify and characterize the genes responsible for thaueran synthesis and export through bioinformatics, transposon mutagenesis, gene clone and expression, reverse transcriptase quantitative real time PCR, and genome sequencing and annotation. Ultimately, this knowledge will contribute to control of viscous bulking and sludge dewatering problems. However, a broader range of important environmental biotechnical processes may be forthcoming from understanding thaueran synthesis. They may include thaueran remedial solutions for heavy metal and radionuclide immobilization, anaerobic carbon channeling and sequestration, greenhouse gas mitigation through acetate incorporation into thaueran, and novel applications such as thaueran-mediated wound healing.

Sequencing of MZ1T genome has been accomplished through collaboration with the Joint Genome Institute (JGI). The genome size is 4.5 Mbp, GC content is 68.3%, and there are 4,092 protein coding genes. Three putative thaueran gene clusters were found within the genome. One tight cluster with a size of 20.67kb encoding 14 genes may contain most necessary genes for thaueran formation and export. Another two clusters are loosely organized. Through transposon mutagenesis, mutants not producing abundant thaueran and not flocculating have been obtained and verified, and were further used in reverse transcriptase quantitative real time PCR to compare the differential expression levels of the presumable EPS genes among mutants, wild type MZ1T and under different growth conditions. The results indicated a correlation of the expression level of the test genes and the abundance of EPS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

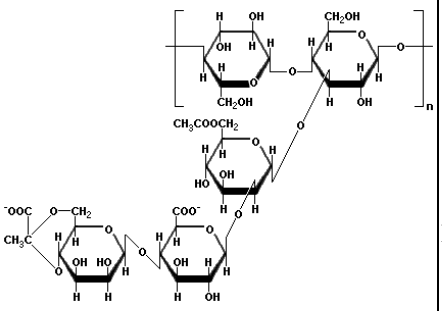
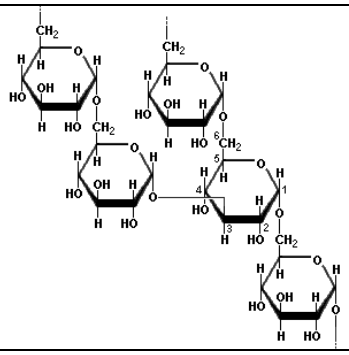
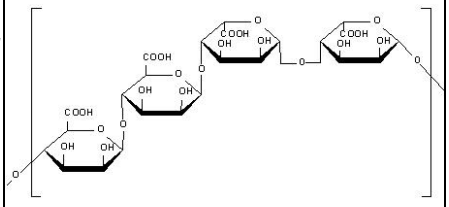
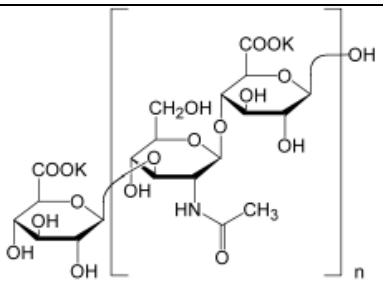
Thauera aminoaromatica MZ1T was first isolated from the wastewater treatment facility of Eastman Chemical Company, Kingsport, Tennessee, in the process of identifying the causal agent of abundant zoogleal clusters which were contributing to poor sludge dewatering(60). During the isolation, zoogleal clusters were isolated from the active sludge using a micromanipulator mounted on an inverted microscope. MZ1T was found forming small colonies on agar plates and floc in liquid medium. Through 16S RNA sequencing, MZ1T was found to belong to the *Thauera* genus and closely related to the genera *Azoarcus* and *Zoogloea*(60). The lineage of this bacterium is proteobacteria, β-proteobacteria, Rhodocyclales and Rhodocyclaceae.

Production of EPS is a very common feature for many Gram-negative bacteria such as α-proteobacterium *Rhizobium* sp.(23), *Agrobacterium* sp.(73), β-proteobacteria *Azoarcus* sp.(48), *Chromobacterium violaceum*(68) and *Zoogloea sp.*(82), and γ-proteobateria *Halmonas Maura*(5), *Azotobacter vinelandii*(36), *Haemophilus* sp.(76), *Xanthomonas campestris*(51) and *Pseudomonas sp*(18). Less Gram-positive bacteria than Gram negative bacteria have been found producing EPS, such as *Leuconostoc mesenteroides*(90), *Lactobacillus* sp.(39, 112, 122), *Streptococus bovis*(80).

In general, the repeating units vary from 3 to 8 monosaccharides(94). The monosaccharides often found in bacterial EPS are D-glucose, D-galactose, L-rhamnose, N-acetyl-gucosamine, N-acetyl-galactosamine, fucose, glucuronic acid, noncarbohydrates. Some EPS are linear and some are branched; more EPS have been found composed with heterogeneous monosaccharides (HePS) than homogenous monosaccharides (HoPS)(107). For example, Dextran is a branched HoEPS made up of glucose molecules(78). Xanthan is a branched HeEPS composed from the pentasaccharide repeat units(44). Gellan is a linear HeEPS with the tetrasaccharide repeat unit, [**D-Glc**(β1→4)**D-GlcA**(β1→4)**D-Glc**(β1→4)**L-Rha**(α1→3)](52). Culture conditions and the composition of the media greatly influence the yield and the molecular characteristics of EPS(69).

The following steps are usually found in EPS synthesis and export: carbohydrate substrate uptake, carbohydrate metabolism, nucleotide-sugar synthesis, such as UDP-glucose and essential building blocks of EPS, last step is polymerization and export(89). Two major models for synthesis and export of polysaccharides in G-negative bacteria are ABC transport dependent and Wyz-dependent pathways(89). All the genes responsible for EPS synthesis and export are usually clustered into large fragment in the genome(37). Table 1.1 showed the name, structure, genetic determinants and host organisms of some common bacterial exopolysaccharide, which suggested that EPS gene clusters are usually 10 to 20 kb in length and export genes and length determinant genes are found at the end of the cluster, multiple glycosyltransferase genes are involved.

Table 1.1 Common bacterial EPS names, structures, host organisms and genetic determinants

EPS name	Species	Chemical Composition and Structure	Genetic determinants/ Organization	GenBank ID and Ref
Xanthan	*Xanthomonas campestris*		Gene number: 15 gum genes (*gumBCDEFGHIJKLMNOP*) Cluster size: 18.0kb 1 glycosyltransferase gene **By Bielefeld University** Germany	AM920689 (51, 113)
Dextran	*Streptococcus mutans*		Gene number: 7 genes Gene names: *rgpABCDEF* Cluster size: 7.2kb 2 glycosyltransferase genes ABC transporter By University of Oklahoma	AE014133 (1, 78)
Alginate	*Pseudomonas aeruginosa, Azotobacter vinelandii.*		Gene number: 11 genes Cluster size: 13.0kp 2 glycosyltransferase gene Wzy-dependent By University of Laval Quebec Canada	CP000744 (36, 91)
Hyaluronic acid	*Streptococcus equi*		Gene number: 15 genes Cluster size: 17.3kb 4 glycosyltransferase genes ABC transporter By Wellcome Trust Sanger Institute Cambridge, United Kingdom	FM204883 (46, 61)

However, EPS production had not been found in *Thauera* sp. until the discovery of *Thauera aminoaromatica* MZ1T. When compared with other *Thauera* species such as

Thauera selenatis(65) and *Thauera aromatica*(100), only *Thauera aminoaromatica* MZ1T produce abundant EPS and cause cellular flocculation in liquid culture.

In previous studies, thaueran was purified through filtration, lyophilization and liquid chromatography (LC)(2). Its size was determined by gel-permeation chromatography. The structure and composition was determined by gas chromatography –mass spectrometry (GC-MS), nuclear magnetic resonance (NMR) and Fourier-Transform Infrared (FTIR) spectroscopy. Thaueran has an approximate molecular weight of 260 KDa with the repeating units containing four monosaccharides: rhamnose, galacturonic acid, N-acetyl-glucosamine, and N-acetyl-fucosamine(2) (Figure 1.1). The specific organization of the monsaccharide within the structure of the polymers has not yet been determined.

Since thaueran production is unique for *Thauera aminoaromatica* MZ1T in the genus of *Thauera* and the thaueran composition is unique among the bacterial EPS, we are interested in identifying the genes responsible for thaueran production and export and its biosynthesis pathway. These discoveries will contribute to a better understanding of carbon flow specifically from acetate to thaueran and sequestration by *Thauera aminoaromatica* MZ1T. As acetate is a central metabolite in anaerobic carbon cycling with 50% of carbon flow from acetyl-CoA to CH_4(30), acetate incorporation into thaueran is applicable to greenhouse gas emission control.

The abundant thaueran on the surface of MZ1T presents some challenges for this investigation. Firstly, its interaction with DNA causes very low yield during chromosomal DNA extraction. Secondly, its protecting function prevents the cell from taking up heterogeneous DNA which causes very low efficiency of the electroporation of transposons. Third, wild type and less thaueran producing mutants (*Thauera aminoaromatica* MZ1T floc+ and floc-deficient strains, respectively) do not show any morphological differences such as smooth and rough, or mucoid and dry phenotype variations when grown on agar plates, which thus requires to further observation in liquid culture and immune detection to screen floc-deficient mutants. Fourth, abundant transposase (IS4 and IS6) genes present throughout the whole genome may be a potential cause of the reversal of the mutants.

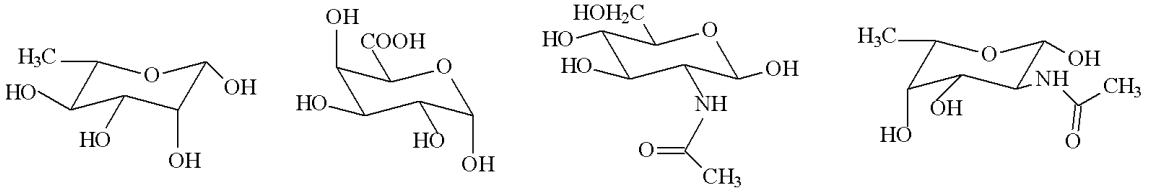


Figure 1.1 Structure of thaueran monosaccharides

Through genome annotation, three putative thaueran gene clusters were found in MZ1T, which is rather unusual,but not unique; for instance, *Azoarcus phosphatis* has two EPS gene clusters responsible for the synthesis of two kinds of different EPS - one more acidic than the other (67, 70). The discovery of genes encoding Wzy family protein and Wzx homolog - polysaccharide-specific-transport (PST) proteins as well as other characterized EPS related proteins in MZ1T genome implies that Wzy-dependent pathway may be used in thaueran biosynthesis(24, 72, 83). This pathway was first described in Salmonella O-antigen synthesis (Figure 1.2)(116, 117, 119) and also found for many other polysaccharides, such as that in *Azoarcus* EbN1(88) and *Rhodospirillum rubrum* (41), and mauran in *Halomonas Maura*(6). In this pathway, the lipid-linked repeating units of the polysaccharide are assembled on the cytoplasmic side of the inner membrane and then transported to the periplasmic side of the inner membrane for polymerization and ligation. After some modification in periplasm, the polysaccharide is translocated across the outer membrane(117, 119).

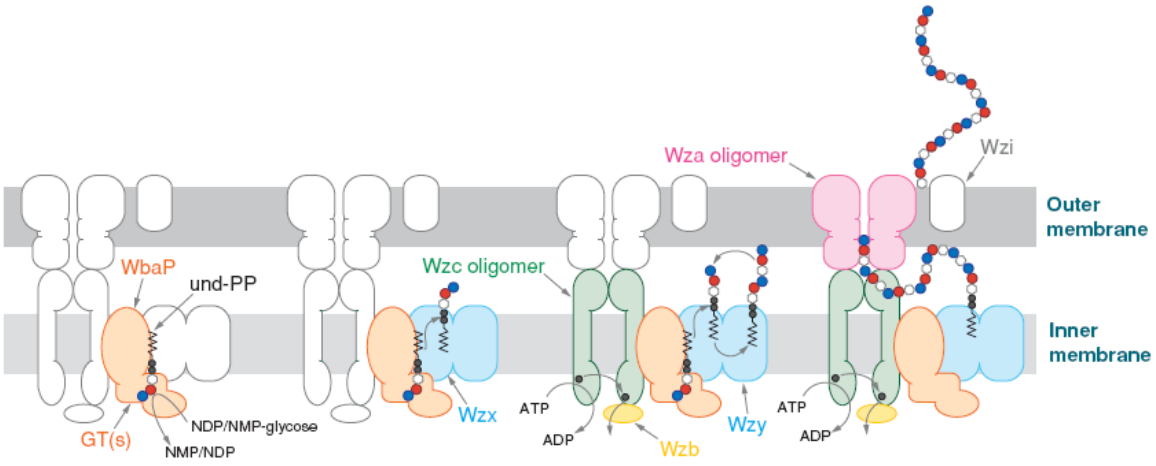


Figure 1.2 Model of Wzy-dependent polysaccharide synthesis and assembly pathway(116)

Another major synthesis pathway involving ABC transporters, found for group II capsular in *Escherichia coli*, *Haemophilus*, *Neisseria Rhizobium* and *Agrobacterium*, has different assembly steps and locations(116). In this pathway, the repeating units are synthesized and polymerized at the inner face of the cytoplasmic membrane, and translocated onto the periplasmic side for ligation with lipid(116). The final product is then translocated from periplasm to the surface of the outer membrane by interacting with export system of periplasm and outer membrane (Figure 1.3) (106).

Proteins involved in this pathway such as the ABC transporter and ligases are also found in the putative thaueran gene clusters of strain MZ1T, which indicates that both major synthesis and export pathways are adopted by strain MZ1T.

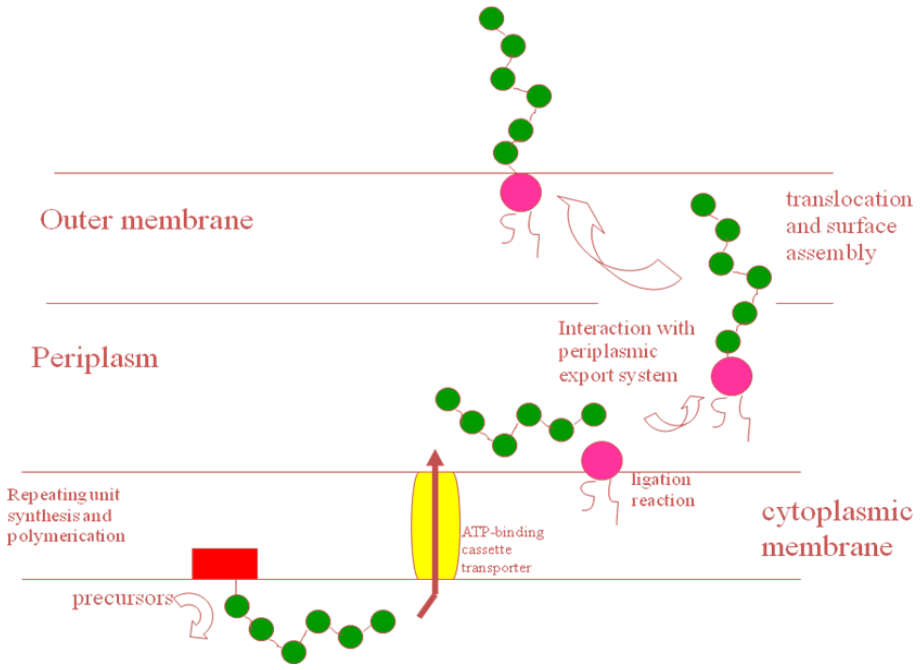


Figure 1.3 Model of ABC transporter involved polysaccharide synthesis and assembly pathway(106).

Our primary goal is to identify the genes responsible for exopolysaccharide production and transport. To have a better understanding of this organism, full identification and characterization of MZ1T based on previous known information have been carried out. This work forms **chapter 2**, which focuses on the comparative study of the morphological, physiological and biochemical characteristics of MZ1T among species in *Thauera* genus as well as its phylogenetic position. In this study, to serve our primary goal, floc formation was also carefully observed in the culture of other *Thauera* species, such as strain B4P and S2, when grown under the same condition as MZ1T.

To gain a comprehensive understanding of MZ1T at the genetic level to possibly lay out the pathways of its EPS formation, the synthesis of the monsaccharides and the repeating units and their polymerization and translocation, the whole genome was sequenced and annotated in corporation with JGI and Oak Ridge National Lab(ORNL). **Chapter 3** is a brief summary of the most important features of the genome in preparation to publish as an announcement and **Chapter 4** is aimed to describe in detail the putative genes involving in carbon metabolism, aromatic compound degradation, global gene regulation and particularly synthesis of the monasaccharides of thaueran repeating units, formation and export of exopolysaccharide in three different regions of the genome, including their physical size, organization and directions.

All the biochemical and molecular work regarding EPS formation in **Chapter 5** complements the discoveries from genome annotation. Although the initial work of transposon mutagenesis is independent from the results from genome annotation, the identification and verification of the interrupted genes are based on the annotation and genome sequence available at the flanking region. Then, gene cloning and expression as well as reverse transcriptase quantitative real time PCR (RT-QPCR) mostly relied on the information from genome sequencing and annotation.

Through the complementation and integration of genomic and molecular analyses, a better understanding of exopolysaachirde and floc formation in MZ1T has been achieved. Further investigation based on present accomplishment will provide even

more detailed and clearer principles and pathways of exopolysaccharide formation in MZ1T and the insight for the control of viscous bulking in the wastewater treatment plant.

Chapter 2

Comparative Description of *Thauera aminoaromatica* MZ1T Necessitates Phylogenetic Reorganization of *Thauera aminoaromatica* species

Author: Ke Jiang[1], Alice C. Layton[1], Michael S. Allen[2], Scott Moser[1], Ying Wang[3], John Sanseverino[4], John R. Dunlap[5], Pinidphon Prombutara[2], Gary S. Sayler[1*]

[1]Center for Environmental Biotechnology, The University of Tennessee, Knoxville, TN 37996

[2]Department of Biological Sciences, University of North Texas, Denton, TX 76203

[3]Center for Renewable Carbon, The University of Tennessee, 2506 Jacob Drive, Knoxville, TN 37996

[4]Tennessee Solar Institute, 2450 E.J. Chapman Drive Suite 112, Knoxville, TN 37996

[5]Advanced Microscopy and Imaging Center, The University of Tennessee, Knoxville, TN 37996

[*]Author for correspondence: Gary S. Sayler sayler@utk.edu

Abstract

A Gram-negative bacterium, strain MZ1T, previously isolated from activated sludge samples from an industrial wastewater treatment plant, is capable of producing abundant exopolysaccharide and degrading various aromatic compounds with nitrate as electron acceptor. Morphologically, cells are short rods with a flagellum. The 16S rRNA gene phylogenetic evidence, G+C content, DNA-DNA hybridization comparison, fatty acid profile, exopolysaccharide production and other physiological characteristics indicate that strain MZ1T, previously characterized strains S2 and B4P are all one species-*T. aminoaromatica* and closely clustered with *T. selenatis* AX39. The cluster of these four strains is well-separated from other *Thauera* species, such as *T. aromatica*, *T. butanivorans T. chlorobenzoica*, and *T. mechenichensis*. Abundant exopolysaccharide production made MZ1T a unique strain in the genus of *Thauera*. However when grown under the same conditions, this feature was found in strain S2 and B4P, but not in *T. selenatis*. These four strains also share a unique fatty acid, 12:0 3-OH, as compared to other *Thauera species*. Based on consideration of this evidence, we recommended strain MZ1T as *Thauera aminoaromatica* MZ1T and proposed the phylogenetic reorganization of *Thauera aminoaromatica* species.

Introduction

Since Macy *et al* described *Thauera selentatis* in 1993(65), many other *Thauera* species were isolated and characterized. Currently, the validly published *Thauera* species are: *Thauera selenatis* (65), *Thauera aromatica*, *Thauera chlorobenzoica*, *Thauera linaloolentis* (32), *Thauera mechernichensis*(95), *Thauera terpenica* (32), *Thauera aminoaromatica* (74), *Thauera phenylacetica* (74) and *Thauera butanivorans* (25). All of these species degrade some aromatic compounds such as benzoic acid under aerobic conditions. Most of them have been isolated from waste water activated sludge after enrichment (4, 32, 33), contaminated water (65) and soil (29, 100).

Strain MZ1T was isolated through micromanipulation from a floc directly obtained from waste water activated sludge without enrichment (60). It was found to be the causal organism forming viscous bulking which resulted in poor dewatering efficiency and increased the costs for dewatering, incineration and disposal (2). Viscous bulking has been attributed to abundant exopolysaccharide production (53). In contrast, most other studies on *Thauera* species were studied on their capacity for aromatic compound degradation (4, 12, 43, 86, 96), and exopolyssacharide production was probably not evaluated in the literature.

Strain MZ1T was identified belonging to *Thauera* genus based on the 16S rRNA phylogenetic analysis in the original work by Lajoioe *et al* (60). Strain MZ1T is capable

of degrading aromatic compounds, such as benzoate and 2-aminobenzoate, but our past study of this organism found an excessive novel exopolysaccharide production hindering activated sludge dewatering in the industrial waste water treatment plant operation.

The exopolysaccharide of MZ1T was purified and its unique composition of monosaccharide units was partially characterized (2) .In this study, morphology, physiology, fatty acid profile and DNA-DNA hybridization revealed that strain MZ1T belongs to the same species as *T. aminoaromatica* S2, and other *T. aminoaromatica* strains, such as b302, NS5 (NCBI GenBack), are distinct from S2 and MZ1T. Therefore, these three strains should not be classified as *T. aminoaromatica*, but probably another independent species outside this cluster. Exopolysaccharide production appears to be a common characteristics of *T. aminoaromatica* species.

Prior isolation and cultivation of bacterial strains

Strain MZ1T was previously isolated directly from a zoogleal cluster using a micromanipulator(60). Unless otherwise stated, strain MZ1T was grown aerobically in Stoke's medium (2) at 30℃ shaking at 150 rpm. Strain *T. selenatis* AX39, *T. aminoaromatica* S2 and *T. phenylacetica* B4P were grown at the same condition as strain MZ1T. Before isolation, S2 was enriched from anoxic ditch sludge on L-tryptophan and B4P from activated sludge of a sewage treatment plant on L-tyrosine (74). The accession

number for strain S2 is DSMZ14742, strain B4P is DSMZ14743, strain AX39 is ATCC 55363, strain MZ1T is in the process of deposit..

Morphological and physiological characterization

MZ1T cell size and shape were determined from scanning and transmission electronic microscopic images (Fig 2.1). Flagellum was stained using Leifson's method (62). The optimal growth temperature was determined using temperatures of 25, 30, 32, 37 and 42 ℃ and the optimal pH for growth was determined using pH 6, 6.5, 7, 7.2, 7.5, 8, 8.5. Standard methods were used to examine the presence of catalase, oxidase and nitrate reductase(98). All experiments were conducted under aerobic conditions in Stoke's medium as described above. Carbon utilization pattern was performed by Biolog using microplates PM1 and PM2 by Biolog (Biolog, Hayward, CA).

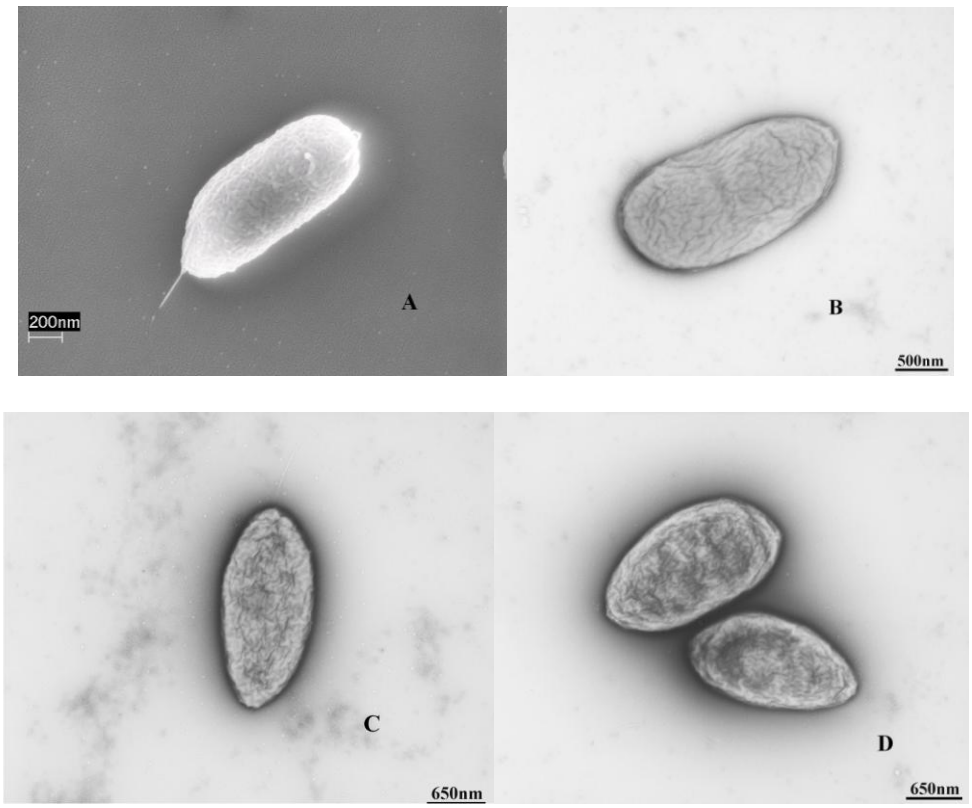


Figure 2.1 Scanning and transmission electronic microscopic images of strain MZ1T (A and B), S2 (C) and B4P (D).

