IDENTIFICATION OF GENES INVOLVED IN FLOCCULATION BY WHOLE GENOME SEQUENCING OF *Thauera aminoaromatica* STRAIN MZ1T FLOC-DEFECTIVE MUTANS

Pinidphon Prombutara

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2015

APPROVED:

Michael S. Allen, Major Professor
Robert C. Benjamin, Committee Member
Dan Kunz, Committee Member
Lee Hughes, Committee Member
Douglas Root, Committee Member
Art Goven, Chair of the Department of Biological Sciences
Costas Tsatsoulis, Interim Dean of the Toulouse Graduate School

Prombutara, Pinidphon. *Identification of genes involved in flocculation by whole genome sequencing of* Thauera aminoaromatica *strain MZ1T floc-defective mutants*. Doctor of Philosophy (Biochemistry and Molecular Biology), December 2015, 112 pp., 9 tables, 27 figures, references, 98 titles.

Thauera aminoaromatica MZ1T, a floc-forming bacterium isolated from an industrial activated sludge wastewater treatment plant, overproduces exopolysaccharide (EPS) leading to viscous bulking. This phenomenon results in poor sludge settling and dewatering during the clarification process. To identify genes responsible for bacterial flocculation, a whole genome phenotypic sequencing technique was applied. Genomic DNA of MZ1T flocculation-deficient mutants were subjected to massively parallel sequencing. The resultant high-quality reads were assembled and compared to the reference genome of the wild type genome. We identified nine nonsynonymous mutations and one nonsense mutation putatively involved in EPS biosynthesis. Complementation of the nonsense mutation located in an EPS deacetylase gene restored the flocculating phenotype. The FTIR spectra of EPS isolated from the wild-type showed reduced C=O peak of the N-acetyl group at 1665 cm-1 as compared to the spectra of MZ1T floc-deficient mutant EPS, suggesting that the WT EPS was partially deacetylated. Gene expression analysis also demonstrated the deacetylase gene transcript increased before flocculation occurred. The results suggest that the deacetylation of MZ1T EPS is crucial for flocculation. The information obtained from this study will be useful for preventing viscous bulking and wastewater treatment system failure, and may have potential applications in the biotechnology sector for the controlled removal of cells.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation and gratitude to my advisor, Dr. Michael S. Allen, for the patient guidance and mentorship he provided to me, all the way from when I was first considering applying to the PhD program in the Biological Department through to completion of this degree. I am truly fortunate to have had the opportunity to work with him. I would also like to thank my committee members, Dr. Dan Kunz, Dr. Lee Hughes, Dr. Tom La Point and Dr. Douglas Root, for the friendly guidance, thought provoking suggestions, and the general collegiality that each of them offered to me over the years.

I am very grateful to Dr. Robert C. Benjamin and his students for their suggestions and providing me a friendly work place in his lab. Special thanks to Dr. Jose Calderon of the Chemistry Department for his assistance on FTIR.

I would like to thank Dr. Ugo Anieto, who as a good friend, was always willing to help and give his best suggestions. It would have been a lonely lab without him. Many thanks to Dr. Sara N. Martinez, Dr. David Visi, Stephanie Simon, and Leslie M. Perry for their assistance and friendship. Finally, I would like to thank my parents for their support and encouraging me with their best wishes.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVATIONS

BOD	Biochemical oxygen demand
cDNA	Complementary deoxyribonucleic acid
ddPCR	Droplet digital polymerase chain reaction
DNA	Deoxyribonucleic acid
EMS	Ethyl methanesulfonate
EPS	Exopolysaccharide
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Km	Kanamycin
LB	Luria broth
NGS	Next generation sequencing
NTG	Nitro-N-nitrosoguanidine
PCR	Polymerase chain reaction
PGM	Personal genome machine
PNAG	Poly-β-D-*N*-acetylglucosamine
PSA	Phenol-sulfuric acid
Rif	Rifampin
RNA	Ribonucleic acid

S.O.C.	Super optimal broth (Catabolite repression)
Tet	Tetracycline
WT	Wild-type

CHAPTER 1

INTRODUCTION

Viscous bulking occurring during the clarification process of activated sludge is responsible for poor sludge settling and dewatering, sometimes leading to failure of the whole wastewater treatment process. This phenomenon is typically caused by non-filamentous floc forming organisms, which produce high levels of exopolysaccharide (EPS). Typically, synthetic polymers are added to neutralize the sludge-surface charge, facilitating flocculation and settling. It is estimated that 25–50 million kg of polymers, costing $130 million, are utilized annually in the United States for the treatment of viscous bulking (Bala Subramanian et al. 2010).

Thauera aminoaromatica strain MZ1T was originally isolated from an industrial wastewater treatment plant experiencing a prolonged episode of viscous bulking. The EPS of this floc-forming bacterium was found to be comprised of four monosaccharides: rhamnose, galacturonic acid, *N*-acetyl glucosamine, and *N*-acetyl fucosamine. The EPS composition of MZ1T was found to be similar to that of *Zoogloea ramigera*, with respect to the presence of aminosugars. Furthermore, FTIR and NMR spectroscopy uncovered the presence of possible non-sugar substitutes esterified to the exopolysaccharide of MZ1T (Allen et al. 2004). Although little is known about the flocculation mechanism of MZ1T, two major mechanisms responsible for flocculation have been proposed. One is that divalent cations, mainly Ca^{2+}, bridge negatively charged functional groups of side chain of EPS on the bacterial surface structures (De Schryver et al. 2008). Another potential mechanism of flocculation is that EPS are bound to

polysaccharide binding proteins (lectins) attached to the cell surface. The latter mechanism has been found to be the primary flocculation mechanism in the yeast *Saccharomyces cerevisiae* (Govender et al. 2008). Interestingly, however, MZ1T flocculation-deficient mutants produced some level of extractable EPS, and some of them contain secondary EPS modification. Additionally, colonies of MZ1T mutants respond differently to stains and dyes when compared to the wild type, and to have a much softer colony texture, indicating greater hydrophilic property on the cell surface (unpublished data). Therefore, multiple mechanisms might participate in the floc formation of MZ1T.

Even though random mutagenesis and phenotype screening provide a powerful method for discovering microbial functions, traditional methods for identifying mutations are labor-and time-intensive since a mutant strain may contain a lot of 50–100 random mutations, requiring extensive experiments to determine which one causes the selected phenotype. In addition, genetic manipulation of the strain MZ1T is difficult due to the massive production of EPS. For a more complete understanding of EPS biosynthesis and floc formation, the genome of *Thauera aminoaromatica* strain MZ1T has been sequenced and annotated. Three putative gene clusters responsible for exopolysaccharide biosynthesis, polymerization, and export were identified, including a Wzx-Wzy dependent pathway of polysaccharide synthesis and export. Moreover, genes involved in signal transduction, a luxR response regulator, and an acyl-acyl-carrier protein synthase were found; however, there has been no discovery of a gene encoding *N*-acyl-homoserine lactone synthetase or its homologue on the genome, so it is unlikely that a quorum sensing mechanism controls the flocculation in MZ1T as has been suggested (Jiang et al. 2012). Nevertheless, the affordability of next generation sequencing is currently transforming the field

of mutation analysis in bacteria. The genetic basis for phenotype alteration can be identified directly by sequencing the entire genome of the mutant and comparing it to the wild-type genome, thus identifying the acquired mutations (Wurtzel et al. 2010).

Here, we propose a ‘‘phenotype sequencing’’ approach in which mutations causing the phenotype can be identified directly from the sequencing of multiple independent mutant strains. In this study, two *T. aminoaromatica* strain MZ1T flocculation deficient mutants were sequenced using the Ion Torrent Personal Genome Machine, an integrated semiconductor-based genome-scale DNA sequencer. The millions of short reads generated were mapped to the reference genome of *T. aminoaromatica* strain MZ1T wild type by NextGENE software; more than 50 mismatches were discovered, including point mutations, deletions, and insertions. Subsequently, candidate mutations involving membrane-associated proteins, EPS biosynthesis, and export were validated by Sanger sequencing and complemented. To enhance our knowledge of the genetic factors and EPS biosynthesis influencing the MZ1T flocculation, we determined how the complemented genes and EPS biosynthesis genes are differentially expressed in floc-forming MZ1T as opposed to the planktonic cells during different growth periods using gene expression analysis by droplet digital PCR.

1.1 Statement of Problems

Overproduction of exopolysaccharide from *T. aminoaromatica* MZ1T results in poor sludge settling and dewaterability of wastewater treatment systems, so called viscous bulking, where this microorganism was originally isolated. In order to prevent the wastewater treatment system failure caused by this phenomenon, addition of synthetic polymer is often

required, which in this case cost over $100,000 per month at a single facility. Furthermore, the use of synthetic chemicals in the wastewater treatment results in a highly polluted environment after incineration and disposal of the sludge. Subsequently, investigation of the causative microorganism discovered a controlled ability to flocculate at stationary phase and extended to the uncovering of a unique secreted exopolysaccharide. However, little is known about the mechanism of flocculation in this organism or its regulation. Understanding the mechanisms of flocculation in MZ1T may ultimately enhance wastewater removal performance. The lessons learned here may also find applications in other biotechnological processes where cells have to be removed by costly centrifugation or filtration post-production.

CHAPTER 2

LITERATURE REVIEW

2.1 Activated Sludge Wastewater Treatment Systems and Problems Associated with Viscous Bulking

The activated sludge system is the most common biological process used in wastewater treatment. A simple activated sludge system consists of one aeration tank and one sedimentation tank or clarifier located downstream from the aeration tank (Figure 2.1). The aeration tank is a biological reactor in which wastes are converted through the activity of microorganisms to less polluting wastes or nonpolluting wastes and more solids or cell mass, mostly bacterial cells (Hogye et al. 2003). Solids in activated sludge systems are known as sludge. Sludge flocculation transforms microbial cells into aggregates, or flocs, and absorbs soluble organic waste for biodegradation. Because the sludge is aerated, the bacteria become very active in the degradation and removal of wastes. Therefore, the term “activated sludge” is used to describe the process in which bacterial solids are active in the treatment or purification of wastes.

The clarifier is a quiescent environment that allows the separation of solids from its suspending medium (water). The clarifier also removes floating foam and scum produced in and released from the aeration tank (Hogye et al. 2003). The settled floc particles represent the settled solids or sludge blanket of the clarifier. Solids in the sludge blanket may be returned to the aeration tank to treat more wastewater or may be removed (wasted) from the activated sludge process for further treatment and disposal. Regardless of the mode of operation or

variation of the activated sludge process used, the activated sludge process relies on floc particles for the degradation of biochemical oxygen demand (BOD) and removal of fine solids and heavy metals. If proper floc formation does not occur, settleability problems and loss of solids usually occur.

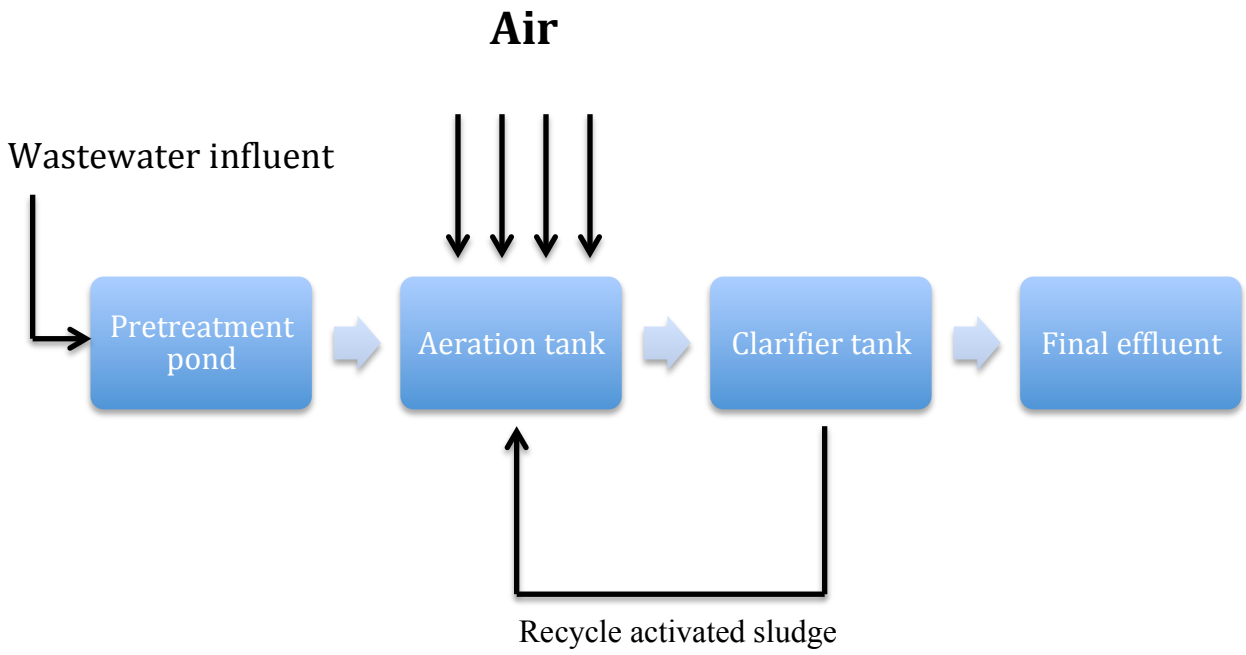


Figure 2.1 Activated sludge wastewater treatment system (adapted from "Typical activated sludge process," Hogye et al. 2003)

Floc formation in the activated sludge process is initiated by a small number of bacteria that are commonly called floc-forming bacteria. Floc-forming bacteria agglutinate or clump together with increasing sludge age (De Schryver et al. 2008). Typically, sludge flocs vary in size from 10 to 1000 µm, and microbial extracellular polymeric substances (EPS) are major components of the activated sludge floc matrix (Jin, Wilén, and Lant 2003). Proper flocs have a balanced mixture of EPS-producing and filamentous bacteria. The latter are thought to act as

the glue that binds cells and filaments together to form sludge flocs. The extracellular polysaccharide creates bridges between microorganisms by interaction with divalent cations such as Ca^{2+} and Mg^{2+} in sludge suspension with negatively charged polymers (De Schryver et al. 2008). Floc strength depends on the integrity of the biopolymer bridging. After the removal of colloidal material and soluble BOD, flocs are separated from the treated water by gravity settling and dewatering in a clarifier. This biosolid-liquid separation is one of the most critical processes in the activated sludge system because failure to settle floc particles results in lost solids (*i.e.* high BOD discharge) and contamination of the final effluent, leading to system failure (Jin, Wilén, and Lant 2003). In fact, it is known that most of the problems of poor activated sludge effluent quality result from the inability of the clarifier to efficiently remove the suspended biomass from the treated water.

Viscous floc or zoogloeal bulking is a common problem affecting the sludge settling ability in the activated sludge process. Viscous bulking is caused by rapid and undesired growth of floc-forming bacteria, which leads to the production of weak and buoyant floc particles. Weak floc particles are easily sheared, resulting in the loss of fine solids, and buoyant floc particles pack poorly in the clarifier. Zoogloeal growth may also appear as a slimy white or grayish-white film on the top of the clarifier (Figure 2.2). These organisms create large quantities of gelatinous, exocellular polysaccharides during rapid growth. The polysaccharides are insoluble in wastewater, less dense than wastewater, and water retentive. Under some circumstances, polysaccharides trap air bubbles and gases resulting in floating sludge and foam. Foam typical of viscous bulking is billowy white (Montoya et al. 2008). The polysaccharides secreted by the floc-forming bacteria have a large and highly charged surface. The net surface

charge of the polysaccharides may be anionic or cationic. Therefore, the addition of an appropriately charged polymer to the clarifier influent may help to trap fine solids and improve solids settle ability. However, use of polymer can be expensive, up to $450 per million gallons treated in clarifier tank (Montoya et al. 2008). In some cases, inorganic coagulants/precipitants such as lime or ferric chloride can be used. These produce a voluminous precipitate that sweeps down the activated sludge, improving sludge settling during a bulking episode. However, sludge production may be significantly increased if these are used (Montoya et al. 2008)

Figure 2.2 Viscous bulking causes unsettling sludge and foam in clarifiers (Retrieved from http://web.deu.edu.tr/atiksu/ana52/4ani.html, May 2015)

2.2 *Thauera aminoaromatica* Strain MZ1T

Thauera aminoaromatica strain MZ1T is a floc forming bacterium, originally isolated from the wastewater treatment plant of Eastman Chemical Company, Kingsport, Tennessee. Subsequently, in the process of identifying the causal agent of abundant Zoogloeal clusters which contributed to a viscous bulking condition at the plant, it was found that MZ1T produces a novel exopolysaccharide (Lajoie et al. 2000). Morphologically, MZ1T cells are Gram negative, short rods. MZ1T cells are motile and possess a polar flagellum. Strain MZ1T grows aerobically in Stoke's medium at the optimal temperature of 30 ºC and pH 7.2, shaking at 150 rpm. Stoke's medium contains citrate as a sole carbon source, polypeptone, salt solution and vitamin solution for MZ1T to meet nutritional requirement (Atlas 2005). Colonies are slimy and creamy white. Interestingly, MZ1T, like other members of the genus *Thauera*, is capable of degrading aromatic compounds such as benzoate and phenol under anaerobic conditions with nitrate as the terminal electron acceptor. In addition, MZ1T EPS can adsorb heavy metals from bulk suspension (Allen et al. 2004). MZ1T produces a large quantity of extracellular polysaccharide from relatively simple short chain fatty acids at the early stationary phase, and floc formation was found to occur during the stationary phase. Also, when growing under laboratory conditions, MZ1T heavily forms large, loose flocs in pure culture. The basic component of the exopolysaccharide is thought to be repeating units containing four monosaccharides: rhamnose, galacturonic acid, *N*-acetyl-glucosamine, and *N*-acetyl-fucosamine (Figure 2.3) (Allen et al. 2004). This composition is novel and unique among *Thauera*; however, the presence of amino sugars, possibly including *N*-acetyl-fucosamine, is similar to that found in *Zoogloea ramigera* (Lu, Lukasik, and Farrah 2001). This latter organism has historically been associated

with zoogloeal cluster formation in wastewater treatment systems. The presence of galacturonic acid in the exopolysaccharide could also allow interactions between its carboxyl groups with divalent cations in sludge suspension, leading to aggregation (Allen et al. 2004).

Besides EPS cross linking with divalent metal cations, another promising potential mechanism of flocculation is the interaction of EPS with polysaccharide binding cell surface proteins such as lectins. For example, *Pseudomonas aeruginosa* and *Azospirillum brasilense* Sp7 lectins located in the outer membrane were found to bind to EPS and to be involved in biofilm formation and aggregation (Latasa et al. 2006; Mora et al. 2008). In addition, IcaB located on the *S. epidermidis* cell surface introduces positive charges to the poly-N-acetylglucosamine polymer (PGA) by deacetylation of GlcNAc moieties. Notably, the presence of deacetylated PGA was essential for biofilm formation, and deacetylation is shown to be possibly involved in biofilm formation in some systems (Vuong et al. 2004).

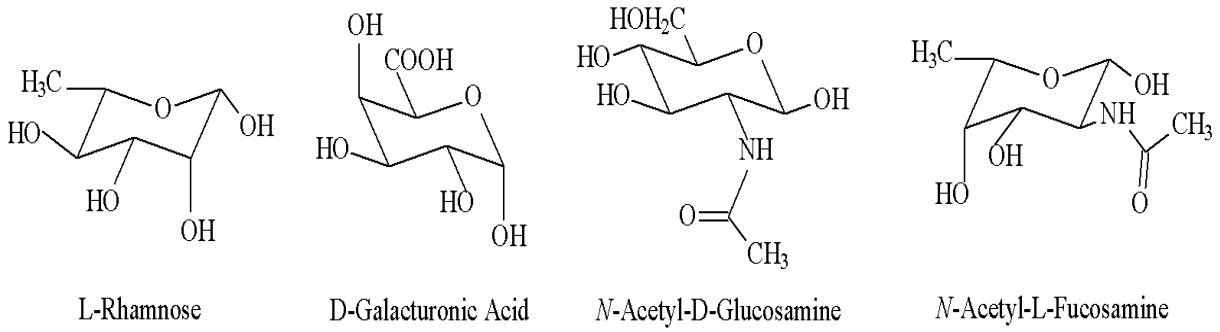


Figure 2.3 The basic components of *T. aminoaromatica* MZ1T exopolysaccharide. (Allen et al. 2004)

Although exopolysaccharide is believed to be responsible for bacterial aggregation, defective floc forming strains still produce some level of the exopolysaccharide. In fact, Allen (2002) isolated defective floc mutants following chemical mutagenesis; however, all mutants were found to produce extractable EPS and to contain the same glycosyl composition found in wild-type EPS. Interestingly, the MZ1T mutants did change their cell surface properties, with increased hydrophilicity. Recently, the genome of *Thauera sp.* MZ1T has been fully sequenced and annotated (Jiang et al. 2012). In it, three putative gene clusters involved in exopolysaccharide biosynthesis, polymerization, and export were identified. One of these is tightly organized, while the other two are loosely clustered. The cluster 1 (20.67 kb) encodes 14 genes from *Tmz1t_1114* to *Tmz1t_1127* and is made up mostly of genes necessary for EPS production, such as glycosyl transferases, UDP-*N*-acetylglucosamine 2-epimerase, and UDP-glucose/GDP-mannose dehydrogenase. In the second putative EPS gene cluster, the discovery of the *wzy* genes implicates a Wzx/Wzy-dependent pathway of polysaccharide synthesis and export may be used in MZ1T. Moreover, 6 out of total 18 genes in this cluster encode proteins that are associated with protein translocation to the cell membrane. Unlike the ABC transporter pathway where the repeating units are synthesized and polymerized at the inner face of the cytoplasmic membrane and translocated onto the periplasmic space for ligation with lipids, in the Wzx/Wyz-dependent pathway repeating units of the polysaccharide are brought by the lipid carrier and put together on the cytoplasmic side of the inner membrane; subsequently, the assembled units are translocated to the periplasmic space for polymerization. After some modification in the periplasm, the polysaccharide is translocated across the outer membrane. The third putative EPS gene cluster includes 45 genes in total. However, polysaccharide

synthesis and export genes are scattered in a wide range of about 54 kb, and some genes encode proteins with unknown function. In addition, three transposon genes, encoding transposase IS4 family proteins (*Tmz1t_3805, Tmz1t_3787 and Tmz1t_3781*), are located within this cluster. Interestingly, many genes within the two latter putative EPS gene clusters have high homology to *Azoarcus sp.*, while none of the genes in the EPS1 cluster is highly related to *Azoarcus sp.*, and no particular organisms have high similarity to multiple genes in this cluster (Jiang et al. 2012). Little is known about regulation of EPS biosynthesis and export in MZ1T. A total of six sigma factors, controlling global gene regulation, were found in the genome, including the housekeeping sigma factor σ^{70}, the nitrogen regulator σ^{54}, the heat shock sigma factor σ^{32}, and three copies of extracytoplasmic function (ECF) sigma factors. Additionally, MZ1T has a large number of genes encoding diverse transporter proteins as well as those involved in chemotaxis. Moreover, it has been shown in some organisms that cell to cell interactions, such as biofilm formation, are mediated by quorum-sensing mechanisms, and genes with potential functions in quorum sensing were found in the MZ1T genome including an acyl-acyl-carrier protein synthase and *lux*R response regulator (12 copies). Nevertheless, no *N*-acyl-homoserine lactone synthetase homologue could be identified in MZ1T. Additionally, MZ1T contains a 78.3 kb pTha01 plasmid, and analysis of this annotated plasmid shows homologs of heavy metal resistance genes. This plasmid consists of large amounts of transposase, integrase and recombinase genes, indicating that a high rate of genetic rearrangement is taking place in this strain (Jiang et al. 2012). Unfortunately, since no systems for wild-type MZ1T manipulation have been successfully developed yet, the results of transposon mutagenesis were not satisfactory in order to retrieve the mutation genotype.

2.3 EPS Biosynthesis

The polysaccharides produced by bacteria can be categorized into the exopolysaccharides (e.g. xanthan, dextran, alginate, cellulose, hyaluronic acid and colonic acid), which can be either secreted or synthesized extracellularly by cell wall-anchored enzymes, the capsular polysaccharides (e.g. K30 antigen), which are secreted but remain attached to the cell and often function as major surface antigens and virulence factors, and the intracellular storage polysaccharides (e.g. glycogen) (Rehm 2010). Further classification divides the polysaccharides into repeat unit polymers (e.g. xanthan and the K30 antigen), repeating polymers, (example.g. cellulose), and non-repeating polymers (example.g. alginate) (Rehm 2010). The establishment of polysaccharides with such varied structures and compositions involves the recruitment of different enzymes and proteins, which is reflected in the varied organizations of the biosynthetic gene clusters. The exopolysaccharide and capsular-polysaccharide biosynthetic gene clusters are likely to be affected by extensive transcriptional regulation involving two-component signal transduction pathways, quorum sensing, alternative RNA polymerase σ-factors and anti-σ-factors, and cyclic di-GMP dependent processes. Induction of exopolysaccharide biosynthesis is often correlated with establishment of biofilm growth mode, during which exopolysaccharides are important matrix components (Sutherland 2001).

The biosynthetic mechanisms fall into two general classes – an isoprenoid lipid carrier dependent mechanism and a lipid carrier independent mechanism. The initial biosynthesis step is activation of monosaccharides through formation of nucleoside diphosphate sugars (such as ADP–glucose), nucleoside diphosphate sugar acids (such as GDP–mannuronic acid) or nucleoside diphosphate sugar derivatives (such as UDP–*N*-acetyl glucosamine). Subsequently,

polymerization and transport across the inner membrane take place which require polymer-specific biosynthesis enzymes and transport mechanisms (Rehm 2010). In the isoprenoid lipid carrier dependent mechanism, highly specific sugar transferases transfer the nucleotide sugars and/or derivatives to the lipid acceptor molecule (bactoprenol, C55-isoprenoid lipid) located in the cytoplasmic membrane creating oligosaccharide repeating units. The oligosaccharide repeating units are then transported by the lipid carrier across the inner-membrane and polymerized into full-length polymer in the periplasm (Rehm 2010). For example, during the assembly of K30 antigen, a well-studied example for repeat unit polysaccharide biosynthesis, transfer of the sugar phosphate from the respective nucleotide sugar to undecaprenyl phosphate is catalyzed by the initiating glycosyl transferase, WbaP, a membrane-anchored polyisoprenyl sugar phosphate transferase. The polymerization reaction happens at the periplasmic side of the cytoplasmic membrane after an undecaprenyl phosphate-linked repeat unit gets transferred across the membrane by Wzx—a putative polysaccharide-specific transport protein (the so-called flippase) that also interacts with WbaP (Figure 2.4). In addition, the integral membrane protein Wzy has been proposed to be the polymerase that catalyses the transfer of the nascent polymer from its undecaprenyl phosphate carrier to the new lipid-linked repeat unit. The control of polymer length and guiding of the nascent polymer chain through the periplasm to the outer-membrane also requires transphosphorylation of C-terminal tyrosine residues in the Wzc oligomer and dephosphorylation by the Wzb phosphatase. The polymer is then translocated through outer membrane channel protein, Wza. Wzi is responsible for surface attachment of the capsular polysaccharide (Whitfield 2006) (Figure 2.4). Also, gumD (ortholog of WbaP) and gumK have been identified as glycosyl transferases involved in

transferring xanthan activated sugar precursors to the lipid carrier, and study of the xanthan polymerization process indicated that xanthan chains grow as described in the Wzy-dependent manner of K30 antigen biosynthesis (Rehm 2010).

