Molecular mechanisms of planktonic cellular aggregation in *Burkholderia multivorans*

Mirela Rodrigues Ferreira

Thesis to obtain the Master of Science Degree in

Microbiology

Supervisor: Doctor Leonilde de Fátima Morais Moreira

Examination Committee:

Chairperson: Doctor Jorge Humberto Gomes Leitão

Supervisor: Doctor Leonilde de Fátima Morais Moreira

Members of the Committee: Doctor Arsénio do Carmo Sales Mendes Fialho

November 2018

ACKNOWLEDGEMENTS

This work was financed by Programa Operacional Regional de Lisboa 2020 (LISBOA-01-0145-FEDER-007317).

First of all, I would like to thank my supervisor Prof. Leonilde Moreira for all support and guidance during my master thesis.

I would like to acknowledge Prof. David P. Speert from the University of British Columbia, Canada, for providing the clinical isolate used in this work.

I would also like to thank every member of the Biological Sciences Research Group (BSRG), who have contributed in some way to make this work possible to achieve.

Obviously, I cannot forget to thank all my friends, who were present during this year: Sara Gomes, for the adventures in the laboratory, for having been there when despair appeared, when giving up seemed easier. Marta Mota for all the support, especially in the last month of master thesis in which none of us gave up. Inês Oliveira, because, although we have not seen each other every day, you have always been present throughout this year to cheer me up in the most difficult times. And finally, António Seixas, Bernardo Balugas and Nuno Melo, thank you for the most fun moments I spent in the lab.

Finalmente, queria agradecer a minha família, aos meus pais e à minha irmã em especial, por mais uma vez estarem presentes e por sempre me ter em dado a força e a motivação para continuar a fazer o meu melhor.

RESUMO

As bactérias podem proliferar como células livres ou em comunidade. Um exemplo destas comunidades são os biofilmes, os quais estão associados a uma superfície ou os agregados celulares planctónicos que se encontram livres. A formação de agregados celulares planctónicos é relevante em ambientes naturais, mas também durante a interação com células hospedeiras resultando muitas vezes em patogenicidade. É o caso de infeções respiratórias crónicas em doentes de fibrose quística (FQ), muitas vezes causadas por *Pseudomonas aeruginosa* em bactérias pertencentes ao complexo *Burkholderia cepacia* (*Bc*c) e que são capazes de formar tais agregados celulares. Embora se conheçam alguns mecanismos moleculares envolvidos na formação de agregados celulares planctónicos por certas bactérias, no complexo *Bc*c, estes permanecem desconhecidos. Por essa razão, usou-se um isolado clínico de *Burkholderia multivorans* proveniente de um doente com FQ para construir uma biblioteca de mutantes usando um transposão. Analisaram-se cerca de 900 mutantes, tendo 15 deles sido selecionados com base no seu fenótipo de agregação. Caracterizou-se cada um desses mutantes quanto à capacidade de formação de biofilme associados a superfícies, à motilidade, à produção de exopolissacáridos, à suscetibilidade a antibióticos, à taxa de crescimento em meio sintético e à virulência em *Galleria mellonella*. Considerando os resultados obtidos, foi impossível estabelecer uma relação linear entre cada fenótipo e a capacidade de agregação. Contudo, foi possível identificar prováveis intervenientes envolvidos na formação deste tipo de agregados, entre os quais estão genes envolvidos em sinalização, reguladores da transcrição, diversas proteínas membranares e proteínas envolvidas no metabolismo de lípidos.

PALAVRAS CHAVES

Agregados celulares planctónicos, Fibrose quística,
Burkholderia multivorans, Biblioteca de mutantes

ABSTRACT

Bacteria can proliferate as planktonic cells or as sessile communities. One example of these communities are biofilms, in which cells are attached to a surface or can form non-attached planktonic cellular aggregates. The formation of planktonic cellular aggregates is of relevance in natural environments, but also during interaction with host cells resulting in pathogenicity. In patients with cystic fibrosis (CF), microorganisms from the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* can form such planktonic cellular aggregates and often establish a chronic infection. Although some molecular mechanisms involved in the formation of planktonic cellular aggregates in bacteria are known, in *Burkholderia cepacia* complex*,* these mechanisms remain poorly understood. For that reason, it was used a *Burkholderia multivorans* cystic fibrosis isolate to construct a transposon mutant library. About 900 mutants were screened and 15 of them selected due to their distinct aggregation phenotype. Each of these mutants was characterized concerning phenotypes such as surface-attached biofilm formation, motility, exopolysaccharide production, antibiotic susceptibility, growth rates in synthetic cystic fibrosis medium, and virulence in *Galleria mellonella*. Although it was impossible to establish a linear relationship between each phenotype and the aggregation phenotype, it was still possible to identify probable players involved in the formation of aggregates. Among these players are intracellular messengers, transcriptional regulators, ribosomal proteins, membrane proteins, and proteins involved in lipid metabolism.

KEYWORDS

Planktonic cellular aggregates, Cystic fibrosis,
Burkholderia multivorans, Transposon mutant library

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CF - Cystic Fibrosis

FQ – Fibrose Quística

CFTR - Cystic Fibrosis Transmembrane Conductance Regulator

EPS - Extracellular Polymeric Substances

eDNA - Extracellular DNA

SAB - Surface-Attached Biofilm

PCA - Planktonic Cellular Aggregates

TCS - Two-component system

RR - Response Regulator

TCP - Toxin Co-regulated Pilus

PIA - Polysaccharide Extracellular Adhesin

HCD - High Cell Densities

AHL - Acylated homoserine lactone

AI-2 - Autoinducer 2

ABC - ATP-Binding Cassette

NBD - Nucleotide Binding Domains

ENaC - Epithelial Na^+ Channel

ASL - Airway Liquid Surface

Bcc – *Burkholderia cepacia* complex

pO_2 - Partial Pressure of Oxygen

ROS - Reactive Oxygen Species

RNS - Reactive Nitrogen Species

LdhA - D-Lactate Dehydrogenase A

HSL - Homoserine Lactone

LTTR - LysR - Type Transcriptional Regulator

BvlR - Biofilm and Virulence LysR Regulator

LdhR - Lactate Dehydrogenase Regulator

SCFM – Synthetic Cystic Fibrosis Medium

LB -Lennox Broth

BLAST – Basic Local Alignment Search Tool

NCBI – National Center for Biotechnology Information

YEM – Yeast Extract Mannitol

(p)ppGpp - Guanosine Tetraphosphate and Guanosine Pentaphosphate

GMP - Guanosine Monophosphate

GTP - Guanosine Triophosphate

c-di-GMP - Cyclic Dimeric Guanosine Monophosphate

UMP – Uridine Monophosphate

EBP – Enhancer Binding Protein

1 INTRODUCTION

Bacteria can live as free cells or as sessile cells. The transition from the free-swimming cells to microbial complexes mode of growth happens due to changes in environmental conditions. The microbial complexes formed by bacteria are known as biofilms (1). Biofilms can be described as a community of microorganisms enclosed within a matrix of exopolysaccharides (2). In biofilms, cells can be attached to a surface or can form planktonic cellular aggregates. The first one corresponds to the most studied biofilm model; the second is formed when bacteria adhere to each other and not to a surface (3).

1.1 Bacterial biofilms

The existence of microbial biofilms appeared for the first time documented in the seventeenth century, when Antoine Van Leeuwenhoek observed in his own teeth *"animacules"* deposited. Now, we know that these creatures were bacteria of dental plaque (4).

Nowadays, the advances of microscopy allow us to define microbial biofilms as a community of interacting unicellular organisms within an enclosed matrix (1). In 1994, Costerton *et al*. (2) described the basic structure of a biofilm as the local where bacteria grow in microcolonies within the exopolysaccharide matrix. Between the microcolonies there is a "circulatory system" formed by filled-water channels. This system provides nutrients to microcolonies and allows the efflux of metabolic products from the biofilm (Fig. 1). Bacterial biofilms can be found in different contexts. From the perspective of man, biofilms can represent a useful tool at the biotechnological and industrial level (5) and a major problem at the clinical level (6).

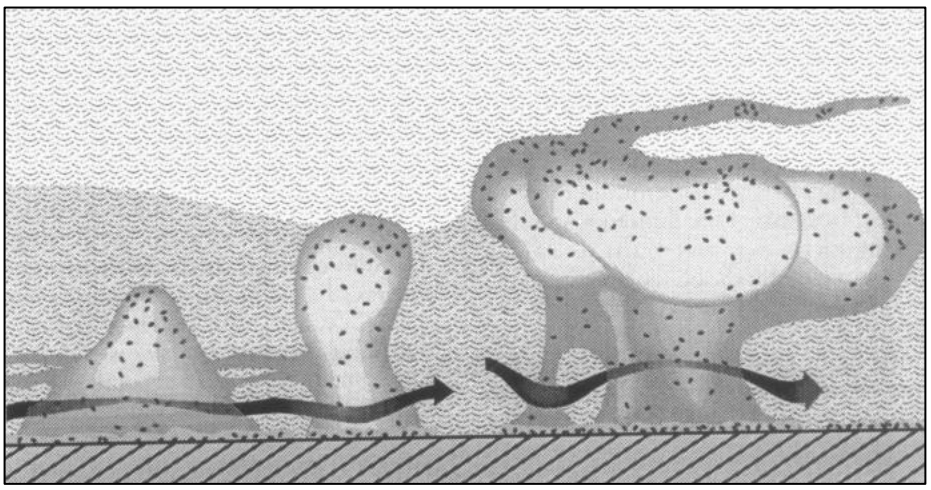

Fig. 1 The basic structure of a biofilm. Bacteria grow in microcolonies within the exopolysaccharide matrix. Between the microcolonies there is a "circulatory system" formed by filled-water channels (retrieved from (2)).

As said before, biofilms are a community of microorganisms embedded in a self-produced matrix of extracellular polymeric substances (EPS). This matrix is mainly composed of polysaccharides, proteins, lipids, ions and extracellular DNA (eDNA) and confers the architecture and stability to

the biofilm (7). When compared to planktonic cells there are some properties that emerge when a biofilm is formed. Some of them include: the ability for horizontal gene transfer; the intracellular communication through chemical signals, i.e., molecules used in quorum sensing or electrical signals; the resistance to dissection, and the protection from antimicrobial agents, among others (1, 7).

In the literature, there is some divergence about biofilms definition, but in this work, it will be considered two types of biofilms: (1) surface-attached biofilm (SAB) and (2) planktonic cellular aggregates (PCA), a non-attached biofilm also known as tight microcolonies.

1.1.1 Surface-attached biofilms

The surface-attached biofilms are formed by cells attached to a solid biotic or abiotic surface. This type of biofilm is the most studied one and the current conventional formation model is shown in figure 2. This model includes five classical stages of development: the first one begins with cells adhesion to the surface, followed by the irreversible attachment of the cells. At that time, cells start to grow and divide, until the formation of a mature microcolony. A group of mature microcolonies forms a biofilm. This biofilm allows bacteria to cope with environmental stresses such as the presence of antibiotics, the lower concentration of nutrients, among others. Some of the cells that compose the biofilm, when sensing better conditions, are released and the process of dispersion occurs (8).

Although, this is the basic model of biofilm formation, and therefore the best known and studied, there are authors defending this should be reviewed. These authors argue that the basic model of biofilm formation should consider both planktonic cellular aggregates and planktonic (free) cells as initiators of biofilm formation and as forms of dispersion (Fig. 2) (8, 9). However, both studies confirmed, through computational simulations, that not always is more advantageous the biofilm to be seeded by planktonic cellular aggregates. This is explained by the fact that relative fitness of aggregates depends on the density of surrounding planktonic cells. Planktonic cells work better than aggregates when the competition is low, because the cells have access to more quantity of nutrients. But, when the competition is high, i.e., when there are a lot of surrounding unaggregated cells, cellular aggregates exhibits higher fitness because of extending vertically above the surface, and this allows cells at the top of aggregates to have better access to growth resources (8, 9).

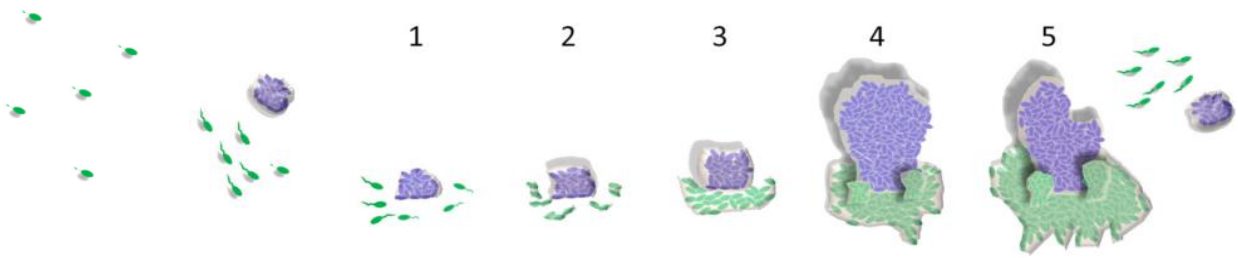

Fig. 2 The formation of surface-attached biofilms. (Green) The current conventional biofilm formation model comprises 5 different steps: 1. Planktonic cells adhere to the surface; 2. Irreversible attachment of the cells; 3. Cells start to grow and divide; 4. Form a mature microcolony; 5. Biofilm dispersion by planktonic cells. **(Purple)** The proposal biofilm formation's model comprises 5 different steps: 1. Planktonic cellular aggregates adhere to the surface; 2. Irreversible attachment of the aggregates; 3. Cells in the aggregate

start to grow and divide; 4. Form a mature microcolony; 5. Biofilm dispersion by planktonic cellular aggregates (retrieved from (8)).

In addition to the higher fitness of planktonic cellular aggregates under particular conditions, these aggregates, as initiators of biofilm formation, have other advantages, namely, the fact that the aggregates already present resistance to antibiotics and some phenotypic adaptations that are usually associated with biofilms (8).