Strain MZ1T cells were motile rods with dimensions of 0.5 x 1.1-1.8 μm. Each cell has a monotrichous flagellum, which is consistent with *T. selenatis*, *T. mechernichensis*, *T. linaloolentis* and *T. terpenica* and differs from *T. aromotica* and *T. chlorobenzoica*, which have peritrichous flagella. The optimal growth temperature and pH were 30 ℃ and pH 7.2. Catalase, oxidase and nitrate reductase tests were all positive. When grown in liquid culture, MZ1T produces abundant extracellular polysaccharide and forms flocs at early stationery stage. Under aerobic conditions, benzoate, succinate, aspirate, glutamate, proline, leucine, serine and alanine are utilized.

A summary of the differentiating characteristics of all named *Thauera* species is shown in Table 2.1.

Table 2.1 Comparison of the differentiating characteristics of the published *Thauera* species with MZ1T

***Thauera* Species**	**Cell morphology**	**Cell size(μm)**	**Mobility**	**Flagellation**	**Optimal pH**	**Optimal temperature ℃**	**G+C (mol%)**	**Type strain**	**Year**
T. selenatis	Rods	0.56 X 1.4	+	monotrichous	8	28	66	AX39, ATCC55363	1993
T. aromatica	Short rods	0.5-1.5 X 1.0-2.5	+	peritrichous	7.2-7.8	28-37	66-67	K172, DSM 6984	1995
T. mechernichensis	Short rods	0.75 X 1.5-2.0	+	monotrichous	ND	40	65	TL1, ATCC700857	1999
T. linaloolentis	Rods	0.7 X 1.4-2.7	_	n/d	7.0-7.3	32	66	47LoI, DSM 12138T	1998
T. terpenica	Rods	0.9 X 1.6-2.2	+	monotrichous	7.9-8.8	32	64	58Eu, DSM 12139	1998
T. chlorobenzoica	Rods	0.85-0.97 X 1.2-2.7	+	peritrichous	7.5-8.0	30-37	69	3CB-1, ATCC700723	2001
T. aminoaromatica	Rods	0.5-0.75 X 2.0-3.0	+	n/d	7.5-8.6	28	68.4	S2 DSMZ 14742	2002
T. phenylacetica	Short rods	0.75-1.0 X 1.5-2.0	+	n/d	7.0-7.5	28	70.6	B4P, DSMZ 14743	2002
T. butanivorans	Rods	0.6-0.8 X 1.1-2.4	+	monotrichous	n/d	30	67.3	Bu-B1211, ATCC43655	2009 (1980)
MZ1T strain	Rods	0.5 X 1.1-1.8	+	monotrichous	7.2	30	68.3	ATCC????	This manuscript

16S rRNA gene phylogenetic analysis

The complete MZ1T 16S rRNA gene sequences were obtained from gene cloning in phylogenetic analysis (60). 16S rRNA gene sequences of other *Thauera* species and related *Azoarcus* species were retrieved from the GenBank database. Multiple alignments were performed in the program CLUSTAL X (111). A phylogenetic tree was created with the neighbour-joining (93) and maximum-parsimony (31) methods in the MEGA4 program (109).

The phylogenetic tree based on 16S rDNA gene sequences of these 16 *Thauera* strains showed that MZ1T is closely grouped with 99% 16S rDNA sequence similarity to strains *T. phenylacetica* B4P , *T. aminoaromatica* S2, *T. selenatis* AX39 and *T. aminoaromatica* a241 (Fig2). The cluster of these five strains is well-separated from strains in *T. aromatica, T. chlorobenzoica, T. mechernichensis, T. butanivorans, T. linaloolentis* and *T. terpenica.* Two other *T. aminoaromatica* strains, NS5 and b302 are not closely clustered with the type strain S2 and indicate some discrepancies in the classification of *T. aminoaromatica* species. Due to the similarity of MZ1T with strains S2, B4P, a241 and AX39 and based on *Bergey's Manual of Systematic Bacteriology*, DNA-DNA hybridization studies were used to complement the results from 16S rDNA sequence analysis.

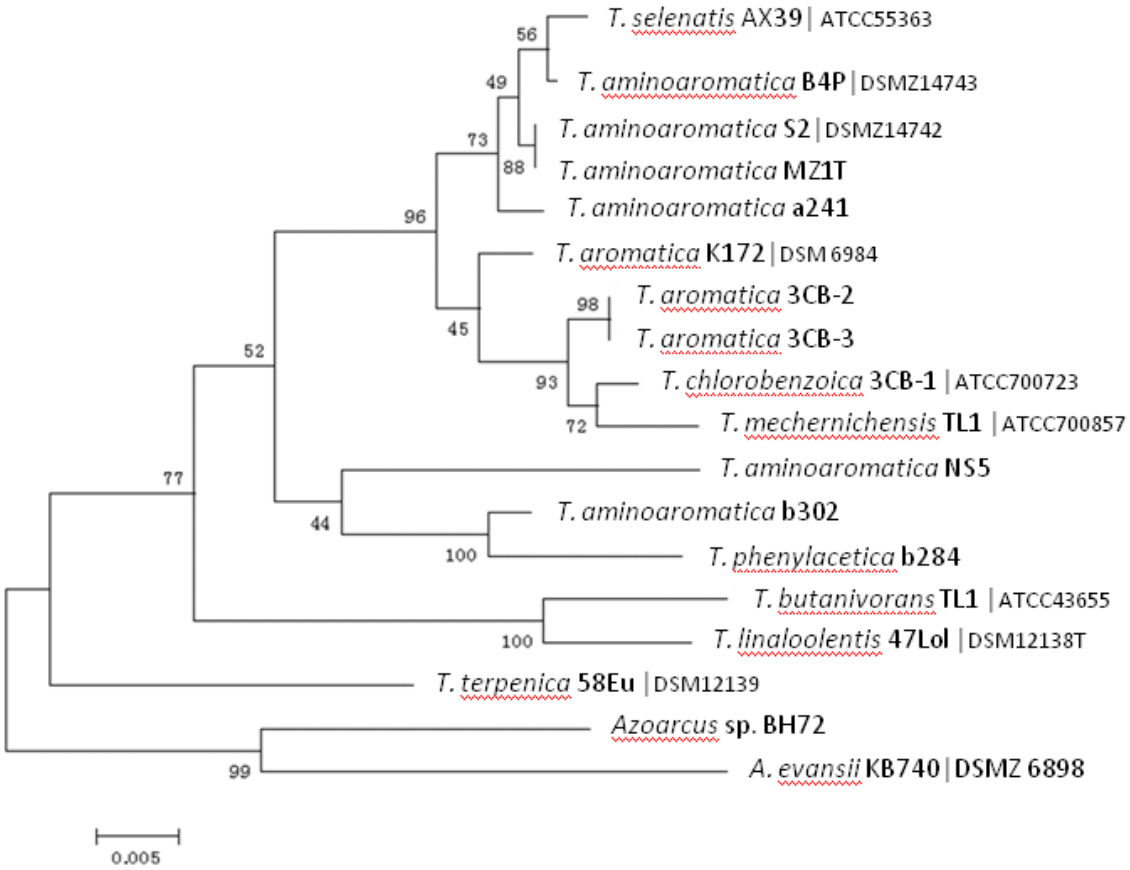


Figure 2.2 16S rDNA based phylogenetic tree showing the position of MZ1T among Thauera genus and the type strains of Azoarcus evansii KB740 and Azoarcus sp. BH72.

The tree was constructed by the neighboring-joining method and rooted by using *Azoarcus evansii* and *Azoarcus* sp. BH72 as the outgroup. Numbers at nodes represent bootstrap values (based on 1000 resamplings.) Fifteen recognized species belonging to the genera *Thauera* were analysed while constructing the phylogenetic tree. Bar, 0.005 nucleotide substitutions per nucleotide position. The GenBank accession numbers for the 16S rDNA sequences for the strains used in this study are as follows (in parentheses): *T. aminoaromatica* b302 (EU434482), *T. aminoaromatica* a241 (EU434467) *T. aminoaromatica* NS5 (FJ609701), *T. selenatis* AX39 (X68491), *T. mechernichensis* TL1(Y17590), *T. linaloolentis* 47Lol (AJ005817), *T. terpenica* 58Eu (AJ005817), strain MZ1T (AF110005), *T. chlorobenzoica* 3CB-1 (AF123264), *T. aromatica* K172 (X77118), *T. butanivorans* Bu-B (AB021377), *T. phenylacetica* b284 (EU434481), *T. aromatica* 3CB-3 (AF229882), *T. aromatica* 3CB-2 (AF229881), *T. phenylacetica* B4P (AJ315678), *A. evansii* KB740 (X77679), *Azoarcus* sp. BH72 (AF011344).

DNA-DNA Hybridization

DNA-DNA hybridization was performed between strain MZ1T and *T. selenatis* ATCC 55363, *T. phenylacetica* B4P DSM 14743 and *T. aminoaromatica* S2 DSM 14742 by Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ,) (Braunschweig, Germany). Briefly, all strains were grown in Stoke's medium and harvested at 24 hrs. Biomass for each strain was suspended in iso-propanol/water (1:1, v/v) and sent to DSMZ. DNA was isolated at DSMZ using a French pressure cell (Thermo Spectronic) and purified by chromatography on hydroxyapatite (14). Hybridizations were performed using a model Cary 100 Bio UV/Vis spectrophotometer equipped with a Peltier temperature-controlled 666 multicell changer with an in-situ temperature probe (Varian) in 2X SSC and 10% formamide at 70 ℃ (21, 49). The results were based on the renaturation curves.

DNA-DNA hybridization studies showed that MZ1T was 100% similar to strain S2, 78.9% to strain B4P and 59.6% to *T. selenatis* ATCC55363 respectively (Table 2.2). When the recommended threshold value of 70% DNA-DNA similarity is used for the definition of bacterial species [*ad hoc* committee] (114), these results indicate that MZ1T does not belong to the same species as *T. selenatis* ATCC55363, but does belong to the same species as strains S2 and B4P. Although a previous study based on DNA-DNA hybridization suggested that strain S2 and B4P are different species (74), the results obtained in this study of 84.6% DNA-DNA hybridization suggest they are the same

species and are consistent with the 16S rRNA analysis. Therefore, both S2 and B4P should actually be considered as the same species on the basis of genomic DNA-DNA similarity and 16S rRNA sequence identity. Taken together these results suggest MZ1T should be classified as the same species as both S2 and B4P and we recommend MZ1T be classified as *Thauera aminoaromatica* strain MZ1T.

Determination of G+C content

Genomic DNA was isolated from a 100 mL overnight culture before floc formation using phenol-chloroform method(87). Pulse-Field Gel Electrophoresis (PFGE) was performed to estimate the genome size and melting temperature was measured (21) to calculate the G+C content of MZ1T. These results indicate that MZ1T has a genome size of 4.5 Mbp and a G+C content of 68%. All *Thauera* species have very high G+C content, ranging from 64 to 70.6% (4, 25, 32, 65, 74, 99), and the G+C content of MZ1T is particularly close to that of *T. aminoaromatica* strain S2 (74).

Table 2.2 DNA-DNA hybridization based genomic similarities between strain MZ1T and *T. selenatis* ATCC55363, *T. aminoaromatica* S2DSMZ14742 and *T.phenylacetica* B4P DSMZ14743.

Strain	***Similarity (%)***
***T. selenatis* ATCC55363**	59.3(59.8)
***T. aminoaromatica* S2DSMZ14742**	100.8(100.6)
***T.phenylacetica* B4P DSMZ14743**	78.8(78.9)

(Values in parentheses are duplicate results.)

Extracellular Polysaccharide Production

Extracellular polysaccharide of MZ1T was previously extracted and purified using a Pellicon 2 Mini Holder filtration apparatus equipped with a 100-kDa MWCO membrane (Millipore) (2). Previously, none of the other *Thauera* species were reported to produce abundant extracellular polysaccharide. However, when *T. aminoaromatica* strain S2 and *T. phenyacetica* strain B4P were grown in Stoke's medium at 30ºC under the same condition as MZ1T, moderate amounts of extracellular polysaccharide were observed, with S2 strain produces more than B4P strain.

Cellular Fatty Acid Composition

The cellular fatty acid profile of strain MZ1T was determined by MIDI labs (Newark, DE) using the MIDI Sherlock® Microbial Identification System. Cells grown on Stoke's plate for 24 hrs at 30ºC were harvested and saponified. Samples were methylated, extracted and cleaned up according to MIDI standard sample preparation protocol. The converted fatty acid methyl esters (FAMEs) were then separated with an Agilent Technologies Gas Chromatograph (GC). The peaks of each fatty acid are automatically named and quantified using the Sherlock Microbial Identification System (MIS) with a FAME profile generated by Sherlock software.

The predominant fatty acids of MZ1T are 16: 1w7c (50.65%), 16:0 (25.81%), 18:1 w7c (9.37%) 12:0(6.3%), 10:0 3-OH (3.87%) and 12:0 3-OH (3.16%) (Table 2.3).

The fatty acid 12:0 3-OH is generally not found in *Thauera* genus but only found in *T. selenatis* (34, 99). Although *T. mechernichensis* and *T. aromatic* 6964 were described to contain hydroxylated fatty acids (95) , the actual data did not show any 12:0 3-OH but only 10:0 3-OH. Therefore, MZ1T is close to *T. selenatis* based on membrane fatty acid composition. Strain S2 and B4P also showed the presence of 12:0 3-OH with the percentage of 3.22 and 3.23, respectively. These results suggested that these four strains distinguish themselves from all other characterized *Thauera* species based on their fatty acid profile which separate them from the remaining tested *Thauera* species. This proposal is supported by the argument that *T. aromatica* and *T. chlorobenzoica* are phylogenetically and physiologically differentiated from other species at the generic level(99).

Table 2.3 Major fatty acids of MZ1T, *T. selenatis* ATCC 55363, *T. phenyacetica* strain B4P (DSMZ 14742) and *T. aminoaromatica* strain S2 (DSMZ 14743)

Fatty Acid(%)	MZ1T	*T. selenatis*	B4P	S2
16:1w7c	50.65	44.57	46.14	45.41
16:0	25.81	22.84	25.72	25.91
18:1 w7c	9.37	8.86	11.23	12.6
12:0	6.3	7.73	7.2	6.7
10:0 3-OH	3.87	5.49	4.48	4.26
12:0 3-OH	3.16	4.18	3.23	3.22

Description of *Thauera aminoaromatica* MZ1T

Cells are rod-shaped (0.5 μm wide and 1.1-1.8 μm long), Gram-negative, catalase- and oxidase positive, and motile by means of a polar flagellum. The optimal growth temperature and pH are 30℃ and pH 7.2 respectively. No growth factors are required when succinate is used as the carbon source. It has a G+C content of 68%. MZ1T is able to degrade aromatic compounds, such as benzoate and 2-aminobenzoate. Fatty acid profile consists of 16:1w7c (50.65%), 16:0 (25.81%), 18:1w7c (9.37%) 12:0 (6.3%), 10:0 3-OH (3.87%) and 12:0 3-OH (3.16%). It produces abundant exopolysaccharide at the early stationary phase and forms flocs in liquid culture. When grown on agar plates, no obvious exopolysaccharide is observed.

Features shared by *T. aminoaromatica* and *T. selenatis* species

Strain MZ1T was identified as a *T. aminoaromatica* species in this study, and several forms of evidence showed that this species is well-separated from other *Thauera* species, such as *T. aromatica, T. chlorobenzoica* and so forth. Based on the literature, none of the *Thauera* species were reported to produce abundant exopolysaccharide, but MZ1T, S2 and B4P were found to do so in this study and may serve as a unique feature for the species *T. aminoaromatica. T. selenatis* did not produce exoplysaccharide. The presence of the hydroxylated fatty acid 12:0 3-OH in *T. aminoaromatica* and *T. selenatis* species is unique within this genus.

Phylogenetic Reorganization of *T. aminoaromatica* and *T.phenylacetica.*

A previous study indicated several characteristics of B4P that distinguished itself from other *Thauera* species. For instance, the shape of B4P cells is oval rod, instead of straight rods as other *Thauera* strains. It is the only species that requires *p*-aminobenzoate as a growth factor, has the highest G+C content (70.6%), and its substrate spectrum as well as protein profile are different from other *Thauera* species (74). However, the DNA-DNA similarity of B4P with S2 and MZ1T, at 84.5% and 78.9%, respectively, suggests that B4P is the same species as S2 and MZ1T when the threshold value of 70% is used(114) .

From 16S rRNA gene phylogenetic analysis, other *T. aminoaromatica* strains, such as NS5 and b302 as well as *T. phenylacetica* b284 are separated from their type strains S2 and B4P. This inconsistency revealed the problem of the classification of these two species and supports our proposal to regroup strains in species *T. aminoaromatica* and *T. phenylacetica* and strain NS5, b302 and b284, which do not cluster with strain S2, should not be classified as *T. aminoaromatica.*

Acknowledgements

We would like to thank Dr. Georg Fuchs at University of Freiburg for generously providing strain S2 and B4P.

Chapter 3

Genome sequence of the Exopolysaccharide Producing Bacterium *Thauera aminoaromatica* MZ1T

Author: Ke Jiang[1], John Sanseverino[2], Ying Wang[3], Alice C. Layton[1], Michael S. Allen[4], Jennifer M. DeBruyn[5], Scott Moser[1], Dan Close[1], Gary S. Sayler[1*]

[1]Center for Environmental Technology, The University of Tennessee, Knoxville, TN 37996

[2]Tennessee Solar Institute, 2450 E.J. Chapman Drive Suite 112, Knoxville, TN 37996

[3]Center for Renewable Carbon, The University of Tennessee, Center, 2506 Jacob Drive, Knoxville, TN 37996

[4]Department of Biological Sciences, University of North Texas, 1155 Union Circle, # 305220, Denton, TX 76203

[5]Department of Biosystems Engineering and Soil Science, 2506 EJ Chapman Drive, The University of Tennessee, Knoxville, TN 37996

[*]Author for correspondence: Gary S. Sayler sayler@utk.edu

Abstract

We report the first complete genome sequence for a member of the *Thauera* genus. *Thauera aminoaromatica* MZ1T is an exopolysaccharide producing bacterium isolated from a chemical industry activated-sludge plant in Tennessee. The genome annotation revealed that MZ1T has diverse metabolic potential indicating benzoate degradation, heavy metal resistance, exopolysaccharide biosynthesis and transportation and a possible quorum sensing regulation system.

Species in the *Thauera* genus degrade various aromatic compounds under anaerobic and aerobic conditions and have been isolated from waste water activated sludge after enrichment (4, 25, 32, 33, 95), contaminated water (65, 66) and soil (29, 100). *Thauera* spp. MZ1T was directly isolated from a viscous bulking floc sample without enrichment (60) and shown to be a member of *Thauera* by phylogenetic analysis using 16S rRNA gene sequence showed that strain MZ1T belongs to *Thauera* genus (60) and more specifically a *T. aminoaromatica* species by DNA-DNA hybridization (54). Like other *Thauera* species, MZ1T degrades aromatic compounds such as benzoate and 2-aminobenzoate (74).

Genomic DNA was extracted using a phenol chloroform method (87) from overnight culture in the Center for Environment Biotechnology at the University of Tennessee. Prepared genomic DNA was sequenced, assembled and annotated at Joint Genome Institute and Oak Ridge National Lab. The genome contains one chromosome (Contig 151) and one plasmid (Contig 150) for a total genome size of 4.5 Mb with a GC content of 68.2% and 3,978 protein-encoding genes. The circular chromosome is 4,496,212 bp in length with a coding density of 89%, a GC content of 68%, 3,903 protein coding genes, 74 structural RNA genes, 89 pseudo genes and 4 copies of 5S, 16S and 23S rRNA genes. The plasmid (pTha01) is 78,374 bp in size and has a GC content of 62%, 77% coding density, 75 protein coding genes , 4 pseudo genes and nonstructural RNA genes.

The annotation of the MZ1T genome indicated the complete glycolytic pathway and citric acid cycle are present as well as two complete acetate assimilation systems with the key enzymes being acetate-CoA ligase and acetate kinase - phosphate acetyltransferase, respectively, allowing MZ1T to utilize acetate as a carbon source(9). Three putative gene clusters responsible for exopolysaccharide biosynthesis, polymerization and export were found. The discovery of the *wzy* gene in one of the cluster implicates a Wzy-dependent pathway of polysaccharide synthesis and export in MZ1T (24, 118, 119). Unlike other related *Thauera* spp. (29, 43, 96), MZ1T does not appear to have genes for anaerobic toluene or phenol degradation; however, genes for both anaerobic and aerobic benzoate degradation are present. The genome of MZ1T contains a total of six sigma factors controlling global gene regulation. They are the housekeeping sigma factor σ^{70}, the nitrogen regulator σ^{54}, the heat shock sigma factor σ^{32}, as well as three copies of *e*xtra*c*ytoplasmic *f*unction (ECF) sigma factor (40). More than ten copies of two component regulatory systems and three copies of toxin and antitoxin pairs were found in the genome. Genes encoding efflux pumps for heavy metal resistance to arsenic, cadmium, lead, silver, zinc but not for selenium have been found. MZ1T has a probable quorum sensing pathway based on the finding of more than 10 copies of *lux*R response regulator genes and the genes encoding N-acyl-homoserine lactone synthetase and an acyl-acyl-carrier protein synthesis enzyme (11). We speculate that MZ1T may utilize the quorum sensing pathway to induce exopolysaccharide production and "clumping" when the cells reach sufficient density. Also encoded are adhesion related

proteins which may be related to exopolysaccharide production, quorum sensing or "clumping."

Nucleotide sequence accession number. The annotated genome sequence of *Thauera aminoaromatica* MZ1T was deposited in GenBank under the accession number CP001281 for the chromosome and CP001282 for the plasmid. The annotation data can also be accessed at Computational Biology at Oak Ridge National Lab (http://genome.ornl.gov/microbial/thau/)

The initial efforts of sequencing, assembly and annotation were made by Joint Genome Institute and Oak Ridge National Lab. The preparation of genomic DNA and subsequent data analysis were performed in the Center for Environmental Biotechnology at University of Tennessee. This work was supported by the Center for Environmental Biotechnology and the University of Tennessee Waste Management Research and Education Institute.

Chapter 4

The Genome Sequence of *Thauera aminoaromatica* MZ1T, an Exopolysaccharide Producing Bacterium, reveals the Carbon Flow and its Potential Role in Carbon Sequestration

Author: Ke Jiang[1], John Sanseverino[2], Ying Wang[3], Alice C. Layton[1], Michael S. Allen[4], Jennifer M. DeBruyn[5], Scott Moser[1], Dan Close[1], Gary S. Sayler[1*]

[1]Center for Environmental Biotechnology, The University of Tennessee, Knoxville, TN 37996

[2]Tennessee Solar Institute, 2450 E.J. Chapman Drive Suite 112, Knoxville, TN 37996

[3]Center for Renewable Carbon, The University of Tennessee, Center, 2506 Jacob Drive, Knoxville, TN 37996

[4]Department of Biological Sciences, University of North Texas, 1155 Union Circle, # 305220, Denton, TX 76203

[5]Department of Biosystems Engineering and Soil Science, 2506 EJ Chapman Drive, The University of Tennessee, Knoxville, TN 37996

[*]Corresponding Author: Gary S. Sayler sayler@utk.edu

Abstract

Thauera aminoaromatica MZ1T is a floc forming bacterium isolated from the wastewater treatment plant of a major industrial chemical manufacturer. It is the first bacterium from the *Thauera* genus whose genome was sequenced. In previous research, MZ1T was identified as a significant component of viscous clusters that resulted in poor sludge dewaterability. In pure culture, MZ1T produces copious quantities of EPS from relatively simple short chain fatty acids. Analysis of the 4.52MB genome indicated the complete glycolytic pathway and citric acid cycle are present as well as the three key enzymes for assimilation of acetate: acetate-CoA ligase and acetate kinase - phosphate acetyl transferase. Three major putative exopolysaccharide gene clusters were found within the genome. One tight cluster with a size of 20.67 kb encoding 14 genes may contain most necessary genes for thaueran formation and translocation. In this cluster, genes responsible for polysaccharide biosynthesis, polymerization, chain length determinant and export have been identified. Another two clusters are loosely organized. Amino acid similarity analysis suggests that the tight cluster may be a result of a later horizontal gene transfer than the other two clusters.

Introduction

In *Thauera* genus, no exopolysaccharide and floc formation in other species has been reported in literature before this study(4, 32, 65, 95). *T. aminoaromatica* MZ1T is the first species in this genus has been observed to produce abundant exopolysaccharide that causes floc formation. The combination of the monosaccharides in the repeating unit identified in thaueran is also unique among bacterial EPS(107).

Strain MZ1T was first isolated from the wastewater treatment facility of Eastman Chemical Company, Kingsport, Tennessee, in the process of identifying the causal agent of abundant zoogleal clusters which were contributing to poor sludge dewaterability(60). During the isolation, zoogleal clusters were isolated from the active sludge using a micromanipulator mounted on an inverted microscope. MZ1T was found forming small colonies on agar plates and floc in liquid medium. Through 16S rRNA sequencing, MZ1T was found to belong to the *Thauera* genus and closely related to the genera *Azoarcus* and *Zoogloea* (60). The lineage of this bacterium is proteobacteria, β-proteobacteria, Rhodocyclales and Rhodocyclaceae. Further characterization of MZ1T through DNA-DNA hybridization and phylogenic analysis suggested that MZ1T belong to *Thauera aminoaromatica* species(54).

Production of EPS is a very common feature for many Gram negative bacteria such as α-proteobacteria *Rhizobium* sp.(23), β-proteobacteria *Azoarcus* sp.(48) and γ-

proteobacterium *Halmonas Maura*(5). However, EPS production had not been found in *Thauera* sp. until the discovery of *Thauera* sp. MZ1T. When compared with other *Thauera* species such as *Thauera selenatis*(65) and *Thauera aromatica*(12), only *T. aminoaromatica* sp. cause cellular flocculation and to produce abundant EPS in liquid culture.