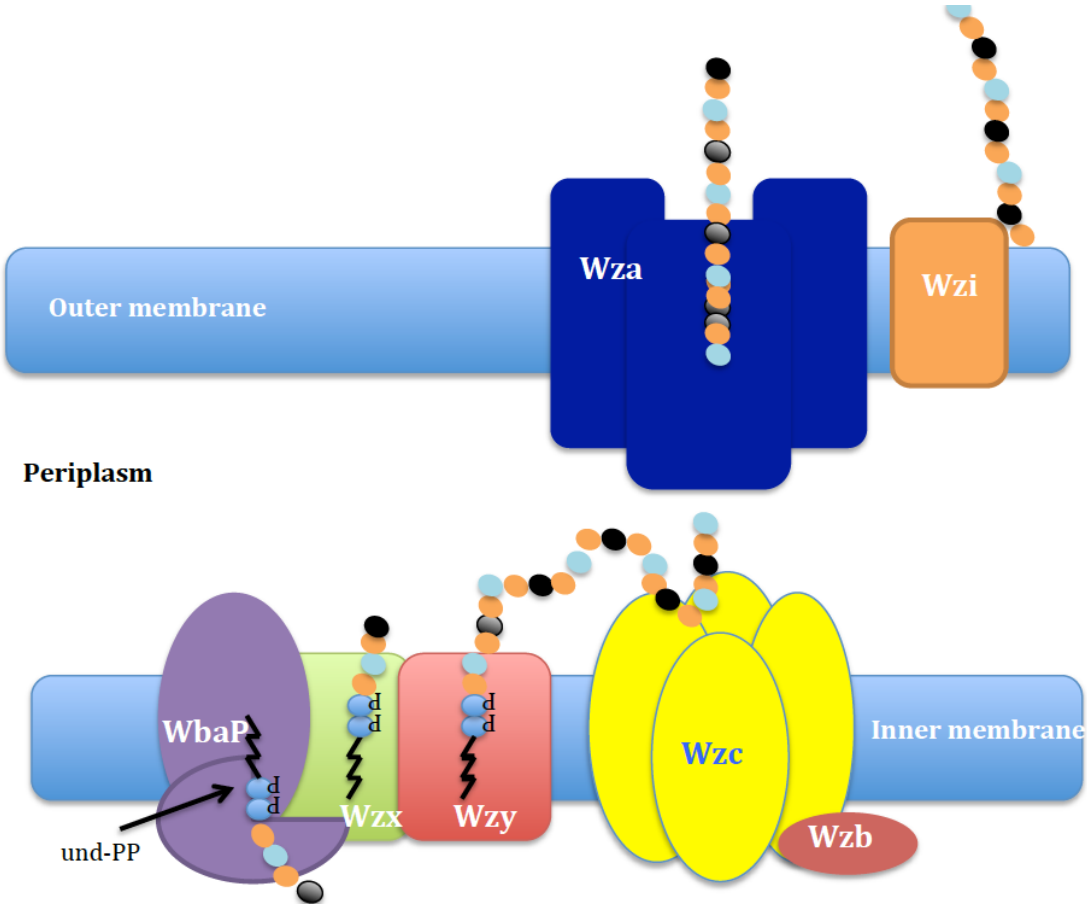


Figure 2.4 Mechanism of Wzx/Wzy-dependent group 1 capsular polysaccharide secretion systems in *E. coli*. Oligopolysaccharide repeat unit of the capsular polysaccharide is assembled on undecaprenyl lipid carrier phosphates (und-PP) by glycosyl transferase, WbaP. Wzx, flippase, facilitates transport of the und-PP-linked repeat units across the inner membrane. Elongation requires Wzy-dependent polymerization via transfer of the developing polymer from its lipid carrier to the new lipid-linked repeat unit. Transphosphorylation of C-terminal tyrosine residues in the Wzc oligomer and dephosphorylation of the Wzb phosphatase are necessary for the control of polymer length and guiding of the polymer chain through the periplasm to the outer-membrane. The polymer is then translocated through the outer membrane channel protein, Wza. Wzi is responsible for surface attachment of the capsular polysaccharide. This figure is reproduced from figure 4 “A model for biosynthesis and assembly of group 1 and 4 capsules, ” (Whitfield 2006), with permission from Annual reviews publishing.

On the other hand, no isoprenoid lipid carrier has been observed as an intermediate for alginate, cellulose, and poly-β-D-*N*-acetylglucosamine (PNAG) biosynthesis. For these bacterial exopolysaccharides, it seems that a membrane-embedded glycosyl transferase plays a role in polymerization and translocation of the polymers across the inner membrane simultaneously (Figure 2.5) (Whitney and Howell 2013).

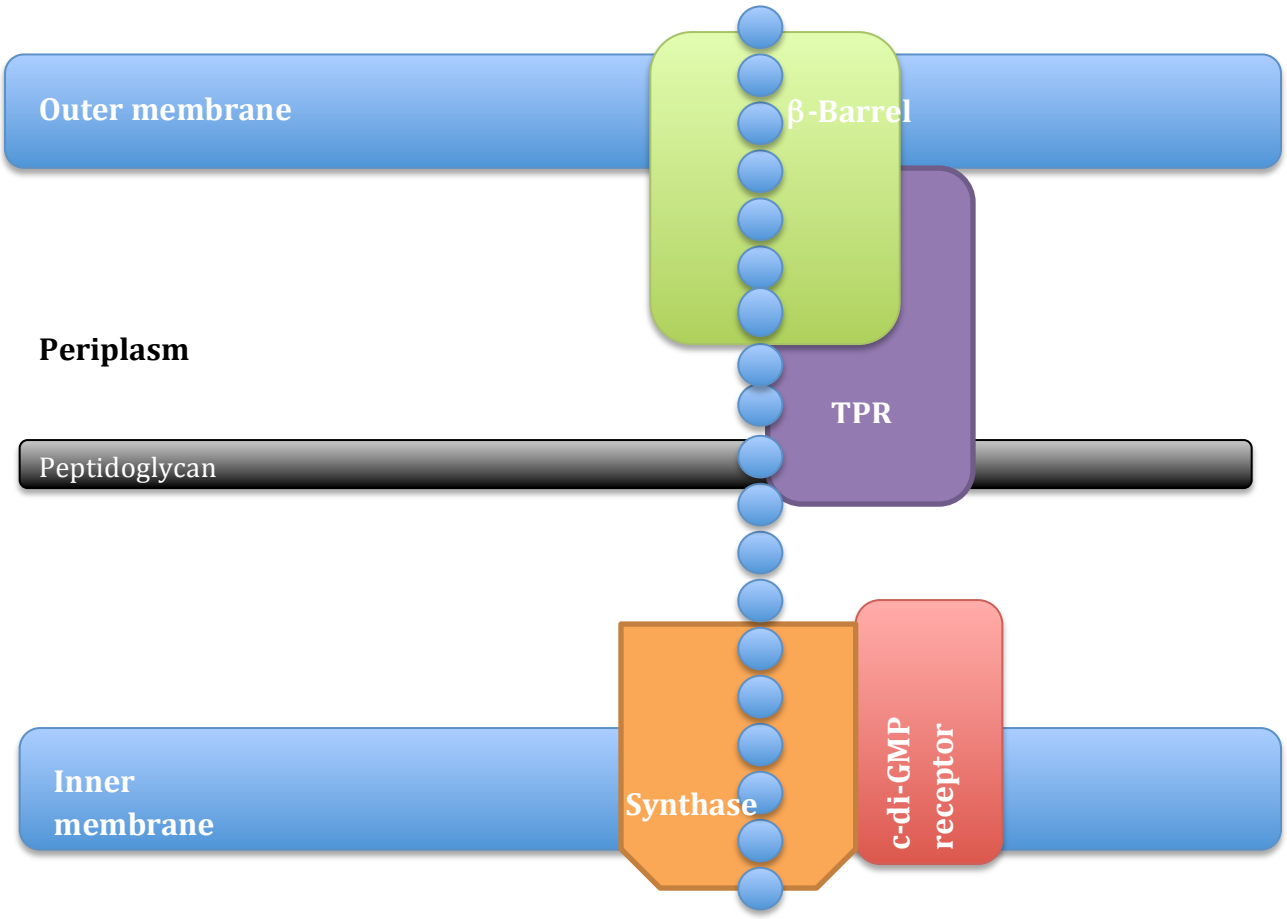


Figure 2.5 Mechanisms of lipid carrier independent polysaccharide biosynthesis. An inner-membrane synthase protein is involved in polymerization and translocation of polysaccharide across the inner membrane. In some of Gram-negative undecaprenyl lipid carrier independent polysaccharide biosynthesis, the polysaccharide biosynthesis is activated by cyclic dimeric guanosine monophosphate (c-di-GMP) that binds to an inner-membrane receptor. Once the polymer gets to the periplasm, a tetratricopeptide repeat (TPR)-containing scaffold protein guards it from degradation before it is transported across the outer membrane through a β -barrel porin. This figure is reproduced from figure 1 "Mechanisms of polysaccharide secretion, " (Whitney and Howell 2013), with permission from Elsevier Publisher.

Alginate is believed to be polymerized and directly transported across the inner membrane using its glycosyltransferase, Alg8, in conjunction with Alg44, a bis-(3'–5')- cyclic-dimeric guanosine monophosphate (c-di-GMP) binding protein which has been shown to be required for alginate polymerization. The polymer is then modified in the periplasm through selective O-acetylation by the concerted action of AlgI, AlgJ, and Alg,F and epimerized by AlgG. Alginate is secreted by a specific porin, AlgE. In fact, AlgE, a homopolymeric 18-stranded β-barrel porin, is an outer membrane protein that is capable of spontaneously incorporating into planar lipid bilayers and forming a highly anion-specific channel. AlgK also contains a lipid moiety that anchors the protein to the outer membrane (Whitney and Howell 2013). Deletion mutants of *algK* suggest that the encoded protein may play a role in the localization of the outer membrane porin protein AlgE, which led to the hypothesis that AlgK/AlgE interact to form a novel type of secretin that differs structurally from other bacterial capsular polysaccharide secretion systems (Keiski et al. 2010).

In cellulose biosynthesis, BcsA, an inner-membrane protein with multiple transmembrane domains and a cytoplasmic family 2 glycosyl transferase domain, facilitates cellulose polymerization from UDP-glucose and translocation of the newly formed polymer across the inner membrane similar to alginate synthase, Alg8. However, unlike alginate biosynthesis, in which Alg8 and Alg44 are responsible for the polymerization and c-di-GMP binding activities, separately, BcsA also processes c-di-GMP binding activity at its C terminus using a PilZ domain. Moreover, *bcsC* is thought to encode a large outer-membrane protein containing a N-terminal domain that resides in the periplasm and a C-terminal porin domain that assists cellulose export across the outer membrane. These two domains appear to be like

AlgK and AlgE from the alginate secretion system (Whitney and Howell 2013). Lastly, the polymerization of UDP-*N*-acetylglucosamine precursor and translocation of β-1,6-linked *N*-acetyl-D-glucosamine (PNAG) across the inner membrane in *E. coli* are catalyzed by the putative PNAG synthase, PgaC . The PgaC protein is predicted to contain multiple transmembrane domains and a large cytoplasmic domain that shares homology with family 2 glycosyl transferases as found in Alg8 and BcsA (Whitney and Howell 2013). In addition, PNAG is an important component of the biofilm matrix produced by various bacteria, and PgaC is conserved among these bacteria; for example, HmsR and BpsC found in *Yersinia pestis* and *Bortella bronchiseptica* contain family 2 glycosyl transferase activity (Whitney and Howell 2013). Once the PNAG polymer is in the periplasm, it is partially (~22%) deacetylated by the carbohydrate esterase PgaB. The degree of deacetylation observed among PNAG-producing bacteria is varied depending on the enzyme activity of their PgaB homolog. PgaA, a predicted outer-membrane protein, appears to have a domain arrangement resembling BcsC/AlgK and AlgE and carries out PNAG export. Moreover, in the absence of PgaB, PNAG accumulates in the periplasm of *E. coli*, leading to speculation that the putative export function of PgaA is specific for partially deacetylated PNAG (Whitney and Howell 2013). The exact function of the predicted inner-membrane protein PgaD is unknown. However, it is thought to assist the polymerization process because *pgaD* gene deletion also stops the production of PNAG (Whitney and Howell 2013).

2.4 EPS Gene Organization

Exopolysaccharide syntheses in many bacteria is a multiple step process involving the interconnected activity of many enzymatic proteins. In fact, genes employed in this process are

usually grouped in large clusters located on chromosomes or megaplasmids (Dimopoulou et al. 2014; Finan et al. 2001; Rehm 2010). Some of the EPS biosynthesis gene clusters have been cloned and sequenced, and found to form long operons with similarities in their genetic organization (Rehm 2010). Among these genes encode enzymes essential for the synthesis of nucleotide sugar precursors, enzymes engaged in unit assembly and modification, proteins responsible for polymerization of repeating or homopolymer units and transport of EPS outside the bacteria. For example, xanthan biosynthesis of *Xanthomonas campestris* was controlled by the *gum* genes with 14 genes involved in the process (Vorholter et al. 2008). Hay et al. (Hay et al. 2010) identified 24 genes in *Pseudomonas aeruginosa* responsible for the production of alginates. Marvasi et al. (Marvasi, Visscher, and Casillas Martinez 2010) also reported that 16 genes were involved in the levan biosynthesis of *Bacillus subtilis*. In addition, biofilm was shown to be regulated by 15 genes and 12 genes in *Bacillus subtilis* and *Burkholderia cenocepacia*, respectively (Fazli et al. 2013; Kearns et al. 2005). Moreover, one of the most comprehensive studies on the genetics and biochemistry of EPS biosynthesis in Gram-negative bacteria has been made by Walker and his colleagues on Rhizobium strains. In *S. meliloti*, the genes responsible for the synthesis of EPS I form a large exo/exs cluster (~35 kb) located on the pSymB megaplasmid (Reuber and Walker 1993). In this region, 28 exo/exs genes organize in several operons containing the genes encoding enzymes for the synthesis of nucleotide sugar precursors (*exoB* and *exoN*), enzymes involved in unit assembly (*exoY*, *exoF*, *exoA, exoL, exoM, exoO, exoU* and *exoW*) and modification (*exoZ*, *exoH* and *exoV*), and proteins necessary for polymerization of repeating units and transport of EPS I (*exoP*, *exoT, exoQ* and *exsA*) (Glucksmann, Reuber, and Walker 1993). However, genes crucial for sugar precursor synthesis

(*exoC*) and regulation of EPS I production (*exoD*, *exoR*, and *exoS*) are not connected to this region but scattered throughout the chromosome of *S. meliloti* (Reed and Walker 1991; Uttaro et al. 1990; Yao et al. 2004). The synthesis of the second *S. meliloti* exopolysaccharide, named galactoglucan (EPS II), is directed by *exp* genes located in a 27-kb cluster on the pSymB plasmid, at a distance of 160 kb from the *exo/exs* genes (Becker et al. 1997). This cluster comprises of 22 genes organized into five operons: *wga* (*expA*), *wgcA* (*expC*), *wggR* (*expG*), wgd (*expD*) and *wge* (*expE*) (Bahlawane et al. 2008). Among them, four genes (*wgaG*, *wgaH*, *wgaI* and *wgaJ*) are involved in the synthesis of deoxythymidine diphospho-sugar precursors (dTDP-rhamnose and dTDP-glucose), and six genes encode potential glycosyltransferases: WgaB and WgeB - glucosyltransferases and WgaC, WgcA, WgeD and WgeG galactosyltransferases. Other genes of this cluster are potentially engaged in the polymerization (*wgdA* and *wgdB*) and regulation of EPS II synthesis (*wggR*) (Bahlawane et al. 2008). In contrast to the well-studied model of the synthesis of *S. meliloti* EPS I, the data concerning EPS biosynthesis in *R. leguminosarum* are deficient in quantity. Genes involved in the synthesis of nucleotide sugar precursors as well as genes engaged in the synthesis and export of EPS are located on the chromosome of *R. leguminosarum*, and the majority of them are grouped in a large region known as the Pss-I gene cluster. This 33.7-kb long region encompasses more than 20 genes (Król et al. 2007).

Nevertheless, the *pssA* gene encoding a protein activating the first step of the synthesis of the octasaccharide subunit is organized in a single open reading frame and located at a long distance from other *pss* genes (Ksenzenko et al. 2007). *pssA* is a highly conserved gene present in all *R. leguminosarum* biovars and also in other closely related species, such as *Rhizobium etli* (Janczarek, Kalita, and Skorupska 2009). Mutations in *pssA* totally block EPS production and

result in the induction of empty (devoid of bacteria) non-nitrogen-fixing nodules on roots of host plants (Ivashina et al. 1994; Janczarek and Rachwal 2013; van Workum et al. 1997). In the *S. fredii* NGR234 genome, a 28-kb region containing *exo* genes located in pNGR234 has been identified (Streit et al. 2004). These genes are highly homologous to the *exoA, exoB, exoY, exoL, exoM, exoN and exoP* genes of *S. meliloti*. Large parts of *exo* gene clusters of *S. fredii* NGR234 and *S. meliloti* species are closely related, especially their exoX-exoY regions, which are almost identical. Additionally, the *exoG* gene, not linked with this *exo* region, has been discovered in the *S. fredii* NGR234 genome (Streit et al. 2004; Zhan et al. 1990). The existence of similar *exo* genes in *S. fredii* NGR234 and *S. meliloti* might be explained by the fact that both of these rhizobial species produce EPS of very similar structures. However, some differences in the genetic organization of these *exo* clusters have been distinguished. For example, a non-functional homologue of *S. meliloti exoH*, which is responsible for succinylation of EPS by this bacterium, is altered in its location from the *exo* cluster in plasmid pNGR234b to *S. fredii* genome (Schmeisser et al. 2009). These finding explain why, in contrast to *S. meliloti* EPS, EPS of *S. fredii* NGR234 is not succinylated. Moreover, a region involved in EPS biosynthesis has also been identified in the genome of *Bradyrhizobium japonicum*. This cluster comprises six ORFs organized into at least four different operons. One gene seems to be homologous to *exoB* encoding a UDP-galactose 4'-epimerase. Other ORFs were identified as UDP-hexose transferases and one ORF identical to *S. meliloti exoP*, which has been suggested to be involved in EPS chain-length determination (Becker et al. 1998).

2.5 Identification of Genes Involved in Bacterial Exopolysaccharide Production

2.5.1 Strategies for Identifying Structural Genes Involved in EPS Biosynthesis

The structural genes involved in polysaccharide biosynthesis are clustered in a large number of Gram-negative organisms, such as *Escherichia coli, Xanthomonas campestris, Rhizobium meliloti, Zooglea ramigera, Pseudomonas sp.* and *Acetobacter xylilum* (Barrere, Barber, and Daniels 1986; Easson, Sinskey, and Peoples 1987; Harding et al. 1987; Rehm 2010; Standal et al. 1994; Thorne, Tansey, and Pollock 1989). Thus, strategies that enable the identification and isolation of one gene facilitate identification of others. This is particularly useful for the identification and isolation of genes for which no or only tedious screening methods can be devised. In order to find genes involved in polysaccharide biosynthesis a number of different strategies have been employed. One of which is mutation-based approaches. Mutations can arise spontaneously or can be induced by the use of a variety or agents such as chemicals, ultraviolet light or transposons. In addition, induced mutations occur at higher frequencies than spontaneous mutants; therefore, unless a spontaneous mutant is already available, most researchers will induce mutations in the host. However, a mutant will have to be screened to identify those displaying the appropriate phenotype. Since the process of mutation by chemical agents and transposon mutagenesis is not selective, it is first essential to design a suitable screening procedure for mutant isolation and identification. Mutagenesis with EMS (Barrere, Barber, and Daniels 1986; Thorne, Tansey, and Pollock 1989) and NTG (Harding et al. 1987) were used to generate mutants defective in xanthan biosynthesis of *X. campestris*. Restoration of the mucoid phenotype by complementation of EPS defective mutants with cloned genomic DNA was used to identify fragments containing the

polysaccharide biosynthetic genes (Barrere, Barber, and Daniels 1986; Harding et al. 1987; Thorne, Tansey, and Pollock 1989). Moreover, transposon mutagenesis of the insert DNA with Tn5 (Barrere, Barber, and Daniels 1986) and mini-Mu (Harding et al. 1987) was used to isolate the complementation regions of the insert DNA. Marker exchange of Tn5 insertions from cloned DNA into the *X. campestris* genomic (Barrere, Barber, and Daniels 1986) or the complementing plasmid (Harding et al. 1987) provided evidence that the genes involved in xanthan biosynthesis were clustered. In addition, mutants of *Burkholderia cepacia* IST408 unable to produce exopolysaccharide were isolated using a random plasposon mutagenesis strategy. Transconjugants were tested for EPS production using the lipophilic dye Sudan Black B, and 58 mutants, producing reduced or undetectable amounts of EPS, were selected. The retrieved recombinant plasmids incorporated chromosomal insertions stretching from 1 to 11 kb. Subsequently, the DNA fragment flanking the TnMod-KmO insertion of all the plasmids was sequenced to distinguish EPS biosynthetic genes (Moreira et al. 2003). Another study showed that screening of 5000 randomly mutagenized colonies with the transposon Tn*5* for defects in social-motility and EPS in *Myxococcus xanthus* led to identification of two genetic regions essential for EPS biosynthesis: the EPS synthesis (*eps*) region and the EPS-associated region (Easson, Sinskey, and Peoples). Mutants with insertions in the *eps* and *eas* regions were defective in social motility and fruiting body formation (Lu et al. 2005).

In another study, mini-Tn5 insertion was used to identify the biosynthesis of the EPS mauran by *H. maura* strain S-30 by initially isolating an EPS-defective mutant that carried a single insertion of mini-Tn5 in its genome and analyzing the regions located both upstream and downstream of the insertion site. The analysis of the flanking regions of the insertion site in the

H. maura mutants resulted in the identification of five ORFs (*epsABCDJ*), which 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 (Arco et al. 2005).

Recently, an extracellular polysaccharide bioflocculant containing a neutral sugar, amino sugar, and an uronic acid isolated from *Bacillus licheniformis* has been applied to treat sugarcane-neutralizing juice to remove colloids, suspended particles, and coloring materials in a sugar refinery factory. To investigate genes involved in this bioflocculant synthesis, a fosmid library consisting of 1,824 recombinant clones was generated from *Bacillus licheniformis* genomic DNA and screened for the production of the bioflocculant. A two-pooling scheme was used to isolate the positive clones based on flocculating activity and four positive clones with the highest flocculating activity were selected for sequence analysis. A fragment of 30-kb was identified with 26 hypothetical genes in the bioflocculant producing clone. Most of the predicted proteins encoded by the inserted genes showed significant homology with enzymes involved in the biosynthesis of polysaccharides (Yan et al. 2013).

However, a major drawback of this approach is that screening can pick up mutants whose changed phenotype is not caused by mutations in structural genes. Distinguishing between the targeted mutants and those in genes exerting pleiotropic effects or those in genes involved in regulation of polysaccharide biosynthesis can be a labor intensive task. In addition, it can also be very difficult to isolate genes for which no reasonable screening procedure can be devised. For example, it could be impossible to screen for mutations in genes encoding the transfer of non-sugar substituents, such as acetyl groups, to the polymer. Polysaccharide from such mutants would have to be isolated from the culture medium, purified and possibly

characterized in detail. Once the desired mutant has been isolated, however, it can be complemented with plasmids containing cloned wild-type DNA to identify host fragments containing the gene(s) of interest. These DNA fragments can then be genetically characterized, subcloned and used to complement a variety of mutants, and/or directly sequenced.

2.5.2 Identifying Mutations by Whole-Genome Sequencing

The study of mutants to elucidate gene functions has a long and successful history; however, to discover causative mutations in mutants that were generated by random mutagenesis often takes years of laboratory work and requires previously generated genetic and/or physical markers, or resources like DNA libraries for complementation. Recently, next-generation sequencing technologies have revolutionized the field of microbial genomics and genetics (Dark 2013). Next-generation sequencing technology is massively parallel. It has great advantages over the Sanger sequencing method because all reactions occur at the same time, and independently. Basically, next-generation sequencing platforms require library or template DNA preparation starting with amplicon/genomic DNA size selection. Then, DNA fragments are linked with an adaptor sequence and attached to beads or a glass surface followed by clonal PCR amplification (emulsion or bridging PCR) in order to adequately amplify the sequencing signal for the detectors during the sequencing reaction. Finally, tens of thousands to many millions of DNA sequencing reactions occur independently, and the reads are detected at the same time according to the sequence mechanism/chemistry of each platform. This technology provides higher accuracy and cost reduction, leading to vast numbers of sequencing reads (Pareek, Smoczynski, and Tretyn 2011). Therefore, the rapid development of next-generation sequencing platforms has enabled the use of sequencing for an alternative to the classical

genetic mapping of mutations. By aligning short reads produced by NGS from a mutant to *a priori* sequenced reference genome of the wild type, mutations are inferred from the differences between the WT reference and the sequenced mutant (Harper et al. 2011). Also, the cost per megabase (i.e. one million bases) of sequence has dropped dramatically from $100 to $0.05 during time from 2008 to 2015 (Figure 2.6). On the other hand, one could sequence a whole human genome for $4200 with average 30X coverage depth using Illumina Hiseq 2500 platform (Wetterstrand 2015). Consequently, it is now affordable to sequence an entire prokaryotic genome in order to identify acquired mutations.

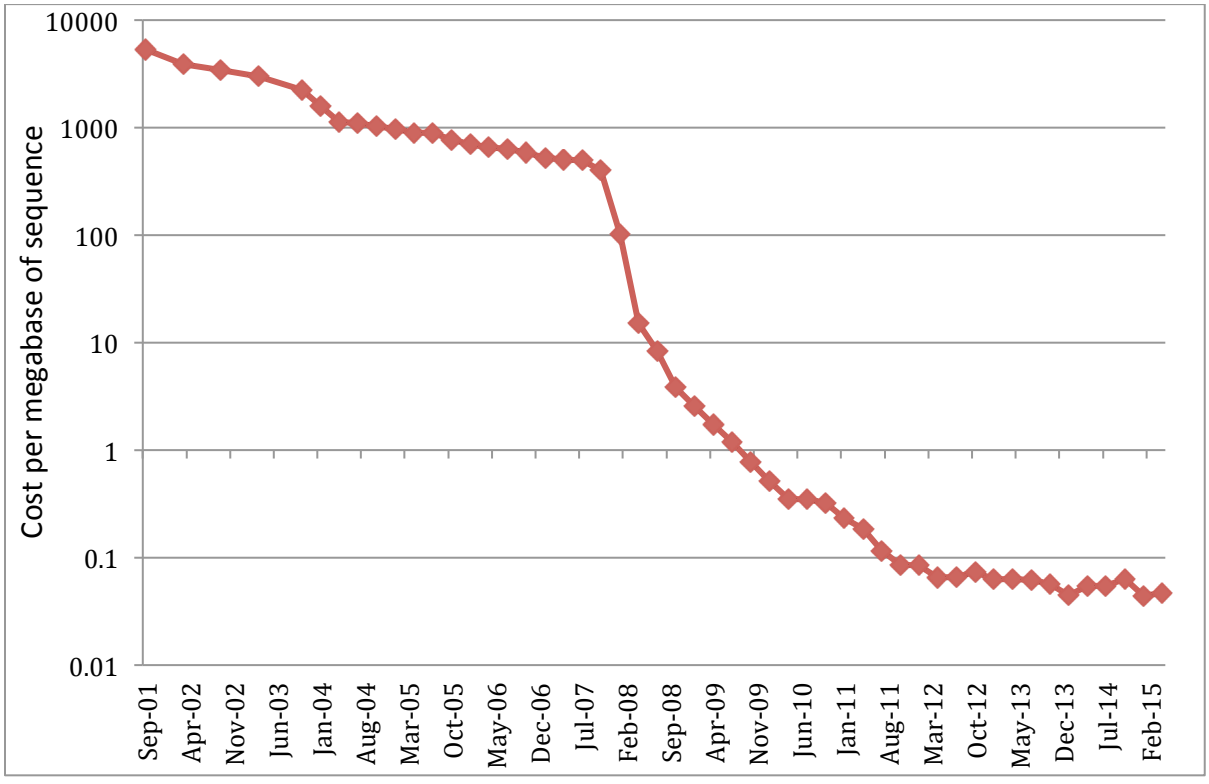


Figure 2.6 Cost per megabase of sequence, from 2001 to 2015. Adapted from the NIH NHGRI Genome Sequencing Program website.