1.1.2 Planktonic cellular aggregates

Planktonic cellular aggregates are non-attached biofilms also known as tight microcolonies in which bacteria adhere to each other but not to a surface (3). In wastewater, for example, some bacteria can form this type of aggregates, however, in these cases the aggregates are called clumps and/or flocs (5). Usually clumping is a prerequisite to flocculation, and flocculation requires the encapsulation of the cells in a matrix of exopolysaccharides (10).

The formation of planktonic cellular aggregates is of relevance in natural environments, but also during interaction with host cells resulting in pathogenicity. As previous studies reported the ability of *Burkholderia cepacia* complex bacteria and *Pseudomonas aeruginosa* to form such planktonic cellular aggregates (11), and knowing these microorganisms often establish chronic infection in cystic fibrosis patients, it is relevant and important to discuss what is known about the formation of planktonic cellular aggregates in bacteria (12, 13).

Similar to biofilms, aggregates represent some advantages such as, protection against antibiotics and the hostile environment. Nevertheless, there are evidences that cells in aggregates differ in some properties from cells in biofilms new phenotypic properties, such as higher metabolic activity than planktonic cells and cells in surface-attached biofilms, arise when cells form cellular aggregates, but in terms of mutation rate, this one is similar to the mutation rate of surface-attached biofilms when compare to planktonic cells (6).

Taking into account the information available in the literature, the model presented in figure 3 could be an example of the formation of planktonic cellular aggregates. The PCA formation initiates with the cell-to-cell approach through motility or Brownian motion. When occur the contact between bacterial cells through physical forces and biological mechanisms, starts the formation of a planktonic cellular aggregate (14). This aggregate matures and then the dispersion phase takes place.

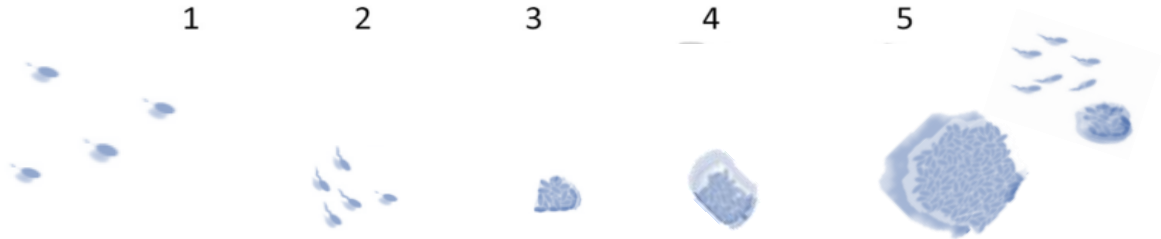

Fig. 3 The formation of planktonic cellular aggregates. 1. Cell-to-cell approach through motility or Brownian motion; 2. Physical and chemical interaction between planktonic cells with the formation of planktonic cellular aggregate; 3-4. Growth of planktonic cellular aggregate; 5. Dispersion of planktonic cellular aggregates by smaller planktonic cellular aggregates or planktonic cells (adapted from (8)).

1.2 Molecular mechanisms of planktonic cellular aggregates

The formation of planktonic cellular aggregates involves several molecular mechanisms. Some of them are exclusive to some species of bacteria and others are common to all species. In this section, the objective is to provide an overview of what is known at the genetic level about this process.

Successful cellular aggregation depends on contact between bacterial cells through physical forces and biological mechanisms. However, for this contact to occur, a prior approximation of the cells via Brownian motion or swimming motility is necessary (14). There are two mechanisms in this approximation. The first mechanism corresponds to the initial formation of “seeding” aggregates by random collisions of motile cells or by Brownian motion (14, 15). The Brownian motion is the random motion of a particle when it is deflected by the force of the impact of individual molecules of the liquid. Since bacteria are very small organisms, they can be compared to these particles and can, thus, suffer this type of movement (16). However, it is more advantageous to be flagellate, thus, bacteria can actively search their favorable environment (17). The second mechanism of cells approximation allows bacteria to search for ideal conditions, and is usually controlled by chemotaxis (15). The reason for this is that each bacterium has specific medium conditions that allow them to survive, namely, the type of carbon source, nitrogen and oxygen availability, nutrient availability, and others. Through chemo-attractants and repellents, bacterial cells can be moved to the same sites, thus occurring their approximation. Thus, since chemotaxis guides the movement of bacteria to the favorable environment by controlling flagella and pili, transient cell-cell contacts between motile cells depend not only on the efficient motile apparatus, but also on chemotaxis mechanisms (18, 19).

Bacteria have several mechanisms that allow their movement in both liquid and solid surfaces. These mechanisms include swimming motility, swarming motility, gliding motility, twitching motility and floating motility (20). The movement to the site where the biofilm will be formed happens in both SAB and PCA. For that reason, and because there is lack of information about the need of motility in the initial formation of planktonic cellular aggregates, the search in literature was extended up to studies about SAB.

The motility of bacteria is provided by flagella and pili. Flagella is the most studied motility structure and it is composed of 3 substructures: the basal body, the filament and the hook (Fig. 4). Pili is used in twitching motility and gliding motility (20). Both flagella and pili biogenesis, assembly and regulation require the action of a highly complex machinery that comprises the interaction of numerous proteins and, therefore, the regulation of a large number of genes (20, 21).

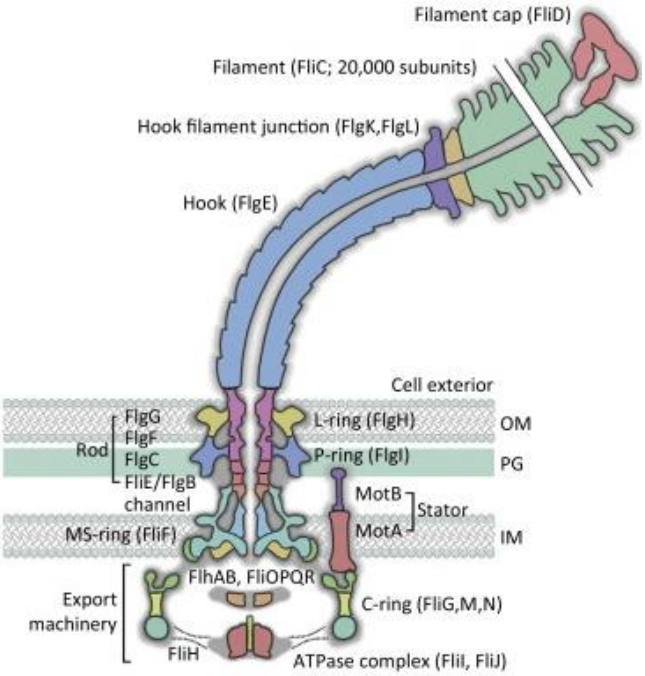

Fig. 4 Structure of flagella. Flagella is composed of 3 substructures: the basal body, the filament and the hook (retrieved from (22)).

There are several reports in the literature that highlight the crucial role of motility and chemotaxis in the formation of biofilms in distinct species of bacteria such as *Escherichia coli* (23), *Helicobacter pylori* (24), among others. There are three categories of mutants related to motility: *fla* mutants, *che* mutants and *mot* mutants. *Fla* mutants are aflagellate due to the disruption of flagellar structural components or regulators required for structural component expression. The mutants that have *che* phenotypes are defective in chemotaxis due to the disruption of the cytoplasmic signal transduction system or the chemoreceptors. Lastly, *mot* mutants exhibit mutations on MotA, MotB or FliG components and therefore are nonmotile flagellate mutants (25).

In the chemotaxis pathway, signals are transmitted to the flagellar motor by the chemoreceptor, through a two-component system (TCS). TCS consists in a sensor kinase and a response regulator (RR) that interact to each other. In case of chemotaxis (Fig. 5) there is a chemoreceptor sensor protein, where a ligand (chemo-attractant or repellents) will bind, generating a change of conformation of the chemoreceptor. This one recruits two more proteins CheW and CheA. The first is a transducer protein and CheA acts like a sensor kinase, autophosphorylating. Autophosphorylation causes the interaction between CheA and the response regulator proteins (CheB and CheY), causing the transfer of the phosphate group (P) to an aspartate residue of response regulators, making them active. Active CheY interacts with components of the flagellar switch complex, affecting rotation of the motor and CheB mediates adaptation to the stimulus with CheR. Once the balance in the system is restored, a phosphatase (CheZ) removes the phosphate group from the CheY (dephosphorylation), leaving the regulatory system ready to retain another stimulus (26, 27).

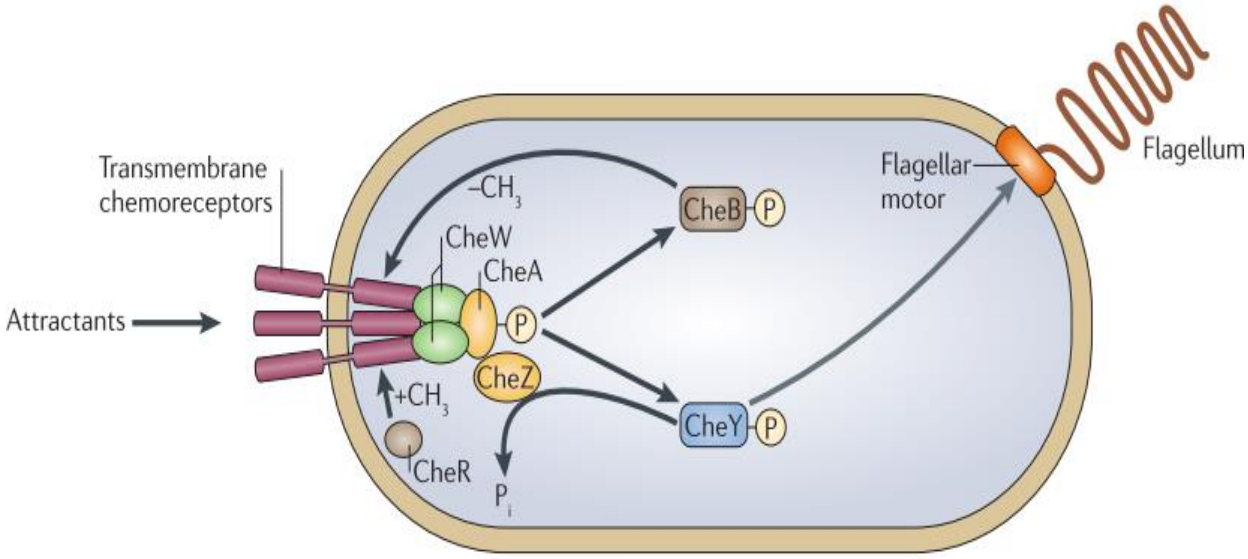

Fig. 5 *Escherichia coli* chemotaxis pathway. In the chemotaxis pathway signals are transmitted to flagellar motor by chemoreceptor, through a two-component system (retrieved from (26)).

In 1998, Pratt L. and Kolter R. (23), used *Escherichia coli* as a model organism to determine which genes would be involved in the initial formation of biofilms. In this specific case, they have been able to prove that the lack of some genes related to the flagellum determines the inability of these mutants to form biofilms. In this study, they have tested the three types of mutants: *fla, mot* and *che.* It was revealed that nonchemotactic mutants appear to form biofilms indistinguishable from the wild-type. On the other hand, the aflagellated mutants and *mot* mutants were severely defective in biofilm formation. In view of these results, the authors concluded that chemotaxis in these conditions and in this organism, is completely dispensable for normal biofilm formation. However, on the contrary, the motility is critical for normal biofilm formation.

In a different study with *Helicobacter pylori,* the results differ somewhat, since chemotaxis proteins CheY1 and CheAY2 are essential in the gastric mucosa colonization process. Chemotaxis seems to be important because *che* mutants, although they have the ability to colonize the gastric mucosa, are unable to remain in the mucosal layer and are, therefore, removed by gastric flow (24). Thus, depending on the organism and conditions, chemotaxis may or may not be crucial in the initial biofilm formation process.

After the cell-to-cell approach of the bacteria, the physical and chemical interactions that allow the maintenance of the cellular aggregates structure begin to occur. The flagellum and pili, in addition to the role they play as force-generating movement, they also act as a way of interaction between bacterial cells. In *Vibrio cholerae* it was determined that Type IV pilus is required for microcolony formation mediating pilus-pilus interaction between bacterial cells. *Vibrio cholerae* uses TCP (toxin co-regulated pilus), the Type 4b Pilus system. Each TCP filament is composed of TcpA pilin subunits. In the study carried out by Kirn *et al.* using directed mutagenesis, the existence of mutations in charged amino acids present in the C-terminal domain of TcpA, known as the domain of interactions, led to a decrease in pilus-pilus interactions between bacteria and, consequently, to a decrease in cellular aggregation (28).

In addition to the flagellum and pili, cellular aggregation is also mediated by bacterial products, more specifically, the extracellular polymeric substances where are included polysaccharides,

extracellular DNA and ions. *Aquincola tertiaricarbonis* RN12 was used as a model to explore the biosynthesis of extracellular polysaccharides and the regulation of floc/aggregates formation. Through transposon mutagenesis and subsequent complementation, a gene cluster related to exopolysaccharide biosynthesis and gene *rpoN1* encoding an alternative sigma factor, RpoN1 (σ^{54}), were identified as disrupted in mutants unable to flocculate (5).

One way of regulating gene expression is through transcriptional regulators, namely through alternative sigma factors like RpoN1. Sigma factors recruit the core RNA polymerase to recognize promoters with specific DNA sequences (29). Thus, the authors of the previous study thought that the identified sigma factor could regulate the biosynthesis of extracellular polysaccharides, since RpoN1 mutants presented more released exopolysaccharides than bound polysaccharides (polysaccharides that allow the cells to be bound). However, it was verified that the sigma factor did not directly control the expression of that cluster of genes despite regulating unknown genes involved in the formation of flocs (5).