In previous study, EPS of *Thauera* sp. MZ1T was partially purified through filtration, lyophilization and liquid chromatography (LC). Its size was determined by gel-permeation chromatography. The structure and composition was determined by gas chromatography – mass spectrometry (GC-MS), nuclear magnetic resonance (NMR) and Fourier-Transform Infrared (FTIR) spectroscopy. The EPS has an approximate molecular weight of 260 KDa with the repeating units containing four monosaccharides: rhamnose, galacturonic acid, N-acetyl-glucosamine, and N-acetyl-fucosamine(2).

Genes encoding EPS are often found organized in large clusters. In *Pseudomonas solanacearum*, a EPS cluster was found has 7 genes (*epsAPBCDEF*) with a size of 18 kb(47). *Lactococcus lactis* has a plasmid encoding a 12kb EPS gene cluster with 14 genes(*epsRXABCDEFGHIJKL*) (57). *Streptococcus thermophilus* Sfi6 was found has a EPS gene of 14.5kb encoding 13 genes (*epsABCDEFGHIJKLM*) (103). In closely related organisms, such as *Azoarcus* sp. BH72, four major putative EPS, LPS and CPS gene clusters with 18, 16, 13 and 33 genes have been identified(58) and two putative EPS gene clusters were found in *Azoarcus* sp. Ebn1(88).

Since EPS production is unique for MZ1T in the genus *Thauera* and the EPS composition is unique among all the bacterial EPS, we are interested in identifying the genes responsible for EPS production and its biosynthesis pathway, the monosaccharide synthesis, degradation of aromatic compounds, global regulation and acetate assimilation as well as the central carbon metabolism in the genome. These findings will be beneficial for a better understanding of the carbon flow in MZ1T and its potential ability for carbon sequestration through incorporation of acetate into thaueran which prevents the emission of carbon dioxides and methane into the atmosphere.

Materials and Methods

Genome sequencing and annotation

T. aminoaromatica MZ1T is a strain isolated in the Center for Environment Biotechnology at the University of Tennessee. When preparing genomic DNA, cells were grown aerobically in Stoke's medium(2) at 30℃ shaking at 150 rpm. Thaueran was washed off by using 1M NaCl, and genomic DNA was isolated from a 100 mL overnight culture before floc formation using phenol-chloroform method(87). Pulse-Field Gel Electrophoresis (PFGE) was performed to estimate the genome size and melting temperature(21) The genome of MZ1T was sequenced by the Joint Genome Institute (Walnut Creek, CA) using the standard shotgun method and Sanger sequencing(16). Coding sequences (CDS) were identified by Critica and Glimmer gene finders through

the Oak Ridge National Laboratory Genome Analysis Pipeline(7, 22). CDS identification was verified through amino acid similarity analysis in the basic local alignment search tool for proteins (BLASTP) against the nonredundant database of GenBank(3). Ribosome binding sties were identified using Artemis(92). Genes encoding tRNAs were located with tRNAScan-SE tool(64). The functional assignment of CDs was performed by comparing the CDs with the KEGG, InterPro, TIGEFarm, PFarms, and Clusters of Orthologous Groups of Proteins (COGs) databases(42, 77, 81, 101, 110).

Identification of the physical location of thaueran related genes and other genes

Known genes related to polysaccharide synthesis from other organisms, such as *Xanthomonas campestris*(113), *Pseudomonas aeruginosa*(104) and *Nitrosomonas europaea*(15), were blasted against *Thauera aminoaromatica* MZ1Tgenome and direct searching in the computational annotation. Genes related to aromatic compound degradataion, monosaccharide biosynthesis, acetate metabolism and so forth were searched in genome. The identified putative genes were verified in Artemis(92).

Identification of the specific functions and homologous organisms

The identified thaueran related genes were blasted against the NCBI Genebank database and the highest related homolog and Pfam domain were investigated. If the function of the highest related protein is uncharacterized or not well characterized, other

homologs with lowest E-score and identities were utilized to designate thaueran related genes.

Results and Discussion

General features of the genome and mobile elements

T. aminoaromatica MZ1T contains a circular chromosome and a plasmid with a total genome size of 4,518,936 bp, including 3,978 protein-encoding genes. The overall GC content of the genome is 68.2 %. The chromosome is 4,496,212 bp in length with a coding density of 89%, GC content of 68%, 3,903 protein coding genes, 74 structural RNA genes, 89 pseudo genes and 4 copies of 5S, 16S and 23S rRNA genes, two copies in the positive strand and two copies in the negative strand(Figure 4.1). The plasmid (pTha01) is in size of 78,374 bp and has a GC content of 62%, 77% coding density, 75 protein coding genes, 4 pseudo genes and none structural RNA genes(Figure 4.2).

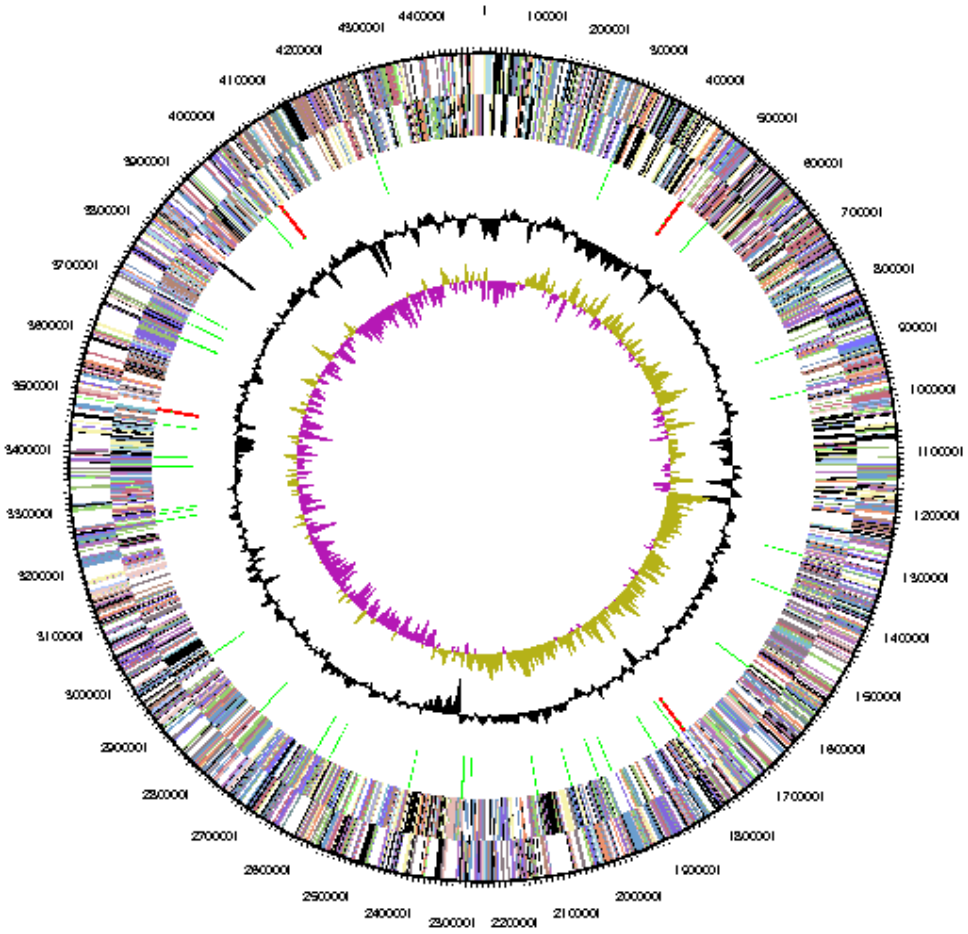


Figure 4.1 Diagram of the *T. aminoaromatica* MZ1T genome.

The outrermost two circles (circles 1 and 2) show the genes in the forward and reverse strans, respectively; different colors indicate different function categories. The next circle (circle 3) shows transposon and IS elements (green) and prophage (red); circles 4 shows the GC content, and circle 5 shows the GC skew.

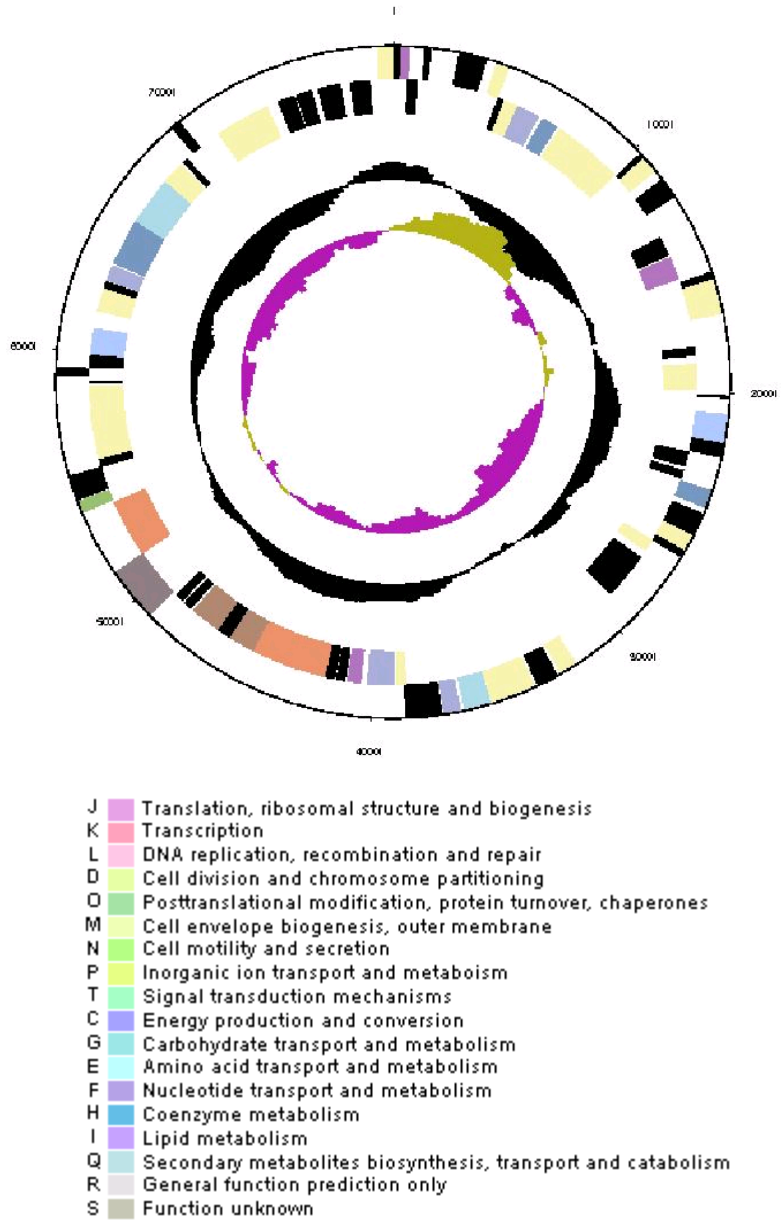


Figure 4.2 Diagram of the *T. aminoaromatica* MZ1T plasmid pTha01.

The outmost two circles (circle1 and 2) show the genes on forward strand and reverse strand (color by COG categories with tRNAs in green, sRNAs in red and other RNAs in black); circle 3 shows GC content and circle 4 shows GC skew.

MZ1T has at least 46 genes related to transposases and several genes encoding phage proteins. Three putative prophage regions were identified in the genome and one was confirmed in prophinder on ACLAME website (http://aclame.ulb.ac.be/) (63). This putative prophage covers 11.9 kb and located in the genome between 3,445,157 and 3,457,057 bp, which corresponds to genes from Tmz1t_3171 to Tmz1t_3179.

Since the genome of *T. aminoaromatica* MZ1T is the first that has been sequenced and annotated in *Thauera* genus, a comparison study was performed with its most related species in the genus of *Azoarcus*, *Azoarcus* sp. strain BH72 as shown in table 4.1 . The genome size, GC content, number of coding sequences and other features are similar for these two organisms, while the hypothetical proteins in MZ1T is more than 5 times than that in *Azoarcus* sp. BH72, which indicates *Azoarcus* sp. BH72 can be a useful reference strain for further study in MZ1T and MZ1T has even more abilities and functions than the current annotation could exhibit.

Table 4.1 Genome features of *Thauera* sp. MZ1T in comparison to the N2-fixing endophyte *Azoarcus* sp. strain BH72.

Feature	*Thaeura* MZ1T	*Azoarcus* BH72
Size of chromosome (bp)	4,518,946	4,376,040
Plasmids	1	0
GC content	68.3	67.9
Coding Sequences	4,054	3,992
Function assigned	3,094	3,418
Conserved hypothetical	510	517
Hypothetical protein	333	57
% of genome coding	89	91.2
Average length (bp)	993	999
Maximal length (bp)	14,006	6,330
rRNA operons	4	4
tRNAs	53	56

Carbon metabolism and signal transduction

Since MZ1T was originally isolated from acetate-rich wastewater, acetate metabolism in MZ1T is of particular interest. There are two acetate assimilation systems in microorganisms. One is AMP-ACS (acetate-CoA ligase) pathway, in which AMP-ACS first converts acetate and ATP to the intermediate acetyladenylate and then reacts acetyl-AMP with CoASH to form acetyl-CoA(8, 121). This is a high-affinity pathway and can function in low concentration acetate environments. The other is reversible PTA-ACKA (acetate kinase and phosphate acetyltransferase) pathway which also can

assimilate acetate in high concentrations of acetate(121). Both pathways are identified in the annotation of MZ1T genome. With respect to AMP-ACS system, three genes (gene Tmz1t_1412, 2652, 3467) are predicted in MZ1T to encode acetate-CoA ligase. The product encoded by gene Tmz1t_2652 shares high similarity with its counterpart in *Azoarcus sp.* BH72; gene Tmz1t_1412 product is closely related to CoA ligase in *Ralstonia pickettii*; and gene Tmz1t_3467 product presents similarity with the enzyme in *Aromatoleum aromatium* EbN1. Regarding PTA-ACKA system, the genes encoding acetate kinase and phosphate acetyltransferase in MZ1T form a gene cluster (gene Tmz1t_3607-3608). It displays high similarity with the corresponding enzymes in *Aromatoleum aromatium* EbN1 and *Azoarcus sp.* BH72. Although gene Tmz1t_1093 and 1580 also encode phosphate acetyltransferase in MZ1T, gene 1093 and 1580 products do not display significant similarity with the enzyme encoded by gene Tmz1t_3607. To complete the acetate assimilation through either AMP-ACS or the reversible PTA-ACKA pathway, glyoxylate bypass (GB) is required to permit the net accumulation of four-carbon biosynthesis during growth on acetate substrate (123). Two unique enzymes necessary for glyoxylate cycle, isocitrate lyase (ICL) and malate synthase (MAS), are present in MZ1T genome. Isocitrate lyase is encoded by gene Tmz1t_1247 and has 95% identity with ICL in *Azoarcus sp.* BH72; and malate synthase is encoded by gene Tmz1t_3129 and is related to MAS in *Aromatoleum aromatium* EbN1. All these above findings indicate that *Thauera* MZ1T has the capacity

to utilize acetate as its carbon source, which is consistent with the growth experiments when grown anaerobically.

Lipopolysaccharide, capsule polysaccharide and exopolysaccharide

Although EPS related genes are dispersed throughout the whole genome of MZ1T, three major clusters of genes encoding proteins for the formation of EPS, capsule polysaccharide and LPS have been identified in MZ1T genome.

Lipopolysaccharide (LPS) is composed by the O side chain and core oligosaccharide and lipid A which are covalently linked and located at the outer membrane of Gram-negative bacteria. Genes encoding for the enzyme responsible for the first step of lipid A synthesis (LpxA Tmz1t_2165), disaccharide synthetase (LpxB Tmz1t_2166), acylase (LpxD Tmz1t_2168), UDP-2,3-diacylglucosamine hydrolase (LpxH Tmz1t_3049) and tetraacyldisaccharide 4'-kinase (LpxK Tmz1t_3215) have been identified in MZ1T genome(Table 4.2).

Three thaueran related gene clusters were found in MZ1T, one tightly organized and the other two loosely clustered, which is rather unusual but not unique; for instance, *Azoarcus phosphatis* has two *eps* gene clusters responsible for the synthesis of two different EPSs - one more acidic than the other(67, 70).

More than half genes within the two latter putative thaueran gene clusters have high similarity to *Azoarcus* sp., none of the genes in putative thaueran gene

cluster_1(EPS1) is highly related to *Azoarcus* sp. Furthermore, no particular organisms have high homology to multiple genes in this cluster. Cluster_1 is from Tmz1t_1114 to Tmz1t_1127 with a size of 20.67 kb and encodes 14 genes which include most of the genes necessary for EPS production and transport(Figure 4.3). Genes responsible for polysaccharide biosynthesis, polymerization, and export as well as chain length determinant have been identified. No regulatory elements have yet been identified in these cluster; however, since tyrosine kinase was found to be able to control the initiation of EPS production in *Streptococcus thermophilus* (75), the tyrosine kinase gene identified in putative thaueran gene cluster_1 may be involved in the regulation of thaueran production.

Most genes directing polysaccharide biosynthesis, encoding glycosyl transferase, UDP-N-acetylglucosamine 2-epimerase, UDP-glucose/GDP-mannose dehydrogenase, have high homology (52 to 81% identity) to corresponding genes in various EPS producing bacteria such as *Pseudomonas aeruginosa*(104) and *Nitrosomonas europaea*(15). Two genes involved in EPS chain length determination and export also have high homology (46 to 47% identity) to related genes in *Nitrosomonas europaea* ATCC 19718. The gene responsible for EPS polymerization has a 28% identity with the EPS polymerase gene in *Bacillus licheniformis* ATCC 14580 (Table 4.3). Based on the intergenic distances between the genes within this cluster, there are about 4 putative operons and genes in this cluster are all located on the positive strand.

In the second putative thaueran gene cluster, there is a core cluster (EPS2) encoding 18 genes with a size of 22 kb which is probably a single operon according to the intergenic distances between genes in this core cluster(Figure 4.4). Genes in EPS2 are all on the negative strand. Six genes encode proteins that are associated with the PEP-CTERM/EpsH system. The characteristics of this system is that they all possess Pro-Glu-Pro C-terminal domain and are involved in exopolysaccharide associated protein sorting in Gram-negative bacteria which is corresponding to the LPXTG/sortase system in gram positive bacteria(41). The discovery of *wzy* gene (Tmz1t_3268) in this cluster implicates that the Wzy-dependent pathway of polysaccharide synthesis and export has been utilized by MZ1T. Other gene homologs in this pathway have also been identified: the chain length determinant protein gene, *Wzz* gene (Tmz1t_3281), polysaccharide export protein gene, *Wza* gene (Tmz1t_3282), protein-tyrosine kinase gene *wzc* gene (Tmz1t_3280) (Table 4.4).

The third putative thaueran gene cluster (EPS3) has 45 genes in total. In this cluster, genes related to polysaccharide synthesis and export are not tightly linked but are scattered in a wide range of about 54 kb(from Tmzlt_3769 to 3770) and on both positive and negative strands. Like the second putative thaueran gene cluster, more than half of the genes in EPS3 also have high identities with *Azoarcus* sp.. Some genes encoding proteins with unknown functions are scattered in it. Several transposon genes are identified within and close to this cluster. They all encode transposase IS4 family proteins(Tmz1t_3805, Tmz1t_3787 and Tmz1t_3781) and all have high similarity with

Azoarcus sp., which indicates that this cluster may be a product of horizontal gene transfer from *Azoarcus* sp. and the existence of lipopolysaccharide synthesis genes in this cluster suggested that EPS3 may be responsible for lipopolysaccharide or capsular polysaccharide rather than exopolysaccharide formation in MZ1T.

None of the genes from EPS1 are highly related to *Azoarcus* species and its GC content is 59%, which is much lower than the overall GC content, 68.2%, and the GC content of EPS2, 69%. Based on its lower GC content and lacking high similarities of amino acid with *Azoarcus* spp., EPS1 is probably a newer genetic element than the other two putative thaueran gene clusters and the rest of the gene compositions of MZ1T due to a recent horizontal gene transfer event(50).

The dramatic differences between EPS1 and EPS2 as well as the fact that the composition, yields and other factors of EPS may be influenced by the growth conditions, medium and environment, MZ1T may activate different clusters to adopt different environment and nutrient conditions.

Exopolysaccharide formation and transport pathways

The discovery of genes encoding Wzy family protein and Wzx homolog - polysaccharide-specific-transport (PST) proteins as well as other characterized EPS proteins in MZ1T genome imply that Wzy-dependent pathway may be used in MZ1T EPS biosynthesis(41). This pathway was first described in *Salmonella* O-antigen

synthesis(116) and also found for many other polysaccharides, such as that in *Azoarcus* EbN1 and *Rhodospirillum rubrum*(41), and mauran in *Halomonas Maura*(6). In this pathway, the lipid-linked repeating units of the polysaccharide are assembled on the cytoplasmic side of the inner membrane and then transported to the periplasmic side of the inner membrane for polymerization and ligation. After some modification in periplasm, the polysaccharide is translocated across the outer membrane(117).

The synthesis pathway involving ABC transporters has been found for group II capsular polysaccharide in *Escherichia coli*, *Haemophilus*, *Neisseria Rhizobium* and *Agrobacterium*. It has different assembly steps and their specific accruing locations(119). In contrast to the Wyz dependent pathway, in ABC transporter pathway, the repeating units are synthesized and polymerized at the inner face of the cytoplasmic membrane, and translocated onto the periplasmic side for ligation with lipid(116). The final product is then translocated from periplasm to the surface of the outer membrane by interacting with export system of periplasm and outer membrane(117). Proteins involved in the ABC transporter dependent pathway, such as the ABC transporter and ligases, are also found in the putative thaueran gene clusters in MZ1T, which suggests that this pathway may be also adopted by strain MZ1T.

Table 4.2 Genes involving LPS production in *T. aminoaromatica* MZ1T

Gene number	Product	Gene Name	Best hits Organisms
Tmz1t_2165	lipid-A-disaccharide synthase	*lpxB*	*Azoarcus* sp. BH72
Tmz1t_2166	acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	*lpxA*	*Azoarcus* sp. BH72
Tmz1t_3421	UDP-3-0-acyl N-acetylglucosamine deacetylase	*lpxC*	*Azoarcus* sp. BH72
Tmz1t_2168	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	*lpxD*	*Azoarcus* sp. EbN1
Tmz1t_3049	UDP-2,3-diacylglucosamine hydrolase	*lpxH*	*Azoarcus* sp. BH72
Tmz1t_3215	tetraacyldisaccharide 4'-kinase	*lpxK*	*Azoarcus* sp. BH72

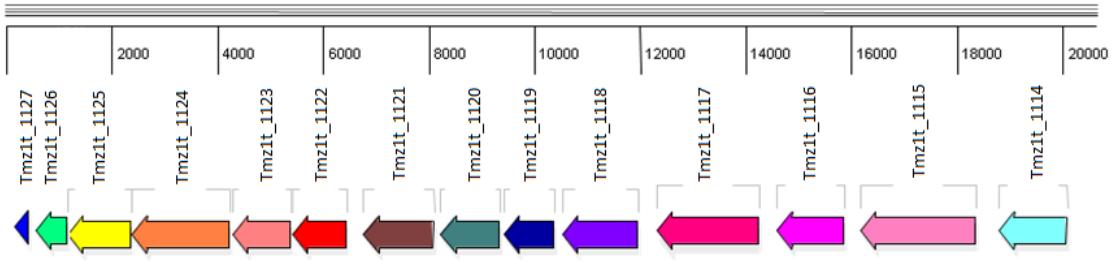


Figure 4.3 Thaueran gene cluster1 map with strand ruler (bp).

This cluster has a size of 20.67kb encoding 14 genes for EPS production and export.

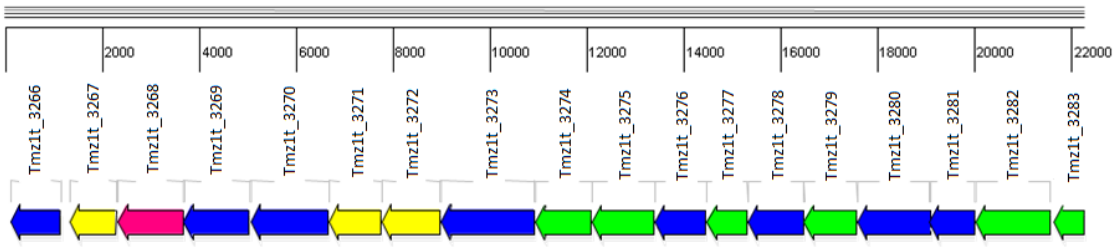


Figure 4.4 Thaueran gene cluster_2 core cluster gene map with strand ruler (bp).

This core cluster has 18 genes with a size of 22kb found in the second putative thaueran gene cluster_2. The small intergenic distance(<250bp) between genes in this cluster suggested that this core cluster may be a single operon. Two glycosyl transferase genes are in yellow; six genes related to PEP-CTERM sorting system are in green; *wzy* gene is in pink.

Monosaccharide biosynthesis

As expected from their diverse roles in gluconeogenesis, LPS and EPS biosyntheisis, energy transduction, and central metabolism, the genes for monosaacharide biosynthesis are widely distributed across the genome. One exception to this appears to be a rhamnose operon containing several components of the rhamnose synthase holoenzyme. The encoded proteins are responsible for the conversion of D-glucose-1-phosphate to dTDP-L-rhamnose. These include the dTDP-glucose-4,6-dehytdratase (Tmz1t_2139) which converts D-glucose-1-phosphate to dTDP-4-dehydro-6-deoxy-D-glucose, and the enzyme catalyzing the last step in the pathway, dTDP-4-dehydrorhamnose reductase (Tmz1t_2138) that catalyzes the conversion of dTDP-L-4-oxo-rhamnose to dTDP-L-rhamnose.