(Data accessed at http://www.genome.gov/sequencingcosts/)

There are several examples of the use of short read WGS for mutation detection. Srivatsan et al. (Srivatsan et al. 2008) sequenced a selection of *Bacillus subtilis* laboratory strains using Illumina WGS and identified a variety of mutations including two synthetic *relA* suppressing mutations, each residing in a separate *relA* homolog and each having only a partial suppressing effect. Davis and Waldor (Davis and Waldor 2009) sequenced *rnaE* mutants of *Vibrio cholera*, using Illumina WGS in a search for *rnaE* suppressors and reported single-nucleotide substitutions and single-nucleotide indels compared to the reference sequence. Many other strategies for mutation identification through whole-genome sequencing have been applied to several model organisms, including *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Arabidopsis thaliana, Neurospora crassa*, the rodent malaria parasite *Plasmodium chabaudi*, and humans (Flibotte et al. 2010; Irvine et al. 2009; Lupski et al. 2010; McCluskey et al. 2011; Pomraning, Smith, and Freitag 2011; Puente et al. 2011; Roach et al. 2010; Sarin et al. 2008; Zuryn et al. 2010). This indicates that this approach is feasible even for large eukaryotic genomes. In this research project, we chose the Ion Torrent PGM next generation technology as a tool to reveal the factors/genes involved in MZ1T flocculation. The Ion Torrent PGM couples semiconductor technology with a simple sequencing chemistry in which a nucleotide is incorporated into a strand of DNA by a polymerase, resulting in a release of a hydrogen ion as a byproduct (Figure 2.7). The semiconductor device (ion chip) uses a high-density array of micro-machined wells to carry out this sequencing process in a massively parallel way with each well holding a different DNA template. The Ion Torrent Personal Genome Machine (PGM™) sequencer sequentially floods the chip with one nucleotide after another. If nucleotide A, for example, is added to a DNA template and incorporated into a strand of DNA,

then a hydrogen ion will be released. The charge from that ion will change the pH of the solution and can be detected (in bulk from clonally amplified DNA residing in the well) directly by the ion sensor beneath the wells (Figure 2.7). In addition, if the next nucleotide that floods the chip is not a match, no pH change will be recorded, and no base will be called. If there are two identical bases on the DNA strand, the pH change will be double, and the chip will record two identical bases. PGM is the first commercial sequencing machine that does not require fluorescence and camera scanning, resulting in higher speed, lower cost, and smaller instrument size (Rothberg et al. 2011).

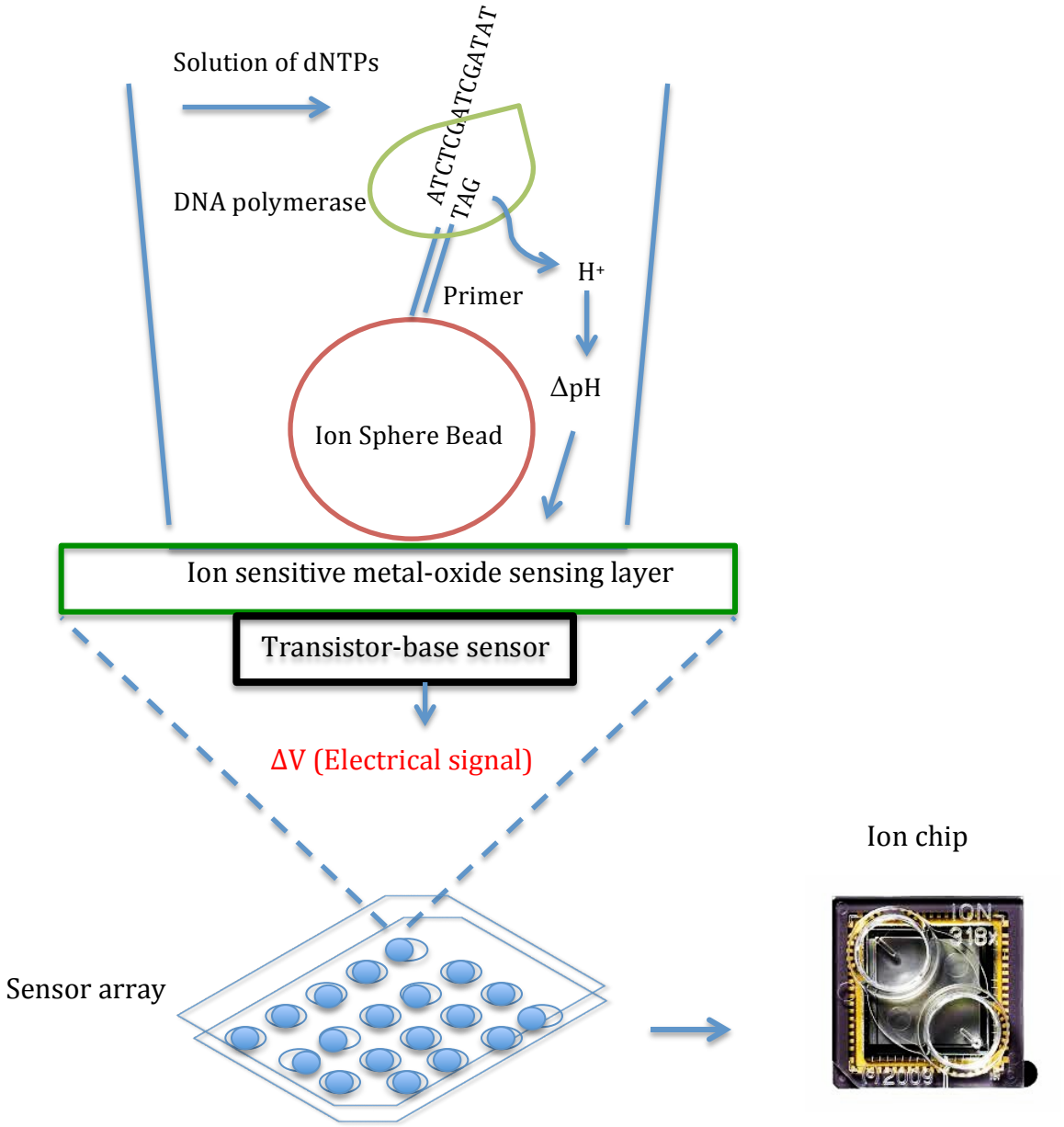


Figure 2.7 Principles and elements of semiconductor sequencing. The Ion Torrent platform executes DNA sequencing as the complimentary strand of DNA is synthesized. In fact, clonal DNA immobilized on a bead is synthesized by polymerase in the presence of a pure solution of one nucleotide. Nucleotide incorporation releases a hydrogen ion. Changing the pH of the well (ΔpH) alters the surface potential of the ion-sensitive metal oxide layer, and it is turned to a voltage signal by transistors. The wells are washed and flooded sequentially with pure solutions of other nucleotides. Sequencing chips are fabricated with sensor array, containing 1.2 M, 6.1 M and 11 M micro-wells (Ion chip 314, 316 and 318). This figure is reproduced from figure 1 "Sensor, well and chip architecture " (Rothberg et al. 2011), with permission from Nature publishing group.

Currently, it enables 400 bp reads in 5 h using Ion chip 316 v2 (Table 2.1) and the sample preparation time is less than 10 hours for 8 samples in parallel. However, to detect the length of homopolymer runs, the sensor must detect the magnitude of the pH change to determine how many nucleotides were incorporated. Thus, errors on the Ion Torrent platform are mostly insertions and deletions in homopolymer runs, and the error rate increases as the homopolymer length progresses (Figure 2.8), resulting from difficulties in evaluating the magnitude of signal when several dNTPs are incorporated in one cycle (Bragg et al. 2013). Depending on the throughput demands of the experiment and application, one can select among the Ion 314, Ion 316, or Ion 318 sequencing chips. The only difference among the chips is the number of interrogating wells, ranging among 1M for the Ion 314 sequencing chip, 6 M for the Ion 316, and 12 M for the Ion 318. These sequencing chips can create from hundreds of thousands to up to 5.5 M reads (Table 2.1).

Table 2.1 An overview of Ion torrent sequencing technology adapted from "An overview of current sequencing technologies" (Dark 2013), with permission from Dove press limited.

Platform Ion torrent	Runtime	Sequence yield per run	Reported accuracy	Mean read length	Template DNA required	Reads per run
Ion 314 chip	2.3–3.7 h	30–100 Mb	Q30	200–400 bp	100 ng	400–550 k
Ion 316 chip	3–4.9 h	300 Mb–1 Gb	Q30	200–400 bp	100 ng	2–3 M
Ion 318 chip	4.4–7.3 h	600 Mb–2 Gb	Q30	200–400 bp	100 ng	4–5.5 M

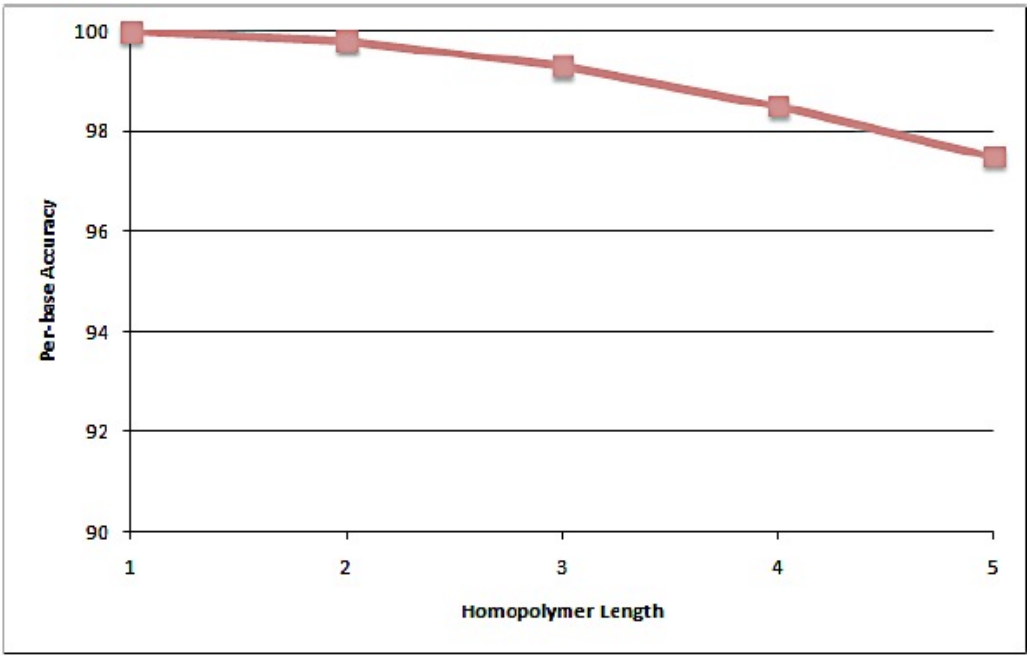


Figure 2.8 Homopolymer accuracy. The error rate increases as the homopolymer length progresses. A 5-homopolymer is currently called with 97.5% per base accuracy. Data generated from a single run of an Ion 314 chip using *E. coli* DH10B genome and Torrent Suite software v1.3.0. Data supporting of this figure can be obtained within the Ion Community (http://ioncommunity.iontorrent.com). Copyright, 2012 by Life technology

The overall experimental workflow for the semiconductor sequencing is outlined in Figure 2.9. After cell and total DNA isolation, construction of sequencing libraries from a genomic DNA sample is achieved by genomic DNA fragmentation and ligation of ion adaptors. Subsequently, fragmented DNA with ligated adaptors is size selected and purified. Each individual DNA fragment is immobilized on an Ion Sphere Particle (ISP) and clonally amplified. The process is automated with a supplementary OneTouch System (Figure 10 a). The resulting beads with amplified (emulsion PCR), individually cloned DNA fragments are then enriched to eliminate "empty" beads— this process is also carried out by a robotic enrichment system (ES) of the Ion OneTouch system (Figure 10 b). Finally, after being loaded on a selected sequencing chip, the beads containing clonal populations of the DNA from an experimental sample are laid out in wells and incubated serially with pure, unmodified nucleotides of DNA. Incorporation of nucleotides are continuously detected by measuring changes in the hydrogen ion concentration during the sequencing process on the PGM machine, and simultaneously processed on a server for further analysis and assembly. The Ion Torrent PGM is a bench top sequencer that can be put in any individual laboratory and not necessarily a sequencing center or core facility. Price including Ion Torrent PGM, server, OneTouch and OneTouch ES sample automation systems is about $80,500, and approximate cost per run of Ion chip 314 v2, 316 v2 and 318 v2 is $225, $425 and $625 respectively. However, sample prices do not include the cost of generating the initial fragmented genomic DNA library with adaptors which have an additional cost of between $75−200 depending on method used (Loman et al. 2012). In addition, initial training and practical experience can be obtained within a couple of weeks for most laboratories.

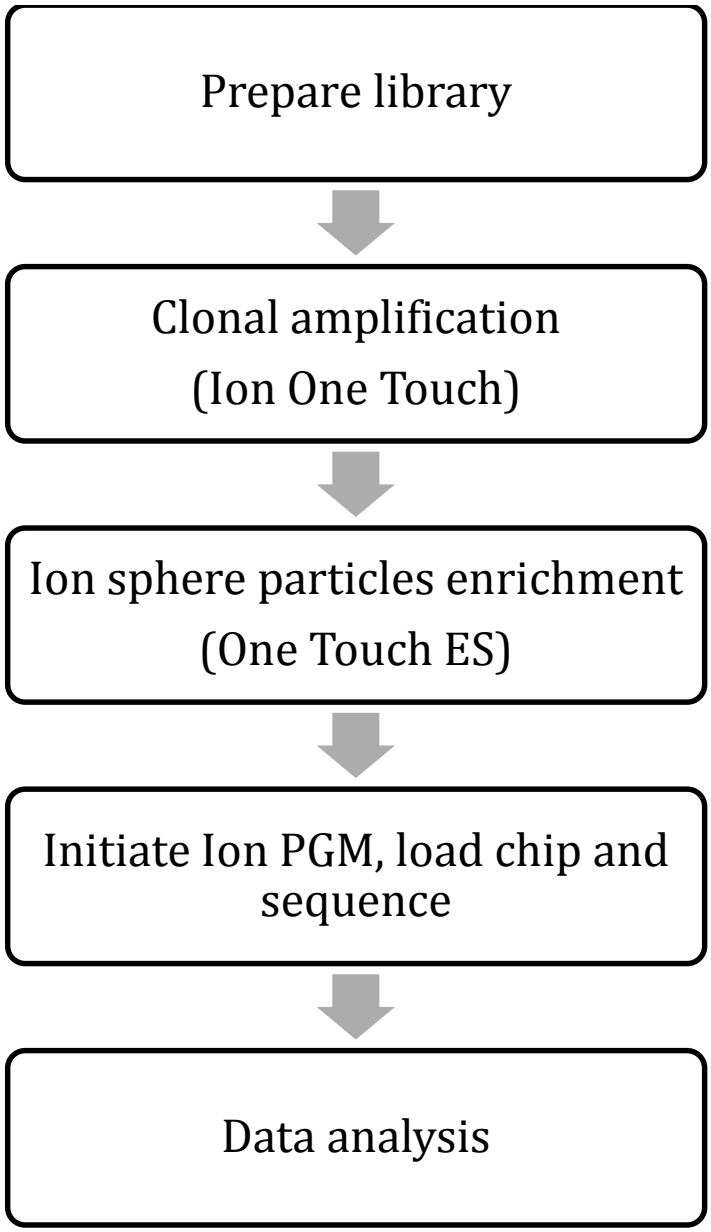


Figure 2.9 Experimental workflow for the semiconductor sequencing (Retrieved and adapted from “Ion PGM sequencer protocol” within the Ion Community (http://ioncommunity.iontorrent.com). Copyright, 2012 by Life technology

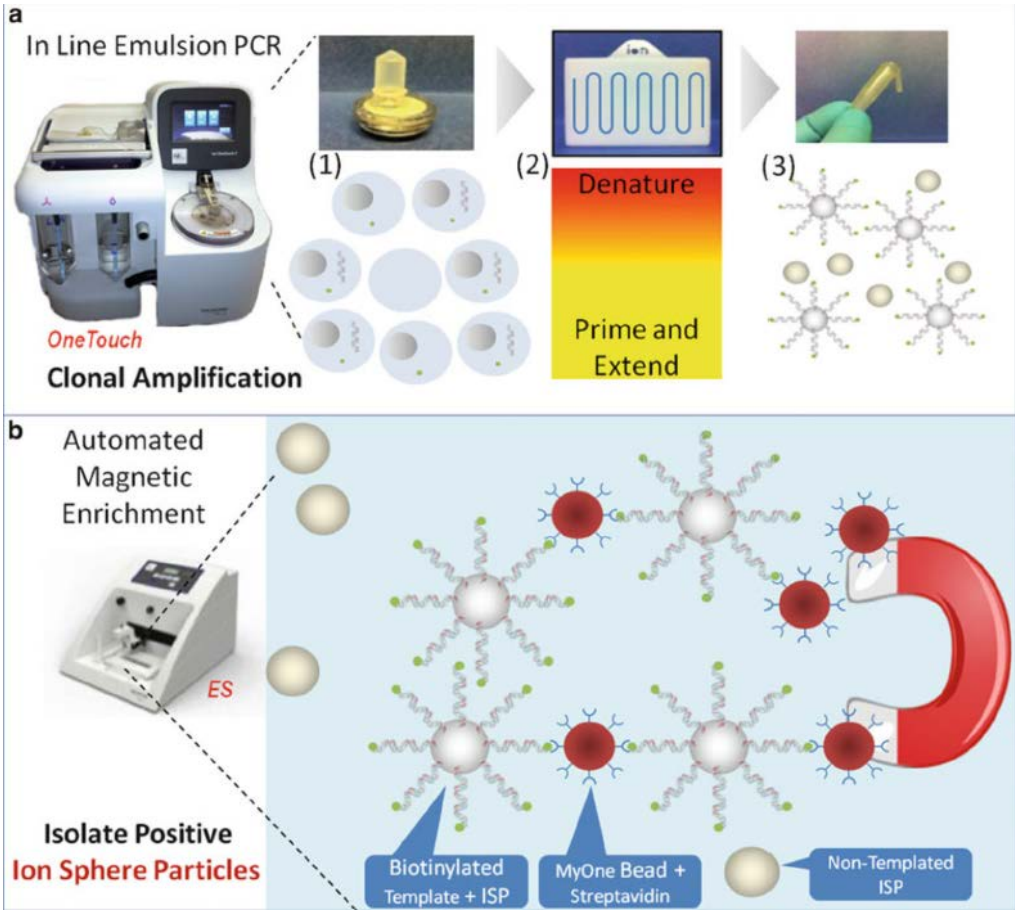


Figure 2.10 Automatic emulsion PCR technology. (a) The Ion OneTouch™ Instrument provides breakthrough technologies that automatically create templated Ion Sphere™ particles. First, millions of micro-size emulsion oil sphere beads are created, and a library DNA fragment attaches to the beads. Second, The fully integrated thermal cycler and disposable path amplification plate system enables robust thermal cycling of the templated Ion Sphere™ particles. Third, the templated Ion Sphere™ particles are recovered by high-speed centrifuge. The green dots appear as biotin that has been incorporated on the primer 5' -end of the template or DNA molecule during the emulsion PCR process. (b)The Ion OneTouch™ ES automated magnetic enrichment machine employs Streptavidin-linked C1 magnetic beads to bind to the biotinylated templated Ion Sphere™ particles that can be loaded directly onto the Ion semiconductor chip, delivering automated, highly reproducible enrichment with every run. This figure is reproduce from figure 4 "The schematic diagram illustrates key steps in the process using OneTouch™ instrumentation" (Kohn et al. 2013), with permission from Springer publisher.

An outstanding example application of whole genome sequencing by the Ion Torrent PGM sequencer is the identification of microbial pathogens. In May and June of 2011, an ongoing outbreak of exceptionally virulent Shiga-toxin- (Stx) producing *Escherichia coli* O104:H4 centered in Germany occurred, where more than 3000 people became infected. The whole genome sequencing on the Ion Torrent PGM and HiSeq 2000 sequencers helped the scientists to identify the type of *E. coli* and directly apply the information to deternine its antibiotic resistance. The strain appeared to be a hybrid of two *E. coli* strains—enteroaggregative *E. coli* and enterohemorrhagic *E. coli*—, which may help explain why it has been particularly pathogenic. Moreover, reference-guided draft assemblies of both strains were achieved within 62 hours using the newly presented PGM™ (Mellmann et al. 2011).

2.6 Research Objectives

The goal of this research is to understand the mechanism of cell-cell interaction, in this case bacterial flocculation, in *T. aminoaromatica* strain MZ1T and the factors regulating its expression. Elucidation of these factors will have important ramifications for environmental and engineered systems. Our central hypothesis is that flocculation in *Thauera sp*. MZ1T is mediated by 1) cross linking of the EPS with a surface expressed protein containing an EPS binding lectin domain. Additionally, we speculate that 2) secondary modification of the EPS may be involved. Finally, we hypothesize that EPS biosynthesis is regulated by entry into stationary phase. Thus, the Specific Aims of this research are to:

1. Identify and complement mutations leading to the floc$^-$ or floc-deficient phenotypes in MZ1T-20A and MZ1T-39A, respectively.

2. Determine the conditions influencing EPS production by Droplet digital PCR detection of the mRNAs of selected EPS genes.

Previously, two flocculation deficient mutants had been isolated, one of which still forms tiny, reduced flocs (39A mutant), while the other is completely devoid of floc formation ability (20A mutant) (Allen 2002). DNA has now been isolated from both mutant and sequenced using the Ion Torrent PGM semiconductor sequencer. In order to identify factors involved in flocculation, the genome sequence reads of the mutants were compared to the MZ1T wild-type genome from the NCBI database, and a mismatch sequence list was sorted and scored for potential relevant mutations. Next, possible mutant genes (with high coverage score and predicted to result in protein structural change) potentially relevant to the flocculation process were identified and both the mutant version of the gene along with the parent gene of wild type were amplified by PCR and sequenced using Sanger sequencing to confirm the mutation. The wild-type putative flocculation genes were amplified, cloned and used for complementation of the MZ1T floc$^-$ mutant strains using broad host range plasmids pRK415 and its derivatives and introduced by tri-parental mating. Finally, we further investigate conditions influencing expression of the complemented gene linked to EPS production and floc formation.

CHAPTER 3

MATERIALS AND METHODS

3.1 Bacterial Plasmids and Strains

Thauera aminoaromatica MZ1T wild type strain was originally isolated from the wastewater treatment plant of Eastman Chemical Company in Kingsport, Tennessee (Lajoie 2000). Two mutant stains, *Thauera* sp. MZ1T 39A and 20A, were previously created by chemical mutagenesis using *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG) from a spontaneeous *Thauera aminoaromatica* MZ1T rifampicin resistant mutant of the wild-type strain (Allen 2002). All bacterial strains and plasmids used in this study are listed in Table 3.1.

Table 3.1 Bacterial strains and plasmids used in this study

Stain and plasmid	Relevant Genotype/Characteristics	Source
MZ1T wild type	Floc^{+}, Rifampicin resistant (rif^{r}) MZ1T	Allen, et al. 2004
MZ1T 39A	Floc-reduced NTG mutant of MZ1T wild type	Allen, et al. 2004
MZ1T 20A	Floc^{-} NTG mutant of MZ1T wild type	Allen, et al. 2004
E. coli Top10	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR)	Invitrogen

Stain and plasmid	Relevant Genotype/Characteristics	Source
E. coli DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA*-*argF*)U169, hsdR17(r_K^- m_K^+), λ–	Invitrogen
pCR2.1	TA cloning PCR vector	Invitrogen
pRK2013	RK2-transfer gene-containing helper plasmid, tc^r	Figurski, et al. 1979
pRK415	Broad-host-range vector for Gram negative bacteria, tet^r	Keen, et al. 1988
pCR2.1: *tmz1t_0834*	TA cloning PCR vector harboring *tmz1t_0834* gene	This work
pCR2.1: *tmz1t_1376*	TA cloning PCR vector harboring *tmz1t_1376* gene	This work
pCR2.1: *tmz1t_1383*	TA cloning PCR vector harboring *tmz1t_1383* gene	This work
pCR2.1: tmz1t_1679	TA cloning PCR vector harboring *tmz1t_1679* gene	This work
pCR2.1: *tmz1t_2095*	TA cloning PCR vector harboring *tmz1t_2095* gene	This work
pCR2.1: *tmz1t_3143*	TA cloning PCR vector harboring *tmz1t_3143* gene	This work
pCR2.1: *tmz1t_3249*	TA cloning PCR vector harboring *tmz1t_3249* gene	This work
pCR2.1: *tmz1t_3637*	TA cloning PCR vector harboring *tmz1t_3637* gene	This work

Stain and plasmid	Relevant Genotype/Characteristics	Source
pCR2.1: *tmz1t_3801*	TA cloning PCR vector harboring *tmz1t_3801* gene	This work
pCR2.1: *tmz1t_3810*	TA cloning PCR vector harboring *tmz1t_3810* gene	This work
pRK415: *tmz1t_0834*	Broad-host-range vector harboring *tmz1t_0834* gene	This work
pRK415: *tmz1t_1376*	Broad-host-range vector harboring *tmz1t_1376* gene	This work
pRK415: *tmz1t_1383*	Broad-host-range vector harboring *tmz1t_1383* gene	This work
pRK415: *tmz1t_1679*	Broad-host-range vector harboring *tmz1t_1679* gene	This work
pRK415: *tmz1t_2095*	Broad-host-range vector harboring *tmz1t_2095* gene	This work
pRK415: *tmz1t_3143*	Broad-host-range vector harboring *tmz1t_3143* gene	This work
pRK415: *tmz1t_3249*	Broad-host-range vector harboring *tmz1t_3249* gene	This work

Stain and plasmid	Relevant Genotype/Characteristics	Source
pRK415: *tmz1t_3637*	Broad-host-range vector harboring *tmz1t_3637* gene	This work
pRK415: *tmz1t_3801*	Broad-host-range vector harboring *tmz1t_3801* gene	This work
pRK415: *tmz1t_3810*	Broad-host-range vector harboring *tmz1t_3810* gene	This work

3.2 Culture Conditions and Storage

All *Thauera* MZ1T strains were grown at 30°C in either Stokes' medium (Atlas 2005) or *Thauera* Defined Medium (TDM) (Rabus and Widdel 1995). All *E. coli* strains were grown at 37°C except in mating experiments with MZ1T, which were cultured at 30°C. Liquid cultures were grown shaking in 250 mL flasks at 200 r.p.m. Freezer stocks were prepared by adding 0.5mL samples from an overnight culture (48 h. culture for MZ1T) with appropriate antibiotics to sterilized tubes containing 0.5 mL sterile 50% (v/v) glycerol. Tubes were maintained frozen at –80°C until ready for use.