In *Staphylococcus aureus* aggregation is due to production of polysaccharide intracellular adhesin (PIA). PIA is a poly-β(1-6)-N-acetylglucosamine produced by the enzymes encoded by the *ica*ABC operon. To evaluate the role of PIA in cellular aggregation, Haaber *et al.* created a *Δica* mutant and confirmed that aggregation does not occur (6). In case of *Staphylococcus aureus,* in addition to PIA, fibronectin binding proteins A and B promote surface-attached biofilm formation, but these authors proved that, in case of aggregation, these two proteins do not contribute to aggregation (6).

In addition to polysaccharides, the presence of extracellular DNA is often essential for the formation of aggregates (30). The eDNA is released by various biological mechanisms, which differ depending if bacteria are Gram-positive or Gram-negative. In the case of Gram-negative bacteria, eDNA is released during lysis of cells due to the action of prophages and metabolites such as phenazines. In the case of Gram-positive bacteria, is due to the action of lytic proteins, enzymes and peptides in the cells, which allow the release of eDNA. Some chemical and physical stresses may also be associated with the release of extracellular DNA (31).

During the initial encounter of two bacteria, van der Waals forces and repulsive electrostatic forces determine whether two cells can get close enough for acid-base interactions to occur. These interactions are the most predominant ones in bacterial aggregation. These types of interactions are possible due to the action of extracellular DNA. Long chains of extracellular DNA interact with other components of the biofilm matrix or with the cell surface components penetrating the barrier between neighboring cells and thus facilitate interactions between cells that otherwise are repelled by electrostatic forces (32).

In the literature, it has been proven that extracellular DNA is often one of the major components of cellular aggregates (30). Given the lack of information at the genetic level of the release of extracellular DNA in PAC, it was extrapolated the knowledge regarding SAB, thus determining the importance of identifying the mode of regulation of eDNA release and the genes involved in production of components that bind to eDNA to stabilize and form biofilms.

In *Staphylococcus epidermidis* it was shown that the extracellular DNA present in its biofilms is very similar to the chromosomal DNA of this species. Through the analysis of the *ΔatlE* mutant it

was found that the extracellular DNA was generated through the activity of the autolysin AtlE in a subpopulation of *S. epidermidis.* Thus, the lack of activity of the autolysin AtlE was considered an indirect cause of the failure in the formation of biofilms (33).

As said before, when a biofilm is formed some properties emerged, namely, the ability of bacteria to communicated through chemical signals, i.e. molecules used in quorum sensing. Quorum sensing is a cell-cell communication process that bacteria use to regulated gene expression in order to coordinate and synchronize behavioral responses at high cell densities (HCD) (34). The molecules used in quorum sensing are signaling molecules known as autoinducers. Gram-negative bacteria use acylated homoserine lactones (AHL), Gram-positive bacteria use peptide signals, and both Gram-negative and Gram-positive bacteria use autoinducer-2 (AI-2) (34).

Gram-negative bacteria, AHLs after being synthesized, are diffused out of the cell, where they accumulate. Once they reach a critical concentration outside the cell, the AHLs are internalized again as autoinducers. The interaction between AHL and the LuxR protein becomes favorable, resulting in the formation of the AHL + LuxR complex which, in turn, will interact with the DNA altering the regulation of the target genes. In the case of Gram-positive bacteria, they use peptides that, after being produced by the cell, are actively transported to the extracellular environment. As soon as the concentration of the peptide reaches a certain threshold, these are "sensed" by a two-component regulatory system. The response regulator of this system will interfere with the regulation of quorum sensing target genes (34).

Quorum sensing regulates several mechanisms involved in the formation of cellular aggregates, namely, chemotaxis (15), (possibly) eDNA release (30), and perhaps others.

Escherichia coli only produces a known quorum sensing molecule - autoinducer 2. In 2016, Laganenka *et al*. (15), have shown that this molecule functions as a chemoattractant inducing aggregation in this species. In *E. coli*, the AI-2 molecule is produced by the enzyme LuxS. The uptake of this compound is made by a carrier consisting of 4 different proteins: LsrB, LsrA, LsrD and LsrC. The AI-2 undergoes phosphorylation by LsrK and then binds to LsrR, activating it. In turn, the active LsrR leads to derepression of the *lsr* operon and possibly other genes. When the transport of AI-2 occurs, AI-2-bound LsrB interacts with the sensory domain of a major chemoreceptor of *E. coli*, triggering a chemotactic response. In that study, *ΔluxS* and *ΔlsrB* mutants showed a decrease in aggregation due to failure in AI-2 synthesis and perception, respectively. The *ΔluxS* mutant also had decreased motility, making interpretation difficult, but the *ΔlsrB* mutant showed normal motility and chemotaxis failure, which was comparable to the *ΔcheY* mutant failure. Thus, the *ΔlsrB* mutants did not present chemotaxis for AI-2 and therefore had a decrease in the formation of cellular aggregates. This study provided the evidence that chemotaxis towards a self-secreted attractant mediates autoaggregation of bacteria.

In 2009, Suzuki et al. (30) investigated whether quorum sensing influenced the formation of flocs / aggregates in *Rhodovulum sulfidophilum*. To that end, and considering that the cyclodextrins associate with the AHL, inhibiting the quorum sensing signal, these authors added α-cyclodextrin to the medium and analyzed the influence of quorum sensing inactivation on the formation of flocs / aggregates. There was a marked decrease in the rate of flocculation and an

increase of the extracellular DNA in the medium, an important factor, considering that, as also demonstrated in this article, eDNA functions as a cell-to-cell interconnecting compound. However, in that work article, the authors failed to prove that the amount of extracellular DNA in the medium was controlled by this quorum sensing system. With their study it was proved that the inactivation of this quorum sensing system led to a decrease in the rate of aggregation (30). For this reason, and although an inhibitory compound has been used, the possibility of inactivating that system at the genetic level has to be considered.

As reported in many studies, the biological mechanisms that influence the formation of planktonic cellular aggregates depends on the medium conditions i.e. the type of carbon source, nitrogen and oxygen availability, nutrient limitation, antibiotic concentrations, and, consequently, the metabolic adaptations by bacteria in each situation. *Azospirillum brasiliense* are microaerophilic bacteria that look for places with low oxygen concentration by aerotaxis. However, when cells are exposed to high concentrations of oxygen, these bacteria form cell-to-cell interactions which leads to the formation of clumps. Flocculation happens if beyond oxygen stress, bacteria also need to support nitrogen limitation (10). Bible *et al*. (10), using proteomics determined which proteins are expressed in the wild-type strain, a nonclumping strain and an hyperclumping strain. Through the results obtained they identified which metabolic pathways were involved in the process of clumping. In the hyperclumping strain, the proteins upregulated are related to the amino acid metabolism, metabolism of cofactors and vitamins, motility and signal transduction, protein turnover (chaperones), and protein secretion. It was possible to prove that clumping cells adjust their metabolism by increase carbon storage and conserve amino acids. These adaptions allow clumping cells to flocculate. In the flocculation deficient strain, the authors prove that cells modified the properties of the cell surface, which include an increase of EPS production. This work prove that bacteria must adapt their metabolism and behavior to survive under stress condition (10).

As shown in figure 3, the last step of the dynamic life of planktonic cellular aggregates is their dispersion. This can occur by the release of small size planktonic cellular aggregates or the release of planktonic cells (8). Regardless of the mode of dispersion of planktonic cellular aggregates, the main objective of this process is the progression of colonization and dissemination of bacteria into the environment (35).

Like the molecular mechanisms involved in planktonic cellular formation described above, this step is a complex process that depends on environmental signals, signal transduction pathways and effectors (35). However, these mechanisms will not be reviewed in this introduction.

1.3 Planktonic cellular aggregates in Cystic Fibrosis related microorganisms

To underline the great need, in clinical terms, for knowledge on the molecular mechanisms involved in the formation of planktonic cellular aggregates, this introduction will provide the example of cystic fibrosis, in which microorganisms from the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* can form such planktonic cellular aggregates and often establish chronic infections.

1.3.1 Cystic Fibrosis

Cystic fibrosis is an autosomal recessive genetic disorder that follow the Mendel's laws despite all the phenotypic variation among patients. In 1938, Dorothy Andersen (36) described for the first time this disease. Since that time, the scientific community has been trying to understand the mechanisms underlying the pathophysiology of the CF disease and the possible therapies that may be implemented.

Cystic fibrosis is caused by mutations in a gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) protein. This regulator belongs to the ATP-binding cassette (ABC) family of membrane proteins (37). CFTR is a glycoprotein consisting of 1480 amino acids. Like all ABC transporters and as shown in figure 6, this contains two transmembrane domains and two nucleotide binding domains (NBD1 and NBD2), which bind and hydrolyze ATP. In addition, it has a regulatory domain (R), a unique feature of this protein (37). Most of CFTR mutations are located in NBD1, such as Phe508del, and only a few are found in nucleotide binding domain 2 (38).

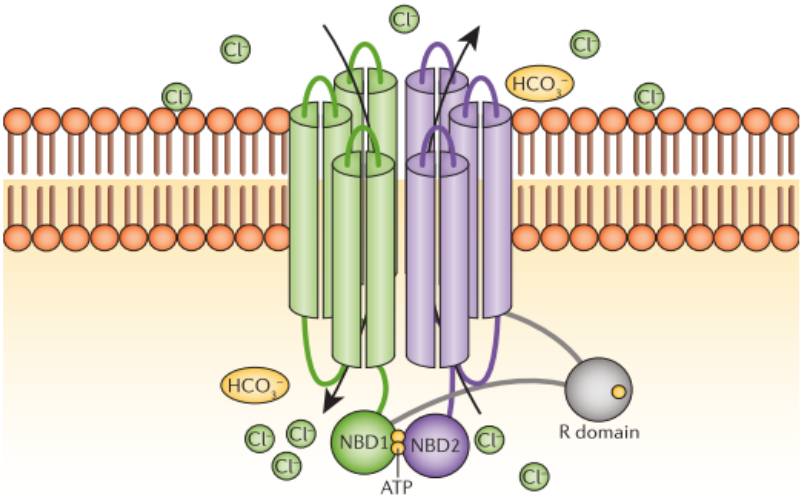

Fig. 6 Structure of the cystic fibrosis transmembrane conductance regulator protein. CFTR contains two transmembrane domains, two nucleotide binding domains (NBD1 and NBD2), which bind and hydrolyze ATP and a regulatory domain (R). This protein is a channel with specificity for anions (chloride and bicarbonate) (retrieved from (39)).

The main function of the cystic fibrosis transmembrane conductance regulator is to be a channel with specificity for anions, more specifically for chloride and bicarbonate. However, CFTR functions also as regulator of other ion channels and transporters, e.g. ENaC (epithelial Na^+

channel), of vesicle trafficking, ATP release, and the expression of inflammatory epithelial mediators (38, 40). CFTR protein is found in the apical membranes of epithelial cells and, for that reason, is expressed in multiple tissues (lung, pancreas, intestine, vas deferens and others), leading to the characteristic of multiorgan disease associated with cystic fibrosis (41).

Epidemiology

Cystic fibrosis is referred as the most common lethal genetic disease in Caucasians. More than 1500 mutations have been identified, but just a few are associated with the disease. The prevalence of disease worldwide varies largely and, in some populations, a small number of CFTR mutations predominate while other areas display greater mutational diversity (42).

In 2015, the European Cystic Fibrosis Society Patient Registry had 42.054 patients listed with CF and, in the same year, in United States, there was 28.983 patients registered (43, 44). In Europe, Portugal is one of the countries with low birth prevalence, about 1:6000 to 1:8000 (42, 45). A recent study made in Portugal describes the Portuguese adult CF population. In this study, the average age of patients was 31.2 years which reflects the increased survival in this pathology. Unlike many countries with well-established newborn screening programs, in this study the median age at diagnosis was 13 years and 38% of the patients were diagnosed only in adulthood, evidencing the lack of such programs in Portugal (45).

The average life expectancy has increased during the past six decades. Nowadays, about 50% of patients with cystic fibrosis are adult, and infants born in 2015 have an estimated average life expectancy of 45.2 years (44).

Physiopathology and the disease spectrum

As said before, cystic fibrosis is known to affect several organs (41, 42). However, the onset of disease-related symptoms varies from patient to patient and some only appear throughout life (42).To understand all symptoms, it's necessary to know how each organ is affected by this disease.

In cystic fibrosis, pulmonary diseases are the major cause of mortality and morbidity. In 2015, it was reported as the cause of death of 64.5% of individuals who died from cystic fibrosis-related causes (44). For this reason, clinical manifestations are often subdivided into two groups: pulmonary manifestations and extrapulmonary manifestations. Therefore, due to its relevance, this introduction will focus mainly on the pulmonary diseases associated with cystic fibrosis.

The environmental conditions of the lungs are the biggest change in people with this disease. In healthy people, only the upper respiratory tract is colonized by microorganisms, and the lower tract, due to the existing immunity mechanisms, is considered practically sterile. Since the environment is the factor that dictates the interactions that occur between the host and microorganisms, CF patients often have microorganisms in the lower respiratory tract (38). However, based on recent studies, it has been found that this disease not only alters lung environmental conditions but also causes anatomic changes in the respiratory system, an essential factor for the proper functioning of the lungs (46).

As shown in figure 7, the human respiratory tract is highly complex anatomically, being divided into two distinct parts: upper respiratory tract and lower respiratory tract. The first part includes paranasal sinuses, nose and larynx. The lower respiratory tract consists of the air conduction zone and the respiratory zone, where gas exchange occurs (38, 47).