Biosynthesis dTDP-D-N-acetylfucosamine typically occurs as an offshoot of the aforementioned pathway from dTDP-4-oxo-6-deoxy-D-glucose. Two genes potentially involved in this transformation are Tmz1t_1130 and Tmz1t_1127, homologs of fdtB and fdtC respectively, which have been identified in *Aneurinibacillus thermoaerophilus* strain

L420-91T S-layer glycan biosynthesis and encode the last two steps in the synthesis of dTDP-D-N-acetylfucosamine (85). The two encoded proteins are responsible for the transfer of an amine from L-glutamate and its subsequent acetylation. In MZ1T, these genes are contained in a large cluster of EPS-related biosynthetic genes. However, it should be noted that MZ1T is predicted to contain the L stereoisomer of FucNAc in its EPS, which may explain the failure to detect a homologue of fdtA as is found in *A. thermoaerophilus*.

Global gene regulation

The genome of *T. aminoaromatica* MZ1T contains a total of six sigma factors controlling global gene regulation. These include the housekeeping sigma factor σ70 (*rpoD*, Tmz1t_1189), the nitrogen regulator σ54 (*rpoN*, Tmz1t_0708), the heat shock sigma factor σ32 (*rpoH*, Tmz1t_1062), as well as three extracytoplasmic function (ECF) sigma factors (40). Of the ECF sigma factors, Tmz1t_2321 is most likely a stress-responsive homolog of σE (*rpoE*) given the co-localization of homologs of *rseABC* (Tmz1t_2320-2318), which regulate σE turnover in *Escherichia coli* (19, 20). The second ECF σ, Tmz1t_1854, is located immediately upstream of an arsenical resistance protein and regulator, suggesting a role in the response to heavy metals. The final ECF σ factor, Tmz1t_1920, appears to be located within a large operon containing several conserved unknown proteins. This predicted operon is adjacent to a second operon containing a ferric reductase and two putative cytochromes. Given the lack of any identifiable fecI

homologs in the genome MZ1T, this might indicate a potential role in the regulation of iron uptake for this sigma factor.

Aromatic compound degradation

MZ1T contains homologs to enzymes involved in degradation of monoaromatics, suggesting a capacity for degradation of benzoate (aerobic and anaerobic), toluene (aerobic) and phenol (aerobic).

Homologs to proteins involved in the aerobic degradation of benzoate and phenol, proceeding via a catechol intermediate, are present in MZ1T. Benzoate 1,2 dioxygenase subunit genes(Tmz1t_3096 and Tmz1t_3097) and associated ferredoxin reductase(Tmz1t_3098) and dehydrogenase(Tmz1t_3099) are most similar to BenABCD genes in *Dechloromonas aromatica* RCB, with protein identities between 83% and 89%. Genes Tmz1t_3011 to 3016 encode subunits of phenol (or toluene) monooxygenases, most similar to proteins from *Azoarcus* strain BH72 (74 – 88% identity). Adjacent to these genes is a putative catechol 2, 3 dioxygenase gene (Tmz1t_3109) with 83% identity to BH72, catalyzing the further degradation of the catechol intermediate for these aerobic monoaromatic pathways.

Thauera aromatica strains, as well as several *Azoarcus* and *Dechloromonas* spp. are capable of coupling anaerobic toluene degradation with nitrate respiration (17). This pathway is well characterized, proceeding via benzylsuccinic acid to benzoyl-CoA (102).

Despite its close relationship to these strains, MZ1T does not have homologs of the genes in this pathway. It does, however have a putative genes for anaerobic degradation of benzoate: Tmz1t_3196 is a benzoate CoA ligase homolog (79% identity to *Azoarcus* sp. BH72) suggesting a capability to covert benzoate to benzoyl-CoA. MZ1T also has homologs to the suite of enzymes catalyzing the further degradation of benzoyl-CoA through ring hydrolysis, including a benzoyl-CoA reductase (Tmz1t_2948 – 2951), Enoyl CoA-hydratase (Tmz1t_2952), hydroxyl acyl-CoA dehydrogenase (Tmz1t_2953), and 3-oxoacyl-CoA hydrolase (Tzm1t_2954), exhibiting high similarity (86% to 96% identity) and synteny to the benzoyl degradation operon characterized in *T. aromatica* (13).

Virulence and interaction factors

MZ1T has 3 pairs of toxin and antitoxin encoded by 12 genes. At least 10 pairs of two component regulatory systems were found. Five out of ten of them are homologues of *Azoarcus* sp. two are homologous to *Burkholderia* sp. Five of them are constituted of PAS sensor protein and LuxR pairs.

Iron-transport related proteins

There are at least 10~14 TonB-dependent receptors and TonB family proteins, which implies that MZ1T may have more than one TonB-dependent regulatory systems. Therefore, it is possible that TonB-dependent transducers may not only interact with their

cognate anti-sigma factor but also have cross-signaling between different systems, thus resulting in complicated regulatory networks (56).

Secretion and communication

Efflux pumps for heavy metal resistance to arsenic, cadmium, lead, silver, zinc but not for selenium are identified in the genome. MZ1T have type I, II and III secretion systems(Table 4.5). Secretion systems III, usually found in plant pathogens, was found enable the organism to inject toxins directly into the host cells(59). The existence of this system in MZ1T supports again that MZ1T is closely related to a plant pathogen, *Azoaracus* sp. Ebn1, in which secretion systems III is indispensible for direct the toxin into the host plant(88).

MZ1T has a probable quorum sensing pathway. There are at least 10 luxR response regulator genes and a N-acyl-homoserine lactone synthetase and Tmz1t_2328 is an acyl-acyl-carrier protein synthesis enzyme. Therefore, MZ1T may utilize the quorum sensing pathway to induce exopolysaccharide production and "clumping" when the cells reach sufficient density. Also, it has adhesion related proteins which may be related to exopolysaccharide producing, quorum sensing or "clumping."

Motility

It has genes for both a flagellum and type IV pilus(fimbriae) for flagellum independent twitching motility(Table 4.5).

In conclusion, MZ1T was found have genes for various carbon metabolic pathways, aromatic compound degradation pathways, three putative exopolysaccharide synthesis and export clusters, various heavy metal efflux pumps, channels and four types of secretion systems to adopt environment that may be harsher than its original habitat. The existence of the two acetate metabolism pathways enables MZ1T to lead the carbon in acetate into the substrate for monosaccharide synthesis and in this way, acetate in the habitat of MZ1T could be converted and sequestrated into thaueran. Since the latter is much more stable than carbon dioxide and methane that are usually produced by acetate, and thaueran has the potential application for heavy metal absorption as well as providing rare natural monsaccharide, such as frucosamine, this carbon flow from acetate to thaueran through MZ1T will not only reduce the green house gas emission but also provide useful strategies to clean up heavy metal contaminated sites and offer valuable sugar sources. Therefore, this genome study reveals a potentially important role of MZ1T in environmental carbon flow and sequestration.

Nucleotide sequence accession number

The annotated genome sequence of *T. aminoaromatica* MZ1T has been deposited in the GenBank Database under accession number CP001281 for the chromosome and CP001282 for the plasmid.

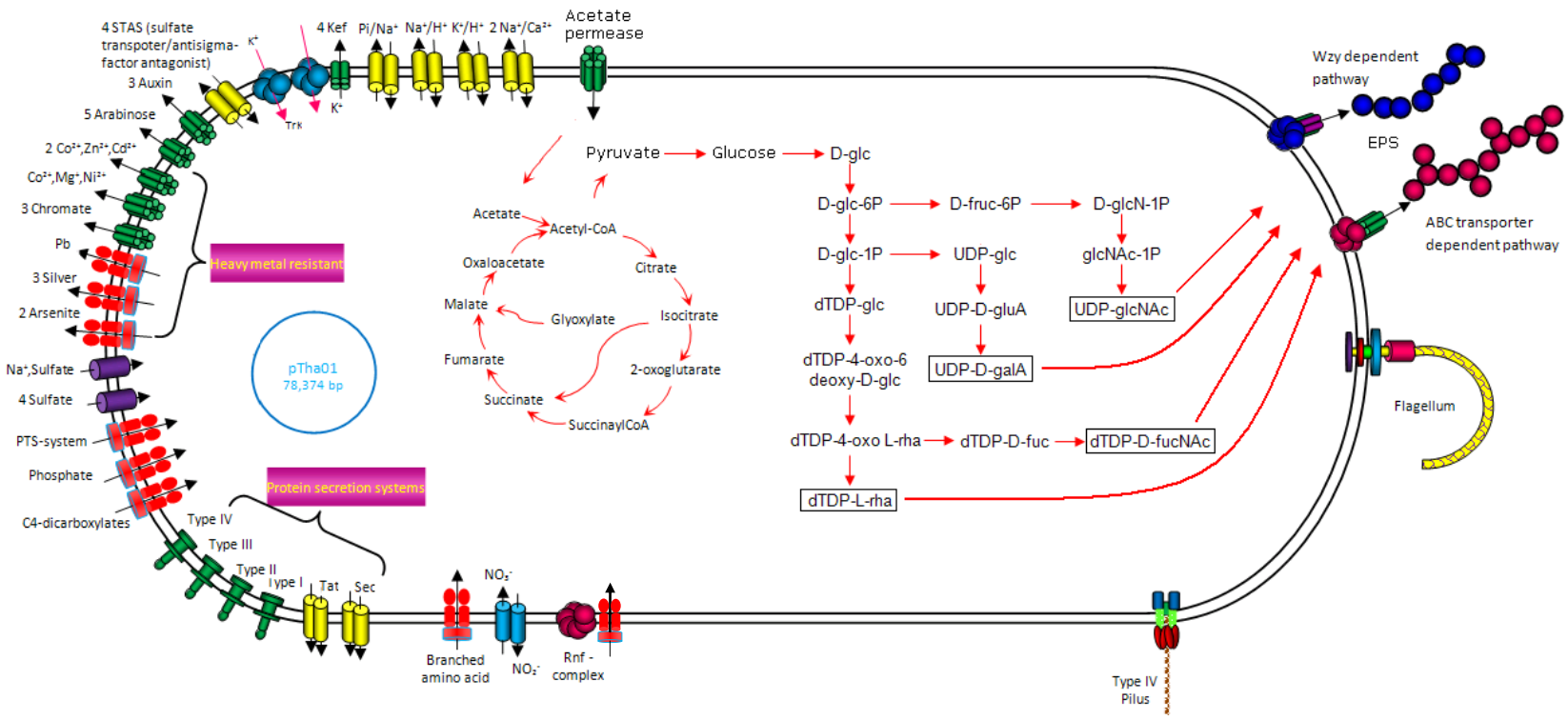


Figure 4.5 A preliminary diagram of the general physiological features of MZ1T

Table 4.3 Summary of MZ1T thaueran gene clulster_1(EPS1)

Gene	*Strand*	*Start*	*End*	*AA Length*	*Intergenic distance*	*Gene Name*	*Gene Product*	*Identities (%)*	*Organism*
Tmz1t_1114	f	1216366	1217661	432	437	*yccz*	Polysaccharide export protein	47	*Nitrosomonas europaea*
Tmz1t_1115	f	1218098	1220290	731	298	*epsB*	Chain length determinant protein	46	*Nitrosomonas europaea*
Tmz1t_1116	f	1220588	1221865	426	336	*wbpO*	Polysaccharide biosynthesis protein	81	*Bordetella bronchiseptica*
Tmz1t_1117	f	1222201	1224138	646	335	n/a	Acyltransferase 3	58	*Shewanella woodyi* 51908
Tmz1t_1118	f	1224473	1225927	485	45	n/a	Polysaccharide biosynthesis protein	42	*Parvibaculum lavamentivorans*
Tmz1t_1119	f	1225972	1227033	354	61	n/a	Hypothetical protein		*n/a*
Tmz1t_1120	f	1227094	1228230	379	116	*YveQ*	Capsular polysaccharide biosynthesis protein	28	*Bacillus licheniformis* DSM 13
Tmz1t_1121	f	1228346	1229689	448	310	ugd	UDP-glucose 6-dehydrogenase	66	*Methylococcus capsulatus*
Tmz1t_1122	f	1229999	1231027	343	27	n/a	Putative glycosyltransferase	31	*Polaribacter irgensii* 23-P
Tmz1t_1123	f	1231054	1232160	369	61	n/a	UDP-N-acetylglucosamine 2-epimerase	68	*Syntrophus aciditrophicus* SB
Tmz1t_1124	f	1232221	1234071	617	-3	n/a	HeparinaseII/III family portein	58	*Pseudomonas aeruginosa*
Tmz1t_1125	f	1234068	1235285	406	-3	n/a	Glycosyl transferases group 1	78	*Pseudomonas aeruginosa*
Tmz1t_1126	f	1235282	1235887	202	79	n/a	UDP-galactose phosphate transferase	76	*Vibrio vulnificus* YJ016
Tmz1t_1127	f	1235966	1236628	221	n/a	n/a	Putative acetyl transferase protein	52	*Acidovorax sp*. JS42

Table 4.4 Summary of MZ1T thaueran gene cluster_2 core cluster(EPS2)

Gene	*Strand*	*Start*	*Stop*	*Size (bp)*	*A.A length*	*Intergenic Distance (bp)*	*Product*	*Identities (%)*	*Organisms*
Tmz1t_3266	r	75670	74621	1050	350	181	acyltransferase 3	32	*Aeromonas salmonicida* A449
Tmz1t_3267	r	76814	75852	963	321	31	glycosyl transferase family 2	33	*Haloarcula marismortui*
Tmz1t_3268	r	78204	76846	1359	453	-13	wzy family polymerase, exosortase system type 1 associated	49	*Thiobacillus denitrificans* ATCC 25259
Tmz1t_3269	r	79565	78192	1374	458	16	putative CapK protein	67	*Nitrosomonas europaea*
Tmz1t_3270	r	81192	79582	1611	537	-4	exosortase 1	44	*Rhodoferax ferrireducens*
Tmz1t_3271	r	82286	81189	1098	366	-4	glycosyl transferase group 1	57	*Dechloromonas aromatica* RCB
Tmz1t_3272	r	83494	82283	1212	404	21	glycosyl transferase group 1	68	*Dechloromonas aromatica* RCB
Tmz1t_3273	r	85444	83516	1929	643	15	exosortase 1 system-associated amidotransferase 1	86	*Azoarcus sp*. BH72
Tmz1t_3274	r	86626	85460	1167	389	-4	sugar transferase, PEP-CTERM/EpsH1 system associated	63	*Azoarcus sp*. BH72
Tmz1t_3275	r	87912	86623	1290	430	4	sugar transferase, PEP-CTERM/EpsH1 system associated	53	*Alkalilimnicola ehrlichei* MLHE-1
Tmz1t_3276	r	88978	87917	1062	354	16	conserved hypothetical protein	79	*Azoarcus sp*. BH72
Tmz1t_3277	r	89834	88995	840	280	1	polysaccharide deactylase family protein, PEP-CTERM locus subfamily	80	*Azoarcus sp*. BH72
Tmz1t_3278	r	91002	89836	1167	389	-4	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14).	79	*Azoarcus sp*. BH72
Tmz1t_3279	r	92093	90999	1095	365	11	secretion ATPase, PEP-CTERM locus subfamily	72	*Azoarcus sp*. BH72
Tmz1t_3280	r	93622	92105	1518	506	-38	conserved hypothetical protein	44	*Azoarcus sp*. BH72
Tmz1t_3281	r	94529	93585	945	315	13	Non-specific protein-tyrosine kinase (EC 2.7.10.1)	68	*Azoarcus sp*. BH72
Tmz1t_3282	r	96087	94543	1545	515	70	polysaccharide chain length determinant protein, PEP-CTERM locus subfamily	66	*Azoarcus sp*. BH72
Tmz1t_3283	r	96787	96158	630	210	n.d.	polysaccharide export protein, PEP-CTERM sytem-associated	81	*Azoarcus sp.* EbN1

Table 4.5 Gene clusters of *T. aminoaromatica* MZ1T for polysaccharide synthesis, aromatic compound degradation, urease, virulence and flagella synthesis

Gene no.	*Gene name*	*Comments/function of gene product*	*Best hits organism*
Acetate uptake			
Tmz1t_3489	*actP*	acetate permease	*Azoarcus* sp. (strain BH72).
Tmz1t_0556	*actP*	acetate permease	*Azoarcus* sp. (strain BH72).
EPS gene cluster 2			
Tmz1t_3209		glycosyl transferase family 2	*Azoarcus* sp. (strain BH72).
Tmz1t_3216		Biopolymer transport protein ExbD/TolR	*Azoarcus* sp. (strain BH72).
Tmz1t_3217		MotA/TolQ/ExbB proton channel	*Azoarcus* sp. (strain BH72).
Tmz1t_3244		polysaccharide biosynthesis protein	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3246		lipopolysaccharide biosynthesis protein-like protein	*Geobacter uraniumreducens* Rf4.
Tmz1t_3249		glycosyl transferase family 2	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3250		polysaccharide deacetylase	*Azoarcus* sp. (strain BH72).
Tmz1t_3251		glycosyl transferase family 2	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3252		conserved hypothetical protein	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3252		glycosyl transferase family 2	*Gloeobacter violaceus*
Tmz1t_3253		sulfatase	Thiobacillus denitrificans (strain ATCC 25259).
Tmz1t_3254		4-amino-4-deoxy-L-arabinose transferase and related glycosyltransferase of PMT family-like	*Rhodospirillum rubrum* (strain ATCC 11170 / NCIB 8255).
Tmz1t_3255		hypothetical protein	
Tmz1t_3256		TDP-4-keto-6-deoxy-D-glucose transaminase	*Nitrosospira multiformis* (strain ATCC 25196 / NCIMB 11849).

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_3257		GtrA family protein	*Nitrosomonas europaea*
Tmz1t_3258		GCN5-related N-acetyltransferase	*Mariprofundus ferrooxydans* PV-1.
Tmz1t_3259		glycosyl transferase family 2	*Nitrosospira multiformis* (strain ATCC 25196 / NCIMB 11849).
Tmz1t_3260		Tetratricopeptide TPR_2 repeat protein	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3261		hypothetical protein	*Methylococcus capsulatus*
Tmz1t_3262		membrane bound O-acyl transferase MBOAT family protein	*Methylococcus capsulatus*
Tmz1t_3263		glycosyl transferase group 1	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3264		conserved hypothetical protein	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3265		Asparagine synthase (glutamine-hydrolyzing)	*Alkalilimnicola ehrlichei* (strain MLHE-1).
Tmz1t_3266		acyltransferase 3	*Aeromonas salmonicida* (strain A449).
Tmz1t_3267		glycosyl transferase family 2	*Haloarcula marismortui*
Tmz1t_3268		wzy family polymerase, exosortase system type 1 associated	*Thiobacillus denitrificans* (strain ATCC 25259).
Tmz1t_3269		putative CapK protein	*Nitrosomonas europaea*
Tmz1t_3270		exosortase 1	*Rhodoferax ferrireducens* (strain DSM 15236 / ATCC BAA-621 / T118).
Tmz1t_3271		glycosyl transferase group 1	*Dechloromonas aromatica* (strain RCB).
Tmz1t_3272		exosortase 1 system-associated amidotransferase 1	*Azoarcus* sp. (strain BH72).
Tmz1t_3273		sugar transferase, PEP-CTERM/EpsH1 system associated	*Azoarcus* sp. (strain BH72).
Tmz1t_3274		sugar transferase, PEP-CTERM/EpsH1 system associated	*Alkalilimnicola ehrlichei* (strain MLHE-1).

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_3275		FemAB-related protein, PEP-CTERM system-associated	*Azoarcus* sp. (strain BH72).
Tmz1t_3276		polysaccharide deactylase family protein, PEP-CTERM locus subfamily	*Azoarcus* sp. (strain BH72).
Tmz1t_3277		UDP-N-acetylglucosamine 2-epimerase	*Azoarcus* sp. (strain BH72).
Tmz1t_3278		secretion ATPase, PEP-CTERM locus subfamily	*Azoarcus* sp. (strain BH72).
Tmz1t_3279		conserved hypothetical protein	*Azoarcus* sp. (strain BH72).
Tmz1t_3280		Non-specific protein-tyrosine kinase	*Azoarcus* sp. (strain BH72).
Tmz1t_3281		polysaccharide chain length determinant protein, PEP-CTERM locus subfamily	*Azoarcus* sp. (strain BH72).
Tmz1t_3282		polysaccharide export protein, PEP-CTERM sytem-associated	*Azoarcus* sp. (strain EbN1)
EPS gene cluster 3			
Tmz1t_3769		glycosyl transferase group 1	*Congregibacter litoralis* KT71.
Tmz1t_3770		polysaccharide deacetylase	*Congregibacter litoralis* KT71.
Tmz1t_3771		nucleoside-diphosphate-sugar epimerase-like	*Congregibacter litoralis* KT71.
Tmz1t_3772		glycosyltransferase	*delta proteobacterium* MLMS-1.
Tmz1t_3773		glycosyl transferase family 2	*Roseiflexus castenholzii* DSM 13941.
Tmz1t_3774		O-antigen polymerase	*Verminephrobacter eiseniae* (strain EF01-2).
Tmz1t_3775		glycosyl transferase family 25	*Ralstonia eutropha* (strain ATCC 17699)
Tmz1t_3776		transposase, IS4 family	*Azoarcus* sp. (strain EbN1)
Tmz1t_3777		lipopolysaccharide biosynthesis protein	*Azoarcus* sp. (strain EbN1)
Tmz1t_3778		UDP-N-acetylglucosamine 2-epimerase	*Pseudomonas aeruginosa* (strain UCBPP-PA14).

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_3779		polysaccharide biosynthesis protein	*Burkholderia dolosa* AUO158
Tmz1t_3780		glycosyl transferase family 2	*Campylobacter jejuni*
Tmz1t_3781		transposase IS4 family protein	*Azoarcus* sp. (strain BH72).
Tmz1t_3783		Glycosyltransferase-like	*Microscilla marina* ATCC 23134.
Tmz1t_3784		Transferase	*Microscilla marina* ATCC 23134.
Tmz1t_3785		glycosyl transferase, WecB/TagA/CpsF family	*Microscilla marina* ATCC 23134
Tmz1t_3786		transposase IS4 family protein	*Azoarcus* sp. (strain EbN1)
Tmz1t_3787		Undecaprenyl-phosphate glucose phosphotransferase	*Azotobacter vinelandii* AvOP.
Tmz1t_3788		diacylglycerol kinase	*Psychromonas ingrahamii* (strain 37).
Tmz1t_3789		lipopolysaccharide biosynthesis protein	*Rhodoferax ferrireducens* (strain DSM 15236)
Tmz1t_3790		polysaccharide export protein	*Rhodoferax ferrireducens* (strain DSM 15236)
Tmz1t_3791		hypothetical protein	*Burkholderia pseudomallei* Pasteur 52237
Tmz1t_3792		lipoprotein	*Burkholderia pseudomallei*
Tmz1t_3793		protein of unknown function DUF940 membrane lipoprotein putative	*Burkholderia pseudomallei* 305.
Tmz1t_3794		glycosyl transferase family 4	*Halorhodospira halophila* (strain DSM 244)
Tmz1t_3795		NAD-dependent epimerase/dehydratase	*Psychrobacter* sp. PRwf-1
Tmz1t_3798		glycosyl transferase family 2	*Photobacterium* sp. SKA34
Tmz1t_3802		glycosyl transferase group 1	*Photobacterium* sp. SKA34
Tmz1t_3801		polysaccharide biosynthesis protein CapD	*Comamonas testosteroni* KF-1
Tmz1t_3802		NAD-dependent epimerase/dehydratase	*Azoarcus* sp. (strain EbN1)
Tmz1t_3803		UDP-N-acetylglucosamine 2-epimerase	*Azoarcus* sp. (strain EbN1)

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_3804		glycosyl transferase group 1	*Burkholderia phytofirmans* PsJN.
Tmz1t_3805		transposase IS4 family protein	*Azoarcus* sp. (strain BH72)
Tmz1t_3806		conserved hypothetical protein	*Azoarcus* sp. (strain EbN1)
Tmz1t_3807		protein of unknown function DUF820	*Polaromonas naphthalenivorans* (strain CJ2)
Tmz1t_3808		protein of unknown function DUF820	*Halorhodospira halophila* (strain DSM 244 / SL1)
Tmz1t_3810		glucose-1-phosphate thymidylyltransferase	*Psychrobacter cryohalolentis* (strain K5).
Tmz1t_3811		dTDP-4-dehydrorhamnose 3,5-epimerase	*Hahella chejuensis* (strain KCTC 2396)
Tmz1t_3812		lipid A ABC exporter, fused ATPase and inner membrane subunits MsbA	*Azoarcus* sp. (strain EbN1)
Tmz1t_3826		ABC transporter related	*Azoarcus* sp. (strain BH72)
Tmz1t_3827		ABC transporter related	*Azoarcus* sp. (strain BH72).
Tmz1t_0402		inner-membrane translocator	*Azoarcus* sp. (strain BH72).
Tmz1t_0403		inner-membrane translocator	*Azoarcus* sp. (strain BH72).
Tmz1t_0404		Extracellular ligand-binding receptor	*Azoarcus* sp. (strain BH72).
Tmz1t_3838		lipid A biosynthesis acyltransferase	*Azoarcus* sp. (strain BH72).
Tmz1t_3839		lipid A biosynthesis acyltransferase	*Bordetella avium* (strain 197N).
Aromatic compound degradation			
Tmz1t_3111		oxidoreductase FAD/NAD(P)-binding domain protein	*Azoarcus* sp. (strain BH72).
Tmz1t_3112		Phenol hydroxylase conserved region	*Azoarcus* sp. (strain BH72).