3.3 Media and Chemicals

E. coli strains were routinely grown in autoclave sterilized lysogeny (Luria-Bertani, LB) broth (10.0 g tryptone, 5.0 g yeast extract, 10.0 g NaCl per liter, pH 7.0). Filter-sterilized (0.2

μm) antibiotics were aseptically added to LB broth as necessary. Stokes's broth was used for the routine cultivation of MZ1T strains. One liter of Stokes's medium included: 5 g polypeptone, 0.2 g $MgSO_4 \cdot 7H_2O$, 0.15 g $Fe(NH_4)(SO_4)$, 0.1g sodium citrate, 0.05 g $CaCl_2$, 0.05 $MnSO_4$, 0.01 $FeCl_3 \cdot 6H_2O$. The pH of the medium was then adjusted to 7.2 before sterilized by autoclaving at 121°C and 15 p.s.i. for 20 minutes. After the solutions cooled below 50°C, the filter-sterilized vitamin solutions were added to yield final concentrations as indicated: cyanocobalamin, 0.5 mg/L; thiamine hydrochloride, 0.4 mg/L; and biotin, 0.4 mg/L. All vitamin stock solutions were stored in the dark at 4°C

Thauera defined medium (TDM) was applied for the cultivation of MZ1T and mutant strains for all EPS isolation experiments. TDM basal medium contained: 0.3 g NH_4Cl, 0.5 g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7H_2O$, 0.1 g $CaCl_2 \cdot 2H2O$, and 5 g of sodium succinate. This solution was autoclaved at 121°C for 20 min. at 15 p.s.i. and then allowed to cool to below 50°C before the following filter-sterilized stock solutions were aseptically added: 1 mL Trace Element Solution, 1 mL Tungsten solution, 1 mL Vitamin solution, 1 mL Thiamine solution, 1 mL Cyanocobalamin solution, and 30 mL of sodium bicarbonate solution.

Trace Element Solution was composed of: 2.1 g $FeSO_4 \cdot 7H_2O$, 5.2 g Na_2EDTA, 30 mg H_3BO_4, 100 mg $MnCl_2 \cdot 4H_2O$, 190 mg $CoCl_2 \cdot 6H_2O$, 24 mg $NiCl_2 \cdot 6H_2O$, 25 mg $CuSO_4 \cdot 5H_2O$, 144 mg $ZnSO_4 \cdot 7H_2O$, 36 mg $Na_2MoO_4 \cdot 2H_2O$ and 1 L deionized water. The pH was adjusted to 6.0-6.5 and the solution was sterilized in an autoclave as described.

The Tungsten solution contained 200 mg NaOH and 6 mg $Na_2WO_4 \cdot 2H_2O$ in 1 L of deionized water. The solution was sterilized in the autoclave as described.

The Vitamin Solution contained 200 mL 20 mM sodium phosphate buffer(pH 7.1), 8 mg *p*-aminobenzoic acid, 2 mg D-biotin, 20 mg nicotinic acid, 10 mg calcium D-pantothenate, and 30 mg pyridoxin hydrochloride. This solution was filter-sterilized through a 0.2 µm filter and stored in the dark at 4°C

The Thiamine solution was made by adding 20 mg thiamine hydrochloride to 200 mL of sodium phosphate buffer and adjusts pH to 3.4. This solution was filter-sterilized through a 0.2 µm filter and stored in the dark at 4°C.

The Cyanocobalamin (B_{12}) solution contained 5 mg cyanocobalamin in 100 mL of deionized water. This solution was filter-sterilized through a 0.2 µm filter and stored in the dark at 4°C.

The Sodium Bicarbonate Solution contained 84 g of sodium bicarbonate in1 L of deionized water. This solution was filter-sterilized through a 0.2 µm filter and stored at room temperature in a sealed container.

The phosphate-buffered saline (PBS) was prepared by dissolving 8 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4, and 0.24 g of KH_2PO_4 in 800 ml of H_2O. pH was adjusted to 7.0 with HCl and then H_2O added to 1 L. The solution was dispensed into aliquots and sterilized by autoclaving for 15 min at 15 psi. PBS stored at room temperature.

All solid medium agar plates were made by the addition of 15 gL^{-1} of agar prior to autoclave sterilization

3.4 DNA Manipulation Techniques

One milliliter of liquid cultures of MZ1T were inoculated into 9 ml of Stokes's medium overnight at 30^{o}C, and the cells were collected by centrifuged at 5000 X g 4^{o}C for 5 min. Genomic DNA was extracted using a genomic DNA extraction kit from MP Biomedical manufacture. Subsequently, the extracts were treated and purified with RNase I and DNA clean up kit (MO BIO). The quality and the quantity of the genomic DNA were determined by gel electrophoresis and spectrophotometery, (NanoDrop Technologies) respectively. PCR products were purified by QIAquick PCR Purification Kit (QIAGEN). Plasmid mini-preps (5 PRIME FastPlasmid Mini Kit, Fisher Scientific) were used for plasmid DNA isolation. DNA restriction/modification reactions were set up using appropriated buffers and enzymes according to their manufactures, and the final concentration of glycerol in the reaction was kept less than 5% to minimize the possibility of star activity. *E. coli* transformations were carried out using commercial competent cells (Invitrogen), following manufacturer's instruction.

3.5 Next Generation Sequencing

3.5.1 Ion Library and Template Preparation

Library preparation was performed using the NEBNext Fast DNA Fragmentation and Library Prep set for Ion Torrent (New England BioLabs) following the manufacturer's instructions. Briefly, 1 μg of 20A and 39A MZ1T mutants' genomic DNA was enzymatically fragmented and end repaired. Each library fragment was barcoded and ligated to P1 Adapters, enabling subsequent amplification. For optimal sequencing results, a DNA library with a mean

size of 290-330 bp was purified and size-selected using AMPure XP Beads (Beckman Coulter). Products quality and quantity were confirmed and measured by Agilent 2100 Bioanalyzer system (Agilent technologies). Molar equivalents of both 20A and 39A MZ1T DNA libraries were calculated, pooled together and diluted to 8.3 nM (5×10^9 molecules/µL). The sample was amplified by emulsion PCR using the Ion Torrent OneTouch System, following the manufacturer's instructions. The resultant beads were subsequently purified and enriched on the Ion ES automated machine.

3.5.2 Ion Sequencing

The Ion Sequencing Kit was used for sequencing per the manufacturer's instructions. First, the PGM sequencer was cleaned, initialized, and pH calibrated. Enriched templates Ion Sphere Particles (ISP) from above step was annealed to sequencing primer at 95ºC for 2 min and 37ºC for 2 min in a thermal cycler. Subsequently, 3 µl of the sequencing polymerase was added, and the reaction is incubated for 5 min at room temperature. While the enriched template ISPs was being prepared for sequencing, a new Ion v316 chip was tested (chip check) on the PGM sequencer. Then, the ion chip was loaded with the ISPs according to the manufacturer's protocol. Finally, the chip was put on the Ion PGM system and the run was performed.

3.6 Bioinformatics Analysis

After the run was finished (2 h), the Ion data was retrieved and processed. Reads were mapped to the *Thauera sp*. MZ1T reference genome retrieved from NCBI database (RefSeq assembly accession: GCF_000021765.1) using NextGENe software (Softgenetics, State College,

PA); single nucleotide polymorphisms (SNPs) and short insertion/deletions (indels) were called from the consensus sequence, and coverage scores were also calculated

3.7 SNP Validation by Sanger Sequencing

A candidate gene with a single nucleotide polymorphism (SNP) detected by Ion torrent was PCR amplified from their mutant genomes and cloned into pCR2.1 plasmid using the Topo TA cloning kit (Invitrogen). pCR 2.1 plasmids harboring the mutant candidate gene were transformed into Top10 *E. coli* cells (Invitrogen) using heat-shock procedure. Briefly, 2 µL of the pCR 2.1 plasmid was added into the competent cells tube thawed on ice and mixed gently. The competent cell/DNA mixture tube was placed on ice for 30 min. The tube was put into a 42°C water bath for 45 seconds and immediately put it back on ice for 2 min. 250 µL of SOC media was then added into the tube and incubated in 37°C shaking incubator for an hour. Transformants were plated on LB agar plates containing 50 mg/ml kanamycin which were overlaid with 40 µL of 40-mg/mL X-Gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside) to facilitate blue/white screening of colonies. Several of the white colonies generated were inoculated into 2 mL liquid LB medium containing kanamycin. Overnight cultures were pelleted by centrifugation at 10,000 X *g* for 1 min and subjected to plasmid minipreps (5 PRIME FastPlasmid Mini Kit, Fisher Scientific). The resultant plasmid (150ng/uL) was sent to MWG for Sanger sequencing (Huntsville, AL)

3.8 Mutant Complementation

Intact candidate genes ware amplified from the genomic DNA of MZ1T wild type by

PCR. PCR amplification of the full-length genes from the genomic DNA involved 30 cycles of PCR amplification with Q5 proofreading polymerase (NEB) followed by the addition of 2.5 U Taq polymerase and incubation at 72°C for 10 min to add a 3'-A overhangs in order to facilitate cloning into the pCR2.1 TA cloning vector (Invitrogen). The cloned fragments were then transformed into Top10 *E. coli* cells (Invitrogen) as previously described. The resultant plasmid was cut with XbaI and HindIII enzymes (NEB) in a double digest at 37°C for 90 min. The broad host range plasmid pRK415 was also digested in the same manner. The gene fragment purified by gel electrophoresis was ligated into the pRK415 vector using T4 DNA ligase (NEB) per the manufacturer's instructions to yield pRK415: *tmz1t_*<gene locus> (see table 3.1 for specific constructs). Transformation of pRK415 into chemically competent Top10 *E. coli* cells (Invitrogen) was achieved as previously described, and the resultant transformants were screened on LB agar plates containing 15 mg/mL tetracycline.

Transconjugation of pRK415 expression vector and its derivatives into the appropriate MZ1T mutant strains was performed via triparental mating. Briefly, 10–mL cultures of *E. coli* harboring pRK415: *tmz1t_*<gene locus> (see table 3.1 for specific constructs) (donor), MZ1T mutants (recipient), and *E. coli* harboring pRK2013 (helper strains) were grown to an approximate optical density at λ = 600 nm (OD_{600}) of 0.5 in Stokes's broth with 100-µg/mL rifampicin for MZ1T or LB with suitable antibiotics for *E. coli* strains. Cells were collected and the supernatant fraction was decanted. The cell pellets were suspended in an equal volume of phosphate buffered saline (PBS, pH 7.0). This washing was repeated twice and on the third time, the donor, recipient, and helper strains were combined into a single tube and pelleted together by centrifugation at 8000 X g for 1 min The supernatant fraction was decanted, and

the pellet was suspended in 1 mL PBS. Aliquots containing 100 µL of cell suspension were applied to autoclave-sterilized 25 mm, 0.2 µm filter disks on non-selective Stokes's agar plates and incubated at 30°C for 48 h Following incubation, filter disks were aseptically removed and transferred to TDM minimal medium containing 15 g/mL tetracycline and 100-µg/mL rifampicin. Successful transconjugants were screened for flocculation to ensure that proper phenotype had been successfully rescued. All MZ1T strains were similarly transformed with the empty pRK415 as a negative control.

3.9 EPS Purification

1 liter cultures of MZ1T grown in TDM and shaken at 30°C for seven days were centrifuged for 15 min at 8000 X *g* to pellet the cells. The supernatant fraction was decanted through a nylon mesh cloth to remove any cellular debris, and then concentrated to one-tenth its original volume (100 mL) using an automated tangential flow filtration pilot unit equipped with Pellicon 2 Mini Holder filter (Millipore). Membranes were of polyethersulfone having a molecular weight cut-off (MWCO) of 100 kDa. EPS solution was first introduced in a tank then pumped into the membrane. Permeate fraction was collected and retentate fraction was returned to the feeding tank. After removal of the retentate, the resulting desalted polysaccharide solution was then frozen at -80°C and lyophilized.

Following purification, the membrane underwent a cleaning sequence. It was first mechanically washed by two water flushes at fixed pressure (4 bars). Then, 0.1 M solution of NaOH was used in closed circuit for 1 h still at pressure equal to 4 bars. Finally, the membrane was flushed with deionized water until a neutral pH was achieved. After experiment, the membrane was stored in a 2.5 mM sodium hypochlorite solution at 4ºC (Allen et al. 2004).

3.10 Quantification of EPS

The total sugar quantification was performed by modified Phenol-Sulfuric Acid (PSA) method in microplate format (Masuko et al. 2005). The PSA method involved adding 50 µL of resuspended EPS and 150 µL of concentrated sulfuric acid to a 96-well microplate. Then, the mixture was rapidly shaken and immediately added 30 µL of 5% (w/v) phenol in water. The microplate was heated for 5 min at 90ºC in a static water bath. After cooling to room temperature for 5 min in another water bath, the microplate was wiped dry and $A_{490\ nm}$ was measured by microplate reader using glucose as a standard.

3.11 FTIR Analysis

For FT-IR measurements 100 µL of the desalted crude EPS (10 mg/mL) was dried and clamped against the ATR crystal (Germanium). The absorption spectrum between 750 and 4,000 cm^{-1} was measured by co-adding 100 scans and subtracting both the background and atmospheric water. Spectra were recorded using attenuated total reflectance on a Bio-Rad FTS 6000 FT-IR spectrometer.

3.12 Deacetylation of MZ1T EPS

Twenty mg of MZ1T crude EPS powder was added into 45% NaOH to react for 15 min at 100°C. After cooling and addition of absolute ethanol (1:1), the mixture was centrifuged at 10,000 X *g* for 10 min at 4°C. The obtained sediment was dried and determined by fourier transform-infrared spectrophotometer as described above.

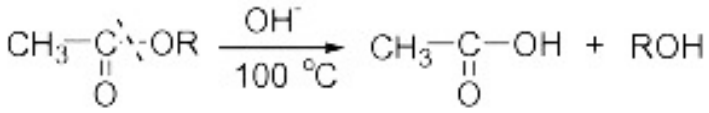


Figure 3.1 Chemical deacetylation reaction occurs in high basic concentration at 100°C

3.13 Gene Expression Analysis

3.13.1 RNA Extraction and cDNA Synthesis

Extraction of total RNA was performed from 18, 24, 36, 48, 72 and 96 h MZ1T wild-type cultures using FastRNA pro kit (MPbio) following the manufacturer's instructions. To remove contaminating genomic DNA, RNA samples were treated using Qiagen's RNeasy on-column DNase I (2.7 U DNase I / 10 µg RNA), followed by Qiagen RNeasy MinElute (for DNase I removal) according to the manufacturer's protocol, before proceeding with cDNA synthesis. The RNA concentration was determined using a Nanodrop ND-1000 instrument (Nanodrop Technologies, Wilmington, DE). cDNA was synthesized from extracted RNA using the iScript™ cDNA synthesis kit (Bio-Rad). Approximately 10 µg of RNA was used for cDNA synthesis in a 20 µL volume. For cDNA synthesis, the following reaction was prepared:

Total RNA:	10 µL
Random primers (3 µg/ µL):	2 µL
RNase-free ddH_2O:	3 µL

Once prepared, the solution was heated to 70°C for 10 min then placed on ice while the following components were added:

5x reverse transcription buffer: 4 µL

Reverse transcriptase: 1 µL

The reaction, totaling 20 µL in volume was then incubated at room temperature for 10 min, followed by incubation at 42°C for 2 h. The reaction was terminated by heating the samples to 65°C for 7 min. Once complete, the samples were diluted with 10X RNase-free ddH_2O and stored at -80°C for later use.

3.13.2 Droplet Digital PCR

Droplet Digital PCR builds on the workflow of quantitative real-time PCR (qPCR), wherein the nucleic acid sample along with primer and/or probe sets are added to a PCR master mix. However, in digital PCR, the sample is first partitioned into hundreds to millions of individual reaction vesicles prior to thermal cycling. In this technique, 8 x 20 µL reaction mixtures are simultaneously divided into tens of thousands of surfactant stabilized droplets using a disposable microfluidic cartridge and a vacuum source (Droplet Generator, Bio-Rad). The resulting droplets are then transferred into a 96-well plate and thermally cycled. Following end-point amplification, the 96-well plate is loaded into a QX-200 Droplet Reader that automatically aspirates the emulsion from each well and assigns droplets as being positive (containing template) or negative (no template) based on the fluorescence emission. Therefore, this methodology removes both the reliance on rate-based measurements (CT values) and the need for the use of calibration curves.

ddPCR mixture was assembled as followed:

cDNA sample	1 µL
primers	2 µL
Bio-Rad ddPCR supermix	10 µL
ddH_2O	7 µL

A 20 µL aliquot was taken from each of the assembled ddPCR mixtures containing primers for MZ1T EPS biosynthesis (*tmz1t* 3801), EPS deacetylase (*tmz1t* 3249) and GAPDH gene transcript measurement and pipetted into each sample well of an eight-channel disposable droplet generator cartridge (Bio-Rad, Hercules, CA, USA). A 70 µL volume of Droplet Generation Oil for EvaGreen (Bio-Rad) was then loaded into each of the eight oil wells. The cartridge was placed into the droplet generator (Bio-Rad) where a vacuum was applied to the outlet wells to simultaneously partition each 20 µL sample into nanoliter-sized droplets. After droplet generation, 40 µL of the generated droplet emulsion was transferred to a new 96-well PCR plate (Eppendorf) and amplified in a C100™ Thermal Cycler (Bio-Rad). The amplification conditions were 10 min DNA polymerase activation at 95°C, followed by 40 cycles of a two-step thermal profile of 30 s at 94°C for denaturation, and 60 s at 60°C for annealing and extension, followed by a final hold of 10 min at 98°C for droplet stabilization and cooling to 4°C. The temperature ramp rate was set to 2.5°C/s, and the lid was heated to 105°C, according to the Bio-Rad recommendations. After the thermal cycling, the plates were transferred to a droplet reader (QX-200, Bio-Rad), and the droplets were streamed in single file on the QX-200 reader, which counted the fluorescent positive and negative droplets to calculate the gene transcripts.

The software package provided with the ddPCR system (QuantaSoft 1.3.2.0, Bio-Rad) was used for data acquisition. Annealing temperatures were experimentally optimized using a temperature gradient of the thermal cycle to avoid non-specific products and primer-dimers. The gene transcript was quantified in triplicate for each cDNA sample.

CHAPTER 4

RESULTS

4.1 Floc-Defective Mutants MZ1T 39A and 20A Genome Sequencing

MZ1T floc-defective stains, 39A and 20A, were generated previously by NTG chemical mutagenesis of the MZ1T wild type (Allen 2002). Flocculation in the 39A mutant is reduced to very low amounts of cell-aggregation at mucher later time than the wild type while the rest of the culture broth stays turbid, whereas MZ1T 20A mutant completely lacks a flocculation phenotype. However, both floc-deficient mutants produced near-wild type quantities of extractable EPS and were found to have the same glycosyl composition of EPS previously detected in extracts from floc^{+} MZ1T (Allen 2002). These findings suggest that flocculation in MZ1T is complex, and there may be more than one factor responsible for the floc formation, which may not be directly related to EPS production. Therefore, in order to understand the mechanism of flocculation and uncover genes involved in this process in MZ1T, high throughput sequencing was used to obtain insight into the genetics of the altered phenotype in these MZ1T mutants relative to the wild type.

In this study, we used a semiconductor-sequencing platform, Ion Torrent, to sequence both mutants. High loading density (76%) of Ion Torrent chips (v316 chip) was accomplished (Figure 4.1) and provided several hundred thousand (134 bp average length) reads (Figure 4.2). The total number of bases was 64.21 (39A) and 102.20 (20A) million (Table 4.1) with average genomic coverages of 15-and 24-fold respectively.

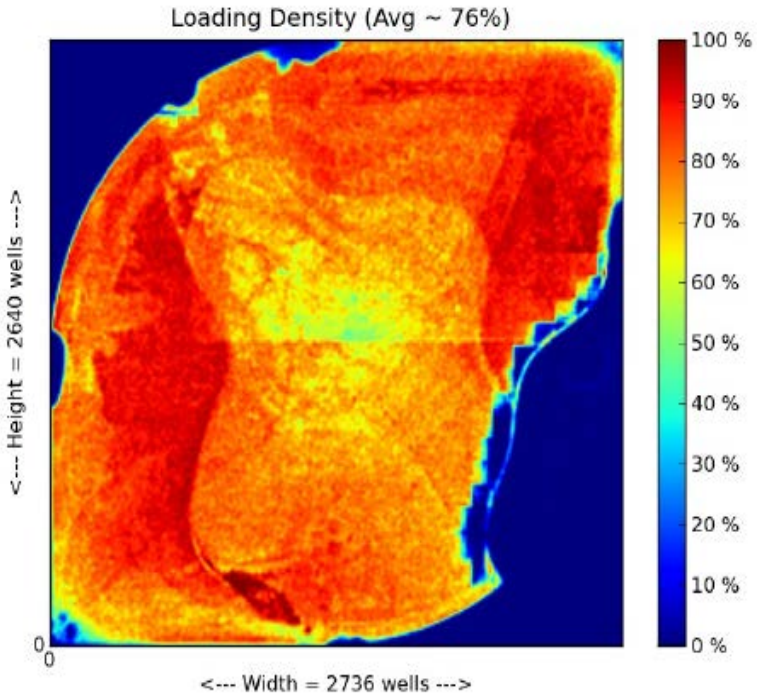
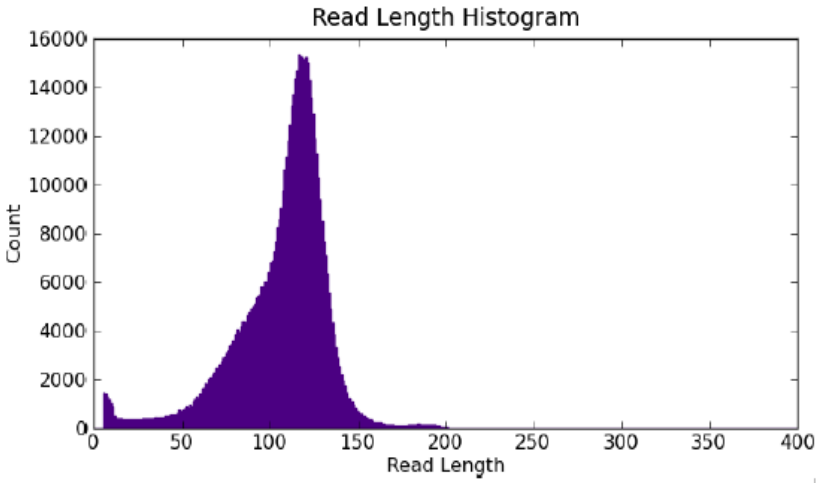
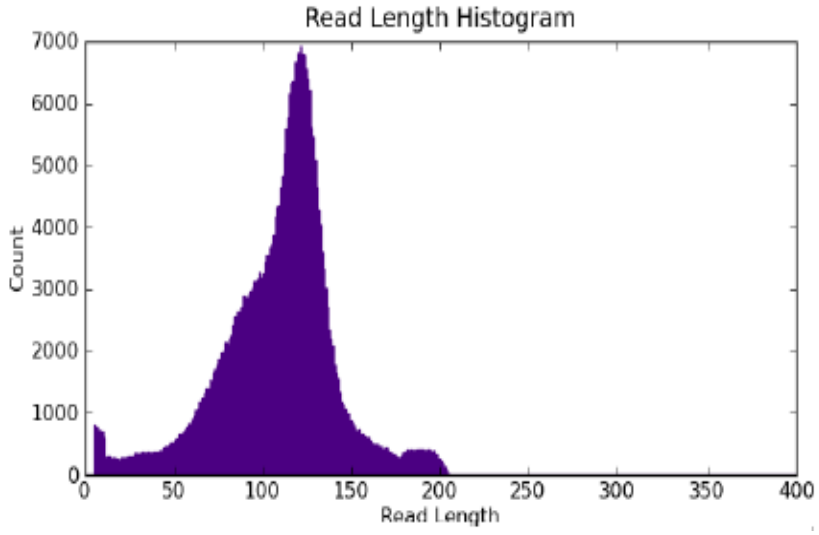


Figure 4.1. Ion torrent chip loading density of MZ1T 39A and 20A mutants



39A



20A

Figure 4.2. Read length histogram of MZ1T 39A and 20A mutant genomic libraries.

Table 4.1 Ion Torrent read data of MZ1T 39A and 20A mutant strains

	MZ1T 39A	MZ1T 20A
Total number of bases (Mbp)	64.21	102.20
Total number of reads	405,962	762,587
Mean length (bp)	134	134

Mean read quality across all the datasets was 32.88 (Figure 4.3). Each quality score, q, generated by the PGM base-caller is Phred-based, where $q = - 10 \times \log_{10}(p_{error})$. A quality score was assigned to each base using a pre-computed quality lookup-table distributed with each version of the PGM software (Life Sciences Technical Note Version 2.20). The reads were compared to the *Thauera aminoaromatica* MZ1T reference genome (RefSeq assembly accession: GCF_000021765.1), and variants were identified via analysis of the mapped reads. The coverage threshold was applied because random verification of several putative mutations as well as coverage and variant frequency analysis indicated that regions with low sequence coverage did not allow for reliable detection of variants. These regions are probably difficult to sequence because of repetitive sequences or the formation of DNA secondary structures (Figure 4.4).