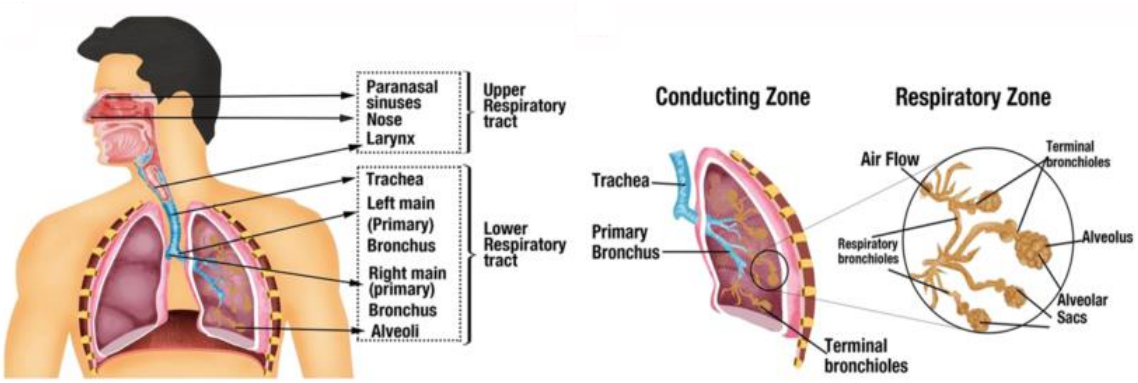

Fig. 7 Human respiratory tract consists in upper respiratory tract and lower respiratory tract. The lower respiratory tract is divided in conducting zone and respiratory zone (retrieved from (38)).

The airway epithelium consists of two types of epithelium: surface epithelium and submucosal gland epithelium. The surface epithelium is composed of three types of cells: the ciliated cells, goblet cells, and basal cells, while the submucosal gland epithelium includes ciliated duct cells, mucus cells, and serous cells (Fig. 8) (40, 48).

To prevent air impurities, as well as possible microorganisms from reaching the places where the gas exchange occurs, the conducting airways are covered by an aqueous film called the airway liquid surface (ASL), which is composed of the periciliary liquid and the mucus gel layer (Fig. 8). The periciliary liquid covers the cilia and allows the ciliary beating. In turn, the mucus layer has in its constitution mucins, produced by goblet cells and submucusal gland epithelia, that traps impurities and microbes inhaled. The mucociliary clearance is the process that allow the removal of impurities of air and microbes from the lungs (40).

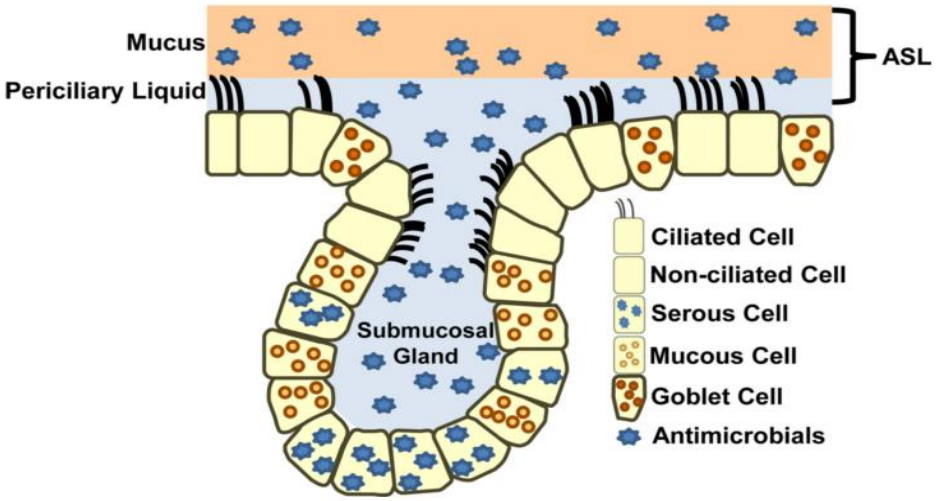

Fig. 8 The airway epithelium. The airway epithelium consists in surface epithelium and submucosal gland epithelium. Each epithelium is composed by different cells. The airway liquid surface is composed of the periciliary layer and the mucus gel layer (retrieved from (48)).

Protein CFTR, although present in the surface airway epithelium, is found mainly in the serous cells of the submucosal glands present in the bronchi (49). Due to this generalized presence, when there is absence or lack of functionality of the CFTR, as in cystic fibrosis, there are several alterations of the pulmonary environment, such as pH decrease, defective mucociliary clearance, decreased antimicrobial activity leading to infection and inflammation, alteration of mucus properties, decreased hydration and volume of ASL, among others (39, 46).

The mucociliary clearance system is the primary innate defense mechanisms for airways. For an efficient mucociliary clearance it is necessary that the ciliary beating, the correct hydration of the ASL and little viscosity of the mucus occurs (50). Normal mucus contains about 95-98% of water in its constitution, and the remaining 2-5% relate to mucins and other important components (38). The presence of CFTR and ENaC, in the surface airway epithelial cells, allows in these cells that both Na^+ absorption and Cl^- secretion occur. In patients with CF, the absence or non-functionality of CFTR leads to the absence of Cl^- secretion and, indirectly, to Na^+ hyperabsorption. Consequently, there is an imbalance of ions leading to a considerable reduction of water in the mucus, thereby decreasing the volume and hydration of ALS (50).

The viscosity of mucus is another property that is modified due to lack of water, but also due to changes in pH. As previously stated, bicarbonate is a pH neutralizer, not being secreted by the CFTR channel, and therefore leading to a reduction in the pH of ALS and secretions of the submucosal glands. The decrease in pH influences the disulfide bonds of the mucins, stabilizing them and leading to an increase of their viscosity and consequently of the viscosity of the ASL (38). The mucus, which is usually above the cilia, approaches and adheres to the surface epithelium forming mucus plaques, due to the modification of its physical and chemical properties (38). The microorganisms that reach the respiratory tract, in the case of CF patients, are not expelled by mucociliary clearance, remain trapped in mucus plaques. At these sites, some microorganisms are able to survive by adapting and beginning to multiply, thus permitting bacterial infection to become established (50).

However, beyond the mucociliary clearance system, the airways have a secondary defense mechanism against microorganisms: the action of the immune system. The immune system defenses include the production of antimicrobials and the action of innate immunity cells, like neutrophils that migrate into the airway lumen, capture and kill bacteria. However, most of the time, in patients with cystic fibrosis, antimicrobial agents like lactoferrin and lysozyme have reduced ability to permeate the typical mucus of CF patients (50).

Both airway epithelial cells and immune cells participate in the inflammatory process characteristic of CF and responsible for lung disease progression and for the decrease of lung function (51). It is still under discussion whether the inflammatory process is necessarily a consequence of an infection. This is because there are evidences that airway epithelial cells are deficient in both signaling and intracellular processes that promote the transcription of inflammatory mediators (39, 51).

Since neutrophils are the dominant inflammatory cells in the lungs, the presence of both pathogens and inflammatory mediators leads to their recruitment (39). The action of neutrophils and macrophages, later recruited to the site of infection, aims to eliminate the pathogens through

their own mechanisms and through the production of more inflammatory mediators that will lead to the recruitment of more immune cells. This process leads to the formation of a cycle of inflammation that perpetuates itself in time (51).

Recurrent infections in the lungs are caused by several microbial species of bacteria, fungi, viruses. Of the microbial community, about 99% are bacteria while the remaining 1% concern viruses and fungi (38). Due to the high percentage and consequent importance, in this introduction will be highlighted the lung infections caused by bacteria.

Children with cystic fibrosis are colonized by *Staphylococcus aureus* and *Haemophilus influenzae,* and when they reach to 5 to 6 years old, *Pseudomonas aeruginosa* starts to be the predominant pathogen in lungs (44). As reported in the Annual Data Report of 2015 of Cystic Fibrosis Foundation Patient Registry there are several other bacteria common in CF patients: *Stenotrophomonas maltophila,* Nontuberculous Mycobacteria*, Achromobacter, Burkholderia cepacia* complex (*Bcc*), among others. The prevalence and the patients median age of first infection varies from species to species (Fig. 9) (44).

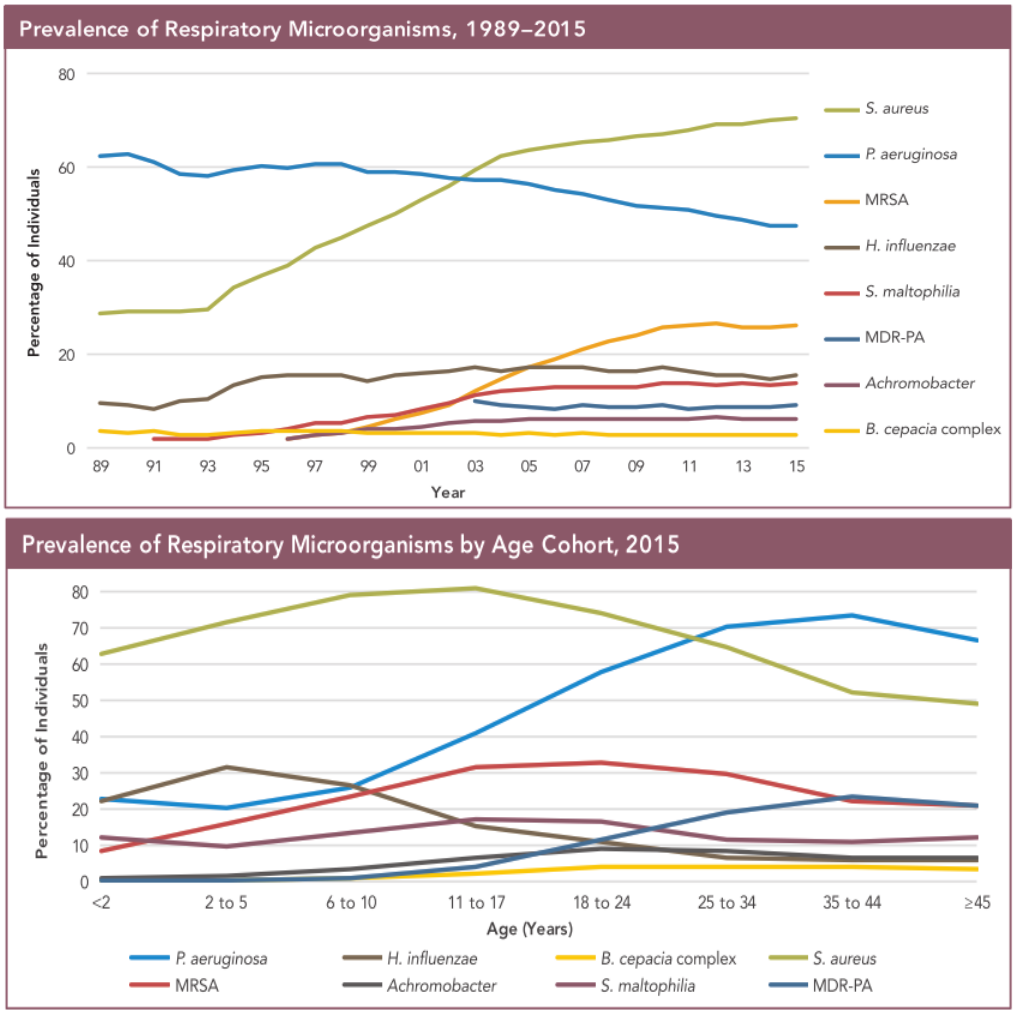

Fig. 9 Prevalence of respiratory microorganisms found in CF patients over the years and in groups of age (retrieved from (44)).

Pseudomonas aeruginosa and *Burkholderia cepacia* complex are considered opportunistic pathogens that cause infections in patients with CF and in immunocompromised patients.

Pseudomonas aeruginosa is the dominant pathogen in cystic fibrosis with 80% of adult patients chronically infected by this pathogen. Chronic infections are associated with declining clinical status (12). *Burkholderia cepacia* complex are present in about 2% to 8% of patient with cystic fibrosis. Patients infected with *Burkholderia cepacia* complex present a highly variable prognosis and evolution of the disease. Some of the patients are considered asymptomatic while others show deterioration of their clinical status over time with a final fatal deterioration occurring within a few days. The cepacia syndrome develops in approximately 1 to 10% of the patients infected with *Bcc*, more often, with the species *B. cenocepacia, B. multivorans* and *B. dolosa*, and is characterized by recurrent episodes of fever, bacteremia, necrotizing pneumonia and often becomes fatal (13).

Considering the effects of the presence of these two pathogens in patients with CF, the following topics will provide the knowledge about the adaptations of these bacteria to CF lung environment, highlighting the formation of planktonic cellular aggregates by these microorganisms.

1.3.2 Adaptation of *Pseudomonas aeruginosa* and *Burkholderia cepacia* complex bacteria to CF lungs

As said before, the microorganisms that reach the respiratory tract of CF patients, are not expelled by mucociliary clearance and remain trapped in mucus plaques (50). This forces bacteria to adapt their lifestyle in order to survive in such different environment, which has a huge amount of sputum composed by mucin, lipids, proteins, amino acids, ions and DNA released by neutrophils (52). Bacterial adaptations to this environment occur at both the genomic and phenotypic levels thus enhancing the chronicity of the infection triggered by pathogens like *Pseudomonas* and *Bcc* (38). The main survival strategy adopted by these bacteria, in this type of conditions, is switch from free-swimming to the biofilm mode of growth (12).

In the next section, it will be provided a brief explanation of some molecular mechanisms behind the formation of biofilms by *Pseudomonas* and *Bcc,* considering the environmental conditions of CF lungs and the adaptations to which these bacteria are subjected to survive under these stress conditions.

In 2002, Worlitsch *et al*. (53) described, for the first time, a schematic model of what could be the biofilm formation model of *P. aeruginosa* in CF lungs, considering the hypothetical events described up to that date (Fig. 10). The model described, with some attenuations, is very similar to the model of the planktonic cellular aggregates previously described in this introduction, since cells in this biofilm are not attached to a surface but to each other, contrary to what happens in the surface-attached biofilm.