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_3113		methane/phenol/toluene hydroxylase	*Azoarcus* sp. (strain BH72).
Tmz1t_3114		monooxygenase component MmoB/DmpM	*Dechloromonas aromatica* (strain RCB)
Tmz1t_3115		Phenol 2-monooxygenase	*Azoarcus* sp. (strain BH72).
Tmz1t_3116		Phenol hydroxylase subunit	*Azoarcus* sp. (strain BH72).
Urease			
Tmz1t_3930		ABC transporter related	*Opitutaceae* bacterium TAV2
Tmz1t_3931		urea ABC transporter, urea binding protein	*Azoarcus* sp. (strain BH72).
Tmz1t_3932		urea ABC transporter, permease protein UrtB	*Azoarcus* sp. (strain BH72).
Tmz1t_3933		urea ABC transporter, permease protein UrtC	*Azoarcus* sp. (strain BH72).
Tmz1t_3934		urea ABC transporter, ATP-binding protein UrtD	*Azoarcus* sp. (strain BH72).
Tmz1t_3935		urea ABC transporter, ATP-binding protein UrtE	*Azoarcus* sp. (strain BH72).
Tmz1t_3936		Urease accessory protein UreD	*Azoarcus* sp. (strain BH72).
Tmz1t_3937		urease, gamma subunit	*Azoarcus* sp. (strain BH72).
Tmz1t_3938		urease, beta subunit	*Azoarcus* sp. (strain BH72).
Tmz1t_3939		urease, alpha subunit	*Azoarcus* sp. (strain BH72).
Tmz1t_3940		UreE urease accessory domain protein	*Azoarcus* sp. (strain BH72).
Tmz1t_3941		Urease accessory protein UreF	*Azoarcus* sp. (strain BH72).
Tmz1t_3942		urease accessory protein UreG	*Pseudomonas stutzeri* (strain A1501)
Virulence			
Tmz1t_3767		addiction module antitoxin, RelB/DinJ family	*Azoarcus* sp. (strain EbN1)
Tmz1t_3768		addiction module toxin, RelE/StbE family	*Xylella fastidiosa* Ann-1

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_2109		addiction module toxin, RelE/StbE family	*delta proteobacterium* MLMS-1
Tmz1t_2746		addiction module toxin, RelE/StbE family	*Acidovorax* sp. (strain JS42)
Tmz1t_2747		prevent-host-death family protein	*Azoarcus* sp. (strain EbN1)
Tmz1t_2146		addiction module toxin, RelE/StbE	IncP-1beta multiresistance plasmid pB8
Flagella synthesis			
Tmz1t_2887		flagellin domain protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2888		flagellar protein FlaG protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2889		flagellar hook-associated 2 domain protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2890	*fliS*	flagellar protein FliS	*Azoarcus* sp. (strain BH72).
Tmz1t_2891	*fliT*	flagellar-like protein FliT	*Azoarcus* sp. (strain BH72).
Tmz1t_2892		conserved hypothetical protein	*Nitrosomonas europaea*
Tmz1t_2893		flagellar related protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2894		type IV pilus assembly PilZ	*Azoarcus* sp. (strain BH72).
Tmz1t_2895		type IV pilus assembly PilZ	*Azoarcus* sp. (strain BH72).
Tmz1t_2896	*fliE*	flagellar hook-basal body complex subunit FliE	*Azoarcus* sp. (strain BH72).
Tmz1t_2897		sigma-54 factor interaction domain-containing protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2898		ATP-binding region ATPase domain protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2899	*fliF*	flagellar M-ring protein FliF	*Azoarcus* sp. (strain BH72).
Tmz1t_2900	*fliG*	flagellar motor switch protein FliG	*Azoarcus* sp. (strain BH72).
Tmz1t_2901		transposase IS4 family protein	*Azoarcus* sp. (strain BH72).

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_2902	*fliH*	flagellar assembly protein FliH	*Azoarcus* sp. (strain BH72).
Tmz1t_2903	*fliI*	flagellar protein export ATPase FliI	*Azoarcus* sp. (strain BH72).
Tmz1t_2904		flagellar export protein FliJ	*Azoarcus* sp. (strain BH72).
Tmz1t_2905		flagellar hook-length control protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2906	*fliL*	flagellar basal body-associated protein FliL	*Azoarcus* sp. (strain BH72).
Tmz1t_2907	*fliM*	flagellar motor switch protein FliM	*Azoarcus* sp. (strain BH72).
Tmz1t_2908	*fliN*	flagellar motor switch protein FliN	*Azoarcus* sp. (strain BH72).
Tmz1t_2909	*fliO*	flagellar biosynthesis protein FliO	*Azoarcus* sp. (strain BH72).
Tmz1t_2910	*fliP*	flagellar biosynthetic protein FliP	*Azoarcus* sp. (strain BH72).
Tmz1t_2911	*fliQ*	flagellar biosynthetic protein FliQ	*Azoarcus* sp. (strain BH72).
Tmz1t_2912	*fliR*	flagellar biosynthetic protein FliR	*Azoarcus* sp. (strain BH72).
Tmz1t_2913		conserved hypothetical secreted protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2914	*flgL*	flagellar hook-associated protein 3	*Azoarcus* sp. (strain BH72).
Tmz1t_2915	*flgk*	flagellar hook-associated protein FlgK	*Azoarcus* sp. (strain BH72).
Tmz1t_2916	*flgJ*	flagellar rod assembly protein/muramidase FlgJ	*Azoarcus* sp. (strain BH72).
Tmz1t_2917	*flgI*	flagellar P-ring protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2918		flagellar L-ring protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2919	*flgG*	flagellar basal-body rod protein FlgG	*Azoarcus* sp. (strain BH72).
Tmz1t_2920	*flgF*	flagellar basal-body rod protein FlgF	*Azoarcus* sp. (strain BH72).
Tmz1t_2921		flagellar basal body FlaE domain protein	*Azoarcus* sp. (strain BH72).
Tmz1t_2922		flagellar hook capping protein	*Azoarcus* sp. (strain BH72).

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_2923	*flgC*	flagellar basal-body rod protein FlgC	*Azoarcus* sp. (strain BH72).
Tmz1t_2924	*flgB*	flagellar basal-body rod protein FlgB	*Azoarcus* sp. (strain BH72).
Tmz1t_2925	*flgA*	flagella basal body P-ring formation protein FlgA	*Azoarcus* sp. (strain BH72).
Pilus synthesis			
Tmz1t_1175	*pilB*	type IV-A pilus assembly ATPase PilB	*Azoarcus sp*. (strain BH72)
Tmz1t_2692		Flp pilus assembly protein CpaB	*Azoarcus sp.* (strain EbN1)
Tmz1t_0834	*pilM*	type IV pilus assembly protein PilM	*Azoarcus sp*. (strain BH72)
Tmz1t_0835		Fimbrial assembly family protein	*Azoarcus sp*. (strain BH72)
Tmz1t_0836	*pilO*	Pilus assembly protein PilO	*Azoarcus sp*. (strain BH72)
Tmz1t_0837	*pilP*	Pilus assembly protein PilP	*Azoarcus sp*. (strain BH72)
Tmz1t_0838	*pilQ*	type IV pilus secretin PilQ	*Azoarcus sp*. (strain BH72)
Tmz1t_2501	*pilZ*	type IV pilus assembly PilZ	*Azoarcus sp*. (strain BH72)
Tmz1t_2663	*pilE*	type IV pilus biogenesis protein PilE	*Janthinobacterium sp.* (strain Marseille)
Tmz1t_2664		fimbrial biogenesis protein	*Methylibium petroleiphilum* (strain PM1)
Tmz1t_2665	*pilV*	type IV pilus modification protein PilV	*Methylobacillus flagellatus* ATCC 51484
Tmz1t_2667	*pilX*	type IV pilus assembly protein PilX	*Janthinobacterium sp.* (strain Marseille)
Tmz1t_0639	*pilZ*	type IV pilus assembly PilZ	*Azoarcus sp.* (strain BH72)
Tmz1t_2684	*pilT*	PilT protein domain protein	*Parvibaculum lavamentivorans* (strain DS-1)
Secretion systems			
Tmz1t_3457		type I secretion outer membrane protein, TolC family	*Azoarcus* sp. (strain BH72)

Gene no.	Gene name	Comments/function of gene product	Best hits organism
Tmz1t_2693		type II and III secretion system protein	*Verminephrobacter eiseniae* (strain EF01-2)
Tmz1t_2694		type II secretion system protein E	*Verminephrobacter eiseniae* (strain EF01
Tmz1t_2695		type II secretion system protein	*Verminephrobacter eiseniae* (strain EF01
Tmz1t_2696		type II secretion system protein	*Verminephrobacter eiseniae* (strain EF01
Tmz1t_1174		type II secretion system protein	*Azoarcus* sp. (strain BH72)
Tmz1t_2529		general secretion pathway protein D	*Azoarcus* sp. (strain EbN1)
Tmz1t_2530		general secretory pathway protein E	*Azoarcus* sp. (strain EbN1)
Tmz1t_2531		general secretion pathway protein F	*Azoarcus* sp. (strain EbN1)
Tmz1t_2532		general secretion pathway protein G	*Azoarcus* sp. (strain EbN1)
Tmz1t_2534		general secretion pathway protein I	*Azoarcus* sp. (strain EbN1)
Tmz1t_2535		general secretory pathway protein J	*Azoarcus* sp. (strain EbN1)
Tmz1t_2536		general secretory pathway protein K	*Azoarcus* sp. (strain EbN1)
Tmz1t_2537		general secretion pathway protein L	*Azoarcus* sp. (strain EbN1)

Chapter 5

Molecular Analysis of the Exopolysaccharide Production in *Thauera aminoaromatica* MZ1T

Abstract

To identify and verify the genes responsible for thaueran formation and export, transposon mutagenesis, gene cloning and expression as well as reverse transcriptase quantitative real time PCR were performed in this study. Through transposon mutagenesis, 5 floc-deficient mutants were generated and their interrupted genes were identified by arbitrary PCR.

Putative thaueran genes were also physically localized in fosmid library clones and induced for expression in non-EPS producing host cells. Immunoblotting tests showed a possible but not significant expression of thaueran in its original host cells.

The mutants were further used in reverse transcriptase quantitative real time PCR to investigate the relationship between the abundance of EPS or floc in the liquid culture and the expression level of putative thaueran genes identified from genome annotation. The data of the treatments under different conditions, such as different temperatures, different media and among the mutants, a nice correlation of the amount of floc in the liquid culture and the expression level of the tested genes was revealed.

Introduction

Since EPS production is unique for *T. aminoaromatic* and the composition of thaueran is unique among the bacterial EPS, particular interest was put on the identification of the genes responsible for thaueran formation and export. Through genome annotation, three putative thaueran gene clusters were identified. To verify the results from the genome annotation, molecular and biochemical experiments were commonly used.

One strategy is to investigate if MZ1T will lose its ability to produce polysaccharide and form floc in liquid culture when the existing putative thaueran genes are interrupted or disrupted through means of random mutagenesis. This idea was based on chemical mutagenesis on MZ1T and transposon mutagenesis on other EPS producing bacterium. For example, through Tn5 transposon mutagenesis, sheath-deficient mutants were successfully generated in *Sphaerotilus natans*. Five of the seven transposon interrupted genes were identified as *sthA*, a homolog to glycosyltransferase, which is involved in linking sugars to lipid carriers during bacterial exopolysaccharide biosynthesis(108). Similar experiments were performed in *Halomonas maura*(5). The mini-Tn5 transposon interrupted genes in EPS-deficient mutants revealed five ORFs (*epsABCDJ*). They form part of a gene cluster (eps) with the same structural organization as others involved in the biosynthesis of group 1 capsules and some EPSs(5). Similarly, EPS-deficient mutants in *Streptococcus thermophilus* Sfi6 were obtained through Tn*916*

mutagenesis and 15 genes involved in polysaccharide biosynthesis and export were identified(103).

In parallel, another approach is to test if non-EPS host bacterium will gain the phenotype of polysaccharide production and floc formation when the putative thaueran genes are introduced in the new system. However, since all three putative thaueran gene clusters are long length and GC-rich, it may create more difficulties for cloning and potentially amino acid preference problem for expression. EPS gene cloning and expression was applied for the expression of a region with a size of 14.52 kb from *Streptococcus thermophilus* Sfi6in in the non-EPS-producing heterologous host, *Lactococcus lactis* MG1363. Through ligation with vector pJIM2799 and transformation, the introduced EPS gene cluster with 13 genes *epsA* to *epsM* is capable of EPS synthesis and secretion in this host(103).

Finally, instead of focusing on the absolute appearance and disappearance of EPS product in the culture to determine if the putative thaueran genes are responsible for thaueran formation, an indirect investigation is to examine relatively if there is a close correlation between thaueran abundance and transcription level of the putative thaueran genes. This method was also used in *Sphaerotilus natans* to determine the relative transcription level of *sthA* in both single cells and filamentous cells(108).

Among all these experiments, the last one showed the most direct and convincing results in testing the hypothesis that the putative thaueran genes identified from genome annotation are responsible for the formation of thaueran in MZ1T.

Methods

A. Transposon mutagenesis

1. Generation of Floc-deficient and thaueran-deficient mutants

Since microscopic images showed MZ1T cells from liquid culture possessing much higher level of thaueran than cells grown on agar plate, wild type *Thauera aminoaromatica* MZ1T collected from Stoke's medium solid agar plates were washed with NaCl solution to reduce thaueran, and then the pellet was washed three times with 10% glycerol to make competent cells for electroporation with the transposon (EZ-TN<R6Kγori/Kan-2>Tnp). To do electroporatioin, 1 μL transposon was added into 50 μL MZ1T competent cells. The voltage used was 2.5KV, currency was 25 μF and resistance was 200Ω. Then cells were recovered and spread on 50 mg/ml Kan Stoke's agar plates for selection.

2. Screening for Floc-deficient and thaueran-deficient mutants

Kanamycin-resistant colonies were inoculated in liquid culture for observation of flocculation. Mutants lacking flocculation were further investigated through immunoblotting. Previously purified thaueran was used to immunize rabbits once every

three weeks for four times to enhance their immune response. The serum of the third time blood sample was collected and used as the primary antibody for this experiment. The mutants were reacted with Rabbit-Anti-thaueran antibody to determine whether they still produce the same thaueran. Briefly, the culture, supernant and the pellet of the samples were loaded onto PVDF membrane that was wetted with methanol and socked in PTB (protein transfer buffer) for 2 min. The blot was then blocked by 3% non-fat milk in PBS for 1.5 hrs at room temperature with gentle rotation. The membrane was then washed twice with TBS-T for 5 min each time. Then, it was incubated with primary antibody with a 1:200 dilution of PAC1672+1682 in TBS-T. After washing three times with TBS-T, the membrane was incubated with secondary antibody with a 1:1000 dilution of HRP-conjugated goat anti-rabbit antibody in TBS-T. After washing three times, the membrane was developed and visualized.

3. Verification of mutants by 16S rRNA analysis

Previous attempts to generate *Thauera aminoaromatica* MZ1T floc-deficient mutants through transposon mutagenesis found false floc-deficient mutants which were characterized as contaminating Gram negative bacteria. To verify that any mutants generated are *Thauera aminoaromatica* MZ1T, genomic DNA of the mutants was extracted and used as the template in PCR to amplify their 16S rRNA gene with both 16S rDNA universal primers, 18f and 1492r and MZ1T 16S rDNA specific primers, 5’ATCCTGGGAACTGCGTTTGTG3’ and 5’CATTGCTGCTCCGAACGACT3’. Then

the sequencing results of the PCR products were blasted against the NCBI GenBank Database.

4. Verification of transposon insertion through PCR with designed primers

A pair of primers, 5'-AAACACGGAAACCGAAGACC-3' and 5'-CATCATTGGCAACGCTACCT-3', designed according to the sequence of the transposon was used in PCR to amplify the possible existing inserted fragment in the genomic DNA of mutants. The expected PCR product is 563 bp, which can be checked through electrophoresis on an agarose gel.

5. Identification of transposon interrupted genes in Floc-deficient mutants

To identify the insertion site, direct sequencing of the flanking region with KAN-2 FP-1 Forward primer and R6KAN-2 RP-1 Reverse primer provided with the transposon kit was carried out but not successful due to poor amplification. Alternative methods of rescue cloning and arbitrary PCR were examined. To do rescue cloning, the mutant genomic DNA is purified using FASTDNA soil spin kit, and sheared with nebulizers or digested with restriction enzymes. Then self-ligation, transformation and selection on Kan plates were performed and selected rescue clones were sequenced with EZ-Tn5 <R6Kγori/Kan-2>Transposon-specific primers. Sequences of the flanking regions are blasted against MZ1T genome sequence and NCBI GenBank database.

For arbitrary PCR, a set of primers for each step was designed according to MZ1T genome sequence and in the guidance of previous studies in other bacteria(55). The sequences of the primers are: ARB6: 5'GGCCACGCGTCGACTAGTACNNNNNNNNNNACGCC3', ARB2: 5'GGCCACGCGTCGACTAGTAC, ARB1:5'GGCCACGCGTCGACTAGTACNNNNNNNNNNNNNNNNGATAT3', Rev EZTN5Out: 5'CCATGAGGGTTTAGTTCTGG3', ForEZTN5Out: 5'GTTGTAACACTGGCAGAGCA3'. Primers ForEZTN5Out, RevEZTN5Out and ARB6 were used for the first round PCR, and the product of the first round PCR was used as the template for the second round PCR with ForEZTN5Out, RevEZTN5Out and ARB2 as the primers. The first round PCR program was 95℃ for 5 min, then 94℃ for 30 seconds, 30℃ for 30 seconds and 72 for 1 min for 5 rounds and 30 rounds of 94℃ for 30 seconds, 55℃ for 30 seconds and 72℃ for 1 min, followed by 72℃ for 5 min. the second PCR program was 30 rounds of 94℃ for 30 seconds, 55℃ for 30 seconds, and 72℃ for 1 min, then followed by 72℃ for 5 min. Amplicons were sequenced and sequences of each mutant were blasted against MZ1T genome sequence and NCBI GenBank database.

B. Cloning and expression of thaueran gene clusters in non-EPS producing host organism

1. Amplification of target gene fragment

To express the putative eps gene clusters in non-EPS producing *Escherichia coli*, the putative thaueran gene cluster, EPS2 with a length of 22.2kb and GC content of 68.3%, was amplified using Herculase from *Stratagene*.

2. Retrieving target sequences from fosmid library

A backup approach to bypass the big challenge of amplifying the 22 kb fragment, fosmid library for Sanger sequencing was used to retrieve the putative thaueran gene clusters of interest. Since the targeted fragment is 22 kb, 40kb fosmid library was chosen for this task. The vector is pCC1Fos and the host cell is *Escherichia coli* DH10B(27).

One advantage of using the gene fragments from clone library is its higher fidelity than PCR products. Since the clones are stored in more than eighty 384-well plates, physical localization of the clones that include the putative thaueran gene clusters is required. To determine the clones containing the complete cluster of the genes of interest, a thorough search of the flanking sequences on both sides of the fragment of interest was performed on NCBI specialized blast website for trace archives. (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&BLAST_SPEC=TraceArchive&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch). The flanking sequences used for this search are about 25 Kb before or after the thaueran gene cluster based on the genome sequence.

The search results provided the accession number of each clone, the insert size, the plate ID and well ID of the clones that meet our interests. Then, the clones that have the complete putative thaueran gene cluster were physically localized according to the plate ID and well ID. The sealer of each targeted well was pierced with 27G needle with 1mL syringe after ethanol wiping. Then, 10 μL of the frozen culture was transferred to 2-YT culture with chloraphenicol. Activated culture was grown in 2-YT culture with chloraphenicol and 0.02% L-arabinose was used to increase the copy number.

3. Verification of the completion of the target sequence

Before cutting out the putative thaueran gene cluster from the clones, the existence and completion of the target sequence were verified through PCR with designed specific primers. A pair of existing primers, EPSFor1 5'-GGAATCGTCCGACCCGCATACCT-3' and EPSRev4 5'-GACAGGGGCAGATGGAGGAGCAC-3', priming at each end of the cluster for amplification experiment were used for part of the primers. Another two primers, EPSR2 5' TCGCCATGTCCGTCCTCACTCTG-3' and EPS3F 5'-GGCGGAATAGTTGTTGCGGGTGA-3', that match primer EPSFor1 and EPSRev4 respectively and could generate a less than 1kb PCR product were designed for this purpose.

4. Inducing the expression of putative thaueran gene cluster in its fosmid

The fosmid vector pCC1Fos has two promotors, T7 and Lac, at each side of the insertion site(Figure 5.1) Therefore, inducing the expression of putative thaueran gene clusters in its fosmid clones, especially those have short distance from the promotors, were attempted.

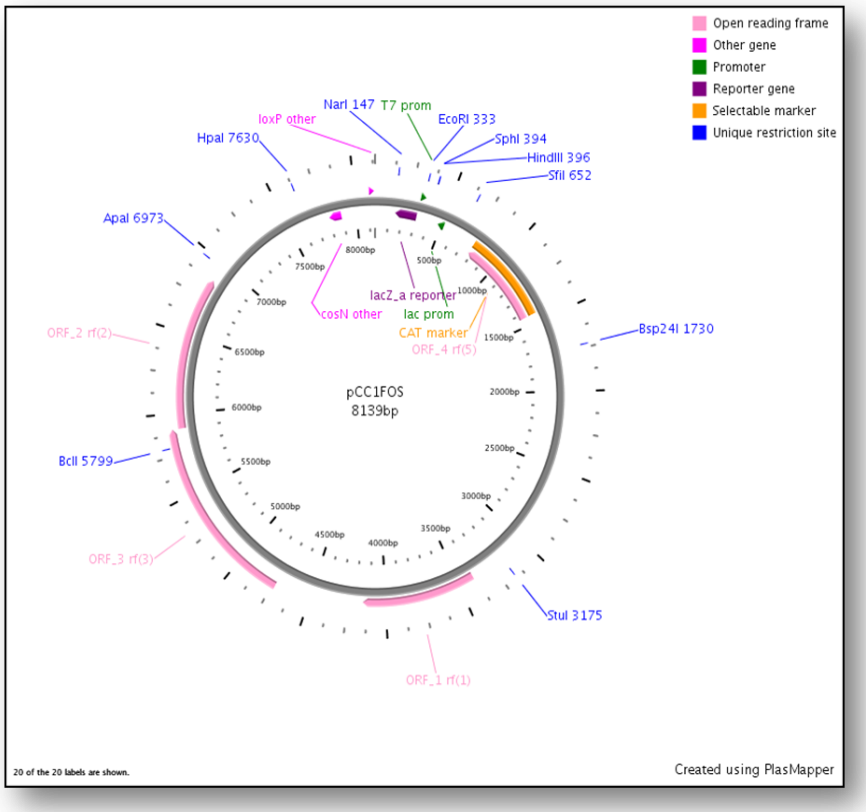


Figure 5.1 The map of fosmid vector pCC1Fos from Epicentre.

T7 promoter and Lac promoter are on each side of the cloning site toin opposite direction

The fosmid clones were grown and stored in 2-YT media. After activation, 10% L-arabinose were used to induce fosmid copy number from one to 100. IPTG was added at OD 0.6 to induce the expression.

5. Immunoblotting

Cultures were then harvested and immune blotting against Rabbit-Anti-thaueran antibody was performed. The positive controls are WT MZ1T and purified thaueran; negative controls are *Escherichia coli* TOPO 10, *Escherichia coli* BL21 and *Escherichia coli* DH10B with fosmid not harboring EPS genes.

The induced culture samples were loaded on the PVDF membrane that was wetted with methanol and socked in PTB (protein transfer buffer) for 2 min. The blot was then blocked by 3% non-fat milk in PBS for 1.5 hrs at room temperature with gentle rotation. The membrane was then washed twice with TBS-T for 5 min each time. Then, it was incubated with primary antibody with a 1:200 dilution of PAC1672+1682 in TBS-T. After washing three times with TBS-T, the membrane was incubated with secondary antibody with a 1:1000 dilution of HRP-conjugated goat anti-rabbit antibody in TBS-T. After washing three times, the membrane was developed and visualized.

C. Reverse transcriptase quantitative real time PCR

1. RNA/DNA extraction

Overnight cultures were harvested. Half of the culture was used for DNA extraction to normalize the biomass of each culture, and half was used for RNA extraction. DNA/RNA Fast spin soil kit was used for purification. After purification, the concentration and purity of each sample was measured using Nanodrop. Briefly, 2 µL of DNase free water was used for blanking and then 2 µL of sample was loaded for reading. Afterwards, a concentration of 10ng/µL was used as the working stock and stored at -70 ℃.

2. Reference genes

Three reference genes from MZ1T genome were used to normalize the result of RT-QPCR. They are genes encoding 16S rRNA (Tmz1t_R0059), DNA-directed RNA polymerase subunit beta (*rpoB*, Tmz1t_3345) and alanyl-tRNA synthetase (*alaS*, Tmz1t_1571).

Initially, 16S rRNA gene was used alone as the reference gene. However, since the concern of its expression under different conditions was raised, other reference genes were chosen to normalize or validate the results. In other organisms(79), genes *map*, *rpoC*, *alaS*, *era*, *gyrA* and *nth* showed less than 1.5 fold variation under all the experimental conditions. In MZ1T, gene Tmz1t_3345 (*rpoB*), encoding DNA-directed

RNA polymerase subunit beta, was used as another reference gene. The primers designed based on the sequence of this gene are 5'-ACCCGGTTCGCATCCACGTA-3' and 5'-GGCTTCGCTGATTCCCTTCC-3', and the expected size of the PCR product is 204bp. Another reference gene is Tmz1t_1571(*alaS*), encoding alanyl-tRNA synthetase. The primers designed for this gene are 5' CCCCGTTCTATGCCGAGTC 3' and 5' TCAGGTGGGTGACCGAGTG 3', and the expected length of the product is 225bp.

3. Target genes

The first two putative thaueran gene clusters are tightly organized on the positive and negative strand respectively. The third cluster is loosed arranged on both strands. Based on the intergenic distances between involved genes, the first two clusters, EPS1 and EPS2, may function as single operons especially for the latter. Therefore, investigation of the transcription behavior of individual genes from them may represent the expression level of the whole cluster. For this reason, only genes in these two clusters were used for this study. EPS1 is more independent from its neighborhood; EPS2 has some satellite *eps* genes around but organized tighter than EPS1 according to the intergenic distance.