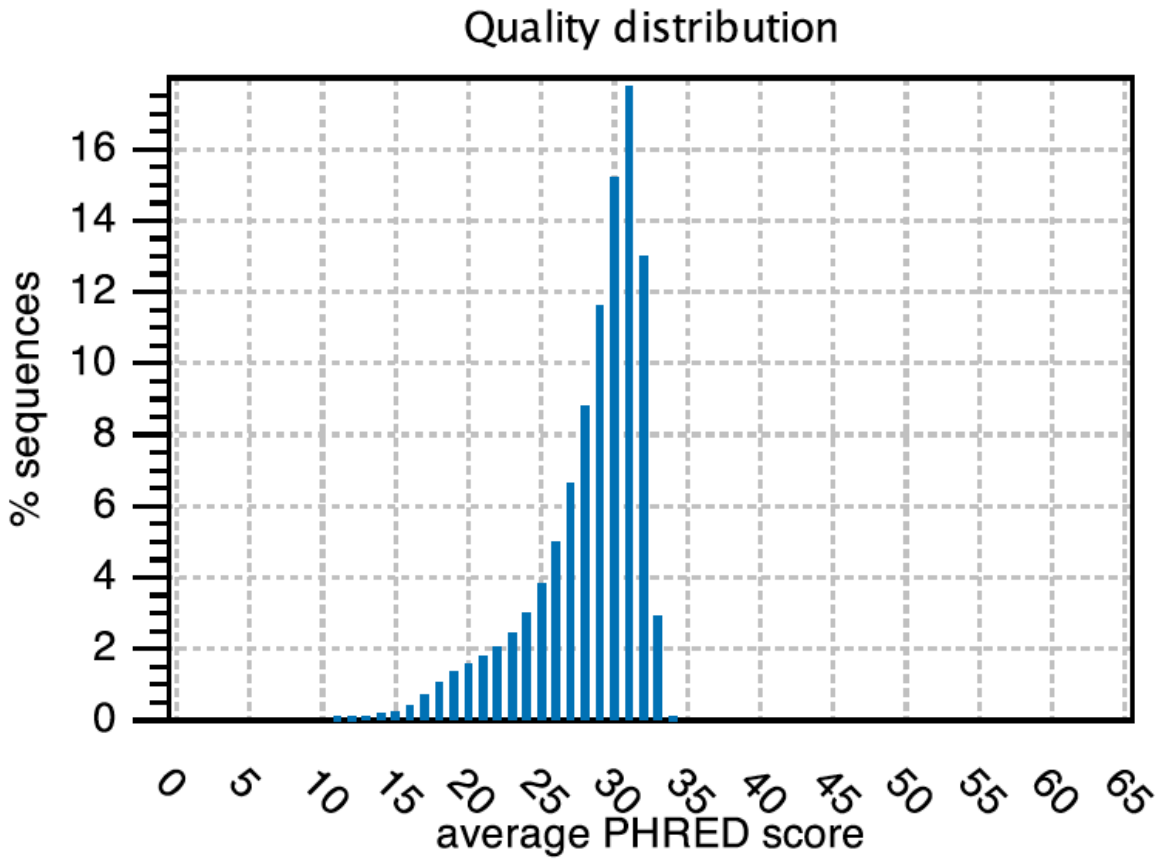
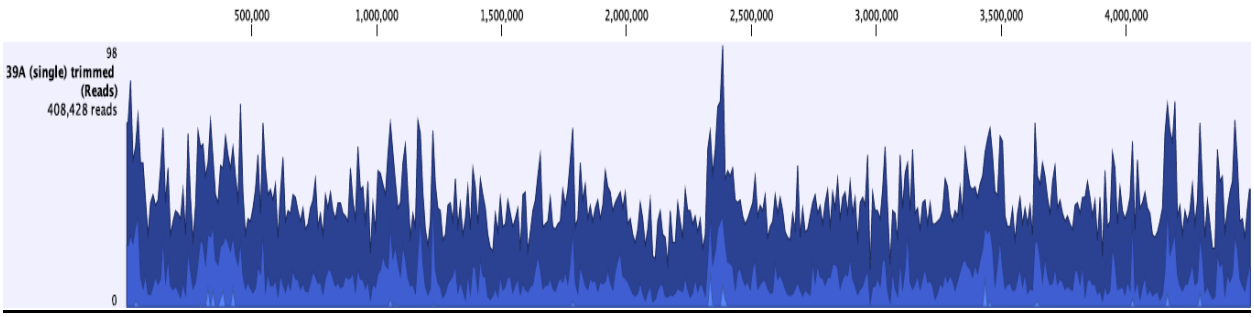


Figure 4.3 Distribution of average sequence quality scores for MZ1T mutants genome sequences. The quality of a sequence is calculated as the arithmetic mean of its base qualities.

39A



20A

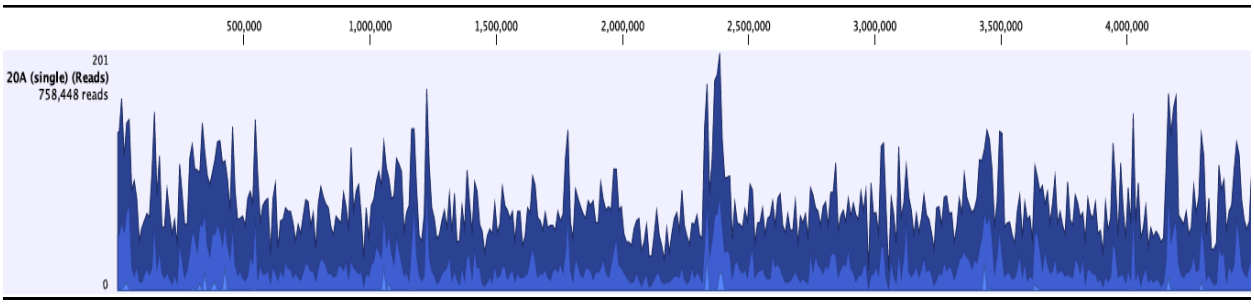


Figure 4.4 Sequence coverage maps of MZ1T 39A and 20A mutants genome sequences.

For most applications, 10–15-fold redundant genome coverage will allow for accurate and cost-effective mutational profiling (Smith et al. 2008). In addition, the SNPs were filtered based on SNP percentages (SNP%). SNPs with a SNP% of ≥75% (for example, number of SNP = 3 of read depth = 4) were chosen. The 75% cutoff for SNP selection was set by considering potential sequencing errors that can be generated by the massively parallel sequencing method. Subsequently, a total of 75 putative point mutations, insertions, and deletions were mapped out from the wild-type genome reference (Table 4.2). Most of the SNPs showed a strong bias to occur at GC sites (GC to AT), which is consistent with the known mutagenic specificity of NTG, and with previous reports on NTG chemical mutagenesis (Ohnishi et al. 2008).

Interestingly, a mutation of the *rhoB* in both MZ1T mutants was identified, caused by a transversion (T to A) leading to an amino acid substitution (glutamine is replaced by leucine) (Table 4.2). The encoded protein is the target of the antibiotic rifampicin. Its identification was suggestive of a successful experimental design, since MZ1T 39A and 20A were derived from a same spontaneous rif^R mutant of the wild type. Therefore, using the phenotype sequencing method by Ion Torrent sequencing proposed here should allow one to discover a genetic mutation responsible for the floc-defective genotype.

Table 4.2 SNP(s) table report; the table shows loci with %SNP > 75% and Coverage > 10X. The percentage of the single most prevalent non-reference base in the aligned column is also shown. Highlighted genes were validated by Sanger sequencing and used for complementation of MZ1T mutants in order to determine genes involved in flocculation.

* Amino acid change predicted to result in mild property change

** Amino acid change predicted to result in severe property change

20A

% SNP	Gene	Gene function	Type of mutation	Base change	Amino acid change/ frame shift position	Coverage
75%	*Tmz1t 0236*	MreB	Frame shift	G/-	G284A	12X
75%	*Tmz1t 0322*	Acriflavin resistance	Frame shift	G/-	A790R**	16X
85%	*Tmz1t 0325*	Glycosyltransferase 28 domain containing protein	Frame shift homopolymer	G/-	G859A*	58X
77%	*Tmz1t 0349*	Phosphate acetyltransferase	Frame shift	G/-	A133R**	44X
83%	*Tmz1t 0432*	TRAP transporter	Frame shift	G/-	G359A*	12X
100%	*Tmz1t 0458*	H (+) transporting two sector ATPase	Non synonymous	G to A	G162D**	61X
100%	*Tmz1t 0472*	ABC transporter	Non synonymous	A to C	C452G*	23X

90%	*Tmz1t 0581*	ABC transporter	Non synonymous	G to A	A272T*	11X
100%	*Tmz1t 0695*	C32 tRNA thiolase	Non synonymous	G to A	T52I*	39X
100%	*Tmz1T 0733*	Acyl-CoA dehydrogenase	Non synonymous	G to A	E8K	29X
100%	*Tmz1t 0811*	Na+ trans locating NADH quinone reductase	Non synonymous	G to A	R119H*	17X
95%	*Tmz1t 0826*	UbiD family decarboxylase	Non synonymous	G to A	R414H*	42X
100%	*Tmz1t 1127*	Acetyl-transferase protein	Non synonymous	A to G	Y9C	22X
83%	*Tmz1t 1130*	Glutamine scyllo-inositol transaminase	Frame shift homopolymer	G/-	G342A*	12X
100%	*Tmz1t 1340*	Uracil-xanthine permease	Non synonymous	G to C	H65D*	20X
78%	*Tmz1t 1369*	GAF sensor signal transduction histidine kinase	Frame shift	C/-	G281A*	18X
85%	*Tmz1t 1548*	UBA/THIF NAD/FAD binding protein	Frame shift	G/-	A207R**	13X
92%	*mogA*		Frame shift	G/-	A184P**	13X
80%	*Tmz1t 1666*	Cyclic nucleotide binding protein	Frame shift	C/-	G130A*	10X
100%	*Tmz1t 1689*	Pseudouridine synthase	Non synonymous	G to A	V297M	10X

75%	*Tmz1t 1710*	Von Willebrand factor A	Frame shift homopolymer	G/-	G248A*	20X
93%	*Tmz1T 1718*	CoA binding protein	Non synonymous	G to A	L36F	15X
100%	*Tmz1t 1755*	NADH dehydrogenase subunit	Non synonymous	C to A	F350L	63X
100%	*Tmz1t 1907*	PAS/PAC sensor containing diquanylate cyclase	Non synonymous	G to A	L221F	15X
78%	*Tmz1t 1994*	Oligoribonuclease	Frame shift	G/-	A43R**	27X
87%	*Tmz1t 2115*	Peptidase M15A	Frame shift homopolymer	-/G -/G	L3P** L20P**	61X 78X
75%	*Tmz1t 2128*	MerR family transcriptional regulator	Non synonymous	G to C	A12P**	12X
77%	*Tmz1t 2129*	Von Willebrand factor type A	Frame shift	C/-	P259R*	13X
75%	*Tmz1t 2217*	Radical SAM protein	Frame shift	G/-	W217C*	12X
92%	*Tmz1t 2242*	Arylesterase	Frame shift	G/-	G155A	12X
100%	*Tmz1t 2560*	Polar amino acid transporter permease	Non synonymous	A to G	V162A*	28X
100%	*Tmz1t 2653*	Hydrolyase Fe-S type tatrate/fumarate subunit alpha	Non synonymous	G to A	E471K	30X

100%	*Tmz1t 2773*	Acetolactate synthase large subunit	Non synonymous	C to T	T140M*	18X
100%	*Tmz1t 2817*	Methanol/ethanol family PQQ dependent dehydrogenase	Non synonymous	C to A	W570C*	11X
98%	*Tmz1t 2951*	Benzoyl CoA reductase subunit C	Non synonymous	G to T	L156I	41X
100%	*Tmz1t 2967*	Phenyl lactate CoA ligase	Non synonymous	G to A	D350N*	13X
100%	*Tmz1t 2994*	Rnf electron transport subunit E	Non synonymous	G to C A to C	M41I I69F	13X 28X
100%	*thrS*	Threonyl-tRNA synthetase	Non synonymous	G to T	R426S	24X
79%	*Tmz1t 3108*	GntR family transcriptional regulator	Frame shift homopolymer	C/-	L69W	19X
77%	*Tmz1t 3143*	Diguanylate cyclase	Frame shift	C/- C/-	R322V** L508W*	13X 11X
100%	*Tmz1t 3207*	Glycine dehydrogenase	Non synonymous	T to C	V99A*	15X
100%	*Tmz1t 3235*	Rnf electron transport subunit D	Non synonymous	G to A	G336D*	28X
100%	*Tmz1t 3249*	Polysaccharide deacetylase	Non sense	G to A	W111 stop	12X
82%	*Tmz1t 3278*	Secretion ATPase	Frame shift	C/-	G249A	17X

100%	*rhoB*	DNA polymerase	Non synonymous	T to A	Q538L**	21X
93%	*Tmz1t 3358*	L-carnitine dehydratase/bile acid induce protein	Non synonymous	G to A	R78W**	27X
100%	*Tmz1t 3486*	Methy-malonyl-CoA mutase	Non synonymous	G to A	A501V*	14X
100%	*Tmz1t 3609*	Gramma-Glutamyltransferase	Non synonymous	G to A	V152M	10X
90%	*Tmz1t 3637*	Family 2 glycosyl transferase	Frame shift	G/-	V162A*	15X
98%	*Tmz1t 3709*	Molybdopterin oxidoreductase	Non synonymous	G to A	A596V*	48X
93%	*Tmz1t 3801*	Polysaccharide biosynthesis CapD	Non synonymous	T to G	L102R**	29X
93%	*Tmz1t 3812*	Lipid A ABC exporter	Non synonymous	G to A	A259V*	27X
88%	*Tmz1t 3840*	S-adenosyl methionine synthetase	Frame shift homopolymer	G/-	G118A	16X
100%	*Tmz1t* 3847	Hpt sensor hybrid histidine kinase	Non synonymous	G to A	R426H*	20X

39A

% SNP	Gene	Gene function	Type of mutation	Base change	Amino acid change/ frame shift position	Coverage
100%	*Tmz1t 0300*	pglZ domain containing protein	Non synonymous	G to A G to A	G454S V646M	32X 34X
80%	*Tmz1t 0312*	Phage/plasmid protein	Frame shift homopolymer	C/-	H312T	10X
76%	*Tmz1t 0325*	Glycosyltransferase 28 domain containing protein	Frame shift	G/-	G859A	29X
100%	*Tmz1t 0472*	ABC transporter	Non synonymous	A to C	C452G*	17X
100%	*Tmz1t 0834*	Type IV pilus assembly protein	Non synonymous	G to A	A36T*	17X
80%	*Tmz1t 0850*	FAD-dependent oxidoreductase	Frame shift	G/-	A343R*	10X
100%	*hisG*		Non synonymous	C to T	A7V*	18X
75%	*Tmz1t 0936*	KAP P-loop domain containing protein	Frame shift homopolymer	C/-	L350S*	20X
90%	*Tmz1t 1084*	Porin	Non synonymous	C to T	T66I*	19X
94%	*Tmz1t 1315*	Response regulator receiver protein	Non synonymous	G to A	G35D**	17X

100%	*Tmz1t 1340*	Xanthine permease	Non synonymous	G to C	H65D*	12X
100%	*Tmz1t 1376*	ABC transporter	Non synonymous	G to A	G170R**	19X
70%	*Tmz1t_1383*	Von willebrand factor A	Non synonymous	G to A	V162A*	17X
100%	*Tmz1t 1385*	ATPase AAA	Non synonymous	G to A	P128L**	24X
100%	*Tmz1t 1679*	PAS/PAC sensor containing diguanylate cyclase/phosphodiesterase	Non synonymous	G to A	T615I*	14X
83%	*Tmz1t 1766*	Recombination factor protein Rar A	Frame shift	G/-	G284A	12X
88%	*Tmz1t 2091*	Cointegrate resolution protein T	Non synonymous	C to T	A104T*	81X
96%	*Tmz1t 3184*	TnsA endonuclease	Non synonymous	G to A	E632K	25X
78%	*Tmz1t 3203*	phosphoribosylformylglycinamidine synthase	Frame shift	C/-	G817A	18X
100%	*SecD*	Transport protein	Non synoymous	G to A	G531D**	12X
95%	*Tmz1t 2115*	Peptidase M15A	Frame shift homopolymer	-/G -/G	L3P** L20P**	21X 22X
100%	*Tmz1t 2095*	CzcA family heavy metal efflux pump	Non synonymous	A to T	V57E**	17X
100%	*Tmz1t 2994*	Rnf electron transport subunit E	Non synonymous	A to C	I69F	13X

90%	*Tmz1t 3143*	Diguanylate cyclase	Frame shift	C/-	L508W*	11X
88%	*Tmz1t 3207*	Glycine dehydrogenase	Frame shift	G/-	G127A	16X
100%	*rhoB*	DNA polymerase	Non synoymous	T to A	Q538L**	21X
100%	*Tmz1t 3810*	Glucose-1-phosphate thymidylyltransferase	Non synonymous	G to A	R128H*	22X
88%	*Tmz1t 3840*	S-adenosyl methionine synthetase	Frame shift homopolymer	G/-	G118A	16X
85%	*Tmz1t 3867*	MltA-interacting MipA Family protein	Frame shift homopolymer	G/-	G11A	20X
100%	*Tmz1t 3869*	PAS/PAC sensor signal transduction histidine kinase	Non synonymous	G to A	R97H*	15X
100%	*Tmz1t 4055*	Winged helix family two component heavy metal response transcription regulator	Non synonymous	G to A	V112I	16X

* Amino acid change predicted to result in mild property change

** Amino acid change predicted to result in severe property change

4.2 SNP Validation Using PCR and Sanger Sequencing

We picked non-synonymous and frame shift mutations in EPS biosynthesis, Type IV pilus assembly, global regulatory and EPS export related genes detected by Ion Torrent, which included *tmz1t_0834, tmz1t _1376, tmz1t _1383, tmz1t _1679, tmz1t _2095, tmz1t _3143, tmz1t_3249, tmz1t _3637, tmz1t _3801*, and *tmz1t _3810*, to test and verify using Sanger sequencing. The candidate genes were PCR amplified from their mutant genomes and cloned in pCR2.1 plasmid using the Topo TA cloning kit and sent to MWG (Huntsville, AL) for Sanger sequencing. The Sanger sequencing results verified these genes to have actual real mutations in their mutant strains (Table 4.3). Of these, EPS deactylase gene of MZ1T 20A mutant was confirmed to have a non-sense mutation introducing a stop codon in the middle of the gene (Figure 4.4 and Figure 4.5). Subsequently, the wild-type version of the verified candidate genes were sub-cloned into the broad host range plasmid pRK415 to further test complementation of the floc-defective MZ1T mutant phenotype.

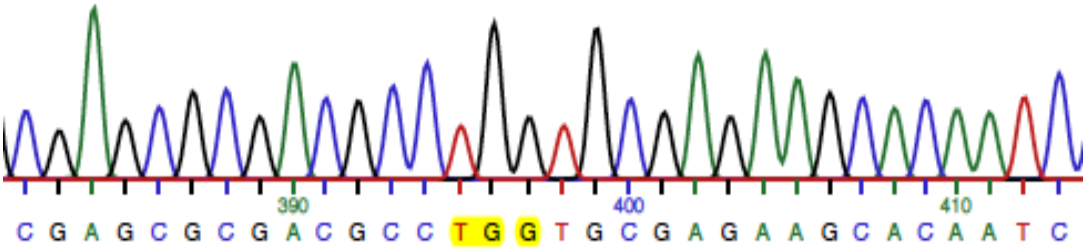
Table 4.3 Candidate genes involved in MZ1T flocculation. Sanger sequencing verified mutations detected by Ion torrent software.

Gene	Gene function	Type of mutation	Base change (NextGENe software)	MZ1T mutant stains	True mutation verified by Sanger sequencing
Tmz1t 3249	Polysaccharide deacetylase	Non sense	G to A	20A	Yes
Tmz1t 3637	Family 2 glycosyl transferase	Frame shift	G/-	20A	Yes
Tmz1t 3801	Polysaccharide biosynthesis CapD	Non synonymous	T to G	20A	Yes
Tmz1t 0834	Type IV pilus assembly protein	Non synonymous	G to A	39A	Yes
Tmz1t 1376	ABC transporter	Non synonymous	G to A	39A	Yes
Tmz1t 1383	Von willebrand factor	Non synonymous	G to A	39A	Yes
Tmz1t 1679	PAS/PAC sensor containing diguanylate cyclase	Non synonymous	G to A	39A	Yes
Tmz1t 2095	CzcA family heavy metal efflux pump	Non synonymous	A to T	39A	Yes
Tmz1t 3143	Diguanylate cyclase	Frame shift	C/-	39A	Yes
Tmz1t 3810	Glucose-1-phosphate thymidylyltransferase	Non synonymous	G to A	39A	Yes

A)

CGAGCGCGACGCC**TGG**TGCGAGAAGCACAATC

B)



C)

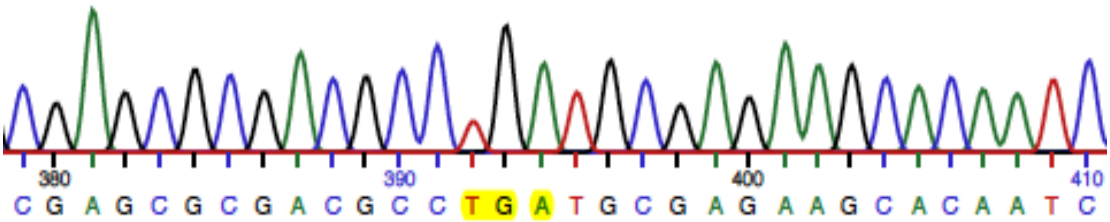


Figure 4.5 Sanger sequencing verification of the Ion torrent sequence. Sequences are shown in the following order: A) the published reference sequence first, followed by the Sanger results for the deacetylase gene in B) MZ1T wild type and C) 20A strains.

A)

Score	Expect	Identities	Gaps
1472 bits(797)	0.0	800/801(99%)	0/801(0%)

```
Query  3554683  ATGACGACCTCCCCCGCCCAACGCGTCCCGGTCCTGATGTACCACCGCATCGGCGCTGCAC  3554742
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  63       ATGACGACCTCCCCCGCCCAACGCGTCCCGGTCCTGATGTACCACCGCATCGGCGCTGCAC  122

Query  3554743  ACAACGACTGGGAACGCAAGTACTGCGTGAGCCCGGAAGCCTTTGCGGCACACATGGACA  3554802
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  123      ACAACGACTGGGAACGCAAGTACTGCGTGAGCCCGGAAGCCTTTGCGGCACACATGGACA  182

Query  3554803  CGCTCGCGCGCGCCGGCTGGAAGGCGGTGTCCATCGACGCCTTCTTCGCCTGGCTGGACG  3554862
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  183      CGCTCGCGCGCGCCGGCTGGAAGGCGGTGTCCATCGACGCCTTCTTCGCCTGGCTGGACG  242

Query  3554863  GCGGAGCGTCCCTGCCGGAGGGCGCCTTCCTGCTCACCTTCGACGACGGCTTCCGCGGCG  3554922
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  243      GCGGAGCGTCCCTGCCGGAGGGCGCCTTCCTGCTCACCTTCGACGACGGCTTCCGCGGCG  302

Query  3554923  TGCACGACCACGCCGGTCCGGTGCTGCGCAGGCTCGGCTGGCCGGCAACGGTCTTCCTGG  3554982
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  303      TGCACGACCACGCCGGTCCGGTGCTGCGCAGGCTCGGCTGGCCGGCAACGGTCTTCCTGG  362

Query  3554983  TCAGCGCGCTCGTCGGCGAGCGCGACGCCTGGTGCGAGAAGCACAATCCGGACGGCCACA  3555042
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  363      TCAGCGCGCTCGTCGGCGAGCGCGACGCCTGATGCGAGAAGCACAATCCGGACGGCCACA  422

Query  3555043  CCTACCCGCTGATGGATCGCGCACAGATCCTCGCGCTGCGCGCCCAGGGCTTCGCCTTCC  3555102
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  423      CCTACCCGCTGATGGATCGCGCACAGATCCTCGCGCTGCGCGCCCAGGGCTTCGCCTTCC  482

Query  3555103  ACTCCCACACGCGCGACCATGCCGACCTGCCGACGCTCGACGACGCGGCGCTCCAGGCCC  3555162
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  483      ACTCCCACACGCGCGACCATGCCGACCTGCCGACGCTCGACGACGCGGCGCTCCAGGCCC  542

Query  3555163  AGCTCGCCGGTGCGCGCGACGACCTCGAGGCCCTGCTCGGTGCGCCGGTCGACTACCTCG  3555222
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  543      AGCTCGCCGGTGCGCGCGACGACCTCGAGGCCCTGCTCGGTGCGCCGGTCGACTACCTCG  602

Query  3555223  CCTACCCGTACGGGCGCTACGACGAACGCGTCCTGCACCAGGCGCGGCAGGCCGGCTACC  3555282
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  603      CCTACCCGTACGGGCGCTACGACGAACGCGTCCTGCACCAGGCGCGGCAGGCCGGCTACC  662

Query  3555283  GCGCCGCGTTCTCGGTGCAGCCCGGGTTCAACCGTCCCGAGGTCGACCGCTTCCGCCTGC  3555342
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  663      GCGCCGCGTTCTCGGTGCAGCCCGGGTTCAACCGTCCCGAGGTCGACCGCTTCCGCCTGC  722

Query  3555343  GCCGGCTCGACGTGTTCGGCACCGACACCCCGGCGATGCTGCGCCGCAAGATCACCCTGG  3555402
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  723      GCCGGCTCGACGTGTTCGGCACCGACACCCCGGCGATGCTGCGCCGCAAGATCACCCTGG  782

Query  3555403  GCAGCAACGACGGCCGCCTGACGGCCGCGCTGCGCTACAACGCCGGCCGGGTCCTCGCCC  3555462
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  783      GCAGCAACGACGGCCGCCTGACGGCCGCGCTGCGCTACAACGCCGGCCGGGTCCTCGCCC  842
```

B)

Score	Expect	Identities	Gaps
1480 bits(801)	0.0	801/801(100%)	0/801(0%)

```
Query  3554682  ATGACGACCTCCCCCGCCCAACGCGTCCCGGTCCTGATGTACCACCGCATCGGCGCTGCA  3554741
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  888      ATGACGACCTCCCCCGCCCAACGCGTCCCGGTCCTGATGTACCACCGCATCGGCGCTGCA  829

Query  3554742  CACAACGACTGGGAACGCAAGTACTGCGTGAGCCCGGAAGCCTTTGCGGCACACATGGAC  3554801
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  828      CACAACGACTGGGAACGCAAGTACTGCGTGAGCCCGGAAGCCTTTGCGGCACACATGGAC  769

Query  3554802  ACGCTCGCGCGCGCCGGCTGGAAGGCGGTGTCCATCGACGCCTTCTTCGCCTGGCTGGAC  3554861
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  768      ACGCTCGCGCGCGCCGGCTGGAAGGCGGTGTCCATCGACGCCTTCTTCGCCTGGCTGGAC  709

Query  3554862  GGCGGAGCGTCCCTGCCGGAGGGCGCCTTCCTGCTCACCTTCGACGACGGCTTCCGCGGC  3554921
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  708      GGCGGAGCGTCCCTGCCGGAGGGCGCCTTCCTGCTCACCTTCGACGACGGCTTCCGCGGC  649

Query  3554922  GTGCACGACCACGCCGGTCCGGTGCTGCGCAGGCTCGGCTGGCCGGCAACGGTCTTCCTG  3554981
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  648      GTGCACGACCACGCCGGTCCGGTGCTGCGCAGGCTCGGCTGGCCGGCAACGGTCTTCCTG  589

Query  3554982  GTCAGCGCGCTCGTCGGCGAGCGCGACGCCTGGGTGCGAGAAGCACAATCCGGACGGCCAC  3555041
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  588      GTCAGCGCGCTCGTCGGCGAGCGCGACGCCTGGGTGCGAGAAGCACAATCCGGACGGCCAC  529

Query  3555042  ACCTACCCGCTGATGGATCGCGCACAGATCCTCGCGCTGCGCGCCCAGGGCTTCGCCTTC  3555101
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  528      ACCTACCCGCTGATGGATCGCGCACAGATCCTCGCGCTGCGCGCCCAGGGCTTCGCCTTC  469

Query  3555102  CACTCCCACACGCGCGACCATGCCGACCTGCCGACGCTCGACGACGCGGCGCTCCAGGCC  3555161
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  468      CACTCCCACACGCGCGACCATGCCGACCTGCCGACGCTCGACGACGCGGCGCTCCAGGCC  409

Query  3555162  CAGCTCGCCGGTGCGCGCGACGACCTCGAGGCCCTGCTCGGTGCGCCGGTCGACTACCTC  3555221
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  408      CAGCTCGCCGGTGCGCGCGACGACCTCGAGGCCCTGCTCGGTGCGCCGGTCGACTACCTC  349

Query  3555222  GCCTACCCGTACGGGCGCTACGACGAACGCGTCCTGCACCAGGCGCGGCAGGCCGGCTAC  3555281
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  348      GCCTACCCGTACGGGCGCTACGACGAACGCGTCCTGCACCAGGCGCGGCAGGCCGGCTAC  289

Query  3555282  CGCGCCGCGTTCTCGGTGCAGCCCGGGTTCAACCGTCCCGAGGTCGACCGCTTCCGCCTG  3555341
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  288      CGCGCCGCGTTCTCGGTGCAGCCCGGGTTCAACCGTCCCGAGGTCGACCGCTTCCGCCTG  229

Query  3555342  CGCCGGCTCGACGTGTTCGGCACCGACACCCCGGCGATGCTGCGCCGCAAGATCACCCTG  3555401
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  228      CGCCGGCTCGACGTGTTCGGCACCGACACCCCGGCGATGCTGCGCCGCAAGATCACCCTG  169

Query  3555402  GGCAGCAACGACGGCCGCCTGACGGCCGCGCTGCGCTACAACGCCGGCCGGGTCCTCGCC  3555461
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  168      GGCAGCAACGACGGCCGCCTGACGGCCGCGCTGCGCTACAACGCCGGCCGGGTCCTCGCC  109
```

Figure 4.6 Blast results of A) *mz1t_3249* EPS deacetylase gene of MZ1T 20A mutant. B) *mz1t_3249* EPS deacetylase gene of MZ1T wild type to referrence data base.