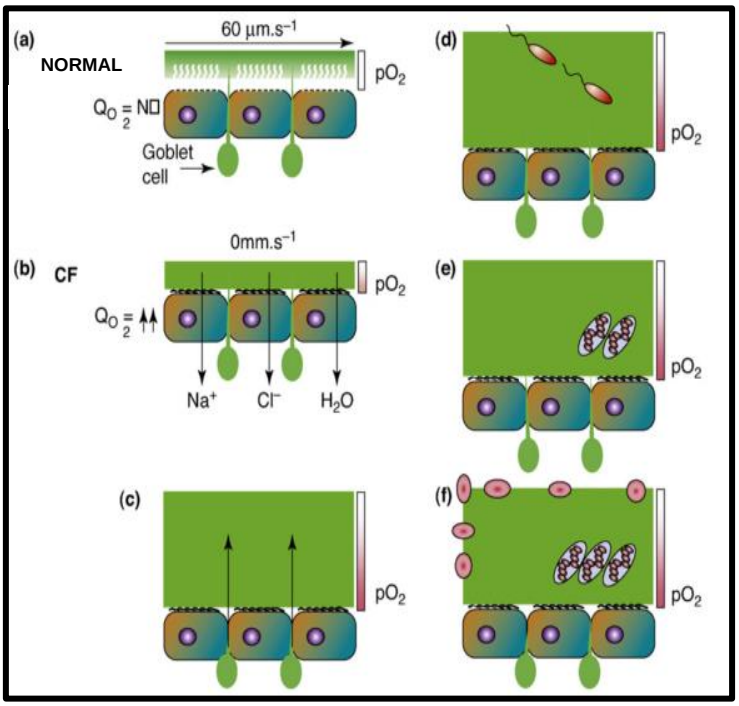

Fig. 10 The formation of biofilms in CF lungs by *Pseudomonas aeruginosa*. pO_2 (partial pressure of oxygen) The color bars to the right of each diagram indicate the pO_2, with increasing red color indicating a progressively more anaerobic airway mucus. **(a)** Normal airway epithelia that enable efficient mucociliary clearance **(b) – (f)** CF airway epithelia**. (b) – (c)** Imbalance of ions causes the reduction of water in the mucus, thereby decreasing the volume and hydration of ALS. The mucus, which is usually above cilia, approaches and adheres to the surface epithelium forming mucus plaques. Due to the O_2 gradient in mucus plaques, some parts of them are hypoxic/anaerobic, as shown by the increasing red color of the color bar. **(d)** *Pseudomonas* reach to the mucus surface and penetrate actively (via flagellar, twitching or swarming motility) or passively (due to mucus turbulence) into the mucus plaques. **(e)** *Pseudomonas aeruginosa* grows in biofilm, setting the stage for persistent and chronic infection**. (f)** Biofilms resist to host defenses, such as, neutrophils, macrophages, antibiotics, allowing the maintenance of a chronic infection by *Pseudomonas* (retrieved from (54)).

Pseudomonas reaches the mucus surface and penetrate actively (via flagellar, twitching or swarming motility) or passively (due to mucus turbulence) into the mucus plaques, where it remains trapped (Fig. 10(d)) (55). Thus, as described in the first step of the formation of planktonic cellular aggregates, bacteria need the motility apparatus to actively move to the sites where they will form biofilms, in this case, in mucus plaques. In *P. aeruginosa* it has been demonstrated the need of flagellum and pili for pathogenesis (56). However, Sriramulu *et al.* (52), when attempting to determine which genetic components could be involved in aggregate formation recognized that the Δ*fliC* mutant was not deficient in the formation of the planktonic cellular aggregates, showing that the flagellum was not entirely necessary for the formation of cell aggregates. In 2012, Caldara *et al.* (57) reported similar conclusions with the tests that they had performed. These authors proved that the presence of mucin reduce the ability of bacteria to produce surface-attached biofilm. Contrariwise, mucins allowed the formation of cellular aggregates, but just in flagellar mutants (Δ*flgE*, Δ*flgK* and Δ*fliD* mutants). Thus, cellular aggregates were formed by nonmotile mutants and in the presence of mucins. This conclusion agrees with the conditions of CF lung and the phenotype of isolates, since the CF lung has mucins (52) and the isolates of later stages of infection do not present motility (54).

Data presented may seem contradictory, however, there may be an explanation for this. Possibly, all the studies are correct, but the former consider the time when bacteria are reaching to the mucus plaques and the last two studies consider the moment when the bacteria are already in the presence of mucin and trapped in the mucus plaques. Additionally, the possible role of the flagellum as a mucin binding element should not be discarded (58).

As reported before, extracellular polymeric substances promote cellular aggregation. In *Pseudomonas* it was extensively reported that strains evolve from non-mucoid morphology to mucoid morphology. Since this mucoid morphology is due to the hyperproduction of alginate (54) and, in 2005, Sriramulu *et al.* verified that a mutant unable to produce alginate (*ΔalgD)* was affected in the formation of tight microcolonies (52), this implicates tris extracellular polysaccharide in cellular aggregation.

Due to the cycle of inflammation, *P. aeruginosa,* when trapped in mucus plaques, are exposed to high levels of reactive oxygen species (ROS) and reactive nitrogen species (RNS) produced by the immune response of host (54). These ROS and RNS react with DNA bases, lipids and proteins. In case of DNA, there are mechanisms that repair the mutations that result from this reaction between ROS/RNS with DNA bases. These mechanisms include the DNA repair system and proof-reading systems. However, even these mechanisms are exposed to these reactive species. When some mutations occur in these mechanisms, they become unable to repair the DNA, so for that reason some populations become hypermutable. In case of *Pseudomonas* and *Bcc* this happens more frequently in later stages of infections and in chronic infection (54, 59). The hypermutable strains present 20-fold higher spontaneous mutation frequency when compared with non-hypermutable strains found in adult CF patients. The mutations and continuously natural selection of the benefic adaptations are essential for the maintenance of pathogens in the stressful and diverse CF lung environment (54).

One of the most common mutation in *P. aeruginosa* happens in the gene *mucA* encoding an inner membrane associated anti-σ-factor. A mutation in this gene leads to hyperproduction of the EPS alginate. Usually MucA binds to AlgT (named AlgU or σ^{22}) preventing alginate biosynthesis. This sigma factor, σ^{22}, when not connected to MucA, can recruit the core RNA polymerase to recognize promoters with specific DNA sequences, more specificity, recognize the DNA sequences of the *algD* operon. Mutations in *mucA* gene modify MucA and, consequently, prevent its interaction with the sigma factor, leading to the alginate biosynthesis (Fig. 11) (29, 54).

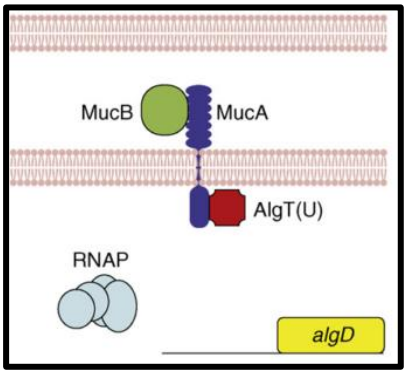

Fig. 11 Molecular basis of alginate production. Sigma factor, σ^{22}, or AlgT(U) when not bound to MucA, can recruit the core RNA polymerase to recognize promoters with specific DNA sequences, more specificity, recognize the DNA sequences of the *algD* operon (retrieved from (54)).

As shown in figure 10(a), normal lungs do not exhibit an oxygen gradient in ALS, but a constant oxygen levels throughout their structure. However, in the case of patients with cystic fibrosis, an oxygen gradient is generated in mucus plaques (Fig. 10(b)-(f)) (55). Due to the O_2 gradient in mucus plaques, some parts of them are hypoxic/anaerobic. However, *Pseudomonas* can growth in these places, because they can switch their metabolism from aerobic to anaerobic (60). The anaerobic metabolism uses an alternative electron acceptors, nitrate (NO^-_3) or nitrite (NO^-_2), but the energy yield is much lower than from aerobic metabolism (61).

In patients with CF, the levels of exhaled NO are lower than in normal people. One reason for this is the oxidation of NO to NO_2^- and NO_3^-. Thus, the high levels of these two components are useful to anaerobic metabolism of microorganisms such as *Pseudomonas.* The genes involved in anaerobic metabolism are increased in isolates of CF chronic patients (62). However, the relation between anaerobic metabolism and the formation of planktonic cellular aggregates was only described in 2005 by Sriramulu *et al*. These authors, through the screen of transposon library, determined that one of the transposon's insertion occurred in *oprF* gene encoding an outer membrane protein. Furthermore, when the authors analyzed the protein expression patterns in TSB medium and in a medium that mimics the CF lung environment, through proteomics, they conclude that, in conditions that mimic the CF lungs, this protein is overexpressed (52). The outer membrane protein, OprF, represents a biomarker for the anaerobic mode of CF lung disease since it is suggested that OprF in *Pseudomonas* functions as permease for NO_3^-/NO_2^- uptake (61).

Some authors also do not exclude the fermentation as a requirement for microcolony formation, since cells inside these structures have a limiting oxygen access but an energy rich-conditions. This consideration had been taken, since, in *P. aeruginosa* a Δ*ldhA* mutant unable to synthesize lactate dehydrogenase produces impaired cellular aggregates (63). This enzyme is responsible for the conversion of pyruvate into lactate.

Quorum sensing is crucial to cell-cell communication and to gene expression regulation in biofilms (34), and in *Pseudomonas* it seems to have a key role in the aggregation process. *Pseudomonas aeruginosa* has two important quorum sensing systems, *las* and *rhl*, that control virulence factors. When the two systems were tested for the formation of tight microcolonies, only the *las* system reveal to be require to the formation of these structures (52).

Another gene promoting the formation of tight colonies in *Pseudomonas aeruginosa is the* LysR regulator (BvlR). The authors of that study compared the ability to form microcolonies by *ΔBvlR* transposon mutant and by the wild-type strain in a synthetic medium that mimics the sputum of CF lungs. The mutant was unable to form tight microcolonies, but after the complementation assay, the tight microcolonies phenotype was restored. Through the obtained results it was possible to recognize the importance of this regulator in the formation of microcolonies, despite in an indirect way. They suggested that it was the repression of cupA-associated fimbrial-based surface attachment by BvlR the cause of the formation of microcolonies, since these structures are not associated with a surface (64).

Regarding the *Burkholderia cepacia* complex less is known about the key player involved in cellular aggregation. Nevertheless, in a report it was demonstrated the need of flagellum and pili to pathogenesis and, consequently, to biofilm formation (65).

As mentioned above, extracellular polymeric substances promote cellular aggregation. *Burkholderia cepacia* complex produce the exopolysaccharide (EPS) cepacian. To understand whether cepacian is important in the formation of cellular aggregates, Silva *et al.* (66) inoculated a mucoid strain of *Burkholderia multivorans* and a nonmucoid derivative with a mutation in *becF* gene (gene required for cepacian biosynthesis), in D-mannitol rich-medium for 3 days. Results of that experiment showed that both strains can form planktonic cellular aggregates, concluding that cepacian wasn't required for the formation of cellular aggregates in this microorganisms (66).

It was reported in the genus *Burkholderia* a possible key role of quorum sensing in aggregates formation. Chandler *et al.* (67) was able to correlate QS with the formation of cellular aggregates. Quorum sensing systems 1, 2 and 3 of *B. thailandensis* were mutated and when grown in minimal medium, the wild-type was able to form aggregates. Contrastingly, the mutant relative to the quorum sensing system 1 wasn't. This mutant was unable to produce the signaling molecule C8-homoserine lactone (C8-HSL), due to the absence of the BtaI1 synthetase. However, the aggregation phenotype was restored when the signaling molecule relative to the quorum sensing system 1, the C8-HSL, was added to the minimal medium. Thus, the authors concluded that the aggregation phenotype is dependent on the quorum sensing system BtaI1-BtaR1 (67).

In *Burkholderia multivorans* it was identified a Lys-type transcriptional regulator named lactate dehydrogenase regulator LdhR involved in the formation of planktonic cellular aggregates (66). Figure 12 shows that both wild-type and *ldhR* mutant are capable to form aggregates, but the wild-type aggregates are in higher number and size than the mutant aggregates (66). Since the *ldhR* gene is co-transcribed with *ldhA* gene encoding a D-lactate dehydrogenase, and *ldhR* mutant complementation only occurred with both *ldhR* and *ldhA* genes or *ldhA* alone, this implicated LdhA in macroscopic cellular aggregates formation (66). Cells inside aggregates have limited oxygen access, but an energy rich environment. To survive, cells must obtain the required energy through fermentation, namely by fermenting pyruvate to lactate. This might explain the formation of aggregates of much smaller sizes if *ldhA* gene is absent.

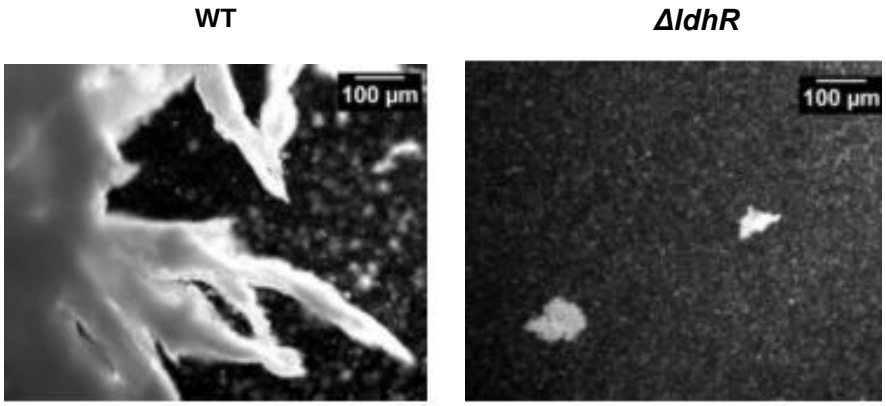

Fig. 12 Planktonic cellular aggregates formed during growth of *B. multivorans* wild-type and *ΔldhR* mutant (retrieved from (66)).