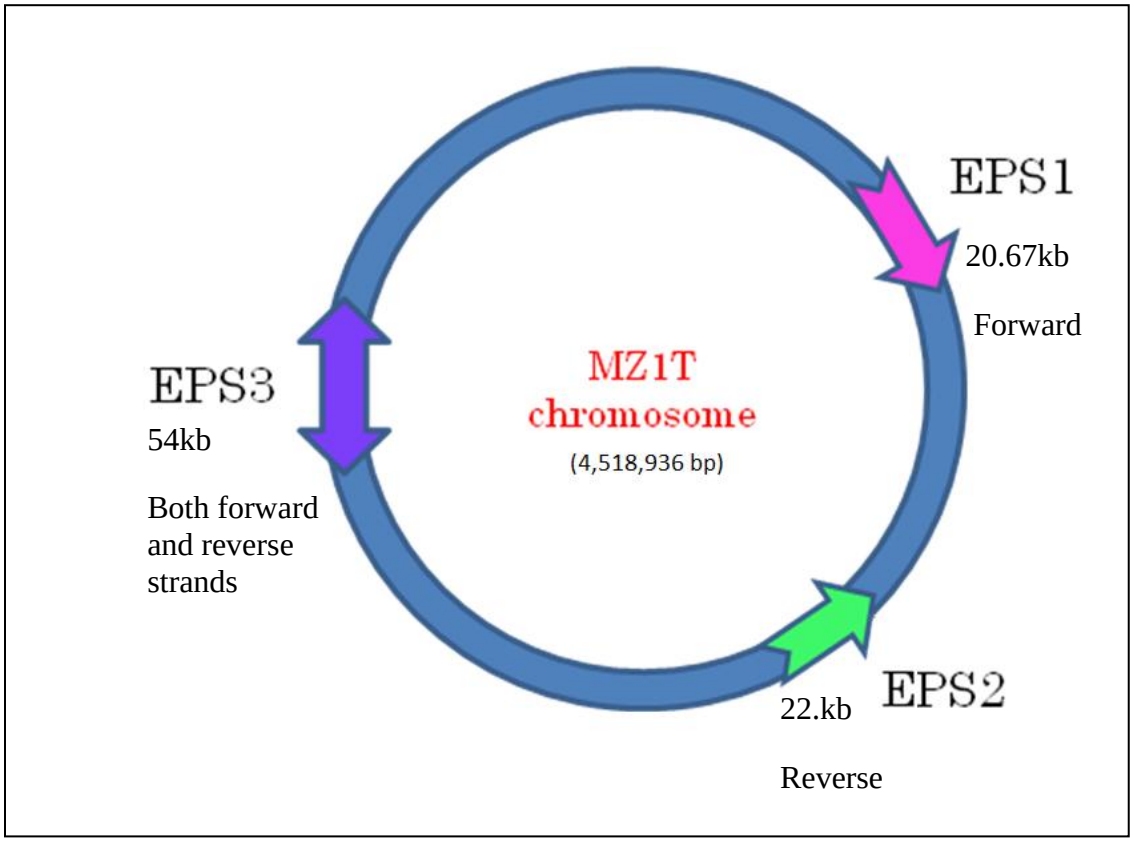


Figure 5.2 The relative physical location of the three putative thaueran gene clusters (EPS1, EPS2 and EPS3) in MZ1T genome.

EPS1, EPS2 and EPS3 represent the location. The direction of the arrow indicates the direction of the involved genes. EPS1 is on the forward strand, EPS2 is on the reverse strand and EPS3 is on both strands.

Five putative *thaueran* genes were chosen as the target genes. In EPS1, two genes Tmz1t_1116 (*wbpO*) and Tmz1t_1118 responsible for encoding polysaccharide biosynthesis proteins were used. In EPS 2, three genes at the beginning (Tmz1t_3268), in the middle (Tmz1t_3237), and at the end (Tmz1t_3283) of this cluster were chosen to test

the hypothesized correlation between putative *thaueran* gene expression level and floc abundance.

4. Treatments and Control of RT-QPCR

To study the transcription of the putative eps gene and the differential expression of them at different conditions or among mutants and wild type MZ1T, relative reverse-transcriptase quantitative real time PCR was carried out. Two series of experiments were designed to test the two hypothses: 1)The putative EPS genes in the floc-deficient mutants and wild type MZ1T have different expression levels; 2) The putative thaueran genes will show differential expression when wild type MZ1T grows in different conditions that result in different thaurean or floc formation levels. For different media, wild type MZ1T was grown in 2YT and Stoke's media at 30 ℃; for different temperature, wild type MZ1T were grown in Stoke's media at 30 and 37 ℃, and for wild type and floc-deficient mutants, wild type MZ1T and transposon mutants as well as chemical mutants were grown in Stoke's media at 30 ℃.

Results and Discussion

Transposon mutagenesis

1. Generation of mutants

Sixteen colonies were formed on Kan-select plates and inoculated in Kan Stoke's liquid media. All mutants grew as normal as the wild type strain and 5 cultures did not

form floc whereas the other 11 flocculated as the wild type *Thauera aminoaromatica* MZ1T after 48 hrs. The five floc-deficient mutants numbered as 2, 4, 6, 11 and 16 were subjected to immune detection.

Duplicate samples of the culture, supernatant and cells of each mutant were tested with Rabbit-Anti-thaueran antibody, the Goat-Anti-Rabbit secondary antibody and dye reagents were used for development. More details were described in previous section. The results are shown in Figure 5.3. The wild type of MZ1T and *Escherichia coli* were the positive and negative controls respectively. The samples of supernatant and the culture showed similar results (data not shown). Mutant 2, 4 and 11 showed obviously less thaueran than the wild type, while mutant 6 and 16 showed almost equivalent amount of thaueran as the wild type. However, the biomass was not quantitatively determined before loading samples. Consequently, these results are qualitation and semiquantitation at best. However, they are a strong indication that some floc-deficient mutants correlate with less abundance of thaueran.

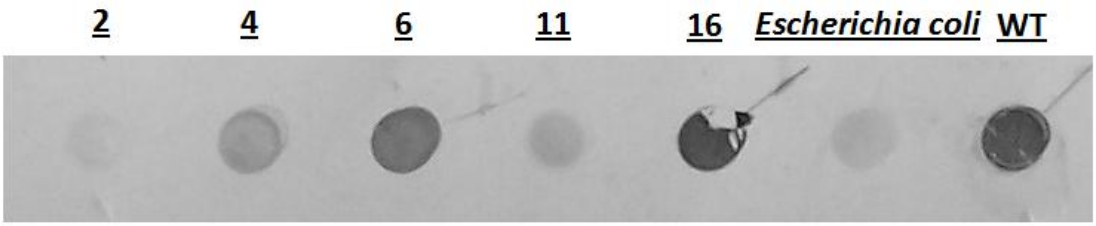


Figure 5.3 Immune detection results of the mutants and wild type *Thauera aminoaromatica* MZ1T.

2. Verification of mutants as *Thauera aminoaromatica* MZ1T through 16S rRNA analysis

To exclude the possibility of false positive mutants, the phylogenetic identities of mutants 11 and 16 were verified by direct sequencing of the 16S rRNA DNA amplified by PCR. Mutants 2, 4, and 6 were verified by sequencing the 16s rRNA DNA cloned in TA clone PCR4.0 vector. Comparison of the sequences with the original MZ1T 16S rRNA DNA sequence showed all the mutants are *Thauera aminoaromatica* MZ1T.

Transmission electron microscope (TEM) and scanning electron microscope (SEM) images of the wild type *Thauera aminoaromatica* MZ1T and mutant 11 were taken and shown in Figure 5.4 and 5.5(courtesy to Dr. John Dunlap). The samples were negatively stained with phosphotungstic acid. After one minute excess stain was removed. The grid was allowed to dry then examined in an Hitachi H800. These micrographs showed that the wild type MZ1T was trapped in the thaueran matrix while the mutant has a very clear background and no obvious thaueran was observed.

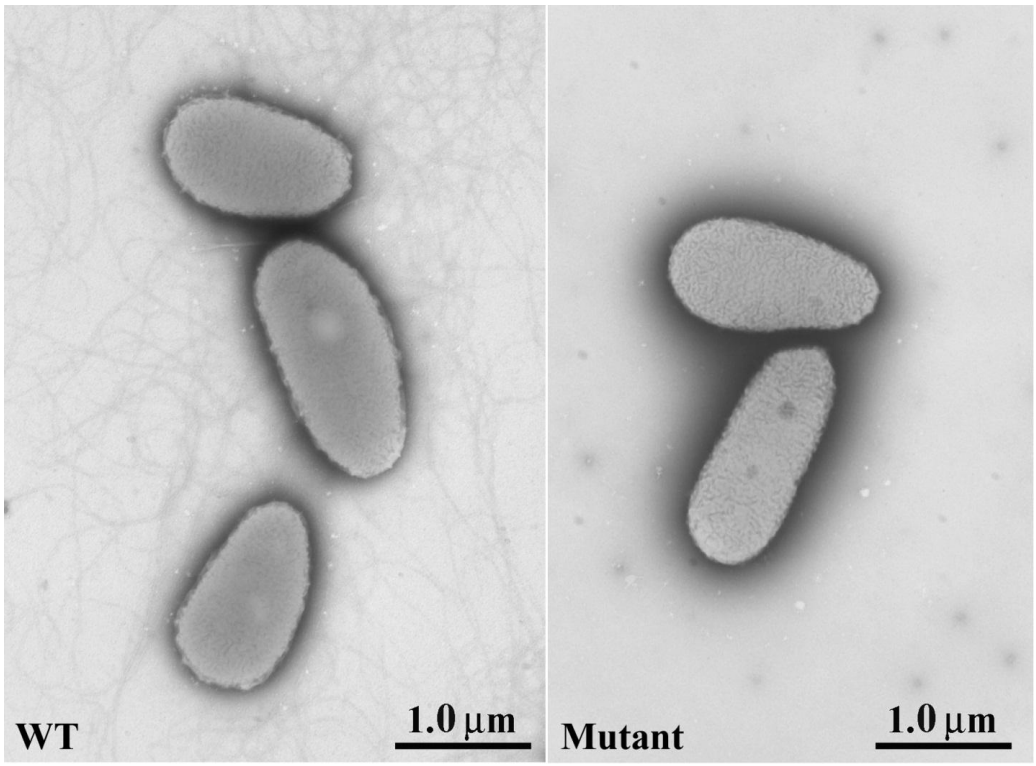


Figure 5.4 TEM images of wild type *Thauera aminoaromatica* MZ1T and mutant 11.

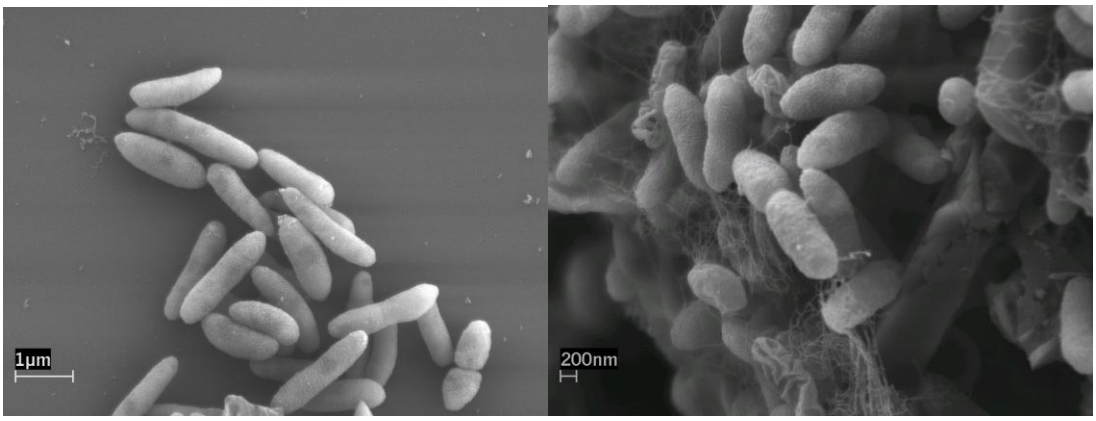


Figure 5.5 Scanning electron micrographs of Floc-deficient mutant 11 and wild type MZ1T

3. Verification of transposon insertion through PCR with designed primers

To verify if the transposon is integrated into the genome of the floc-deficient mutants, a pair of transposon (EZ-TN<R6Kγori/Kan-2>Tnp) specific primers were designed to amplify part of the insertion using the mutant genomic DNA as template.

The primers used are as follows: 5'-AAACACGGAAACCGAAGACC-3' and 5'–AGGTAGCGTTGCCAATGATG -3'. They showed no hits when blasted against *Thauera aminoaromatica* MZ1T genome sequence database. All mutants showed a single band with the correct size while the wild type strain did not generate it (Figure 5.6).

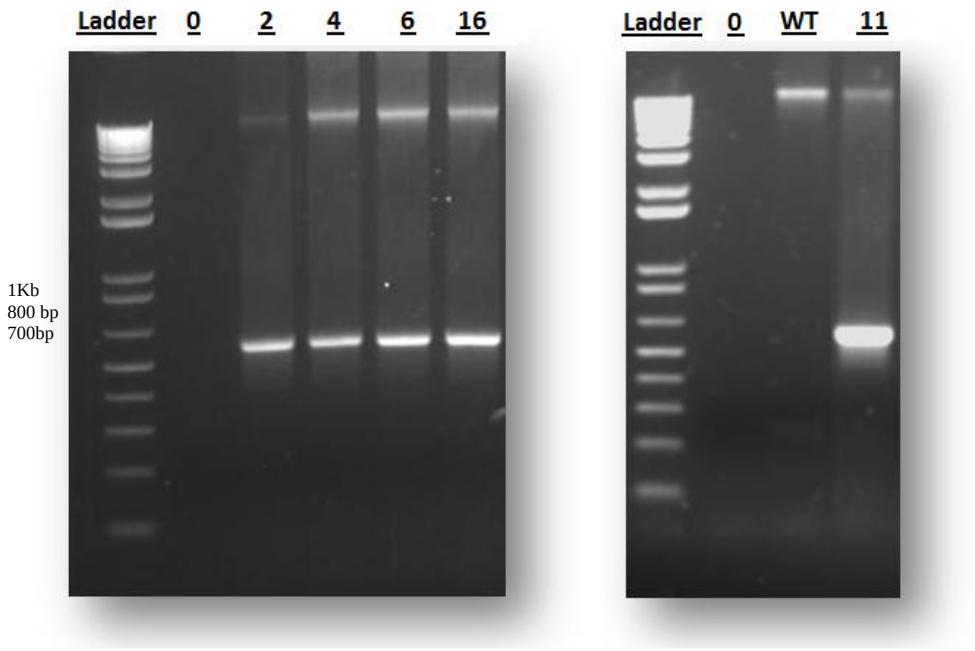


Figure 5.6 PCR products of the mutants and wild type (WT) *Thauera aminoaromatica* MZ1T amplified with the insertion primers on agarose gels.

Mutants number 2, 4, 6, 11 and 16 all showed a single band the amplicon amplified with the transposon specific primers with the expected size of, while the wild type does not have amplification when these primers were used.

Identification of the interrupted gene by arbitrary PCR

Identification of the interrupted gene was attempted by direct sequencing and then rescue cloning, but neither procedure was successful for any of the mutants. Arbitrary PCR was then used and adequate length of flanking sequences was obtained to locate the insertion site of the transposon. For mutant 4, 6 and 16 the insertion is in the intergenic region after the same gene (Tmz1t_0003), encoding DNA polymerase III beta chain (DNA gyraseB) and not the exact same location. DNA gyrase B is a protein that control the DNA supercoiling(35) and is inhibited with novobiocin and norfloxacin(84).The interrupted gene (Tmz1t_1114) of mutant 11 is 552 bp in size and encodes Protoporphyrinogen oxidase (PPOX), which is a membrane glycoprotein responsible for the seventh step in heme production(71). The interrupted gene in mutant 2 is a gene encoding response regulator receiver protein. Therefore, mutants 4, 6 and 16 are global regulatory genes not specific for exopolysaccharide formation. To mimic the impact of the interruption of DNA gyrase B, novobiocin with various concentrations was added in the culture of the wild type MZ1T. However, instead of forming less floc, more floc was observed in the culture. Therefore, the interruption of DNA gyrseB gene may not have relaxed the DNA structure but make it more restrained. However, since mutants were

grown as normally as the wild type strain, the expression of most housekeeping genes was not influenced by the interruption. However, the putative thaueran genes may be more sensitive to the supercoiling structure of DNA than other genes.

Gene cloning and expression

Due to the extremely long length and high GC content of the fragment, the amplification was not successful. After various modifications and optimization of annealing temperature on the basis of the recommended protocol for programming PCR cycles, no amplification was observed either. Other choices of PCR polymerase were considered to amplify the putative EPS gene. Extra-Long PCR Kit from Epicentre, has been successfully used in previous study by other researchers(10, 45). That is amplification up to 40kb from Lambda DNA, 28 kb from human DNA and 30 kb from *Escherichia coli* DNA.

Eleven fosmid *Escherichia coli* clones that contain the complete sequence of the putative thaueran gene cluster_1(EPS1) and seven clones that harbor the entire sequence of putative thaueran gene cluster_2(EPS2) were identified in the library in BLAST trace archive. Two tables that summarize the accession number of the clones, the direction of the insertion in the vector, the size of the insert, the adjacent promoter, the distance from the start codon of the cluster to the promoter, the start and end point of the insert in the genome, and most importantly the plate ID and well ID of the clones were generated (Table 5.1 and Table 5.2).

Table 5.1 Fosmids harboring MZ1T thaueran gene cluster_1 (EPS1 from 1216364 to 1244298 on the forward strand)

Name	NCBI Accession Number	Forward	Start	end	Reverse	end	Start	Insert Length	Adjacent promoter	x bp away	Plate ID	Well ID
EPS1_1	2013537855/6	6033.b1	1202400	1202997	6033.g1	1245531	1246343	43943	lac	13964	61	A22
EPS1_2	*2013539175/6*	7583.b10	1208637	1209073	7583.g10	1244570	1245058	36421	Lac	7727	77	M24
EPS1_3	2013534264/5	2941.b1	1241504	1240790	2941.g1	1204379	1203705	37799	T7	12659	29	I16
EPS1_4	*2013533347/6*	2184.b1	1208433	1208896	2184.g1	1243578	1244368	35935	Lac	7931	21	O18
EPS1_5	201359898/7	864.b1	1212270	1212866	864.g1	1246278	1246744	34474	Lac	4094	9	O23
EPS1_6	2013538542/1	7181.b1	1212987	1213543	7181.g1	1248735	1248881	35894	Lac	3377	73	I20
EPS1_7	2013535512/3	4076.b1	1252215	1251445	4076.g1	1212980	1212194	40021	T7	4170	41	G12
EPS1_8	2013532497/6	1034.b1	1250322	1250968	1034.g1	1214009	1214533	35789	T7	1831	9	C20
EPS1_9	2013538538/7	7171.b1	1209594	1210239	7171.g1	1250055	1250578	40984	Lac	6770	73	E18
EPS1_10	2013536438/9	4731.b1	1247026	1246193	4731.g1	1207620	1207920	39106	T7	8444	49	F7
EPS1_11	2013538334/3	6744.b1	1214470	1215258	6744.g1	1248770	1249378	34908	Lac	1894	69	O6

Table 5.2 Fosmids including Mz1t Exopolysaccharide cluster_2 (EPS2 from 3598316 to 356149 on the reverse strand)

Name	NCBI Accession Number	Forward	Start	end	Reverse	end	start	Insert Length	Adjacent promoter	x bp away	Plate ID	Well ID
EPS2_1	gnl|ti|2013533389/8	2221.b1	3557766	3558306	2221.g1	3598305	3598695	40929	T7	379	21	J4
EPS2_2	gnl|ti|2013534301/2	2985.b1	3604570	3603967	2985.g1	3565423	3564745	39825	lac	6254	29	b4
EPS2_3	gnl|ti|2013533938/7	2641.b1	3600807	3600080	2641.g1	3563726	3563032	37775	lac	2491	25	b14
EPS2_4	gnl|ti|2013539802/1	8009.b1	3571218	3572051	8009.g1	3606344	3607130	35912	T7	8814	81	B12
EPS2_5	gnl|ti|2013539809/10	8013.b1	3616713	3615890	8013.g1	3576991	3576193	40520	lac	18397	81	J12
EPS2_6	gnl|ti|2013535758/9	4263.b1	3566389	3567159	4263.g1	3597565	3598365	31976	T7	49	45	M9
EPS2_7	gnl|ti|2013534764/5	3463.b1	3598453	3597851	3463.g1	3552320	3551736	46717	lac	51	37	M1

Verification of the completion of the target sequence

After the individual clones localized in the 384-well plates were activated and cultured, the designed primers and the cells of each clone as the template were used for PCR. The PCR products were separated and visualized on an agarose gel. The amplicons with the correct size of clones harboring EPS2 were shown on the gel (Figure 5.7), which means that the information provided by searching in NCBI trace archive is accurate and the target sequences in the clones are available for further usage.

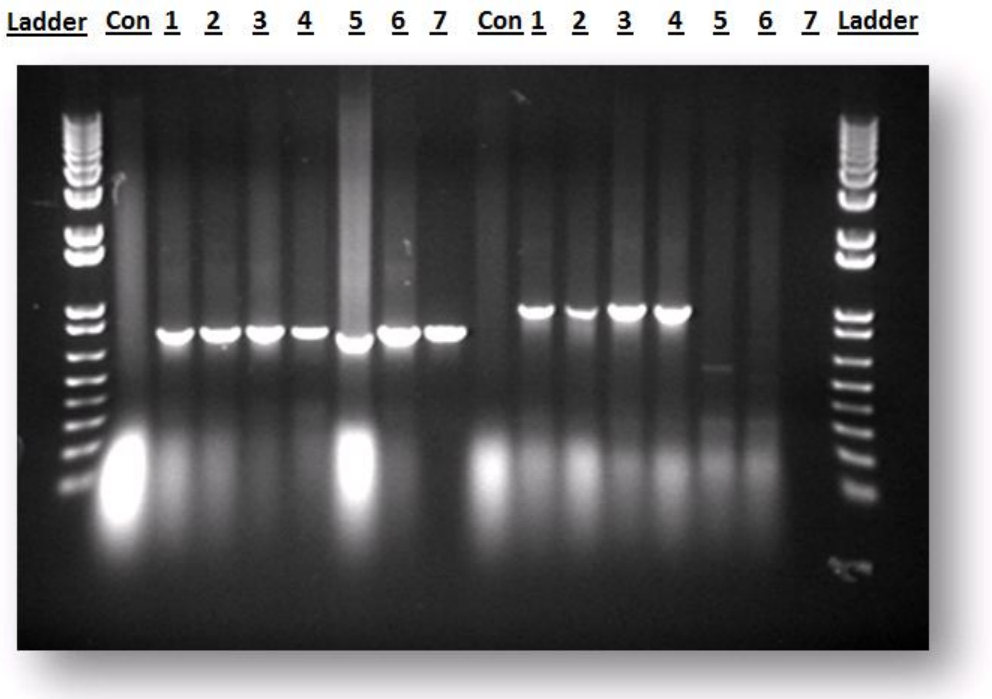


Figure 5.7 Agarose gel of the PCR products from 7 fosmid library clones that harbor putative thaueran gene cluster 2(EPS2) and specific primers.

The right bands are close to 1 kb ladder and the left show slightly less than 800 bp. Both are consistent with the predicted size of 774bp and 980 bp. Number 1 to 7 stands for the 7 fosmid clones that may include the complete EPS2 based on BLAST results in NCBI trace archive.

Inducing the expression of putative thaueran gene cluster and Immunoblotting

Cultures were then harvested and immune blotting with thaueran-antibody was performed. The positive controls are WT MZ1T and purified thaueran; negative controls are *Escherichia coli* TOPO 10, *Escherichia coli* BL21 and library host cells, *Escherichia coli* DH10B, with fosmid not harboring putative thaueran genes.

The results for clones harboring EPS1 and EPS2 in figure 5.8 showed a slightly stronger interaction of the tested clones than some negative control. However, it is much weaker than the positive controls and come negative controls have similar intensity to the tested clones. The results are ambiguous evidence to support the hypothesis that thaueran has been expressed and formed in the *Escherichia coli* host cells, but it can't be ruled out. A possible way to support the possibility of the expression and formation of thaueran in these cells is to quantify the biomass of the loading samples to ensure a fair comparison of the intensity of the interaction between the testing cells and the antibody.

The experiment of cloning and expression of putative thaueran genes did not create a significant positive proof to show that the introduction of putative thaueran gene

sequences endows the expected morphology of thaueran forming in the non-EPS producing host cell, but through the steps of this experiment, the existence of the DNA of the putative thaueran genes were verified. Therefore, the next step of the investigation is to demonstrate the transcription of the putative thaueran genes in the clones or the wild type MZ1T.