4.3 Complementation

Plasmids pRK145 carrying a wild-type candidate gene *tmz1t_3249, tmz1t _3637,* and *tmz1t _3801* were used to complement mutant MZ1T 20A by tri-parental mating. Among these genes, complementation of MZ1T 20A mutant strain with pRK145 plasmid bearing EPS deaectylase gene (*tmz1t_3249)* restored the flocculation phenotype, but no flocculation was observed in MZ1T 20A bearing plasmid pRK415 (Figure 4.7) without the cloned insert. Interestingly, *E. coli* carrying pRK415: *mz1t_3249* formed clumping cells similar to the flocculation of WT MZ1T in contrast to *E. coli* carrying the empty pRK415 vector (Figure 4.8). In addition, MZ1T 39A mutant was also transformed with plasmids pRK415 that carry *tmz1t_0834, tmz1t _1376, tmz1t _3143,* and *tmz1t _3810* genes; however, none of these plasmids could restore flocculation to wild-type level.

A) **B)** **C)**

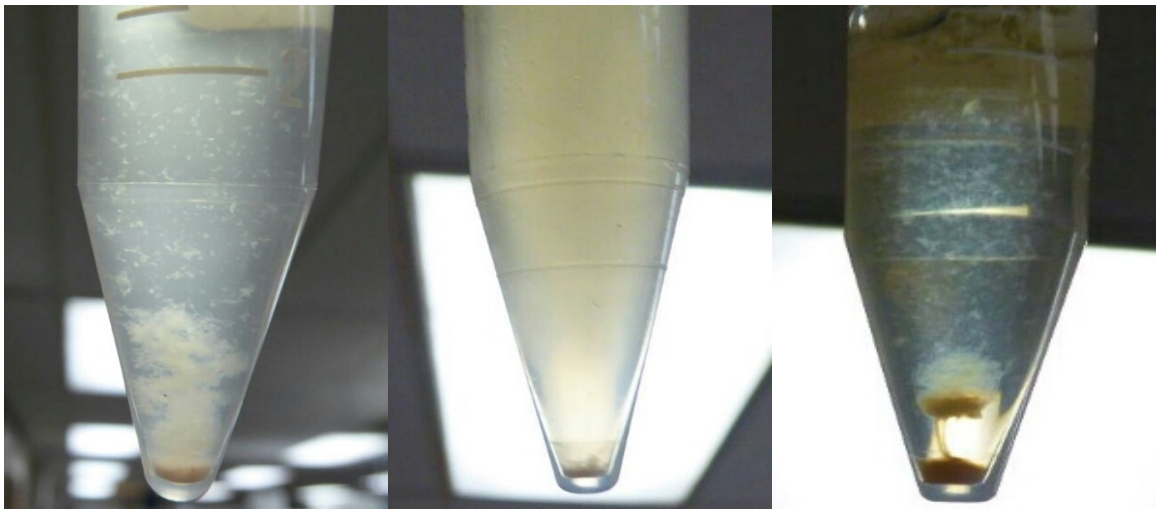


Figure 4.7 Complementation of MZ1T 20A by EPS deacetylase (*mz1t_3249*). A) MZ1T WT B) MZ1T 20A carring pRK415:*mz1t_3249* C) MZ1T 20A carring parental plasmid pRK41

Figure 4.8 Floculation-like cell clumping of *E. coli* carrying pRK415:*mz1t_3249* (Left tube) compared to *E. coli* carrying parental plasmid pRK415 (Right tube).

4.4 EPS Purification and Quantification

As filtration and dialysis gave the best results for EPS purity, Tangential flow filtration (TFF) (Pellicon Mini Cassette Holder, Millipore) was considered as an efficient process to remove salt at larger scale. EPSs from MZ1T strains have been described as having a molecular weight of 260 kDa (Allen et al. 2004); consequently, a membrane of 100 kDa MWCO was chosen. 100 ml of the 1-L initial EPS supernatants were retained and concentrated in the retentate fraction. EPSs from the retentate fractions were subjected to dialysis overnight before being lyophilized. Measurement of total carbon content in the EPSs was done by modified Phenol-Sulfuric Acid (PSA) method in microplate format. The EPS yields of MZ1T WT, 39A and 20A are 82.63 +/- 5.8, 75.09 +/- 6.9, and 72.91 +/- 7.3 mg/L respectively, and these yields are not significantly different from the mean of the wild type at a 95% confidence interval (Figure 4.9).

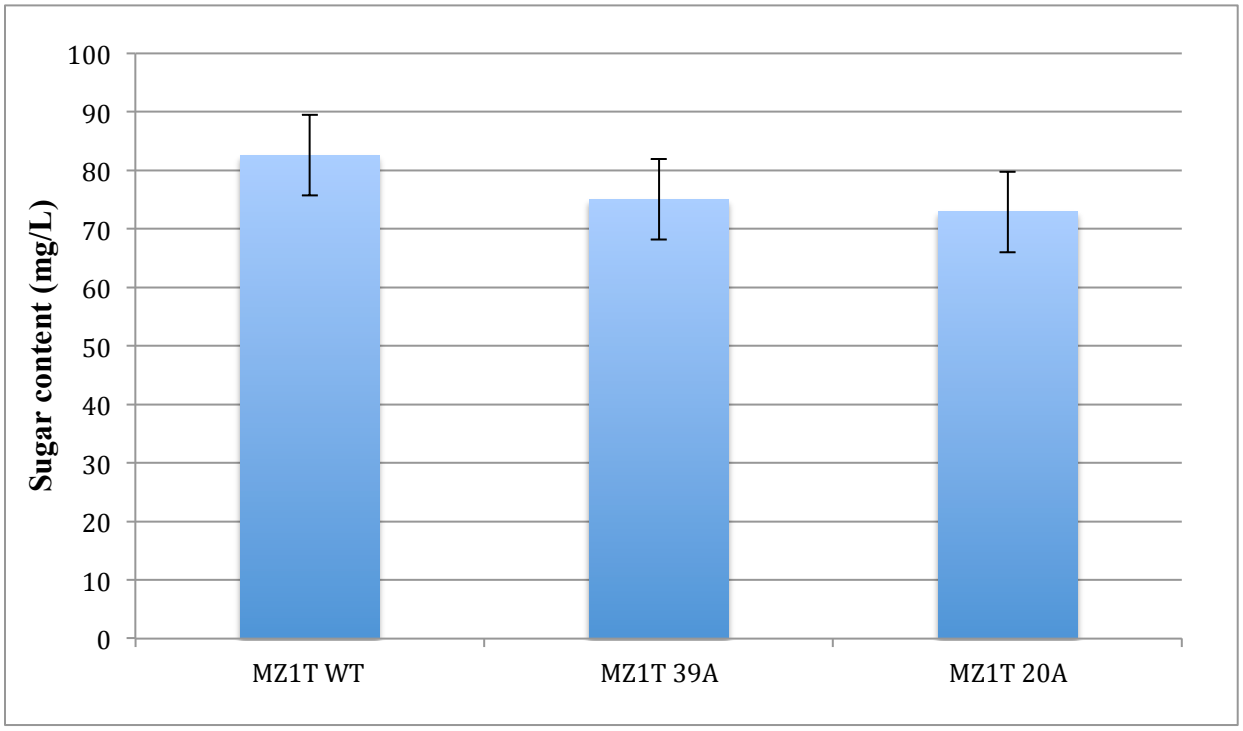


Figure 4.9 Average yields of EPS from floc^{+} and floc^{-} strains. MZ1T EPS was recovered from 1-L cultures of MZ1T wild type (WT) and mutant strains. Yield values are averages of at least three replicates. Error values represent standard deviations.

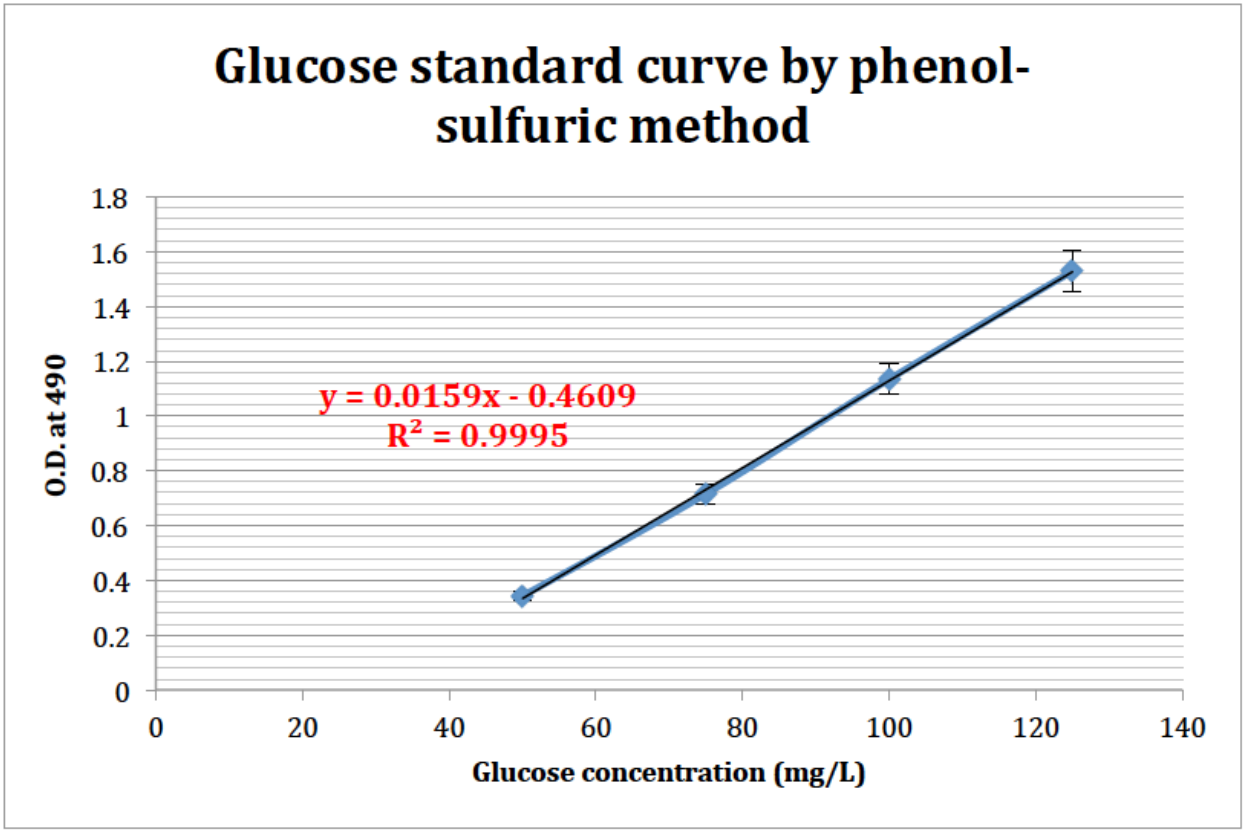


Figure 4.10 Glucose standard curve generated using the by phenol-sulfuric method.

4.5 FTIR Characterization Studies of EPS.

The IR spectra of EPS extracted from MZ1T wild type, 39A, and 20A mutants cultivated in TDM medium for 7 days are shown in Figure 4.10-4.12. FTIR bands of partially purified EPSs from all MZ1T stain consistent with polysaccharide characteristic peaks were detected at 1400 cm^{-1}, corresponding to the symmetrical stretching C=O of COO^- carboxyl groups, and 950-1200 cm^{-1}, the C–H stretching of alcohols, C-OH stretching, and C-O stretching of C-O-C. Moreover, partially purified EPSs from MZ1T 20A, which completely lost flocculation ability, demonstrated interesting altered peaks from MZ1T WT at 1665 cm^{-1} representing reduced C=O stretching vibration peak of the *N*-acetyl group, whereas the N-H deformation vibration peak of a secondary amine group at 1527.84 cm^{-1} increased as compared to the spectra of partially purified MZ1T mutant EPS. Also, chemically deacetylated EPS of MZ1T 20A mutant showed a diminished C=O stretching peak of the *N*-acetyl group, similar to the IR pattern of MZ1T wild-type EPS (Figure 4.13). These findings suggest that secondary modification of functional groups in the EPS could be responsible for MZ1T flocculation. It is also consistent with the discovery of the nonsense mutation in the putative EPS deactylase gene in MZ1T 20A by whole genome sequencing.

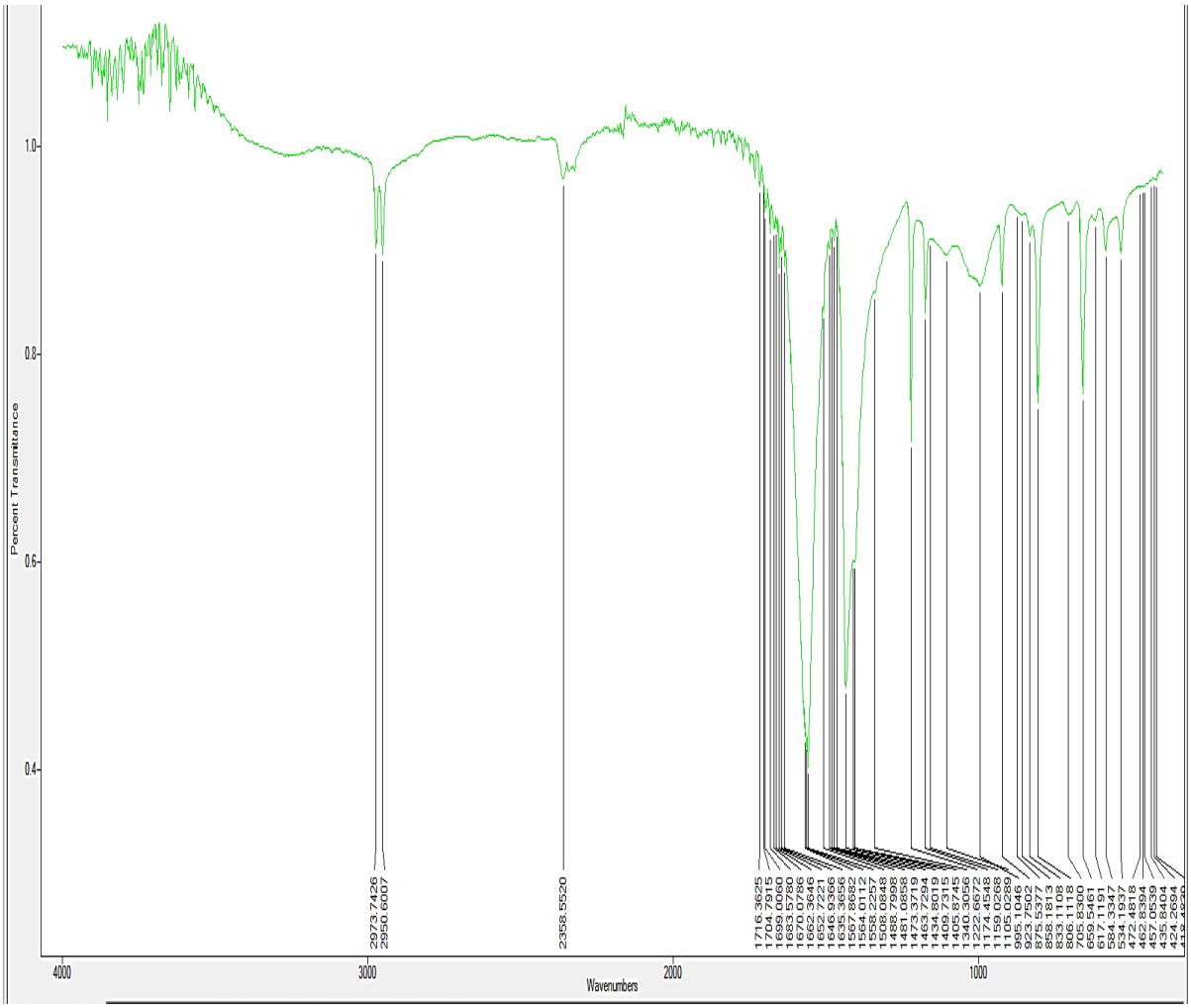


Figure 4.11 FT-IR spectra of MZ1T wild type EPS. Polysaccharide characteristic peaks were detected at 1400 cm^{-1}, corresponding to the symmetrical stretching C=O of COO^- carboxyl groups, and 950-1200 cm^{-1}, the C–H stretching of alcohols, C-OH stretching, and C-O stretching of C-O-C.

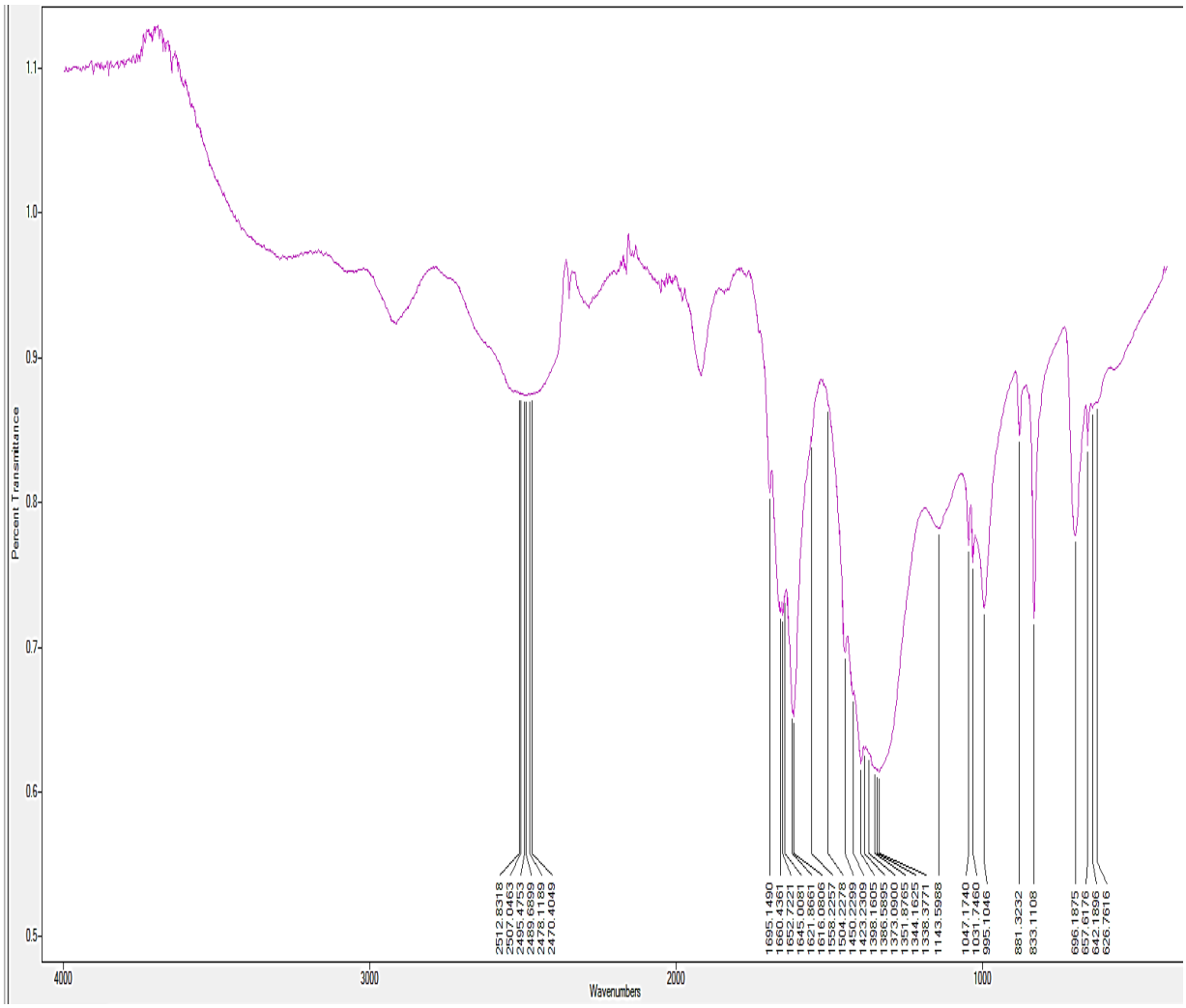


Figure 4.12 FT-IR spectra of MZ1T 39A mutant EPS. MZ1T mutant EPS contains altered peak from MZ1T WT at 1655 cm^{-1} representing reduced C=O stretching vibration peak of the *N*-acetyl group.

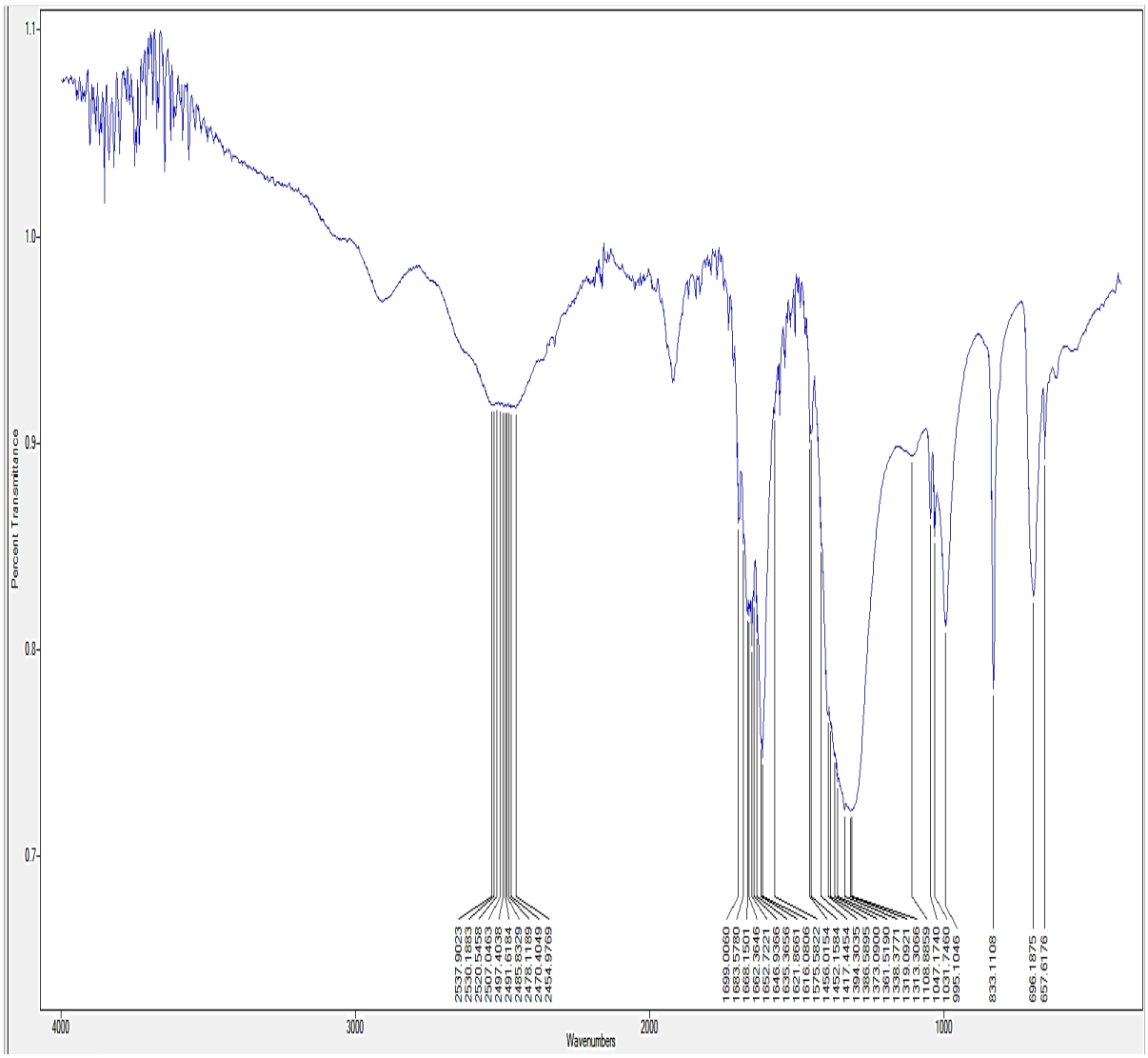


Figure 4.13 FT-IR spectra of MZ1T 20A mutant EPS. MZ1T mutant EPS contains altered peak from MZ1T WT at 1655 cm^{-1} representing reduced C=O stretching vibration peak of the *N*-acetyl group.