1.4 Motivation and thesis outline

The present study focuses on the need to understand the development of chronic infections of the cystic fibrosis lung by pathogens of the *Burkholderia cepacia* complex (Bcc). *Bc*c infections are important risk factors for quality of life and survival of CF patients. While aggressive antibiotic therapy can delay the transition of chronic infection, it can limit future treatment options due to the emergence of resistance. Probably the most promising way is to understand the genetic adaptation involved in establishing chronic infections. Previous work pointed bacterial cell aggregation as a possible mediator of the transition to chronic infection since these structures might resist better to the harsh conditions of the CF lung, including antimicrobials, oxidative stress, osmotic stress, nutrient gradients, the immune system and other microbial competitions.

Therefore, the aim of this work is to identify mechanisms responsible for the formation of cellular aggregates in bacteria. The approach includes the use of a *Burkholderia multivorans* cystic fibrosis isolate able to form cellular aggregates to construct a transposon mutant library, followed by screening to identify different abilities to form cellular aggregates. After identification of mutated genes, it will be characterized the effect of each mutation not only in cellular aggregation, but also in other phenotypes such as surface-attached biofilm formation, motility, exopolysaccharide production, antibiotic susceptibility, growth rates in synthetic cystic fibrosis medium and virulence in *Galleria mellonella*. The aim is to understand not only the genes involved in cellular aggregates formation, but also whether aggregates formation impacts other phenotypes. Identification of key players in this aggregation trait may provide a comprehensive view of genetic pathways to chronic infections, which might lead in the future to different therapeutic intervention and improvement of CF patient's health.

2 MATERIALS AND METHODS

2.1 Bacterial strains and growth conditions

Bacterial strains and plasmids used in this study are listed in table 1. *E. coli* was grown at 37ºC in Lennox Broth (LB) with or without agar, supplemented with kanamycin (50 µg/ml) or chloramphenicol (25 µg/ml) when required to maintain selective pressure. *Burkholderia multivorans* P0213-1, a clinical isolate from a cystic fibrosis patient (patient P0213), and its derivative strains were grown in LB or in SM medium (12.5 g/l $Na_2HPO_4.2H_2O$, 3 g/l KH_2PO_4, 1 g/l K_2SO_4, 1 g/l NaCl, 0.2 g/l $MgSO_4.7H_2O$, 0.01 g/l $CaCl_2.2H_2O$, 0.001 g/l $FeSO_4.7H_2O$, 1 g/l yeast extract, 1 g/l casamino acids, pH 7.2), supplemented with 20 g/l of D-mannitol and kanamycin (500 µg/ml) (only for transposon mutants) when required to maintain selective pressure, at 37ºC with 180 rpm of orbital agitation.

Table 1 Strains and plasmids used in this study.

Bacterial strains or plasmids	Description	Reference or source
Bacterial strains – *Burkholderia multivorans*		
Burkholderia multivorans P0213-1	Cystic fibrosis isolate, Canada Date of isolation 13-02-1996 ID: VC7495	D.P Speert University of British Columbia
B63	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *xdHA_3* gene	This study
B68	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *fadR* gene	This study
B74	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *rpsI* gene	This study
E28	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *pphA_2* gene	This study
E36	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *nemR_2* gene	This study
E50	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in a MMPL family protein gene	This study
E70	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *ppsA* gene	This study
E89	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *cycA_3* gene	This study

Table 1 Strains and plasmids used in this study (continuation).

Bacterial strains or plasmids	Description	Reference or source
Bacterial strains – *Burkholderia multivorans*		
F70	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in an outer membrane porin precursor gene	This study
G2	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *ttgR_3* gene	This study
G38	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *gltB_1* gene	This study
H21	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *clpB_1* gene	This study
H42	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *ypeA* gene	This study
C1-28	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *ttgR_3* gene	This study
C2-22	P0213-1 derivative with the plasposon pTn*Mod*-Km^r, inserted in *ttgR_3* gene	This study
P0213-1/pBBR1MCS	P0213-1 with vector pBBR1MCS	This study
P0213-1/PMF18-7	P0213-1 overexpressed with plasmid PMF18-7	This study
B68/pBBR1MCS	B68 with vector pBBR1MCS	This study
B68/PMF18-7	B68 complemented with plasmid PMF18-7	This study
P0213-1/PMF18-4	P0213-1 overexpressed with plasmid PMF18-4	This study
B74/pBBR1MCS	B74 with vector pBBR1MCS	This study
B74/PMF18-4	B74 complemented with plasmid PMF18-4	This study
P0213-1/PMF18-6	P0213-1 overexpressed with plasmid PMF18-6	This study
E36/pBBR1MCS	E36 with vector pBBR1MCS	This study
E36/PMF18-6	E36 complemented with plasmid PMF18-6	This study

Table 1 Strains and plasmids used in this study (continuation).

Bacterial strains or plasmids	Description	Reference or source
Bacterial strains – *Burkholderia multivorans*		
P0213-1/PMF18-8	P0213-1 overexpressed with plasmid PMF18-8	This study
G2/pBBR1MCS	G2 with vector pBBR1MCS	This study
G2/PMF18-8	G2 complemented with plasmid PMF18-8	This study
Bacterial strains – *Escherichia coli*		
JM109	F´ *traD36 proA*$^{+}$*B*$^{+}$ *lacI*^{q} *Δ(lacZ)M15/ Δ(lac-proAB) glnV44*	Promega
DH5-α	DH5α *recA1* Δ(*lacZYA-argF*)*U169* ϕ80d*lacZ*ΔM15	Gibco BRL
Plasmids		
pTn*Mod*Ωkm	Carrying a Km^{r} Plasposon with pMB1oriR	(68)
pRK600	ColE1 oriV; RP4tra^{+} RP4oriT; Cm^{r}; helper in triparental matings	(69)
pBBR1MCS	4,717-bp broad-host-range cloning vector, Cm^{r}	(70)
pUK21	3089-bp pUC21 derivative, Km^{r}	(71)
PMF18-1	pUK21 derivative containing a 597 bp NdeI/XbaI fragment with *rpsI* gene	This study
PMF18-2	pUK21 derivative containing 995 bp KpnI/XbaI fragment with *rpsI* promoter region and *rpsI* gene from pMF18-1	This study
PMF18-4	pBBR1MCS derivative containing the *rpsI* promoter region and *rpsI* gene from pMF18-2	This study
PMF18-6	pBBR1MCS derivative containing 753 bp KpnI/XbaI fragment with *nemR_2* promoter region and *nemR_2* gene	This study
PMF18-7	pBBR1MCS derivative containing 917 bp KpnI/XbaI fragment with *fadR* promoter region and *fadR* gene	This study
PMF18-8	pBBR1MCS derivative containing 1096 bp KpnI/XbaI fragment with ttg*R_3* promoter region and *ttgR_3* gene	This study

Abbreviations: Cm^{r}, chloramphenicol resistance; Km^{r}, kanamycin resistance

2.2 Transposon mutant library construction

2.2.1 Without enrichment process

A transposon mutant library of the clinical isolate *B. multivorans* P0213-1 was generated by triparental mating with *E. coli* carrying pRK600 (helper) or carrying pTn*Mod*Ωkm (donor) (Fig. S1). The donor strain was inoculated in 3 mL of LB with kanamycin, while the helper strain was inoculated in 3 mL of LB with chloramphenicol and the recipient strain was inoculated in 3 mL of LB. The three cultures were incubated at 37ºC with orbital agitation (250 rpm) for 5 hours. Following incubation, 1 mL of recipient was mixed with 1 mL of each helper and donor. The final mixture was transferred to a LB plate at 30ºC for 24 hours. This conjugation mixture was, then resuspended in 1ml of LB and plated onto 10 LB plates supplemented with 500 µg/ml kanamycin and 40 µg/ml gentamicin, each of them with 100 µl. All were incubated at 37ºC for 24 hours.

The mutant colonies obtained were inoculated into each well of 96-well plates and frozen at -80ºC with 30% of glycerol. The mutant colonies were numbered, as for example, A5 (mutant of cystic fibrosis isolate P0213-1 from well 5 of plate A).

2.2.2 With enrichment process

After performing the conjugation experiment as described, the conjugation mixture was inoculated in 5 mL of 0.9% (wt/vol) NaCl with gentamicin and incubated for 5 hours at 37ºC, in order to kill *E. colli* cells. After incubation, the suspension was centrifuged for 2 min at 8000 rpm of orbital agitation, and the pellet obtained was resuspended in 1 mL of the 30 mL of SM medium with gentamicin and kanamycin. The 30 mL culture was incubated at 37ºC at 180 rpm of orbital agitation for 24 hours. After 24 hours, the culture rested for 20 minutes thus enabling the aggregates in suspension settled in the pellet and 1 mL of the top culture was transferred to fresh SM medium with gentamycin and kanamycin. Five rounds of enrichment were performed before the free cells fraction was serially diluted and plated on LB plates supplemented with kanamycin, gentamicin and Congo red.

The colonies selected exhibited altered staining with Congo red (pattern, intensity) after 72 h of growth compared to the staining of the wild-type on a reference plate inoculated under similar conditions. The mutant colonies obtained were inoculated into each well of 96-well plates and frozen at -80ºC with 30% of glycerol. The mutant colonies were numbered, as for example, C1-5 (mutant of cystic fibrosis isolate P0213-1 from well 5 of plate C1).

2.3 Screening of transposon insertion libraries for differences in planktonic cellular aggregates formation

Each bacterial mutant was grown overnight in 3 mL of SM medium at 37ºC with 250 rpm of orbital agitation. Suspensions with OD_{640nm} 0.1, were prepared in fresh SM medium and incubated at 37ºC with 180 rpm of orbital agitation for 48 hours. After this time, each mutant was observed macroscopically and microscopically for aggregates and compared with wild-type (*B. multivorans* P0213-1) strain.

2.4 Microscopy analysis

B. multivorans strains grown in SM medium for 48 hours were visualized on Zeiss Axioplan microscope, equipped with an Axiocam 503 color Zeiss camera, using a 10x 0.3 NA objective, and controlled with the Zen software.

2.5 DNA manipulation techniques

Genomic DNA from *B. multivorans* strains was extracted using a protocol previously described (72). Plasmid DNA isolation and purification, DNA restriction, agarose gel electrophoresis, DNA amplification by PCR, and *E. coli* transformation were performed using standard procedures (73).

2.6 Detection of the plasposon in the genome of mutants

To identify if the plasposon was inserted into the genome of selected mutants, DNA amplification by PCR followed by agarose gel electrophoresis was performed. Genomic DNA was used as template for 50 µl PCR reactions with 2 µl primers pTn*Mod*ΩKm-fw (5'-GCAGAGCGAGGTATGTAGGC-'3) and pTn*Mod*ΩKm-rev (5'-TTATGCCTCTTCCGACCATC-'3). Amplification occurred as following: initial denaturation at 94ºC for 1:30 minutes; 30 cycles of 30 seconds at 94ºC, 1 minute at 59ºC, and 1:30 minutes at 72ºC; final extension at 72ºC for 7 minutes.

2.7 Identification of the genes disrupted by the plasposon

2.7.1 Sanger sequencing of flanking regions of the plasposon insertion site

The insertion position of the plasposon was determined in selected mutants by digesting genomic DNA with EcoRI (an enzyme not cutting DNA within the plasposon), followed by fragment self-ligation. Then, ligation mixtures were electroporated into *E. coli* electrocompetent cells using a Bio-Rad Gene Pulser II system (400 Ω, 25 µF, 2.5 kV) and grown for 1 hour before being plated on the selective medium LB supplemented with kanamycin. Plasmids were then recovered using ZR Plasmid Miniprep™-Classic kit (Zymoresearch®), following the manufacturer´s instructions. The recovered plasmids were sequenced using primer kmR (5'-CCTTTTTACGGTTCCTGGCCT-'3) and oriR (5'-GTGCAATGTAACATCAGAG-'3), through a Sanger sequencing system at the Instituto Gulbenkian de Ciência (Portugal).

The algorithm BLAST (74) was used to compare sequences of the gene disrupted to database sequences available at the National Center for Biotechnology Information (NCBI) and to genome sequence of the clinical isolate *B. multivorans* P0213-1 available in our group, but not yet deposited in any repository.

2.7.2 Genome sequencing and assembly

Genomic DNA from the *B. multivorans* clinical isolate derivative (B74 mutant) was extracted and purified using the DNeasy® blood and tissue kit (Qiagen®). The DNA sample was sequenced at least to 50-fold coverage using a NextSeq system at the Instituto Gulbenkian de Ciência (Portugal).

All obtained reads were trimmed using Sickle (75) and then aligned by the *de novo* assembler Spades (76). After assembly, using Bandage (77) it was possible to visualize the *de novo* assembly graph and determine how the assembly occurred as well as doing a Basic Local Alignment Search Tool (BLAST) to search for the plasposon location.

Genes were predicted and annotated using Prokka v1.11 (78). Additionally, a visual inspection of alignments using Geneious v.6.1.8 (79) allowed the confirmation of plasposon insertion in genome of the isolate, through the localization of unaligned reads and through search of plasposon genes.

2.8 Growth curves and doubling time estimation

Cells from an overnight culture were inoculated in 50 mL of fresh LB, SM medium or SCFM (synthetic cystic fibrosis medium) (80) (OD_{640nm} 0.1). The culture was incubated at 37ºC with 180 rpm of orbital agitation. OD_{640nm} readings were taken over time for 8 hours and then at 23 and 24 hours. Growth rates were calculated from the exponential phase of growth from at least two independent experiments. The doubling time was calculated from the growth rate of the exponential growth phase.