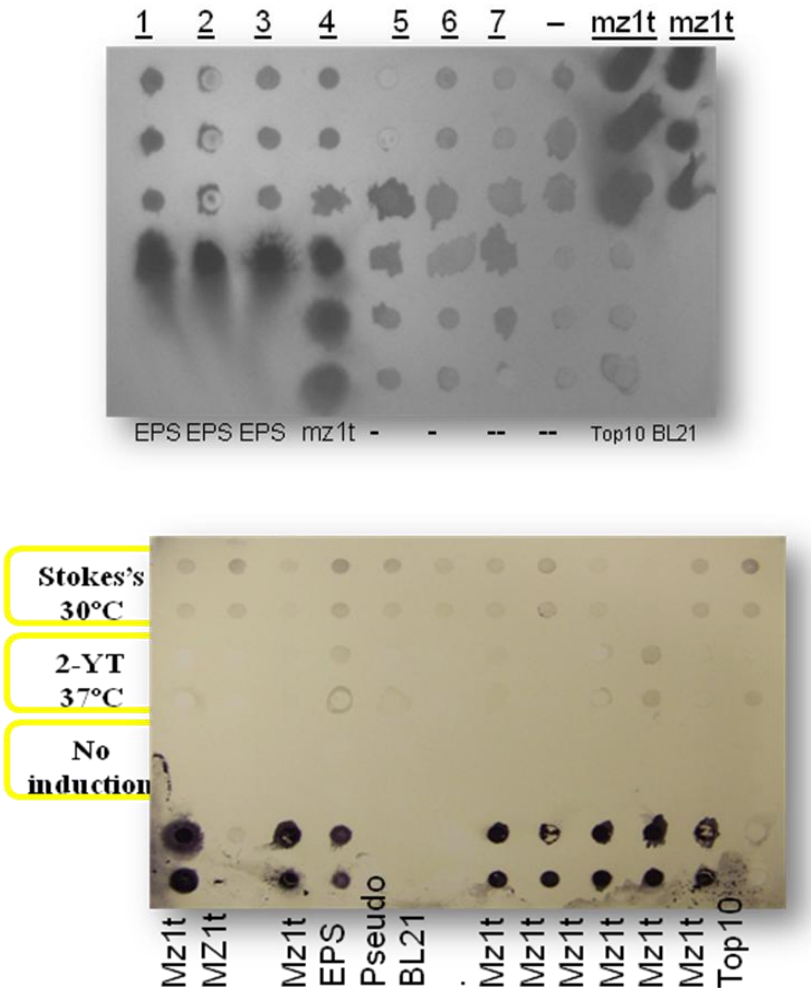


Figure 5.8 Immunoblotting results after inducing the expression of putative thaueran genes in non-EPS producing host cells *Escherichia coli* DH10B.

The upper panel is the results from the seven clones harboring EPS2. The first three rows are triplicates of the 7 clones and wild type MZ1T as the positive control as well as one clone harboring no putative thaueran genes as the negative control. The lower three rows are positive and negative controls. The lower panel is the results from the 11 clones harboring EPS2. The first two rows are clones that were grown in stoke's media at 30 ℃, the second two rows are those grown in 2YT media at 37 ℃, the third two rows are those without induction and grown in 2YT media at 37 ℃, and the last two rows are positive controls as wild type MZ1T and partially purified EPS and negative controls such as *Pseudomonas aeruginosa* and *Escherichia coli* BL21 and TOPO 10 cells.

Reverse-transcriptase Quantitative real time PCR

Primers designed for Real-time Quantitative RT PCR were tested through regular PCR. Single bands of amplicons and expected sizes were visualized by agarose gel electrophoresis when clones 2 and 7 and wild type MZ1T were tested(Figure 5.9). These results indicate that the designed primers are suitable for RT QPCR. Three genes, Tmz1t_3266 (acyltransferase 3), Tmz1t_3273(exosortase 1 system-associated amidotransferase1) and Tmz1t_3282 (polysaccharide export protein), in putative thaueran gene cluster_2(EPS2) were used for this RT QPCR because they are located at the beginning, middle and end of the cluster and encoding proteins that play important roles in EPS synthesis and export. Wild type MZ1T and floc-deficient mutant strain 2 and floc+ strain 7 were used as the template to amplify the expected fragment with the

primers EPS2StaF 5’-CGTGCTCGAATCCTTGCC-3’, EPS2StaR 5’-TCCTCGTTCTCGCTGTGG-3’, EPS2MidF 5’-GGTCTGCTCGCACCACTG-3’, EPS2MidR 5’-CGCCGCTCCTCTACCTCA-3’, EPS2EndF 5’ CGCCGAGGACTTCCATCA-3’ and EPS2EndR 5’-GGCACGGTCGCCACTTTT-3’.

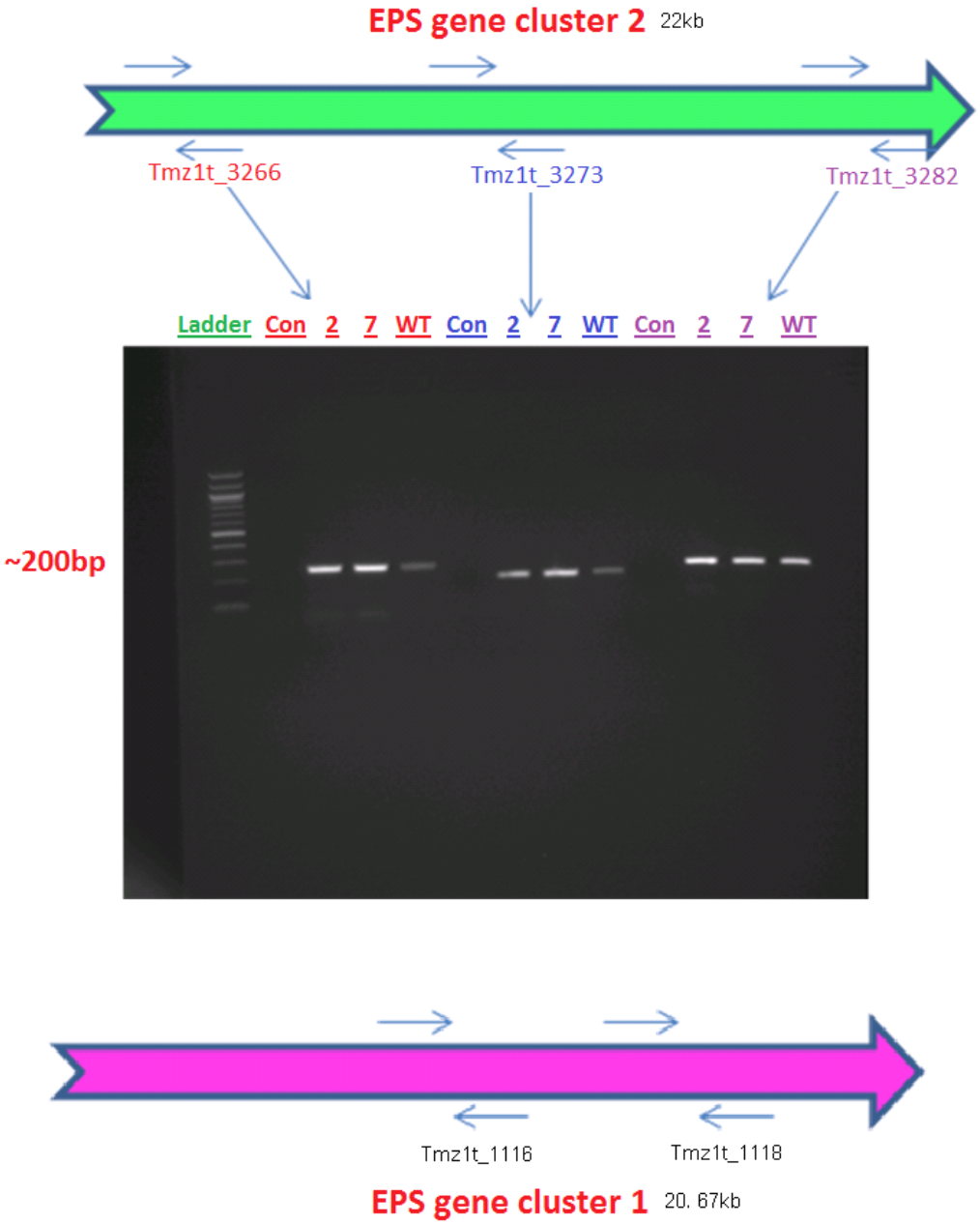


Figure 5.9 The relative location of the target genes tested in RT-QPCR in EPS1 and EPS2 and the size and specificity of the PCR products from EPS2

The size of the bands are around 200bp which is as expected.

Different conditions

For different media, wild type MZ1T grown in Stoke's medium was treated as control and grown in 2-YT as the treated sample. Delta Ct was 5.4, and fold change is 42.9, which shows that the expression level of the putative thaueran gene EPS2_end gene in more nutritious medium is 42.9 times greater than that in less nutritious medium, Stoke's medium. This result is consistent with the observed fact that the culture grown in 2-YT medium formed more flocs than in Stoke's medium. This correlation indicated that a higher expression of the putative thaueran gene is corresponding to a higher production of thaueran.

For different temperature, 30 ℃ was used as the control and 37 ℃ as the treated sample. The Delta Ct of the three putative thaueran genes from EPS2 were 0.392, 0.309 and 0.325, and the fold changes were 1.315, 1.239 and 1.253 (Table 5.3), which indicates the expression levels of putative thaueran genes at higher temperature are slightly higher but still similar to that of the lower temperature. These results also correlate to the observed fact that almost the same level of flocs was observed in the cultures grown at both temperatures.

Table 5.3 The fold change of the three target genes in cluster2

	EPS-end	**EPS-mid**	**EPS-sta**
Delta Ct	0.392	0.309	0.325
Fold Change	1.315	1.239	1.253

Different mutants

For different mutants, the wild type was used as the control and the five mutants generated from transposon mutagenesis the treated samples. Since the three putative thaueran genes from EPS2 behave similarly, gene named as EPS_2_end was first used for this experiment. When only 16S RNA gene was used as the reference gene, the fold changes were shown in Figure 5.10.

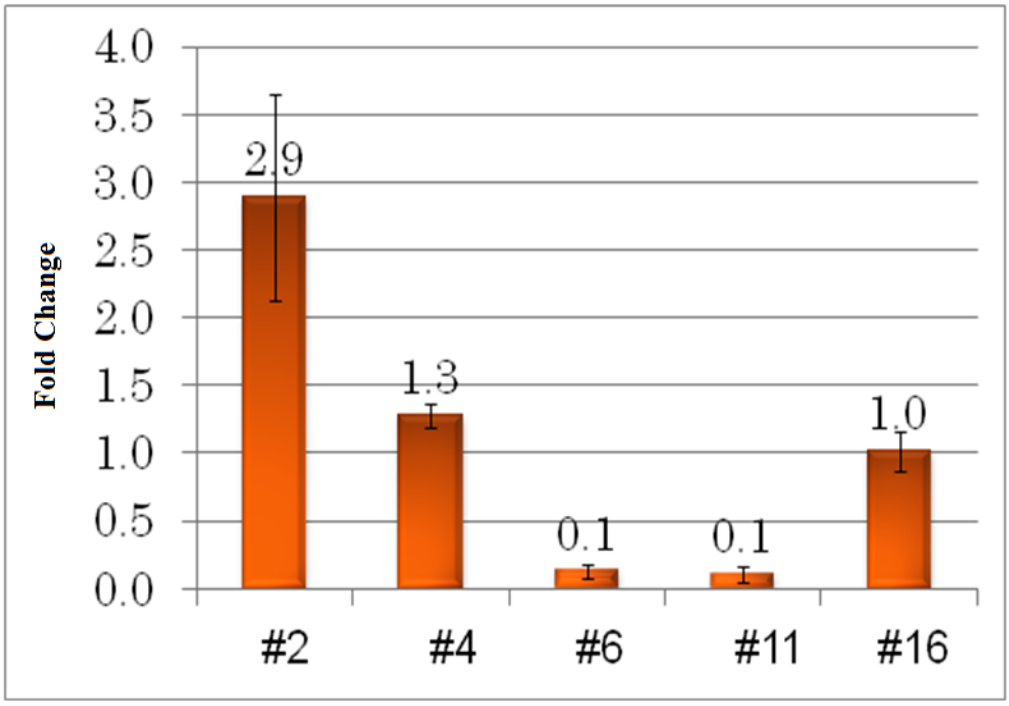


Figure 5.10 The fold change of gene EPS_2_end in transposon mutants v.s. the wild type MZ1T.

The expression level of this putative thaueran gene in MZ1T mutant 6 and 11 is 10 fold lower than that of the wild type MZ1T; the expression level of mutant 16 and 4 is similar to that of the wild type MZ1T, and the expression level of mutant 2 is 3 fold higher than the wild type MZ1T. These results also correlated to the observed floc abundance in the different cultures of the mutants.

When the two putative thaueran genes from EPS1, Tmz1t_1116 and Tmz1t_1118, were tested under the same condition as the previous experiment, EPS_1_mz1t_1116 and EPS_1_mz1t_1118 have similar expression level as gene EPS2-end. Mutant 6 behaves differently from gene EPS2-end. The overall transcription level of the three tested genes is consistent in these five mutants and correlate to the amount of floc observed in the culture; that is the abundance of floc in mutant 2, 4 and 16 is similar to the wild type and that of mutant 11 and 6 have a much lower amount of floc in the culture(Figure 5.11).

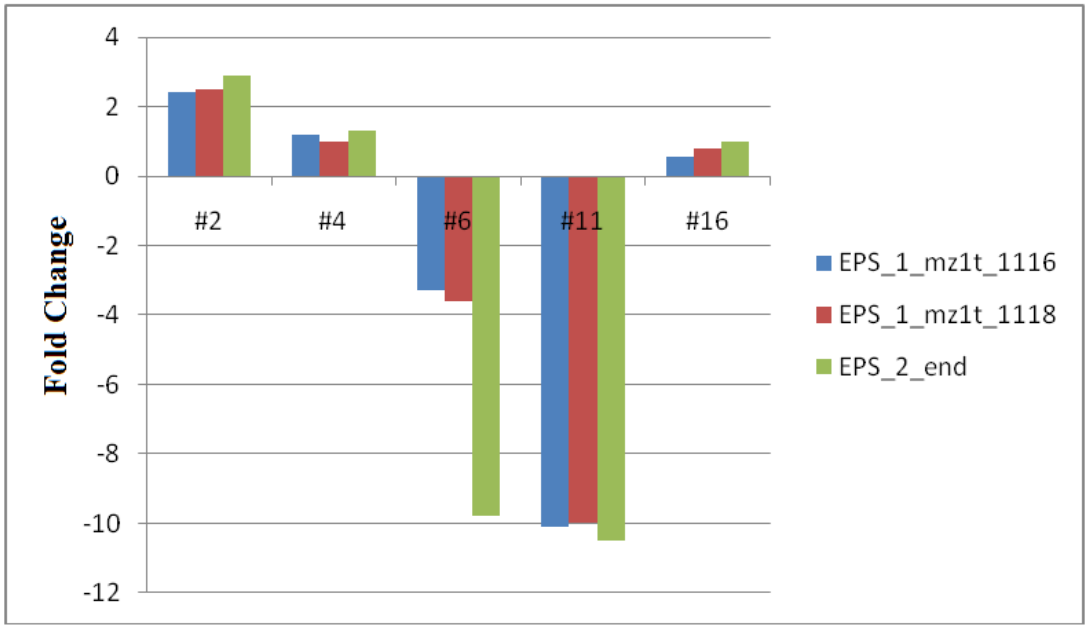


Figure 5.11 The fold changes of EPS_1_mz1t-1116, EPS_1_mz1t_1118 and EPS_2_end in transposon mutants v.s. the wild type MZ1T.

These results suggest that the tested genes are involved in exopolysaccharide formation by MZ1T because the abundance of floc is greatly influenced by the amount of exopolysaccharide. Although the conditions of the culture may play a role in this circumstance, however, since all the tested mutants and wild type strain are all grown under the same condition and the chemical compositions, temperature and pH of medium are all the same so the major parameter that could alter the amount of floc formation is the amount of exopolysacchairde that the organism can produce and export in the culture.

For the previously generated chemically generated mutants, the thaueran yield of each mutant was measured and when compared to the wild type, the amount of thauean produced by mutant 20 is 91% of that of the wild type MZ1T, mutant 26 is 54%, mutant

37 is 15% and mutant 39 is 97%(Table 5.4). Although the fold changes of the mutants do not exactly match the relatively different EPS yield of the chemical mutants, RT-QPCR results showed lower expression level in all the chemical mutants when compared to that of the wild type MZ1T(Figure 5.12).

Table 5.4 Fold change in thaueran production by chemical floc-deficent mutants

Locations	#20	#26	#37	#39	Gene Product
EPS1_1116	4.0	4.6	2.1	2.0	polysaccharide biosynthesis proteins
EPS2_start	2.5	3.9	1.9	1.5	polysaccharide export protein
EPS2_End	3.5	3.7	2.0	1.9	acyltransferase 3
EPS% WT	91	54	15	97	

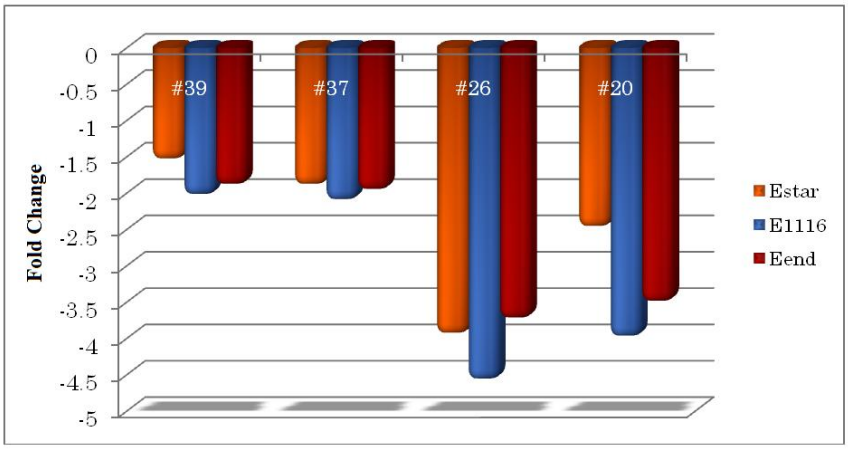


Figure 5.12 The fold change of the gene EPS2_star(Estar), EPS1_1116(E1116) and EPS2-end(Eend) of the chemical mutants and wild type MZ1T

Expression level of EPS2-end gene when normalized to multiple reference genes

Since only one reference gene (16S gene) was used at first, to test if this reference gene has introduced any bias for normalization in previous tests, another two genes (*rpoB*, *alaS*) commonly used for reference genes in QPCR were used to normalize the results from QPCR of EPS-end gene in the five mutants as treated and 2YT medium as treated and wild type as the control. The results based on the normalization of the two new reference genes showed consistence with the results normalized by 16S gene, excepting mutant 6 have a relatively larger differences than the other four mutants. (Figure 5.13).

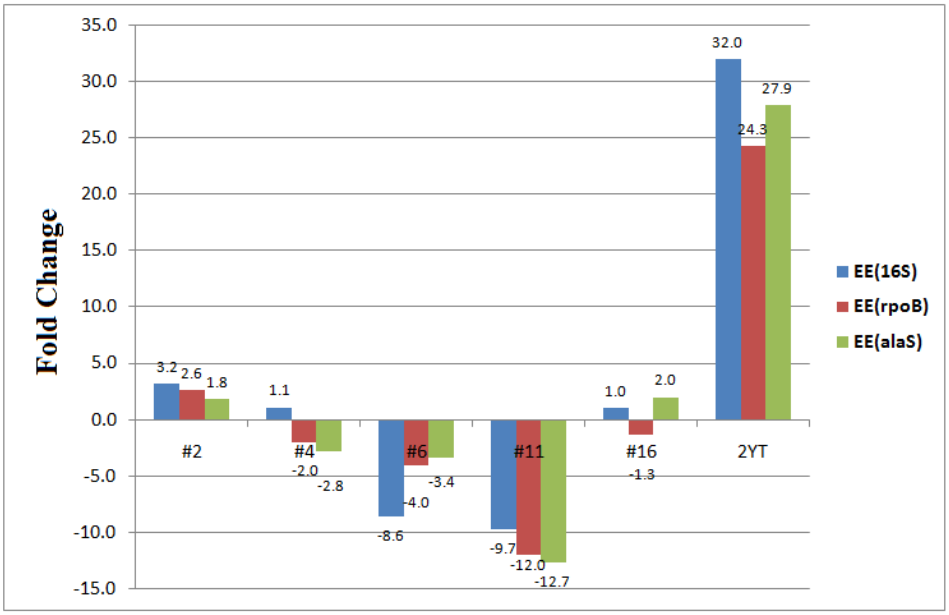


Figure 5.13The fold changes of gene EPS_2_end(EE) in transposon mutants v.s. the wild type MZ1T with three different reference genes or 2YT v.s. Stoke's medium for wild type MZ1T.

In conclusion, through transposon mutagenesis of wild type MZ1T, stable floc-deficient mutants were generated. However, the transpon interrupted genes of the mutants are not directly related to thaueran synthesis and export but more global regulation at the DNA structure and protein regulation levels. These results discovered genes that are indirectly related to the floc-deficient phenotype and provided a broader understanding of the global pathways for thaueran synthesis. Cloning and expression of putative thaueran genes showed a relatively stronger interaction of the immune detection when clones were induced when compared with most of the negative controls. Reverse transcriptase quantitative real time PCR of the selected putative thaueran genes in EPS1 and EPS2 showed a strong correlation between the amount of floc and the transcription level. Therefore, through these molecular and biochemical experiments, the putative thaueran gene cluster_1(EPS1) and cluster_2(EPS2) identified and characterized by genome sequencing and annotation were suggested to be responsible for thaueran synthesis and export for *T. aminoaromatic* MZ1T.

This discovery provides a better understanding of the principles of floc formation and thaueran synthesis and the conditions influencing its abundance and can contribute to solve the dewatering problem in the wastewater treatment plant and also for producing the sugar, such as fucurosamine, as well as directing the carbon flow from acetate or aromatic compounds to thaueran, which will result in enhancing the important role of *T. aminoaromatic* MZ1T in the environment.

Chapter 6

Summary and Future work

Summary

In this study, *Thauera aminoaromatica* MZ1T was fully identified and characterized through biochemical, phyisiological and genetic approaches, such as scanning electron microscopy and transmission electron microscopy analysis, 16S rRNA gene phylogenetic analysis, DNA-DNA hybridization and cellular fatty acid composition analysis. Through genome sequencing and annotation, its general genome features, such as the genome size, GC content, the number of genes encoding proteins and structural RNA, and genes putatively involving in acetate metabolism, aromatic compound degradation, global regulation, monosacharride synthesis, polysaccharide biosynthesis and export were identified.

In order to verify the putative thaueran genes identified in genome annotation, two hypotheses, 1) no thaueran or floc formation will be observed by disrupting the putative thaueran genes using transposon mutagenesis, and 2) expression of one of the entire putative thaueran gene clusters, EPS1 or EPS2, in a non-EPS producing host will result in thaueran and floc formation, were tested.

However, since no systems for MZ1T manipulation is successfully developed yet, the results of transposon mutagenesis are not satisfactory to our expectation. Very few floc-deficient mutants were obtained. The results of cloning and expression of putative thaueran gene clusters in non-EPS producing host through detecting thaueran by using immnoblotting could only weakly support the expression of thaueran in the new hosts.

Therefore, after the verification of the DNA fragments of each putative thaueran gene cluster, reverse transcriptase QPCR was utilized to test the hypothesis that the putative thaueran gene transcription level correlates to the amount of thaueran production in the wild type MZ1T under different growth conditions as well as the floc-deficient mutants and wild type MZ1T. The results showed a clear correlation of the expression level of the putative thaueran genes and the amount of floc formed in the liquid culture under different conditions of the wild type as well as the mutants when compared with the wild type when grown under the same conditions.

Future work

From the results of the current work, there are many places can be improved and further analysis should be performed.

For putative thaueran gene cloning and expression, the fosmid harboring EPS1 and EPS2 will be extracted from the clones using the FosmidMAX™ DNA Purification Kit from Epicentre. From here two strategies can be utilized. The first approach is to

directly introduce the fosmid into expression cells, such as *Escherichia coli* DH10B(27), and induce its expression. The second approach is to look for the restriction enzyme sites at each end of the cluster. Based on this information, the complete fragments of EPS1 and EPS2 can be cut out, In this way, the putative thaueran gene cluster will be introduced into expression vectors, such as pET-30a(105), Copycontrol pCC1BAC(120), CopyRight pEZBAC(97), which are used for maintaining large fragment and are inducible for expression at high levels(120).. After ligation, the vector can be introduced into *Escherichia coli* DH10B or other expression cells that are closely related to MZ1T and maybe have a similar high GC content to EPS2 or relatively lower GC content if EPS1 is introduced.

If complete cluster cloning and expression cannot succeed, part of the cluster can be cloned into separate vectors and introduced into the same host cell to test the production of thaueran. Also, even partial expression of the cluster may render the host cell to produce intermediate products or accumulate the repeating units or polysaccharide between the membranes or inside the cell.

In addition, immnoblotting for known biomass or different dilutions of the same sample should be considered for future test. Moreover, more putative thaueran genes need to be tested for their expression level using RT-QPCR and more test could be done for the chemical mutants. Last, although quantification of polysaccharide is tedious and time consuming(26, 28), a fully quantitative measurement of the amount of thaueran in

different culture could contribute a more precise correlation between thaueran abundance and transcription level.

To identify if putative thaueran gene cluster1(EPS1) and core cluster2(EPS2) constitute an operon respectively, the following research steps will be conducted: 1) analyzing the intergenic distance to see if they are in the range from -50 to 250 bp, 2) searching for *eps* gene consensus promoter sequences, such as JUMPstart and *ops* elements and 3) searching for terminator sequences, such as Rho-dependent terminator.

An alternative method is to use the operon predictor software developed by B.P Westover et al.(115) based on a single probabilistic prediction converted from several types of genomic evidence - intergenic distance, common functional annotation and conservation of gene order.

Since many of the gene functions are not illuminated by the highest related homologous in computational genome annotation, the thaueran synthesis pathway is far from adequate in the existing MZ1T annotation. The missing steps of thaueran synthesis pathway as well as some related extracellular substance synthesis pathways will be filled manually by examining other homologous for good E-scores and relatively high identities or using the pathway hole filler program developed by M.L. Green et al (38).

The specific function of the putative genes could be identified by comparing the sequence with specified genes of that particular function from other organism, or experimentally by interrupting or cloning that gene for further analysis.

In this way, a complete pathway from the uptake of the major carbon source to the biosynthesis and export of thaueran could be depicted so that an easier manipulation of the carbon flow and sequestration through MZ1T as well as the control of floc abundance in wastewater treatment facility could be fulfilled.