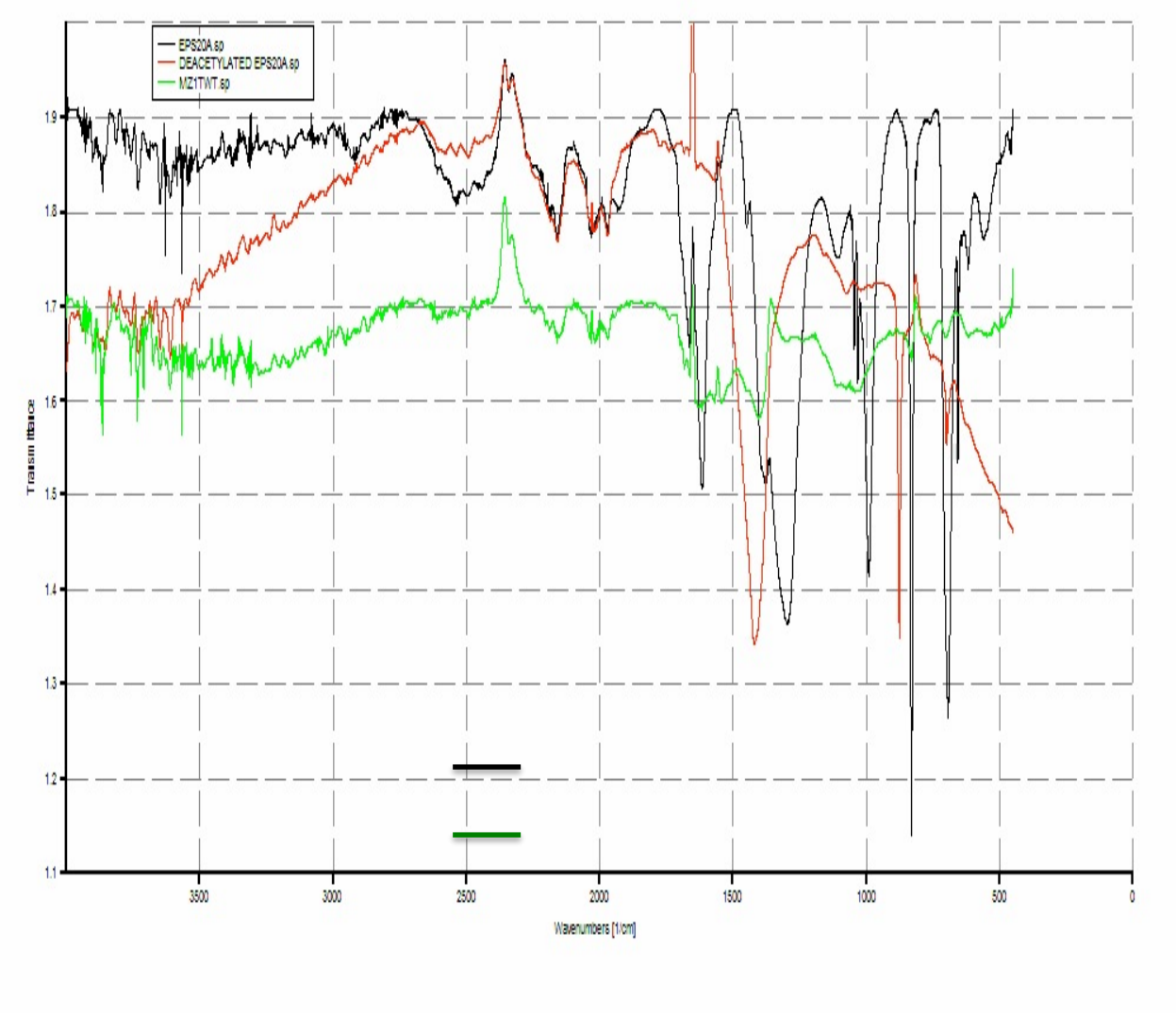


Figure 4.14 Comparison of FT-IR spectra of chemically deaectyled MZ1T mutant 20A EPS, MZ1T mutant 20A EPS and MZ1T wild-type EPS

4.6 Expression of Genes Involved in MZ1T Flocculation

In order to determine how EPS biosynthesis (*tmz1t_3801*) and EPS deacetylase (*tmz1t_3249*) genes influence the differentiation of MZ1T from planktonic to floc-forming growth phases we examined gene expression profiles of EPS biosynthesis (*tmz1t_3801*) and EPS deacetylase (*tmz1t_3249*) genes of MZ1T wild-type culture before and after flocculation using Droplet Digital PCR. RNA extracts were treated with DNase I and subjected to cDNA synthesis as described in section 3.12.1. No genomic DNA contamination was detected from any minus reverse transcriptase reactions. Our results (Figure 4.15 and Figure 4.16) demonstrated an increase in the expression of the EPS biosynthesis and deacetylase genes during 18-24 h. At 36 h, EPS biosynthesis gene expression dramatically decreased and bounded back before floc formation, at 48 h, then decreased again thoughout the 96 h culture growth. In contrast, EPS deacetylase gene expression gradually declined after 24 h culture growth.

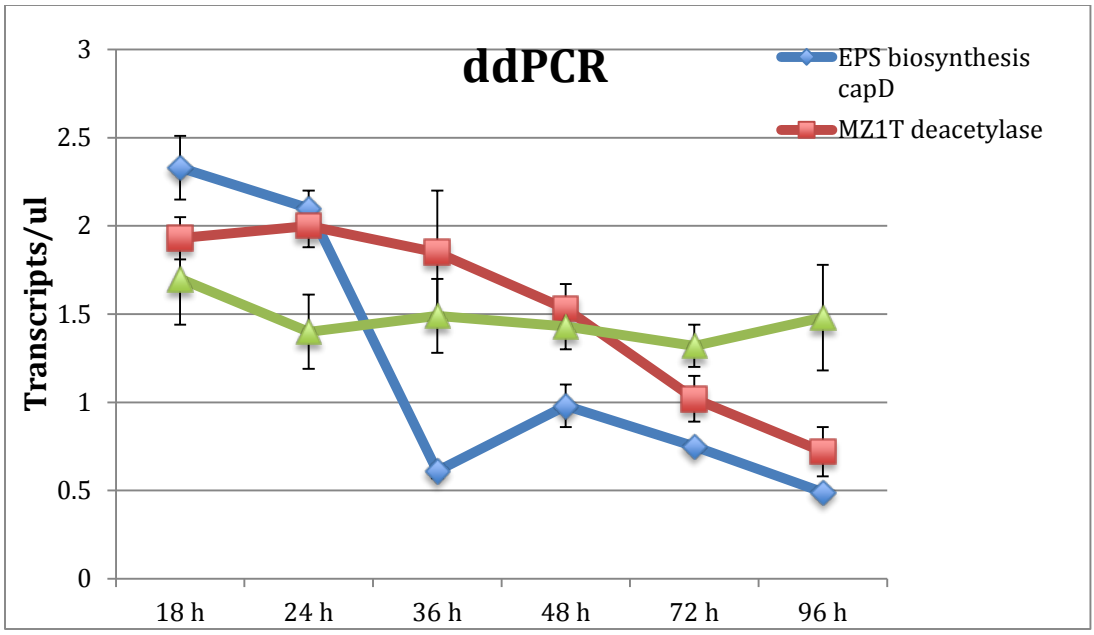


Figure 4.15 Comparison of MZ1T WT EPS biosynthesis *capD*, MZ1T deacetylase, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene expression during 96 h growth culture. Flocculation occurs at 48 h. The gene transcripts quantification was done by droplet digital PCR in triplicate, and standard error bars are included.

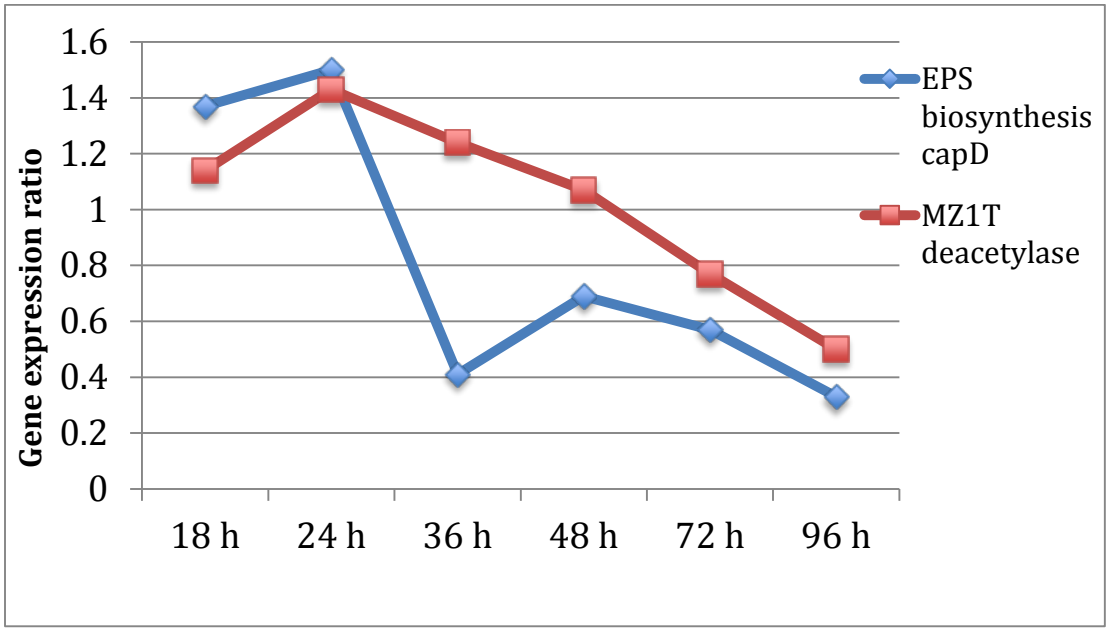


Figure 4.16 Gene expression ratio of EPS biosynthesis *capD* and EPS deacetylase

Table 4.4 Transcripts per µL of EPS biosynthesis *capD* during 96 h growth culture of MZ1T WT by droplet digital PCR

Cultivation time (h)	Transcripts/ µL			Standard deviation	Standard error
18	2.7	2.1	2.2	0.32	0.180
24	2.2	1.9	2.2	0.17	0.100
36	0.61	0.68	0.54	0.07	0.040
48	0.77	1.01	1.17	0.20	0.120
72	0.76	0.75	0.76	0.01	0.005
96	0.46	0.51	0.50	0.03	0.017

Table 4.5 Transcripts per µl of GAPDH during 96 h growth culture of MZ1T WT by droplet digital PCR

Cultivation time (h)	Transcripts / µL			Standard deviation	Standard error
18	1.3	2.2	1.6	0.45	0.26
24	1.8	1.1	1.3	0.36	0.21
36	1.9	1.37	1.2	0.36	0.21
48	1.3	1.7	1.3	0.23	0.13
72	1.2	1.0	1.5	0.20	0.12
96	0.73	1.02	2.7	0.70	0.40

Table 4.6 Transcripts per µl of MZ1T deacetylase during 96 h growth culture of MZ1T WT by droplet digital PCR

Cultivation time (h)	Transcripts / µL			Standard deviation	Standard error
18	2.0	1.7	2.1	0.21	0.12
24	2.0	2.2	1.8	0.20	0.12
36	2.9	1.4	1.22	0.72	0.41
48	1.4	1.2	1.7	0.25	0.14
72	1.15	1.14	0.77	0.22	0.13
96	0.56	0.6	1.0	0.24	0.14

Table 4.7 Gene expression ratio of EPS biosynthesis *capD* and EPS deacetylase

Cultivation time (h)	EPS biosynthesis capD	EPS deacetylase
18	1.37	1.14
24	1.5	1.43
36	0.41	1.24
48	0.69	1.07
72	0.57	0.77
96	0.33	0.50

CHARTER 5

DISCCUSION

5.1 Deep Sequencing of MZ1T Flocculation Mutant Genomes

New technologies for whole-genome sequencing have tremendous potential in aiding the search for mutations of interest. We have demonstrated a proof of concept that next generation sequencing can be a powerful method for identifying NTG induced mutations in *Thauera aminoaromatica* strain MZ1T, which in the past a screen of NTG mutants would require countless hours of work. In fact, more than 15X coverage data reads of MZ1T genome (4.3 Mbp) were generated, sufficient for bacterial genome analysis, and the distribution of read lengths is narrowly distributed indicating a high quantity of usable reads. Moreover, the Ion Torrent sequencing enabled high-confidence detection of a total of 75 single nucleotide polymorphisms (SNPs). The majority of the SNPs identified show characteristics of mismatch mutations induced by NTG chemical mutagenesis (G/C to A/T transitions). In addition, the base-calling software in the Torrent Suite (version 2.0.1) performs two quality assurance steps prior to output of sequences. The first step evaluates the residual between observed flow values and predicted flow values based on a model of the flow cell. Reads with residuals that produce a median absolute value greater than a given threshold are filtered from both the SFF and the FASTQ, as they are assumed polyclonal. The second step scans non-polyclonal reads to identify undesirable regions of the read, which are subsequently trimmed. Undesirable regions are defined as regions containing the adapter sequence (and beyond) as well as low-quality regions. As expected for a new technology, there have been marked improvements in the PGM

since its limited release in January, 2011. The newest kit, used in this study, is the 200 bp OneTouch and has substantially reduced the error rate from that with in the 200 bp Manual kit. Furthermore, the accuracy of PGM quality scores has further increased with sequential sequencing kit releases. The newly released Torrent Suite (2.20) introduced a third trimming approach which clips the read based on high-residual ionogram 1-mer and 2-mer flow values, which may indicate 'noisy' flows. The addition of HRI-based clipping to complement the relatively lax quality trim has proven extremely effective at removing the error-prone end of the reads, albeit at a cost of 20–30 bp of read length (Torrent User Documentation Version 2.2.0).

However, it is widely known that homopolymer stretches are the main sequencing error of the PGM technologies due to overwhelmed flow-based detection. In contrast, the Illumina GAIIx platform uses reversible, fluorescently labeled terminators, which allow each cycle to interrogate only one base at time, and thus, sequencing through homopolymer tracts on that platform is typically not a problem. In this study, we detected some homopolymer reads. In most cases, the sequencing error type had one less base than the reference sequence at the end of homopolymer stretches, resulting in a false-deletion mutant call. (Bragg et al. 2013)

5.2 SNPs Calling of MZ1T 39A and 20A Mutants

Data analysis using NextGENe software revealed promising gene candidates possibly involved in MZ1T flocculation from the mutant's genomic data. In MZ1T 20A mutant, *mz1t_3249* (Polysaccharide deacetylase), *mz1t_3637* (Family 2 glycosyl transferase), and *mz1t_3801* (Polysaccharide biosynthesis, CapD) each contained a mutation leading to amino acid change or in the case of the mutation in MZ1T EPS deacetylase gene, an apparent

nonsense mutation truncating the protein and potentially abolishing its function. These genes are located in EPS cluster 1 and cluster 2 and involved in the biosynthesis and export of the extracellular polysaccharides possibly required for flocculation in *T. aminoaromatica* MZ1T. Moreover, these two gene clusters have highly conserved gene organization in *Rubrivivax gelatinosus* (Steunou et al. 2013). This photosynthetic bacterium can switch from planktonic lifestyle to phototrophic biofilm in mats in response to environmental changes and contain a two-component system EmbRS that negatively controls the polysaccharide synthesis and biofilm formation. Complementation of genes in EPS 1 and EPS 2 clusters of *R. gelatinosus* restored auto-aggregation and fast sinking of the compact mass cells in Δ*EmbRS R. gelatinosus* mutant (Steunou et al. 2013). Therefore, complementation of *mz1t_3249* (Polysaccharide deacetylase), *mz1t_3637* (Family 2 glycosyl transferase), and *mz1t_3801* (Polysaccharide biosynthesis, CapD) would confirm that so-called polysaccharide biosynthesis cluster 1 and cluster 2 are responsible for biofilm formation and flocculation in *R. gelatinous* and in *Thauera* sp. MZ1T.

In MZ1T 39A mutant, due to the reduced floc formation in this strain, we speculate that direct or indirect global regulatory genes are involved in this defective phenotype. We found SNPs in several response regulator receiver protein, PAS/PAC sensor signal transduction kinase, and Diguanylate cyclase genes. It is well known that synthesis of the alginate polymer in *Pseudomonas aeruginosa* is allosterically regulated by the secondary messenger c-di-GMP as binding of c-di-GMP to the inner membrane protein is essential for bacterial exopolysaccharide production leading to auto-aggregation and biofilm formation (Lee et al. 2007). In addition, signal transduction cassettes ArlRS trigger and regulate biofilm formation of *Staphylococcus*

epidermidis in an *ica*-dependent manner (Wu et al. 2012). We also discovered a mutation causing amino acid change in Glucose-1-phosphate thymidylyltransferase gene belonging to EPS cluster 1. Indeed, failure to synthesize an EPS precursor has been demonstrated a diminishing ability to form a biofilm (Kim et al. 2007).

However, none of the other candidate regulatory genes tested could restore floc formation to wild type levels, suggesting that they are not directly linked to regulation of floc formation. Moreover, flocculation is an exceptionally complex process; requiring the coordinate expression and simultaneous regulation of many genes by complicated genetic networks involving all levels of gene regulation. In this study, only complementation of *mz1t_3249* (Polysaccharide deacetylase) in mutant 20A was found to rescue the flocculation phenotype.

5.3 MZ1T Exopolysaccharide deacetylase Plays an Important Role in Flocculation

Complementation of the EPS deacetylase gene restores the wild type flocculation phenotype in MZ1T 20A mutant strain. MZ1T EPS deacetylase belongs to the large and functionally diverse carbohydrate esterase 4 (CE4) superfamily, whose members show strong sequence similarity with some variability due to their distinct carbohydrate substrates. It includes bacterial poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase PgaB, intercellular adhesion proteins IcaB, *Pseudomonas* Pel deacetylase PelA, and many uncharacterized prokaryotic polysaccharide deacetylases. It also includes a putative polysaccharide deacetylase YxkH encoded by the *Bacillus subtilis yxkH* gene, which is one of six polysaccharide deacetylase gene homologs present in the *Bacillus subtilis* genome. Sequence comparison shows all family members contain a conserved domain similar to the catalytic NodB homology domain of rhizobial NodB-like proteins, which consists of a deformed (beta/alpha) 8 barrel fold with 6 or 7

strands. However, within this family most proteins have 5 strands while some have 6 strands. Long insertions are also found in many family members, the function(s) of which remains unknown (Colvin et al. 2013).

Floc$^-$ MZ1T 20A mutant produces significant amounts of extractable EPS and contains the same monosaccharide composition detected in the EPS extracts from floc$^+$ MZ1T; however, FTIR results show different spectra between MZ1T 20A and wild-type EPS. In fact, the C = O peak of the acetyl group of *N*-acetyl-glucosamine and/or *N*-acetyl-fucosamine is greatly reduced in the wild type compared to the MZ1T 20A mutant spectrum indicating modification by possible deacetylation in EPS of MZ1T wild type. This suggests that modification of EPS side chains is involved in MZ1T flocculation. This finding is consistent with that reported for biofilm forming bacteria elsewhere. The adhesive characteristic of exopolysaccharides strongly depends on chain conformation, and is greatly impacted by substituents that change interchain and intrachain interactions (Haag 2006). For example, deacetylation of polysaccharides might promote the conformational transition of the polymer strands from random coils to ordered helices so as to facilitate gel formation, which is mediated by interspersed regions of soluble, hydrated polymer with regions of polymer-polymer interactions (Villain-Simonnet, Milas, and Rinaudo 2000; Rinaudo 2004). In addition, acetyl groups have been shown to be necessary for the stability of bacterial polysaccharides and subsequent biofilm development (Ridout et al. 1997) (Franklin and Ohman 2002) (Tielen et al. 2005).

However, deacetylation in MZ1T wild type appears to occur only partially. Chemical deacetylation of MZ1T mutants EPS using strong base and heat completely removes acetyl groups whereas untreated wild type and 39A mutant still have some degree of reduced C = O

peak for the acetyl groups. Partial deacetylation of poly-beta-1,6-N-acetyl-D-glucosamine (PNAG), which is an intercellular adhesin has been shown to be required for the secretion of the polymer in *E. coli*. In fact, 16.4 ± 9.9% of GlcNAc residues in the wild-type strain were deacetylated, whereas no deacetylation was detected in the *pgaB* mutant strain. These findings indicate that the introduction of deacetylated GlcNAc into PNAG occurs by a dedicated mechanism to deacetylate a polymeric PNAG precursor (Itoh et al. 2008). In addition, deacetylation of poly-β(1-6)-N-acetylglucosamine (PNAG) in *Staphylococcus aureus* and *Staphylococcus epidermidis* introduces positive charges in the otherwise neutral PNAG molecule, as free amino groups are exposed that become protonated at neutral and acidic pH value. The cationic character of PNAG is essential for the attachment of PNAG to the negatively charged bacterial cell surface. By PNAG production, the bacteria, thus, can efficiently change the electrostatic properties of their cell surface (Kropec et al. 2005). Extensive studies on the chitin-chitosan system also suggest that modifying the acetylation state of polysaccharides alters their chemical-physical properties. Partial deacetylation of chitin, a ubiquitous GlcNAc polymer, leads to the production of chitosan, which contains more exposed amine groups and fewer acetyl groups. During the production of chitosan, the intrinsic pKa was found to increase from 6.46 to 7.32 as a function of the degree of deacetylation (Sorlier et al. 2001). The degree of deacetylation influences the physical properties of chitosan by altering electrostatic interactions, hydrogen bonding, and hydrophobic interactions with the surrounding environment (Sorlier et al. 2001).

Therefore, presumably, MZ1T EPS deacetylase creates positive charges on the amino sugars glycosyl component of MZ1T EPS, which attracts the negative charge of the carboxyl

group of galacturonic acid leading to binding of MZ1T EPS and flocculation. Further study using modified polysaccharides will be necessary to completely address this mechanism.

One of the most distinctive features that distinguishe biofilms from planktonic populations is the presence of an extracellular matrix embedding the biofilm bacteria and determining mature biofilm architecture. To date, three exopolysaccharides, β-1,6-N-acetyl-D-glucosamine polymer (PGA), colanic acid, and cellulose, have been detected in the biofilm matrix of *E. coli* and have been shown to be important for biofilm formation, while others such are lipopolysaccharides and capsular polysaccharides may not accumulate significantly in the matrix, but still play an important indirect role in biofilm formation (Starkey et al. 2004). In this study, after introducing expression plasmid pRK415 harboring MZ1T EPS deacetylase gene we observed floculation-like cell behavior in *E. coli.* In 1991, Ogden and Taylor demonstrated flocculation of *E. coli* cells by placing the *pil* operon (type 1 pili genes) under the control of a *tac* promoter-operator (Ogden and Taylor 1991). MZ1T EPS deacetylase are found to have 22% protein similarity to *E. coli* PGA deacetylase (*pgaB*). Presumably, MZ1T EPS deacetylase increase degree of deaetylation of PGA and may interact with other polysaccarrides found in the *E. coli*; for example, deaectylated PGA could bind to Colanic acid a negatively charged polymer of glucose, galactose, fucose, and glucuronic acid and promote flocculation.

5.4 EPS Genes Regulation in MZ1T

Expression of genes involved in EPS biosynthesis are often controlled by complex regulatory networks responding to a variety of environmental and physiological cues, including stress signals, nutrient availability, temperature, etc. (Arciola et al. 2015). Regulation of EPS production can take place at any level, such as transcription initiation, mRNA stability, and

protein activity. For instance, the *vps* genes, involved in EPS biosynthesis in *Vibrio cholerae*, are regulated at the transcription level by the CytR protein, in response to intracellular pyrimidine concentrations (Haugo and Watnick 2002). In our study, expression of the EPS biosynthesis *capD* located in EPS Cluster 1 increases during late exponential phase. It has been proposed that after contact of some bacteria with a surface, altered gene expression induces changes that initiate synthesis of extracellular polysaccharides since alginate, the EPS of *P. aeruginosa* biofilms, is up regulated in recently attached cells in comparison with planktonic cells (Davies and Geesey 1995). The expression of this MZ1T EPS biosynthesis gene dramatically decreases between 24-36 h of growth and surprisingly comes back at 48 h (flocculation time). This phenomenon may involve post-transcriptional regulation. For example, in *E. coli* CsrA post-transcriptionally represses *pga* expression and the production of *N*-acetyl glucosamine polymer by binding to the transcript of the *pgaA* and prevents ribosome binding, affecting *pgaABCD* mRNA stability and accelerating degradation of this transcript (Wang et al. 2005). In addition, it's possible that stress response signal(s) could participate in MZ1T flocculation since increasing of EPS biosynthesis transcript occurs in stationary phase. Evidently, global gene regulation is responsible for regulating MZ1T flocculation rather than specific EPS gene regulation because of the fact that *Thauera sp.* MZ1T lacks homologs of the EmbRS two-component system that negatively controls EPS production and biofilm formation found in *Rubrivivax gelatinosus* (Steunou et al. 2013). One should note that *Rubrivivax gelatinosus* contains EPS clusters fully conserved with *Thauera sp.* MZ1T. This can also imply that the excessive polysaccharide production in *Thauera sp* MZ1T is due to absence of specific EPS gene regulation. Moreover, the EPS deacetylase gene is constantly expressed before the flocculation and gradually reduced

indicating that this gene is necessary for floc formation in *Thauera sp.* MZ1T.

CHAPTER 6

CONCLUSION

Understanding the flocculation mechanism of *Thauera sp.* MZ1T could greatly reduce the cost of wastewater treatment operation by preventing the treatment system failure due to unsettling flocs. In this study, we show that semiconductor-based next generation sequencing (Ion Torrent) has the potential to identify the genes that cause a phenotype variant directly from sequencing of independent mutants. This direct sequencing phenotype identification is particular useful in a bacterial strain that has difficulty in genetic manipulation such as *Thauera sp.* MZ1T. Moreover, omitting cloning steps will greatly reduce cost and time consumption of the classic gene identification techniques. Thus, using a next-generation sequencing approach, future geneticists will effectively be able to merge marker discovery, mapping, and targeted mutagenesis.

We demonstrate that complemented EPS deacetylase gene (*tmz1t_3249*) in *Thauera sp.* MZ1T 20A mutant can rescue the flocculation phenotype. This discovery was facilitated by Ion Torrent read data, identifying nonsense mutations located in the middle of the gene. In addition, we found that purified EPS of the *Thauera sp.* MZ1T wild-type strain was predominately deacetylated, whereas no deacetylation was detected in the *Thauera sp.* MZ1T mutant 20A strain. This finding was confirmed by chemically deacetylation of the EPS from MZ1T 20A mutant, showing an absence of the acetyl group in the IR spectra. Taken together, for the first time, a gene involved in *Thauera sp.* MZ1T flocculation was revealed, and we propose that the EPS deacetylation gene (*tmz1t_3249*) introduce positive charges into the

polysaccharide which may bind to negative charges of carboxyl groups of the galacturonic acid residues in the polysaccharide or bacterial cell surface, promoting cell-to-cell aggregation and flocculation. Also, interestingly, the degree of EPS deacetylation in *Thauera sp.* MZ1T could play an important role in engineering this bacterium to improve settlement properties of sludge and the binding of heavy metals during wastewater treatment.

Further investigation of gene expression shows that EPS biosynthesis gene transcripts rebound during floc formation after sharp declines in late exponential phase, whereas the EPS deacetylase gene (*tmz1t_3249*) is continually expressed until flocculation takes place. Moreover, global and post-transcription regulators are likely to participate in the regulation of floc formation. However, future experiments need to be conducted to pinpoint what factors are responsible for inducing the flocculation in *T.aminoaromatica* MZ1T. This regulation could be related to environmental stress response since an important property of aggregated cells is higher tolerance to stresses such as nutrient limitation, metal toxicity and antibiotics. Again, this breakthrough could lead to optimization of sludge settling processes in wastewater plants.

Finally, identification of a controllable mechanism of flocculation may have commercial relevance in the biotechnology field. The introduction of the EPS deacetylase gene (*tmz1t_3249*) into *E. coli* revealed its ability to induce aggregation in those cells. As *E. coli* is among the most commonly used strains in the biotechnology industry for the production of biological and biochemical compounds, and purification typically first requires removal of cells, a mechanism for controlled autoaggregation and settling could have important industrial applications.

REFERENCES

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', *Environ Microbiol*, 6: 780-90.

Allen, Michael S. 2002. 'Isolation and Characterization of the Exopolysaccharide

Produced by Thauera strain MZ1T and Its Role in Flocculation', The University of Tennessee.

Arciola, C. R., D. Campoccia, S. Ravaioli, and L. Montanaro. 2015. 'Polysaccharide intercellular adhesin in biofilm: structural and regulatory aspects', *Front Cell Infect Microbiol*, 5: 7.