2.9 Transposon mutants complementation

2.9.1 *B. multivorans* electrocompetent cells

Cells from an overnight culture were inoculated in 30 mL of fresh LB (OD_{640nm} 0.1). The culture was incubated at 37ºC with 250 rpm of orbital agitation until reach an OD_{640nm} of 0.8 – 1.0. After incubation, 1.9 mL of culture was transferred to 2 mL Eppendorf tube, followed by a centrifugation at 9000 rpm of orbital agitation for 2 min at 4ºC. The supernatant was removed, and the pellet was washed three times with HEPES 1mM, followed by two wash steps with glycerol 10%. At the end, the cell pellet was resuspended with 40 µl of glycerol 10% and frozen at -80ºC.

2.9.2 Construction of recombinant plasmids

For functional complementation of transposon mutants, parental upstream promoter region and respective gene were cloned into the pBBR1MCS broad-host-range vector. Each gene and respective promotor were amplified from the genomic DNA of the wild-type strain (*B. multivorans* P0213-1), previously extracted and purified using the DNeasy® blood and tissue kit (Qiagen®), by PCR, using a set of forward and reverse primers (Table 2) that included engineered restriction endonuclease sites.

When the upstream promoter region and gene are amplified individually, the PCR products, after purification with DNA Clean & Concentrator™ kit (Zymoresearch®), were first cloned into pUK21 vector. The PCR products, as well as the pUK21 vector, were then digested by the same restriction endonucleases followed by ligation. Then, ligation mixtures were electroporated into *E. coli* electrocompetent cells using a Bio-Rad Gene Pulser II system (400 Ω, 25 μF, 2.5 kV) and grown for 1 hour before being plated on LB supplemented with kanamycin plus X-gal and IPTG for blue/white selection. Plasmids were then recovered using ZR Plasmid Miniprep™-Classic kit (Zymoresearch®), following the manufacturer´s instructions. Recovered plasmids were digested with the restriction enzymes used before, followed by agarose gel electrophoresis to verify the presence of fragments corresponding to the gene and the upstream promoter region.

The DNA fragment that comprises the upstream promotor region and the gene was removed from pUK21 and cloned into the pBBR1MCS vector through the same process described above for pUK21. In this case, selective medium was LB supplemented with chloramphenicol plus X-gal and IPTG for blue/white selection.

Recombinant plasmids obtained were purified, and electroporated into electrocompetent cells of *B. multivorans* P0213-1 and its transposon mutant derivatives using a Bio-Rad Gene Pulser II system (200 Ω, 25 μF, 2.5 kV) and grown for 4 hours before being plated on, LB supplemented with chloramphenicol (250 μg/ml).

An empty pBBR1MCS vector was electroporated into *B. multivorans* P0213-1 and its derivative transposon mutants and used as a control for functional complementation. Restoration of wild-type cellular aggregation phenotype was tested by growing the cells carrying empty control vectors or expressing parental genes in the same conditions described above.

Table 2 List of primers used in the complementation assay

Primer sets	Sequencies	
	Forward	**Reverse**
Pro_rplM	5’ GGGGTACCGAAATGGCGAACGCGCTC 3’	5’ GGAATTCCATATGTGCGGAAAAGCCCTGAATTATAA 3’
rpsI	5’ GGAATTCCATATGATCGGTAACTGGAACTACGGTACG 3’	5’ GCTCTAGAAAGCTTTAGCCGCTATTGTAGGG 3’
P0213_4221	5’ GGGGTACCAGGCTAAACAAGCCCGAG 3’	5’ GCTCTAGATTATGTCCGCACCTCGTC 3’
Bmul3465	5’ GGGGTACCTGTCTCCTCCGTCTCCAGA 3’	5’ GCTCTAGACTATTCTCGTTCCGGCGA 3’
P0213_1617	5’ GGGGTACCCTTCGTCGTACGTGCGTG 3’	5’ GCTCTAGAGCAGTGCGGTAAATCGGG 3’

* Restriction sites are underlined.

2.10 Quantification of cellular aggregates and free cells

The quantification of cellular aggregates and free cells was made based on the protocol previously described (6), with some adjustments. The growth cultures, after incubation in previously described conditions, were transferred to a 50 mL Falcon tube and centrifuged at 1400 rpm of orbital agitation at 25ºC for 30 seconds. After centrifugation, cell suspensions rested for 10 minutes, thus enabling aggregates still in suspension to settled down. Then, the supernatant was removed by pipetting and placed in a new 50 mL Falcon tube. Suspensions containing aggregates were transferred to 2 mL Eppendorf tubes and, then, after a several quick-spins, it was possible to obtain in a 2 mL Eppendorf tube all cellular aggregates of same mutant strain.

The 50 mL Falcon tube with free cells and growth medium was centrifuged for 10 minutes at 4000 rpm of orbital agitation at 25ºC, allowing the separation of free cells from the growth medium. The resulting pellet was resuspended in 5 mL of growth medium and, then, several centrifugations at 13400 rpm of orbital agitation for 2 minutes were performed in 2 mL Eppendorf tubes, in order to collect all free cells of the isolate.

In the end of the procedure, two 2 mL Eppendorf tubes were obtained for each strain. These tubes were opened and placed at 60ºC during at least 72 hours until all cellular aggregates and free cells were dried, presenting a brown color. All 2 mL Eppendorf tubes were weighted before the collection of samples and after samples were dried.

2.11 Exopolysaccharide production

YEM (yeast extract mannitol medium) plates containing 4 g/L mannitol, 0.5 g/L yeast extract and 15 g/L agar were used to evaluate the production of exopolysaccharide (81). After inoculation, YEM plates were incubated for 48 h at 37°C and mucoidy due to exopolysaccharide production was observed by visual inspection.

2.12 Antimicrobial susceptibility

Antimicrobial susceptibility of wild-type and its derivative strains was assessed based on the agar disc diffusion method (82). Müller-Hinton agar (Sigma-Aldrich) plates were inoculated with 100 µl of a suspension at an OD_{640nm} of 0.1 prepared from exponential-phase cells growing on LB medium at 37ºC. Paper discs (BD BBL Sensi-Disc) containing aztreonam (30 µg), piperacillin (75 µg) plus tazobactam (10 µg), piperacillin (75 µg) and tetracycline (30 µg) were applied onto the surface of the inoculated plates. The diameter of the growth inhibition zone was measured after 24 hours of incubation at 37°C. Results are the mean values from five replicates from at least two independent experiments.

2.13 Biofilm formation

Bacteria were grown in LB medium at 37°C to mid-exponential phase and diluted to an OD_{640nm} of 0.05. 200 µL samples of the cell suspensions were used to inoculate 96-well polystyrene microtiter plates. 200 µL of LB medium were used to inoculate some wells functioning as blanks. Plates were incubated at 37°C statically for 48 h, after which the wells were washed three times with 0.9% (wt/vol) NaCl. The biofilm was stained with 200 µL of a 1% (wt/vol) crystal violet solution for 20 minutes at room temperature (83), followed by washing three times with 200 µL of 0.9% (wt/vol) NaCl. The dye was then solubilized with 200 µL of 96% ethanol and the biofilm was quantified by measuring the absorbance of the solution at 590 nm (A_{590nm}) in a microplate reader (Spectrostar nano, BMG LabTech). Results are the means of data from at least six replicates of three independent experiments.

2.14 Motility

2.14.1 Swarming motility

The swarming agar plates containing 0.04% (wt/vol) tryptone, 0.01% (wt/vol) yeast extract, 0.0067% (wt/vol) $CaCl_2$, 0.6% (wt/vol) bacto agar (Difco) (11) were spot inoculated with a 5 µl drop of a culture at an OD_{640nm} of 1.0. After inoculation, swarming plates were incubated for 48 h at 37°C and the diameter of the swarming zone was measured. Results are the means of data from at least ten replicates of two independent experiments.

2.14.2 Swimming motility

The swimming agar plates containing 1% (wt/vol) tryptone, 0.5% (wt/vol) NaCl, 0.3% (wt/vol) noble agar (Difco) (11) were spot inoculated with a 5 µl drop of a culture at an OD_{640nm} of 1.0. After inoculation, swimming plates were incubated for 24 h at 37°C and the diameter of the swimming zone was measured. Results are the means of data from at least ten replicates of two independent experiments.

2.15 Virulence determination in *Galleria mellonella*

Killing assays were performed as described previously (84). *Galleria mellonella* larvae were injected with cell suspensions containing a total CFU of approximately $1x10^4$ in 10 mM $MgSO_4$ with 1.2 mg/ml ampicillin and incubated at 37°C. Survival rates were assessed during the following 3 days post-infection. As a negative control, 10 mM $MgSO_4$ with 1.2 mg/ml ampicillin was used. Ten larvae per isolate were used in at least three independent experiments.

2.16 Statistical analyses

The statistical significance of differences in the data was determined using the one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparisons test or Turkey's multiple comparisons test and using the Mantel-Cox test which were performed using GraphPad Prism software v.5.04 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). Kaplan-Meier survival curves were also performed with GraphPad Prism software v.5.04. Differences were considered statistically significant for *P*-values lower than 0.05.

3 RESULTS AND DISCUSSION

3.1 Screening of *Burkholderia multivorans* plasposon mutants for different abilities to form cellular aggregates

When grown in liquid SM medium, which has high carbon to nitrogen ratio, *B. multivorans* P0213-1 clinical isolate grows both as free cells and cellular aggregates (Fig. 13). These 48 hours aggregates range from microscopic to macroscopic, reaching sizes up to 5 mm.

In order to identify molecular mechanisms involved in the formation of cellular aggregates, it was constructed a plasposon mutant library from P0213-1, using the transposon present in pTn*Mod*ΩKm (Fig. S1). From this plasposon mutant library, about 900 colonies were kept and screened to determine if any of them presented differences in the formation of cellular aggregates. Both mutants and wild-type strain were incubated at 37ºC with 180 rpm of orbital agitation for 48 hours and, then, mutants were visually compared against the wild-type at macroscopic and microscopic levels.

From the 900 mutants screened, 15 mutants were selected due to their distinct phenotype (Fig. 13). These mutants were grown several times to confirm their different ability to form cellular aggregates.

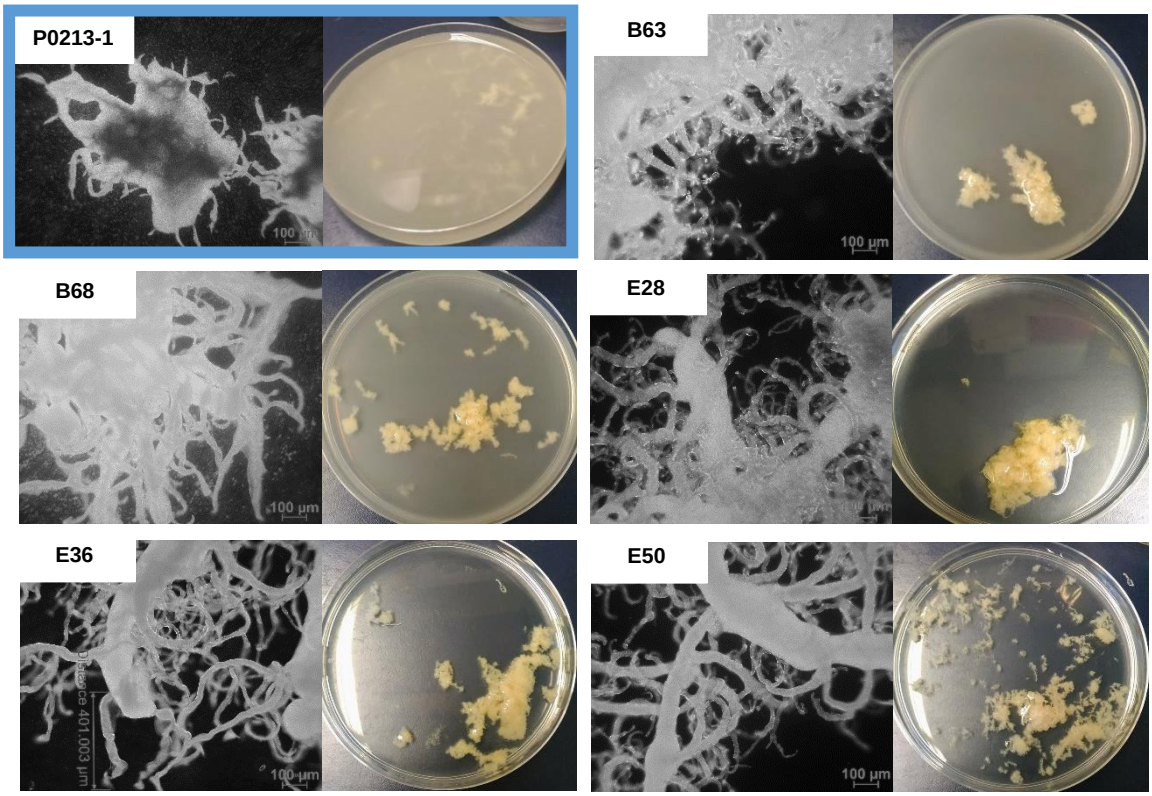

Fig. 13 Screening of the plasposon mutant library of *B. multivorans* P0213-1 for different abilities to form planktonic cellular aggregates. For each strain is shown a microscopy image (on the left) and the liquid culture on a petri dish (on the right). Each culture was grown in SM medium at 37ºC, 180 rpm of orbital agitation, for 48 hours.