References

1. **Ajdi, D., W. M. McShan, R. E. McLaughlin, G. Savi, J. Chang, M. B. Carson, C. Primeaux, R. Tian, S. Kenton, and H. Jia.** 2002. Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen. Proceedings of the National Academy of Sciences of the United States of America **99:**14434.

2. **Allen, M. S., K. T. Welch, B. S. Prebyl, D. C. Baker, A. J. Meyers, and G. S. Sayler.** 2004. Analysis and glycosyl composition of the exopolysaccharide isolated from the floc-forming wastewater bacterium Thauera sp. MZ1T. Environmental Microbiology **6:**780-790.

3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. Journal of molecular biology **215:**403-410.

4. **Anders, H. J., A. Kaetzke, P. Kampfer, W. Ludwig, and G. Fuchs.** 1995. Taxonomic position of aromatic-degrading denitrifying pseudomonad strains K 172 and KB 740 and their description as new members of the genera Thauera, as Thauera aromatica sp. nov., and Azoarcus, as Azoarcus evansii sp. nov., respectively, members of the beta subclass of the Proteobacteria. International journal of systematic bacteriology **45:**327.

5. **Arco, Y., I. Llamas, F. Martinez-Checa, M. Argandona, E. Quesada, and A. Moral.** 2005. epsABCJ genes are involved in the biosynthesis of the exopolysaccharide mauran produced by Halomonas maura. Microbiology **151:**2841.

6. **Arco, Y., I. Llamas, F. Martinez-Checa, M. Argandona, E. Quesada, and A. Moral.** 2005. epsABCJ genes are involved in the biosynthesis of the exopolysaccharide mauran produced by Halomonas maura, p. 2841-2851, vol. 151. Soc General Microbiol.

7. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Molecular biology and evolution **16:**512.

8. **Baldock, M. I., K. Denger, T. H. M. Smits, and A. M. Cook.** 2007. Roseovarius sp. strain 217: aerobic taurine dissimilation via acetate kinase and acetate-CoA ligase.

9. **Baldock, M. I., K. Denger, T. H. M. Smits, and A. M. Cook.** 2007. Roseovarius sp. strain 217: aerobic taurine dissimilation via acetate kinase and acetate CoA ligase. FEMS Microbiology Letters **271:**202-206.

10. **Barnes, W. M.** 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proceedings of the National Academy of Sciences of the United States of America **91:**2216.

11. **Bassler, B. L.** 1999. How bacteria talk to each other: regulation of gene expression by quorum sensing. Current opinion in microbiology **2:**582-587.

12. **Biegert, T., G. Fuchs, and J. Heider.** 1996. Evidence that anaerobic oxidation of toluene in the denitrifying bacterium Thauera aromatica is initiated by formation of benzylsuccinate from toluene and fumarate. European Journal of Biochemistry **238:**661-668.

13. **Breese, K., M. Boll, J. Alt-Morbe, H. Schagger, and G. Fuchs.** 1998. Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium Thauera aromatica. European Journal of Biochemistry **256:**148-154.

14. **Cashion, P., M. Holder-Franklin, J. McCully, and M. Franklin.** 1977. A rapid method for the base ratio determination of bacterial DNA. Analytical biochemistry **81:**461-466.

15. **Chain, P., J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A. Hooper, M. Klotz, and J. Norton.** 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph Nitrosomonas europaea. Journal of Bacteriology **185:**2759.

16. **Chain, P. S. G., P. Hu, S. A. Malfatti, L. Radnedge, F. Larimer, L. M. Vergez, P. Worsham, M. C. Chu, and G. L. Andersen.** 2006. Complete genome sequence of Yersinia pestis strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. Journal of Bacteriology **188:**4453.

17. **Chakraborty, R., and J. D. Coates.** 2004. Anaerobic degradation of monoaromatic hydrocarbons. Applied Microbiology and Biotechnology **64:**437-446.

18. **Christensen, B. E., J. Kjosbakken, and O. Smidsrod.** 1985. Partial chemical and physical characterization of two extracellular polysaccharides produced by marine, periphytic Pseudomonas sp. strain NCMB 2021. Applied and environmental microbiology **50:**837.

19. **De Las Penas, A., L. Connolly, and C. A. Gross.** 1997. The E-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two negative regulators of E. Molecular microbiology **24:**373-385.

20. **De Las Penas, A., L. Connolly, and C. A. Gross.** 1997. SigmaE is an essential sigma factor in Escherichia coli. Journal of Bacteriology **179:**6862.

21. **De Ley, J., H. Cattoir, and A. Reynaerts.** 1970. The quantitative measurement of DNA hybridization from renaturation rates. European Journal of Biochemistry **12:**133-142.

22. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic acids research **27:**4636.

23. **Djordjevic, S. P., H. Chen, M. Batley, J. W. Redmond, and B. G. Rolfe.** 1987. Nitrogen fixation ability of exopolysaccharide synthesis mutants of Rhizobium sp. strain NGR234 and Rhizobium trifolii is restored by the addition of homologous exopolysaccharides. Journal of Bacteriology **169:**53.

24. **Dong, C., K. Beis, J. Nesper, A. L. Brunkan-LaMontagne, B. R. Clarke, C. Whitfield, and J. H. Naismith.** 2006. Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. Nature **444:**226-229.

25. **Dubbels, B. L., L. A. Sayavedra-Soto, P. J. Bottomley, and D. J. Arp.** 2009. Thauera butanivorans sp. nov., a C2-C9 alkane-oxidizing bacterium previously referred to as' Pseudomonas butanovora'. International Journal of Systematic and Evolutionary Microbiology **59:**1576.

26. **Dubois, M., K. A. Gilles, J. K. Hamilton, P. A. Rebers, and F. Smith.** 1956. Colorimetric method for determination of sugars and related substances. Analytical chemistry **28:**350-356.

27. **Durfee, T., R. Nelson, S. Baldwin, G. Plunkett Iii, V. Burland, B. Mau, J. F. Petrosino, X. Qin, D. M. Muzny, and M. Ayele.** 2008. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. Journal of Bacteriology **190:**2597.

28. **Enos-Berlage, J. L., and L. L. McCarter.** 2000. Relation of capsular polysaccharide production and colonial cell organization to colony morphology in Vibrio parahaemolyticus. Journal of Bacteriology **182:**5513.

29. **Evans, P. J., D. T. Mang, K. S. Kim, and L. Y. Young.** 1991. Anaerobic degradation of toluene by a denitrifying bacterium. Applied and environmental microbiology **57:**1139.

30. **Ferry, J. G.** 1992. Methane from acetate. Journal of Bacteriology **174:**5489.

31. **Fitch, W.** 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Systematic zoology **20:**406-416.

32. **Foss, S., and J. Harder.** 1998. Thauera linaloolentis sp. nov. and Thauera terpenica sp. nov., isolated on oxygen-containing monoterpenes (linalool, menthol, and eucalyptol) and nitrate. Systematic and applied microbiology **21:**365-373.

33. **Gallus, C., N. Gorny, W. Ludwig, and B. Schink.** 1997. Anaerobic degradation of a-resorcylate by a nitrate-reducing bacterium, Thauera aromatica strain AR-1. Systematic and applied microbiology**:**540-544.

34. **Garrity, G.** 2005. Bergey's manual of systematic bacteriology.

35. **Gellert, M., K. Mizuuchi, M. H. O'Dea, and H. A. Nash.** 1976. DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proceedings of the National Academy of Sciences of the United States of America **73:**3872.

36. **Gorin, P. A. J., and J. F. T. Spencer.** 1966. Exocellular alginic acid from Azotobacter vinelandii. Canadian Journal of Chemistry **44:**993-998.

37. **Gottesman, S., and V. Stout.** 1991. Regulation of capsular polysaccharide synthesis in Escherichia coli K12. Molecular microbiology **5:**1599-1606.

38. **Green, M., and P. Karp.** 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinformatics **5:**76.

39. **Grobben, G. J., J. Sikkema, M. R. Smith, and J. A. M. Bont.** 1995. Production of extracellular polysaccharides by Lactobacillus delbrueckii ssp. bulgaricus NCFB 2772 grown in a chemically defined medium. Journal of Applied Microbiology **79:**103-107.

40. **Gruber, T. M., and C. A. Gross.** 2003. M ULTIPLE S IGMA S UBUNITS AND THE P ARTITIONING OF B ACTERIAL T RANSCRIPTION S PACE. Annual Reviews in Microbiology **57:**441-466.

41. **Haft, D. H., I. T. Paulsen, N. Ward, and J. D. Selengut.** 2006. Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic. BMC Biology **4:**29.

42. **Haft, D. H., J. D. Selengut, and O. White.** 2003. The TIGRFAMs database of protein families. Nucleic acids research **31:**371.

43. **Harwood, C. S., G. Burchhardt, H. Herrmann, and G. Fuchs.** 1998. Anaerobic metabolism of aromatic compounds via the benzoyl CoA pathway. FEMS Microbiology Reviews **22:**439-458.

44. **Hassler, R. A., and D. H. Doherty.** 1990. Genetic engineering of polysaccharide structure: production of variants of xanthan gum in Xanthomonas campestris. Biotechnology Progress **6:**182-187.

45. **Henke, W., K. Herdel, K. Jung, D. Schnorr, and S. A. Loening.** 1997. Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic acids research **25:**3957.

46. **Holden, M. T., Z. Heather, R. Paillot, K. F. Steward, K. Webb, F. Ainslie, T. Jourdan, N. C. Bason, N. E. Holroyd, and K. Mungall.** 2009. Genomic evidence for the evolution of Streptococcus equi: host restriction, increased virulence, and genetic exchange with human pathogens. PLoS Pathog **5:**e1000346.

47. **Huang, J., and M. Schell.** 1995. Molecular characterization of the eps gene cluster of Pseudomonas solanacearum and its transcriptional regulation at a single promoter. Molecular microbiology **16:**977-989.

48. **Hurek, T., B. Reinhold-Hurek, M. Van Montagu, and E. Kellenberger.** 1994. Root colonization and systemic spreading of Azoarcus sp. strain BH72 in grasses. Journal of Bacteriology **176:**1913.

49. **Huss, V., H. Festl, and K. Schleifer.** 1983. Studies on the spectroscopic determination of DNA hybridization from renaturation rates. System. Appl. Microbiol **4:**184-92.

50. **Jain, R., M. C. Rivera, and J. A. Lake.** 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proceedings of the National Academy of Sciences of the United States of America **96:**3801.

51. **Jansson, P., L. Kenne, and B. Lindberg.** 1975. Structure of the extracellular polysaccharide from Xanthomonas campestris. Carbohydrate Research **45:**275-282.

52. **Jansson, P. E., and A. Lindberg Paul.** 1983. Structural studies of gellan gum, an extracellular polysaccharide elaborated by Pseudomonas elodea. Carbohydrate Research **124:**135-139.

53. **Jenkins, D., M. G. Richard, and G. T. Daigger.** 1993. Manual on the causes and control of activated sludge bulking and foaming. Lewis. Ann Arbor.

54. **Ke Jiang, A. C. L., Michael S. Allen, Scott Moser, Ying Wang, John Sanseverino, John R. Dunlap, Pinidphon Prombutara, Gary S. Sayler.** 2011. Comparative Description of Thauera aminoaromatica MZ1T Necessitates Phylogenetic Reorganization of Thauera aminoaromatica species.

55. **Knobloch, J. K. M., M. Nedelmann, K. Kiel, K. Bartscht, M. A. Horstkotte, S. Dobinsky, H. Rohde, and D. Mack.** 2003. Establishment of an arbitrary PCR for rapid identification of Tn917 insertion sites in Staphylococcus epidermidis: characterization of biofilm-negative and nonmucoid mutants. Applied and environmental microbiology **69:**5812.

56. **Koebnik, R.** 2005. TonB-dependent trans-envelope signalling: the exception or the rule? Trends in microbiology **13:**343-347.

57. **Kranenburg, R., J. D. Marugg, I. I. Van Swam, N. J. Willem, and W. M. De Vos.** 1997. Molecular characterization of the plasmid encoded eps gene cluster essential for exopolysaccharide biosynthesis in Lactococcus lactis. Molecular microbiology **24:**387-397.

58. **Krause, A., A. Ramakumar, D. Bartels, F. Battistoni, T. Bekel, J. Boch, M. Böhm, F. Friedrich, T. Hurek, and L. Krause.** 2006. Complete genome of the mutualistic, N2-fixing grass endophyte Azoarcus sp. strain BH72. Nature biotechnology **24:**1384-1390.

59. **Kubori, T., Y. Matsushima, D. Nakamura, J. Uralil, M. Lara-Tejero, A. Sukhan, J. E. Gal 醬, and S. I. Aizawa.** 1998. Supramolecular structure of the Salmonella typhimurium type III protein secretion system. Science **280:**602.

60. **Lajoie, C. A., A. C. Layton, I. R. Gregory, G. S. Sayler, D. E. Taylor, and A. J. Meyers.** 2000. Zoogleal Clusters and Sludge Dewatering Potential in an Industrial Activated-Sludge Wastewater Treatment Plant. Water Environment Research **72:**56-64.

61. **Laurent, T. C.** 1970. Structure of hyaluronic acid. Chemistry and molecular biology of the intercellular matrix **2:**703?32.

62. **Leifson, E.** 1951. Staining, shape, and arrangement of bacterial flagella. Journal of Bacteriology **62:**377.

63. **Leplae, R., A. Hebrant, S. J. Wodak, and A. Toussaint.** 2004. ACLAME: a CLAssification of Mobile genetic Elements. Nucleic acids research **32:**D45.

64. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research **25:**955.

65. **Macy, J. M., S. Rech, G. Auling, M. Dorsch, E. Stackebrandt, and L. I. Sly.** 1993. Thauera selenatis gen. nov., sp. nov., a member of the beta subclass of Proteobacteria with a novel type of anaerobic respiration. International Journal of Systematic and Evolutionary Microbiology **43:**135.

66. **Mao, Y., X. Zhang, X. Xia, H. Zhong, and L. Zhao.** 2010. Versatile aromatic compound-degrading capacity and microdiversity of Thauera strains isolated from a coking wastewater treatment bioreactor. Journal of industrial microbiology & biotechnology.

67. **Martin, H. G., N. Ivanova, V. Kunin, F. Warnecke, K. W. Barry, A. C. McHardy, C. Yeates, S. He, A. A. Salamov, E. Szeto, E. Dalin, N. H. Putnam, H. J. Shapiro, J. L. Pangilinan, I. Rigoutsos, N. C. Kyrpides, L. L. Blackall, K. D. McMahon, and P. Hugenholtz.** 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotech **24:**1263-1269.

68. **Martin, J. P., and S. J. Richards.** 1963. Decomposition and binding action of a polysaccharide from Chromobacterium violaceum in soil. Journal of Bacteriology **85:**1288.

69. **Marty, N., J.-L. Dournes, G. Chabanon, and H. Montrozier.** 1992. Influence of nutrient media on the chemical composition of the exopolysaccharide from mucoid and non-mucoid Pseudomonas aeruginosa. FEMS Microbiology Letters **98:**35-44.

70. **Mart 韓, H. G., N. Ivanova, V. Kunin, F. Warnecke, K. W. Barry, A. C. McHardy, C. Yeates, S. He, A. A. Salamov, and E. Szeto.** 2006. Wastewater genomics. Nature Biotechnology **24:**1263-1269.

71. **Matringe, M., J. M. Camadro, P. Labbe, and R. Scalla.** 1989. Protoporphyrinogen oxidase as a molecular target for diphenyl ether herbicides. Biochemical Journal **260:**231.

72. **Mazur, A., J. E. Krol, M. Marczak, and A. Skorupska.** 2003. Membrane topology of PssT, the transmembrane protein component of the type I exopolysaccharide transport system in Rhizobium leguminosarum bv. trifolii strain TA1. Journal of Bacteriology **185:**2503.

73. **McIntosh, M., B. A. Stone, and V. A. Stanisich.** 2005. Curdlan and other bacterial (1 3)- -D-glucans. Applied microbiology and biotechnology **68:**163-173.

74. **Mechichi, T., E. Stackebrandt, N. Gad'on, and G. Fuchs.** 2002. Phylogenetic and metabolic diversity of bacteria degrading aromatic compounds under denitrifying conditions, and description of Thauera phenylacetica sp. nov., Thauera aminoaromatica sp. nov., and Azoarcus buckelii sp. nov. Archives of microbiology **178:**26-35.

75. **Minic, Z., C. Marie, C. Delorme, J. M. Faurie, G. Mercier, D. Ehrlich, and P. Renault.** 2007. Control of EpsE, the Phosphoglycosyltransferase Initiating Exopolysaccharide Synthesis in Streptococcus thermophilus, by EpsD Tyrosine Kinase? Journal of Bacteriology **189:**1351-1357.

76. **Morozumi, T., and J. Nicolet.** 1986. Morphological variations of Haemophilus parasuis strains. Journal of clinical microbiology **23:**138.

77. **Mulder, N. J., R. Apweiler, T. K. Attwood, A. Bairoch, D. Barrell, A. Bateman, D. Binns, M. Biswas, P. Bradley, and P. Bork.** 2003. The InterPro Database, 2003 brings increased coverage and new features. Nucleic acids research **31:**315.

78. **Neely, W. B.** 1961. Dextran: structure and synthesis. Advances in Carbohydrate Chemistry **15:**341-369.

79. **Nieto, P. A., P. C. Covarrubias, E. Jedlicki, D. S. Holmes, and R. Quatrini.** 2009. Selection and evaluation of reference genes for improved interrogation of microbial transcriptomes: case study with the extremophile Acidithiobacillus ferrooxidans. BMC molecular biology **10:**63.

80. **Niven, C. F., K. L. Smiley, and J. M. Sherman.** 1941. The polysaccharides synthesized by Streptococcus salivarius and Streptococcus bovis. Journal of Biological Chemistry **140:**105.

81. **Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa.** 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic acids research **27:**29.

82. **Paterson-Beedle, M., J. F. Kennedy, F. A. D. Melo, L. L. Lloyd, and V. Medeiros.** 2000. A cellulosic exopolysaccharide produced from sugarcane molasses by a Zoogloea sp. Carbohydrate Polymers **42:**375-383.

83. **Paulsen, I. T., A. M. Beness, and M. H. Saier Jr.** 1997. Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. Microbiology **143:**2685.

84. **Peter, B. J., J. Arsuaga, A. M. Breier, A. B. Khodursky, P. O. Brown, and N. R. Cozzarelli.** 2004. Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. Genome Biology **5:**R87.

85. **Pfoestl, A., A. Hofinger, P. Kosma, and P. Messner.** 2003. Biosynthesis of dTDP-3-acetamido-3, 6-dideoxy- -D-galactose in Aneurinibacillus thermoaerophilus L420-91T. Journal of Biological Chemistry **278:**26410.

86. **Philipp, B., and B. Schink.** 2000. Two distinct pathways for anaerobic degradation of aromatic compounds in the denitrifying bacterium Thauera aromatica strain AR-1. Archives of microbiology **173:**91-96.

87. **Porebski, S., L. G. Bailey, and B. R. Baum.** 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter **15:**8-15.

88. **Rabus, R., M. Kube, J. Heider, A. Beck, K. Heitmann, F. Widdel, and R. Reinhardt.** 2005. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. Archives of microbiology **183:**27-36.

89. **Reeves, P. R., M. Hobbs, M. A. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, and C. R. H. Raetz.** 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends in microbiology **4:**495-502.

90. **Robyt, J. F., and S. H. Eklund.** 1983. Relative, quantitative effects of acceptors in the reaction of Leuconostoc mesenteroides B-512F dextransucrase. Carbohydrate Research **121:**279-286.

91. **Roy, P. H., S. G. Tetu, A. Larouche, L. Elbourne, S. Tremblay, Q. Ren, R. Dodson, D. Harkins, R. Shay, and K. Watkins.** 2010. Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. PloS one **5:**959-964.

92. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944.

93. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution **4:**406.

94. **Sandford, P. A.** 1979. Exocellular, Microbial Polysaccharides*. Advances in Carbohydrate Chemistry and Biochemistry **36:**265-313.

95. **Scholten, E., T. Lukow, G. Auling, R. M. Kroppenstedt, F. A. Rainey, and H. Diekmann.** 1999. Thauera mechernichensis sp. nov., an aerobic denitrifier from a leachate treatment plant. International Journal of Systematic and Evolutionary Microbiology **49:**1045.

96. **Shinoda, Y., Y. Sakai, H. Uenishi, Y. Uchihashi, A. Hiraishi, H. Yukawa, H. Yurimoto, and N. Kato.** 2004. Aerobic and anaerobic toluene degradation by a newly isolated denitrifying bacterium, Thauera sp. strain DNT-1. Applied and environmental microbiology **70:**1385.

97. **Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon.** 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proceedings of the National Academy of Sciences of the United States of America **89:**8794.

98. **Smibert, R., and N. Krieg.** 1994. Phenotypic characterization. Methods for general and molecular bacteriology **1:**607–654.

99. **Song, B., N. J. Palleroni, L. J. Kerkhof, and M. M. Haggblom.** 2001. Characterization of halobenzoate-degrading, denitrifying Azoarcus and Thauera isolates and description of Thauera chlorobenzoica sp. nov. International Journal of Systematic and Evolutionary Microbiology **51:**589.

100. **Song, B., L. Y. Young, and N. J. Palleroni.** 1998. Identification of denitrifier strain T1 as Thauera aromatica and proposal for emendation of the genus Thauera definition. International Journal of Systematic and Evolutionary Microbiology **48:**889.

101. **Sonnhammer, E. L. L., S. R. Eddy, and R. Durbin.** 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins: Structure, Function, and Bioinformatics **28:**405-420.

102. **Spormann, A. M., and F. Widdel.** 2000. Metabolism of alkylbenzenes, alkanes, and other hydrocarbons in anaerobic bacteria. Biodegradation **11:**85-105.

103. **Stingele, F., J. R. Neeser, and B. Mollet.** 1996. Identification and characterization of the eps (exopolysaccharide) gene cluster from Streptococcus thermophilus Sfi6. Journal of Bacteriology **178:**1680.

104. **Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. L. Brinkman, W. O. Hufnagle, D. J. Kowalik, and M. Lagrou.** 2000.

Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature **406:**959-964.

105. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes* 1. Journal of molecular biology **189:**113-130.

106. **Sutherland, I. W.** 2001. Microbial polysaccharides from Gram-negative bacteria. International Dairy Journal **11:**663-674.

107. **Sutherland, I. W.** 1998. Novel and established applications of microbial polysaccharides. Trends in Biotechnology **16:**41-46.

108. **Suzuki, T., T. Kanagawa, and Y. Kamagata.** 2002. Identification of a gene essential for sheathed structure formation in Sphaerotilus natans, a filamentous sheathed bacterium. Applied and environmental microbiology **68:**365.

109. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution **24:**1596.

110. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research **28:**33.

111. **Thompson, J., T. Gibson, F. Plewniak, F. Jeanmougin, and D. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic acids research **25:**4876.

112. **Van den Berg, D. J. C., G. W. Robijn, A. C. Janssen, M. L. F. Giuseppin, R. Vreeker, J. P. Kamerling, J. F. G. Vliegenthart, A. M. Ledeboer, and C. T. Verrips.** 1995. Production of a novel extracellular polysaccharide by Lactobacillus sake 0-1 and characterization of the polysaccharide. Applied and environmental microbiology **61:**2840.

113. **Vorhölter, F.-J., S. Schneiker, A. Goesmann, L. Krause, T. Bekel, O. Kaiser, B. Linke, T. Patschkowski, C. Rücker, J. Schmid, V. K. Sidhu, V. Sieber, A. Tauch, S. A. Watt, B. Weisshaar, A. Becker, K. Niehaus, and A. Pühler.** 2008. The genome of Xanthomonas campestris pv. campestris B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. Journal of Biotechnology **134:**33-45.

114. **Wayne, L., D. Brenner, R. Colwell, P. Grimont, O. Kandler, M. Krichevsky, L. Moore, W. Moore, R. Murray, and E. Stackebrandt.** 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic and Evolutionary Microbiology **37:**463.

115. **Westover, B. P., J. D. Buhler, J. L. Sonnenburg, and J. I. Gordon.** 2005. Operon prediction without a training set. Bioinformatics **21:**880-888.

116. **Whitfield, C.** 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annu. Rev. Biochem **75:**39-68.

117. **Whitfield, C.** 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annu. Rev. Biochem. **75:**39-68.

118. **Whitfield, C., P. A. Amor, and R. Ko.** 1997. Modulation of the surface architecture of Gram-negative bacteria by the action of surface polymer: lipid A-core ligase and by determinants of polymer chain length. Molecular microbiology **23:**629-638.

119. **Whitfield, C., and I. S. Roberts.** 1999. Structure, assembly and regulation of expression of capsules in Escherichia coli. Molecular microbiology **31:**1307-1319.

120. **Wild, J., Z. Hradecna, and W. Szybalski.** 2002. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome research **12:**1434.

121. **Wolfe, A. J.** 2005. The acetate switch. Microbiology and molecular biology reviews **69:**12.

122. **Yamamoto, Y., S. Murosaki, R. Yamauchi, K. Kato, and Y. Sone.** 1994. Structural study on an exocellular polysaccharide produced by Lactobacillus helveticus TY1-2. Carbohydrate Research **261:**67-78.

123. **Zhu, G., G. B. Golding, and A. M. Dean.** 2005. The selective cause of an ancient adaptation. Science **307:**1279.

Vita

Ke Jiang was born in Deyang, Sichuan province, P. R. China in 1979. She spent 7 years at the Sichuan University at Chengdu where she got Bachelor and Master degrees in Microbiology. Afterwards, she had studied Biochemistry in Louisiana State University from Aug 2004 to July 2005. Then, she had her PhD education at the University of Tennessee, Knoxville from Aug 2005 to May 2011.