Arco, Y., I. Llamas, F. Martinez-Checa, M. Argandona, E. Quesada, and A. del Moral. 2005. 'epsABCJ genes are involved in the biosynthesis of the exopolysaccharide mauran produced by Halomonas maura', *Microbiology*, 151: 2841-51.

Atlas, Ronald M. 2005. *Handbook of Media for Environmental Microbiology, Second Edition* (Taylor and Francis group: 6000 Broken sound parkway, Boca Raton, FL).

Bahlawane, C., B. Baumgarth, J. Serrania, S. Ruberg, and A. Becker. 2008. 'Fine-tuning of galactoglucan biosynthesis in Sinorhizobium meliloti by differential WggR (ExpG)-, PhoB-, and MucR-dependent regulation of two promoters', *J Bacteriol*, 190: 3456-66.

Bala Subramanian, S., S. Yan, R. D. Tyagi, and R. Y. Surampalli. 2010. 'Extracellular polymeric substances (EPS) producing bacterial strains of municipal wastewater sludge: isolation, molecular identification, EPS characterization and performance for sludge settling and dewatering', *Water Res*, 44: 2253-66.

Barrere, G. C., C. E. Barber, and M. J. Daniels. 1986. 'Molecular cloning of genes involved in production of the extracelluar polysaccharide xanthan by Xanthaomonas campestris', *International Journal of Biological Macromolecules*, 8: 372-474.

Becker, A., S. Ruberg, H. Kuster, A. A. Roxlau, M. Keller, T. Ivashina, H. P. Cheng, G. C. Walker, and A. Puhler. 1997. 'The 32-kilobase exp gene cluster of Rhizobium meliloti directing the biosynthesis of galactoglucan: genetic organization and properties of the encoded gene products', *J Bacteriol*, 179: 1375-84.

Becker, B. U., K. Kosch, M. Parniske, and P. Muller. 1998. 'Exopolysaccharide (EPS) synthesis in Bradyrhizobium japonicum: sequence, operon structure and mutational analysis of an exo gene cluster', *Mol Gen Genet*, 259: 161-71.

Bragg, L. M., G. Stone, M. K. Butler, P. Hugenholtz, and G. W. Tyson. 2013. 'Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data', *PLoS Comput Biol*, 9: e1003031.

Colvin, K. M., N. Alnabelseya, P. Baker, J. C. Whitney, P. L. Howell, and M. R. Parsek. 2013. 'PelA deacetylase activity is required for Pel polysaccharide synthesis in Pseudomonas aeruginosa', *J Bacteriol*, 195: 2329-39.

Dark, M. J. 2013. 'Whole-genome sequencing in bacteriology: state of the art', *Infect Drug Resist*, 6: 115-23.

Davies, D. G., and G. G. Geesey. 1995. 'Regulation of the alginate biosynthesis gene algC in Pseudomonas aeruginosa during biofilm development in continuous culture', *Appl Environ Microbiol*, 61: 860-7.

Davis, B. M., and M. K. Waldor. 2009. 'High-throughput sequencing reveals suppressors of Vibrio cholerae rpoE mutations: one fewer porin is enough', *Nucleic Acids Res*, 37: 5757-67.

De Schryver, P., R. Crab, T. Defoirdt, N. Boon, and W. Verstraete. 2008. 'The basics of bio-flocs technology: The added value for aquaculture', *Aquaculture*, 277: 125-37.

Dimopoulou, M., M. Vuillemin, H. Campbell-Sills, P. M. Lucas, P. Ballestra, C. Miot-Sertier, M. Favier, J. Coulon, V. Moine, T. Doco, M. Roques, P. Williams, M. Petrel, E. Gontier, C. Moulis, M. Remaud-Simeon, and M. Dols-Lafargue. 2014. 'Exopolysaccharide (EPS) synthesis by Oenococcus oeni: from genes to phenotypes', *PLoS One*, 9: e98898.

Easson, D. D., Jr., A. J. Sinskey, and O. P. Peoples. 1987. 'Isolation of Zoogloea ramigera I-16-M exopolysaccharide biosynthetic genes and evidence for instability within this region', *J Bacteriol*, 169: 4518-24.

Fazli, M., Y. McCarthy, M. Givskov, R. P. Ryan, and T. Tolker-Nielsen. 2013. 'The exopolysaccharide gene cluster Bcam1330-Bcam1341 is involved in Burkholderia cenocepacia biofilm formation, and its expression is regulated by c-di-GMP and Bcam1349', *Microbiologyopen*, 2: 105-22.

Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, and A. Puhler. 2001. 'The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti', *Proc Natl Acad Sci U S A*, 98: 9889-94.

Flibotte, S., M. L. Edgley, I. Chaudhry, J. Taylor, S. E. Neil, A. Rogula, R. Zapf, M. Hirst, Y. Butterfield, S. J. Jones, M. A. Marra, R. J. Barstead, and D. G. Moerman. 2010. 'Whole-genome profiling of mutagenesis in Caenorhabditis elegans', *Genetics*, 185: 431-41.

Franklin, M. J., and D. E. Ohman. 2002. 'Mutant analysis and cellular localization of the AlgI, AlgJ, and AlgF proteins required for O acetylation of alginate in Pseudomonas aeruginosa', *J Bacteriol*, 184: 3000-7.

Glucksmann, M. A., T. L. Reuber, and G. C. Walker. 1993. 'Genes needed for the modification, polymerization, export, and processing of succinoglycan by Rhizobium meliloti: a model for succinoglycan biosynthesis', *J Bacteriol*, 175: 7045-55.

Govender, P., J. L. Domingo, M. C. Bester, I. S. Pretorius, and F. F. Bauer. 2008. 'Controlled expression of the dominant flocculation genes FLO1, FLO5, and FLO11 in Saccharomyces cerevisiae', *Appl Environ Microbiol*, 74: 6041-52.

Haag, AnthonyP. 2006. 'Mechanical Properties of Bacterial Exopolymeric Adhesives and their Commercial Development.' in AndrewM Smith and JamesA Callow (eds.), *Biological Adhesives* (Springer Berlin Heidelberg).

Harding, N. E., J. M. Cleary, D. K. Cabanas, I. G. Rosen, and K. S. Kang. 1987. 'Genetic and physical analyses of a cluster of genes essential for xanthan gum biosynthesis in Xanthomonas campestris', *J Bacteriol*: 2854-61.

Harper, M. A., Z. Chen, T. Toy, I. M. Machado, S. F. Nelson, J. C. Liao, and C. J. Lee. 2011. 'Phenotype sequencing: identifying the genes that cause a phenotype directly from pooled sequencing of independent mutants', *PLoS One*, 6: e16517.

Haugo, A. J., and P. I. Watnick. 2002. 'Vibrio cholerae CytR is a repressor of biofilm development', *Mol Microbiol*, 45: 471-83.

Hay, Iain D., Zahid Ur Rehman, Aamir Ghafoor, and Bernd H. A. Rehm. 2010. 'Bacterial biosynthesis of alginates', *Journal of Chemical Technology & Biotechnology*, 85: 752-59.

Hogye, S., P. Casey, M. Noah, and E. Winant. 2003. 'Explaining the actived sludge process', *Pipeline*, 14: 2-7.

Irvine, D. V., D. B. Goto, M. W. Vaughn, Y. Nakaseko, W. R. McCombie, M. Yanagida, and R. Martienssen. 2009. 'Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing', *Genome Res*, 19: 1077-83.

Itoh, Y., J. D. Rice, C. Goller, A. Pannuri, J. Taylor, J. Meisner, T. J. Beveridge, J. F. Preston, 3rd, and T. Romeo. 2008. 'Roles of pgaABCD genes in synthesis, modification, and export of the Escherichia coli biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine', *J Bacteriol*, 190: 3670-80.

Ivashina, T. V., M. I. Khmelnitsky, M. G. Shlyapnikov, A. A. Kanapin, and V. N. Ksenzenko. 1994. 'The pss4 gene from Rhizobium leguminosarum by viciae VF39: cloning, sequence and the possible role in polysaccharide production and nodule formation', *Gene*, 150: 111-6.

Janczarek, M., M. Kalita, and A. M. Skorupska. 2009. 'New taxonomic markers for identification of Rhizobium leguminosarum and discrimination between closely related species', *Arch Microbiol*, 191: 207-19.

Janczarek, M., and K. Rachwal. 2013. 'Mutation in the pssA gene involved in exopolysaccharide synthesis leads to several physiological and symbiotic defects in Rhizobium leguminosarum bv. trifolii', *Int J Mol Sci*, 14: 23711-35.

Jiang, K., J. Sanseverino, A. Chauhan, S. Lucas, A. Copeland, A. Lapidus, T. G. Del Rio, E. Dalin, H. Tice, D. Bruce, L. Goodwin, S. Pitluck, D. Sims, T. Brettin, J. C. Detter, C. Han, Y. J. Chang, F. Larimer, M. Land, L. Hauser, N. C. Kyrpides, N. Mikhailova, S. Moser, P. Jegier, D. Close, J. M. Debruyn, Y. Wang, A. C. Layton, M. S. Allen, and G. S. Sayler. 2012. 'Complete genome sequence of Thauera aminoaromatica strain MZ1T', *Stand Genomic Sci*, 6: 325-35.

Jin, Bo, Britt-Marie Wilén, and Paul Lant. 2003. 'A comprehensive insight into floc characteristics and their impact on compressibility and settleability of activated sludge', *Chemical Engineering Journal*, 95: 221-34.

Kearns, D. B., F. Chu, S. S. Branda, R. Kolter, and R. Losick. 2005. 'A master regulator for biofilm formation by Bacillus subtilis', *Mol Microbiol*, 55: 739-49.

Keiski, L. C., H. Michael, J. Sumita, M. N. Ana, Y. Patrick, J. Howard, J. C. Whitney, R. Laura, L. Lori, Burrows, Dennis E., Ohman, and P. Lynne Howell. 2010. 'AlgK is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin', *Structure*, 18: 265-73.

Kim, H. S., M. A. Lee, S. J. Chun, S. J. Park, and K. H. Lee. 2007. 'Role of NtrC in biofilm formation via controlling expression of the gene encoding an ADP-glycero-manno-heptose-6-epimerase in the pathogenic bacterium, Vibrio vulnificus', *Mol Microbiol*, 63: 559-74.

Kohn, Andrea B., Tatiana P. Moroz, Jeffrey P. Barnes, Mandy Netherton, and Leonid L. Moroz. 2013. 'Single-Cell Semiconductor Sequencing', *Methods in molecular biology (Clifton, N.J.)*, 1048: 247-84.

Król, Jarosław E., Andrzej Mazur, Małgorzata Marczak, and Anna Skorupska. 2007. 'Syntenic arrangements of the surface polysaccharide biosynthesis genes in Rhizobium leguminosarum', *Genomics*, 89: 237-47.

Kropec, A., T. Maira-Litran, K. K. Jefferson, M. Grout, S. E. Cramton, F. Gotz, D. A. Goldmann, and G. B. Pier. 2005. 'Poly-N-acetylglucosamine production in Staphylococcus aureus is essential for virulence in murine models of systemic infection', *Infect Immun*, 73: 6868-76.

Ksenzenko, V. N., T. V. Ivashina, Z. A. Dubeikovskaia, S. G. Ivanov, M. B. Nanazashvili, T. N. Druzhinina, N. A. Kalinchuk, and V. N. Shibaev. 2007. 'The pssA gene encodes UDP-glucose: polyprenyl phosphate-glucosyl phosphotransferase initiating biosynthesis of Rhizobium leguminosarum exopolysaccharide', *Bioorg Khim*, 33: 160-6.

Lajoie, A., A. C. Layton, R. I. Gregory, G. S. Sayler, J. Taylor, and A. J. Meyers. 2000. 'Zoogloeal clusters and sludge dewatering potential in an industrial activated sludge weastewater tretment plant', *Water Environ Res*, 72: 56-64.

Latasa, Cristina, Cristina Solano, José R. Penadés, and Iñigo Lasa. 2006. 'Biofilm-associated proteins', *Comptes Rendus Biologies*, 329: 849-57.

Lee, V. T., J. M. Matewish, J. L. Kessler, M. Hyodo, Y. Hayakawa, and S. Lory. 2007. 'A cyclic-di-GMP receptor required for bacterial exopolysaccharide production', *Mol Microbiol*, 65: 1474-84.

Loman, Nicholas J., Raju V. Misra, Timothy J. Dallman, Chrystala Constantinidou, Saheer E. Gharbia, John Wain, and Mark J. Pallen. 2012. 'Performance comparison of benchtop high-throughput sequencing platforms', *Nat Biotech*, 30: 434-39.

Lu, A., K. Cho, W. P. Black, X. Y. Duan, R. Lux, Z. Yang, H. B. Kaplan, D. R. Zusman, and W. Shi. 2005. 'Exopolysaccharide biosynthesis genes required for social motility in Myxococcus xanthus', *Mol Microbiol*, 55: 206-20.

Lu, F., J. Lukasik, and S. R. Farrah. 2001. 'Immunological methods for the study of Zoogloea strains in natural environments', *Water Res*, 35: 4011-8.

Lupski, J. R., J. G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D. C. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D. A. Wheeler, A. L. McGuire, F. Zhang, P. Stankiewicz, J. J. Halperin, C. Yang, C. Gehman, D. Guo, R. K. Irikat, W. Tom, N. J. Fantin, D. M. Muzny, and R. A. Gibbs. 2010. 'Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy', *N Engl J Med*, 362: 1181-91.

Marvasi, M., P. T. Visscher, and L. Casillas Martinez. 2010. 'Exopolymeric substances (EPS) from Bacillus subtilis: polymers and genes encoding their synthesis', *FEMS Microbiol Lett*, 313: 1-9.

Masuko, T., A. Minami, N. Iwasaki, T. Majima, S. Nishimura, and Y. C. Lee. 2005. 'Carbohydrate analysis by a phenol-sulfuric acid method in microplate format', *Anal Biochem*, 339: 69-72.

McCluskey, K., A. E. Wiest, I. V. Grigoriev, A. Lipzen, J. Martin, W. Schackwitz, and S. E. Baker. 2011. 'Rediscovery by Whole Genome Sequencing: Classical Mutations and Genome Polymorphisms in Neurospora crassa', *G3 (Bethesda)*, 1: 303-16.

Mellmann, A., D. Harmsen, C. A. Cummings, E. B. Zentz, S. R. Leopold, A. Rico, K. Prior, R. Szczepanowski, Y. Ji, W. Zhang, S. F. McLaughlin, J. K. Henkhaus, B. Leopold, M. Bielaszewska, R. Prager, P. M. Brzoska, R. L. Moore, S. Guenther, J. M. Rothberg, and H.

Karch. 2011. 'Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology', *PLoS One*, 6: e22751.

Montoya, T., L. Borras, D. Aguado, J. Ferrer, and A. Seco. 2008. 'Detection and prevention of enhanced biological phosphorus removal deterioration caused by Zoogloea overabundance', *Environ Technol*, 29: 35-42.

Mora, Paola, Federico Rosconi, Laura Franco Fragus, and Susana Castro Sowinski. 2008. 'Azospirillum brasilense Sp7 produces an outer-membrane lectin that specifically binds to surface-exposed extracellular polysaccharide produced by the bacterium', *Arch Microbiol*, 189: 519-24.

Moreira, L. M., P. A. Videira, S. A. Sousa, J. H. Leitao, M. V. Cunha, and I. Sa-Correia. 2003. 'Identification and physical organization of the gene cluster involved in the biosynthesis of Burkholderia cepacia complex exopolysaccharide', *Biochem Biophys Res Commun*, 312: 323-33.

Ogden, KimberlyL, and AustinL Taylor. 1991. 'Genetic control of flocculation inEscherichia coli', *Journal of Industrial Microbiology*, 7: 279-86.

Ohnishi, J., H. Mizoguchi, S. Takeno, and M. Ikeda. 2008. 'Characterization of mutations induced by N-methyl-N'-nitro-N-nitrosoguanidine in an industrial Corynebacterium glutamicum strain', *Mutat Res*, 649: 239-44.

Pareek, C. S., R. Smoczynski, and A. Tretyn. 2011. 'Sequencing technologies and genome sequencing', *J Appl Genet*, 52: 413-35.

Pomraning, K. R., K. M. Smith, and M. Freitag. 2011. 'Bulk segregant analysis followed by high-throughput sequencing reveals the Neurospora cell cycle gene, ndc-1, to be allelic with the gene for ornithine decarboxylase, spe-1', *Eukaryot Cell*, 10: 724-33.

Puente, X. S., M. Pinyol, V. Quesada, L. Conde, G. R. Ordonez, N. Villamor, G. Escaramis, P. Jares, S. Bea, M. Gonzalez-Diaz, L. Bassaganyas, T. Baumann, M. Juan, M. Lopez-Guerra, D. Colomer, J. M. Tubio, C. Lopez, A. Navarro, C. Tornador, M. Aymerich, M. Rozman, J. M. Hernandez, D. A. Puente, J. M. Freije, G. Velasco, A. Gutierrez-Fernandez, D. Costa, A. Carrio, S. Guijarro, A. Enjuanes, L. Hernandez, J. Yague, P. Nicolas, C. M. Romeo-Casabona, H. Himmelbauer, E. Castillo, J. C. Dohm, S. de Sanjose, M. A. Piris, E. de Alava, J. San Miguel, R. Royo, J. L. Gelpi, D. Torrents, M. Orozco, D. G. Pisano, A. Valencia, R. Guigo, M. Bayes, S. Heath, M. Gut, P. Klatt, J. Marshall, K. Raine, L. A. Stebbings, P. A. Futreal, M. R. Stratton, P. J. Campbell, I. Gut, A. Lopez-Guillermo, X. Estivill, E. Montserrat, C. Lopez-Otin, and E. Campo. 2011. 'Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia', *Nature*, 475: 101-5.

Rabus, R., and F. Widdel. 1995. 'Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria', *Arch Microbiol*, 163: 96-103.

Reed, J. W., and G. C. Walker. 1991. 'The exoD gene of Rhizobium meliloti encodes a novel function needed for alfalfa nodule invasion', *J Bacteriol*, 173: 664-77.

Rehm, Bernd H. A. 2010. 'Bacterial polymers: biosynthesis, modifications and applications', *applied and industrial microbiology*, 8: 578-92.

Reuber, T. L., and G. C. Walker. 1993. 'Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of Rhizobium meliloti', *Cell*, 74: 269-80.

Ridout, M. J., G. J. Brownsey, G. M. York, G. C. Walker, and V. J. Morris. 1997. 'Effect of o-acyl substituents on the functional behaviour of Rhizobium meliloti succinoglycan', *Int J Biol Macromol*, 20: 1-7.

Rinaudo, M. 2004. 'Role of substituents on the properties of some polysaccharides', *Biomacromolecules*, 5: 1155-65.

Roach, J. C., G. Glusman, A. F. Smit, C. D. Huff, R. Hubley, P. T. Shannon, L. Rowen, K. P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L. B. Jorde, L. Hood, and D. J. Galas. 2010. 'Analysis of genetic inheritance in a family quartet by whole-genome sequencing', *Science*, 328: 636-9.

Rothberg, Jonathan M., Wolfgang Hinz, Todd M. Rearick, Jonathan Schultz, William Mileski, Mel Davey, John H. Leamon, Kim Johnson, Mark J. Milgrew, Matthew Edwards, Jeremy Hoon, Jan F. Simons, David Marran, Jason W. Myers, John F. Davidson, Annika Branting, John R. Nobile, Bernard P. Puc, David Light, Travis A. Clark, Martin Huber, Jeffrey T. Branciforte, Isaac B. Stoner, Simon E. Cawley, Michael Lyons, Yutao Fu, Nils Homer, Marina Sedova, Xin Miao, Brian Reed, Jeffrey Sabina, Erika Feierstein, Michelle Schorn, Mohammad Alanjary, Eileen Dimalanta, Devin Dressman, Rachel Kasinskas, Tanya Sokolsky, Jacqueline A. Fidanza, Eugeni Namsaraev, Kevin J. McKernan, Alan Williams, G. Thomas Roth, and James Bustillo. 2011. 'An integrated semiconductor device enabling non-optical genome sequencing', *Nature*, 475: 348-52.

Sarin, S., S. Prabhu, M. M. O'Meara, I. Pe'er, and O. Hobert. 2008. 'Caenorhabditis elegans mutant allele identification by whole-genome sequencing', *Nat Methods*, 5: 865-7.

Schmeisser, C., H. Liesegang, D. Krysciak, N. Bakkou, A. Le Quere, A. Wollherr, I. Heinemeyer, B. Morgenstern, A. Pommerening-Roser, M. Flores, R. Palacios, S. Brenner, G. Gottschalk, R. A. Schmitz, W. J. Broughton, X. Perret, A. W. Strittmatter, and W. R. Streit. 2009. 'Rhizobium sp. strain NGR234 possesses a remarkable number of secretion systems', *Appl Environ Microbiol*, 75: 4035-45.

Smith, D. R., A. R. Quinlan, H. E. Peckham, K. Makowsky, W. Tao, B. Woolf, L. Shen, W. F. Donahue, N. Tusneem, M. P. Stromberg, D. A. Stewart, L. Zhang, S. S. Ranade, J. B. Warner, C. C. Lee, B. E. Coleman, Z. Zhang, S. F. McLaughlin, J. A. Malek, J. M. Sorenson, A. P. Blanchard, J. Chapman, D. Hillman, F. Chen, D. S. Rokhsar, K. J. McKernan, T. W. Jeffries, G. T. Marth, and P. M. Richardson. 2008. 'Rapid whole-genome mutational profiling using next-generation sequencing technologies', *Genome Res*, 18: 1638-42.

Sorlier, Pierre, Anne Denuzière, Christophe Viton, and Alain Domard. 2001. 'Relation between the Degree of Acetylation and the Electrostatic Properties of Chitin and Chitosan', *Biomacromolecules*, 2: 765-72.

Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang, and R. Chen. 2008. 'High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies', *PLoS Genet*, 4: e1000139.

Standal, R., T. G. Iversen, D. H. Coucheron, E. Fjaervik, J. M. Blatny, and S. Valla. 1994. 'A new gene required for cellulose production and a gene encoding cellulolytic activity in Acetobacter xylinum are colocalized with the bcs operon', *J Bacteriol*, 176: 665-72.

Starkey, Melissa, Matthew R. Parsek, Kimberly A. Gray, and Sung Il Chang. 2004. 'A Sticky Business: the Extracellular Polymeric Substance Matrix of Bacterial Biofilms.' in, *Microbial Biofilms* (American Society of Microbiology).

Steunou, Anne Soisig, Sylviane Liotenberg, Marie-Noêlle Soler, Romain Briandet, Valérie Barbe, Chantal Astier, and Soufian Ouchane. 2013. 'EmbRS a new two-component system that inhibits biofilm formation and saves Rubrivivax gelatinosus from sinking', *MicrobiologyOpen*, 2: 431-46.

Streit, W. R., R. A. Schmitz, X. Perret, C. Staehelin, W. J. Deakin, C. Raasch, H. Liesegang, and W. J. Broughton. 2004. 'An evolutionary hot spot: the pNGR234b replicon of Rhizobium sp. strain NGR234', *J Bacteriol*, 186: 535-42.

Sutherland, Ian W. 2001. 'Microbial polysaccharides fromGram -negative bacteria', *International Dairy Journal* 11: 663–74.

Thorne, L., L. Tansey, and T. J. Pollock. 1989. 'Clustering of mutations blocking synthesis of xanthan gum by Xanthomonas campestris ', *J Bacteriol*, 169: 3593-600.

Tielen, P., M. Strathmann, K. E. Jaeger, H. C. Flemming, and J. Wingender. 2005. 'Alginate acetylation influences initial surface colonization by mucoid Pseudomonas aeruginosa', *Microbiol Res*, 160: 165-76.

Uttaro, A. D., G. A. Cangelosi, R. A. Geremia, E. W. Nester, and R. A. Ugalde. 1990. 'Biochemical characterization of avirulent exoC mutants of Agrobacterium tumefaciens', *J Bacteriol*, 172: 1640-6.

van Workum, W. A., H. C. Canter Cremers, A. H. Wijfjes, C. van der Kolk, C. A. Wijffelman, and J. W. Kijne. 1997. 'Cloning and characterization of four genes of Rhizobium leguminosarum

bv. trifolii involved in exopolysaccharide production and nodulation', *Mol Plant Microbe Interact*, 10: 290-301.

Villain-Simonnet, A., M. Milas, and M. Rinaudo. 2000. 'A new bacterial polysaccharide (YAS34). I. Characterization of the conformations and conformational transition', *Int J Biol Macromol*, 27: 65-75.

Vorholter, F. J., S. Schneiker, A. Goesmann, L. Krause, T. Bekel, O. Kaiser, B. Linke, T. Patschkowski, C. Ruckert, J. Schmid, V. K. Sidhu, V. Sieber, A. Tauch, S. A. Watt, B. Weisshaar, A. Becker, K. Niehaus, and A. Puhler. 2008. 'The genome of Xanthomonas campestris pv. campestris B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis', *J Biotechnol*, 134: 33-45.

Vuong, C., S. Kocianova, J. M. Voyich, Y. Yao, E. R. Fischer, F. R. DeLeo, and M. Otto. 2004. 'A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence', *J Biol Chem*, 279: 54881-6.

Wang, X., A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke, and T. Romeo. 2005. 'CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli', *Mol Microbiol*, 56: 1648-63.

Wetterstrand, Kris. 2015. "DNA Sequencing Costs." In *2001-2015*. http://www.genome.gov/sequencingcosts/: NHGRI Genome Sequencing Program (GSP).

Whitfield, Chris. 2006. 'Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli', *Annu. Rev. Biochem.*, 75: 39-68.

Whitney, J.C., and P.L. Howell. 2013. 'Synthase-dependent exopolysaccharide secretion in Gramnegative Bacteria', *Trends Microbiol*, 21: 63–72.

Wu, Y., J. Wang, T. Xu, J. Liu, W. Yu, Q. Lou, T. Zhu, N. He, H. Ben, J. Hu, F. Gotz, and D. Qu. 2012. 'The two-component signal transduction system ArlRS regulates Staphylococcus epidermidis biofilm formation in an ica-dependent manner', *PLoS One*, 7: e40041.

Wurtzel, O., M. Dori-Bachash, S. Pietrokovski, E. Jurkevitch, and R. Sorek. 2010. 'Mutation detection with next-generation resequencing through a mediator genome', *PLoS One*, 5: e15628.

Yan, S., N. Wang, Z. Chen, Y. Wang, N. He, Y. Peng, Q. Li, and X. Deng. 2013. 'Genes encoding the production of extracellular polysaccharide bioflocculant are clustered on a 30-kb DNA segment in Bacillus licheniformis', *Funct Integr Genomics*, 13: 425-34.

Yao, S. Y., L. Luo, K. J. Har, A. Becker, S. Ruberg, G. Q. Yu, J. B. Zhu, and H. P. Cheng. 2004. 'Sinorhizobium meliloti ExoR and ExoS proteins regulate both succinoglycan and flagellum production', *J Bacteriol*, 186: 6042-9.

Zhan, H. J., J. X. Gray, S. B. Levery, B. G. Rolfe, and J. A. Leigh. 1990. 'Functional and evolutionary relatedness of genes for exopolysaccharide synthesis in Rhizobium meliloti and Rhizobium sp. strain NGR234', *J Bacteriol*, 172: 5245-53.

Zuryn, S., S. Le Gras, K. Jamet, and S. Jarriault. 2010. 'A strategy for direct mapping and identification of mutations by whole-genome sequencing', *Genetics*, 186: 427-30.