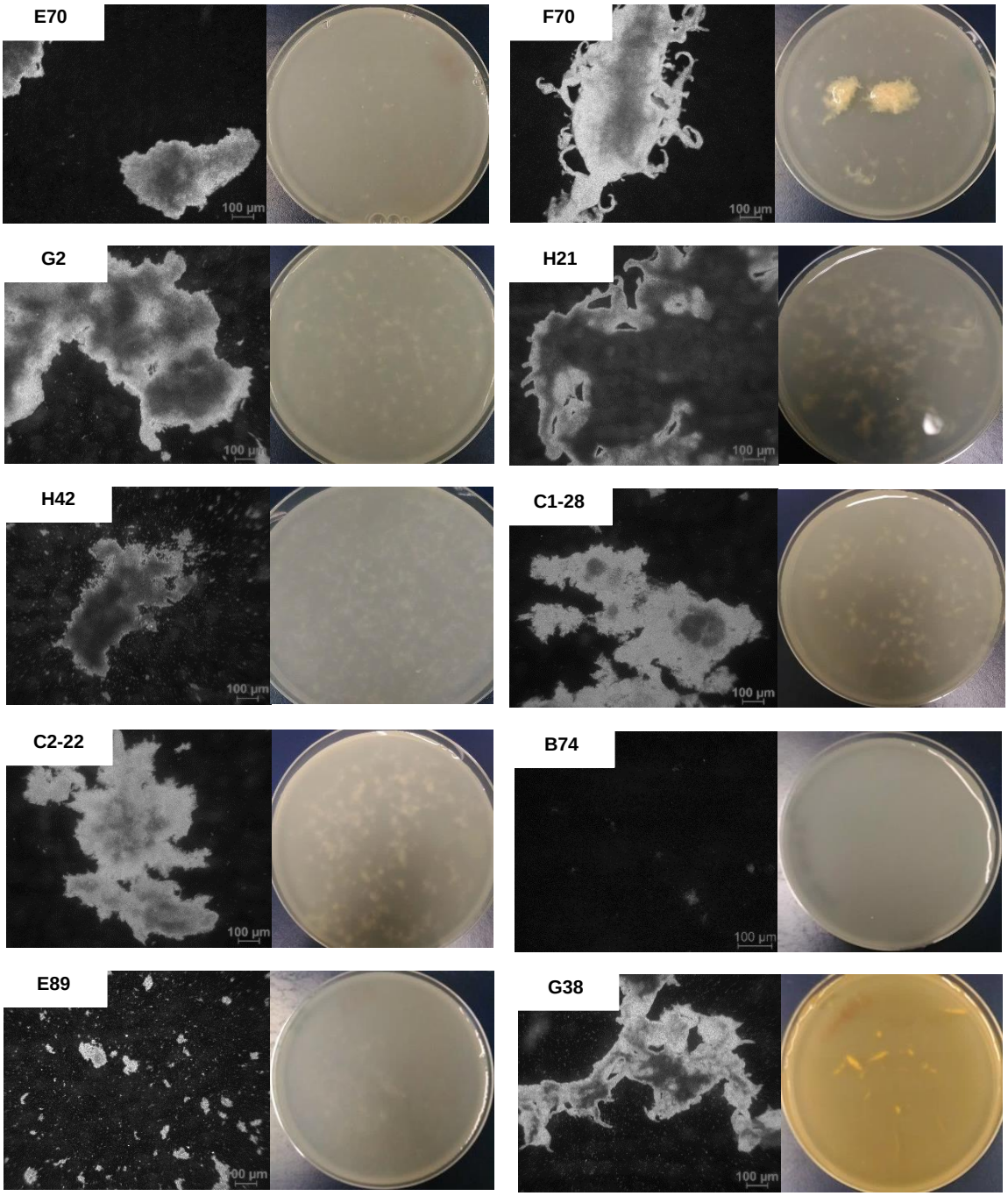

Fig. 13 Screening of the plasposon mutant library of *B. multivorans* P0213-1 for different abilities to form planktonic cellular aggregates (continuation). For each strain is shown a microscopy image (on the left) and the liquid culture on a petri dish (on the right). Each culture was grown in SM medium at 37ºC, 180 rpm of orbital agitation, for 48 hours.

B63, B68, E28, E36 and E50 mutants produce cellular aggregates that comprise the majority of cells in the culture, as it can be seen from the low turbidity of the medium and the low amount of cells present in the microcopy image (Fig. 13). These aggregates seen with 100x magnification are highly structured, showing main branches from where smaller ramifications are protruding.

The next group of selected mutants has an intermediate ability of forming cellular aggregates. In this category are included G2, E70, F70, H21, H42, C1-28 and C2-22 mutants. In general, the size of the aggregates is smaller than the ones of the wild-type strain and they are much less

structured. In this type of aggregates no significant branching is observed, perhaps with the exception of F70 mutant.

The last category of selected mutants includes B74, E89 and G38 mutants. As depicted in figure 13 these mutants produce very few macroscopic aggregates. At the microscopic level, B74 is the one producing smaller aggregates, followed by E89. G38 mutant produces few macroscopic aggregates, but with low structure. Additionally, G38 liquid cultures turn orange while the wild-type and other mutants are white/yellow.

3.2 Confirmation of the presence of the plasposon in the mutant's genome

To confirm that the different abilities to form aggregates by the selected mutants were due to the plasposon insertion and consequently, a disrupted gene, the genomic DNA of the fifteen mutants was extracted and used for plasposon amplification by PCR, followed by agarose gel electrophoresis (Fig. 14). In this experiment the primers were designed to amplify an internal region of the plasposon with a size of 1161 bp.

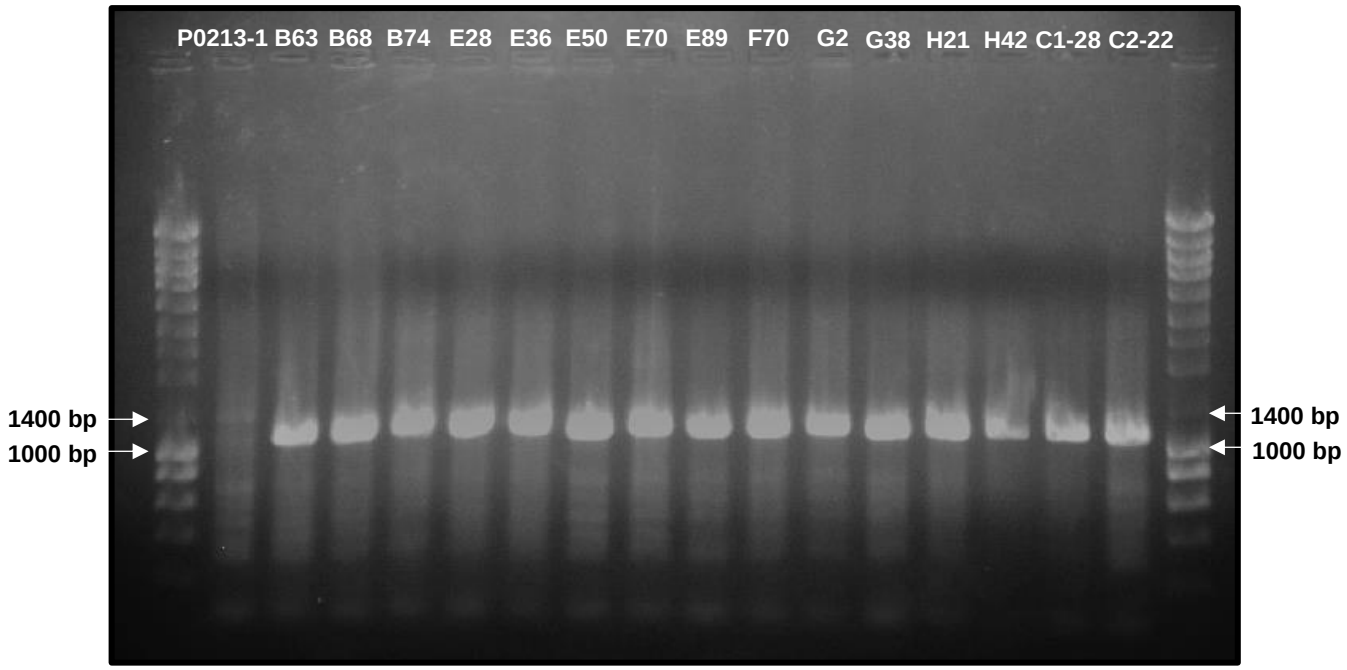

Fig. 14 Electrophoretic separation in 0.8% agarose gel of the PCR products corresponding to an internal region of the pTn*Mod*ΩKm plasposon with an estimated size of 1161 bp. DNA marker is the NZYDNA Leader III.

As shown in figure 14, all selected mutants presented a band with a size between 1000 bp and 1400 bp, which corresponds to the size of the desired fragment of the plasposon. As expected, the wild-type P0213-1 did not present that band, suggesting that the insertion of the plasposon is most likely the cause of the observed phenotypes.

3.3 Identification of plasposon insertion in the genome of the selected mutants

Having demonstrated the presence of the plasposon in the genome of the selected mutants, the next step was to identify their location and consequently the disrupted genes. To do that, two different strategies were tested: sequencing the mutant's genome or sequencing only the flanking regions of the site where the plasposon was inserted.

In the two different strategies used, the genome of the wild-type strain *B. multivorans* P0213-1 was used as the reference genome, since it has been previously sequenced and assembled (85). The assembly of the wild-type genome generated 5 contigs with lengths between 813 and 3298492 bp (Table 3). Through BLAST was possible to identify which contig corresponded to each B. *multivorans* ATCC 17616 chromosome. Contigs 1, 2 and 3 correspond to chromosomes 1, 3 and 2, respectively, and contig 4 also aligned to chromosome 1. Contig 5 did not map onto the reference genome of B. *multivorans* ATCC 17616 and is likely a plasmid.

Table 3 The 5 contigs generated from the assembly of the P0213-1 wild-type genome and the corresponding chromosomes of B. *multivorans* ATCC 17616.

P0213-1 contigs	Contig length (bp)	*B. multivorans* ATCC 17616 chromosome	*B. multivorans* ATCC 17616 chromosome length (bp)
1	3 298 492	1	3 448 421
2	706 259	3	919 805
3	2 464 683	2	2 473 162
4	813	1	-
5	28 489	-	-

In the first strategy, the genome of B74 mutant was sequenced by Illumina short reads technology. After trimming reads and quality assessment, a *de novo* assembly was performed, resulting in 93 contigs. The assembly graph obtained during the *de novo* assembly was visualized with the Bandage bioinformatic tool (Fig. 15). Within this tool it was possible to align the plasposon sequence against the contigs using BlastN and identify which one contains this plasposon.

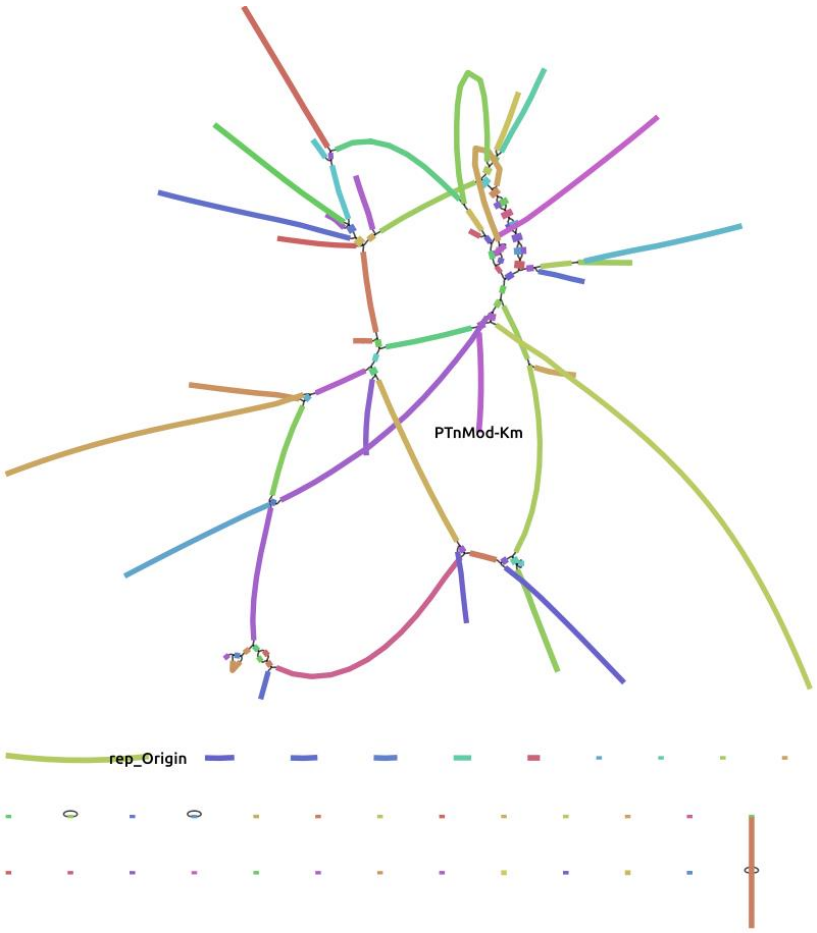

Fig. 15 Assembly graph of B74 mutant's genome obtained after sequencing and *de novo* assembly. Each contig generated from *de novo* assembly is represented through a line. Black lines represent edges that connected two different contigs. PTnMod_Km and rep_Origin indicate the localization of two different regions of plasposon.

In figure 15, each line represents one contig obtained in *de novo* assembly previously performed, and black lines represent edges that connected two different contigs. Based on this figure, it was possible to confirm that *de novo* assembly was successful, due to the high number of dead ends (ends that not connected with others) and due to the low complexity of graph. It is also worth noting that there are just a few disconnected contigs and the connected contigs are present in just one subgraph.

Since this bioinformatic tool also allowed a BLAST search for the plasposon, two different regions of the plasposon sequence were selected to perform the alignment: rep_Origin, the sequence encoding R6K origin of replication and PTnMod-km, the sequence encoding kanamycin phosphotransferase. For the two sequences only one hit was found for each of them. With this result it was possible to conclude that the plasposon was inserted only at one site of the genome (Fig. S2) although it is split into two contigs.

A visual inspection of the alignment using Geneious v.6.1.8 software (79) allowed the confirmation of plasposon insertion in the genome of the mutant, through the localization of unaligned reads and through the search of the plasposon genes (Fig.16).

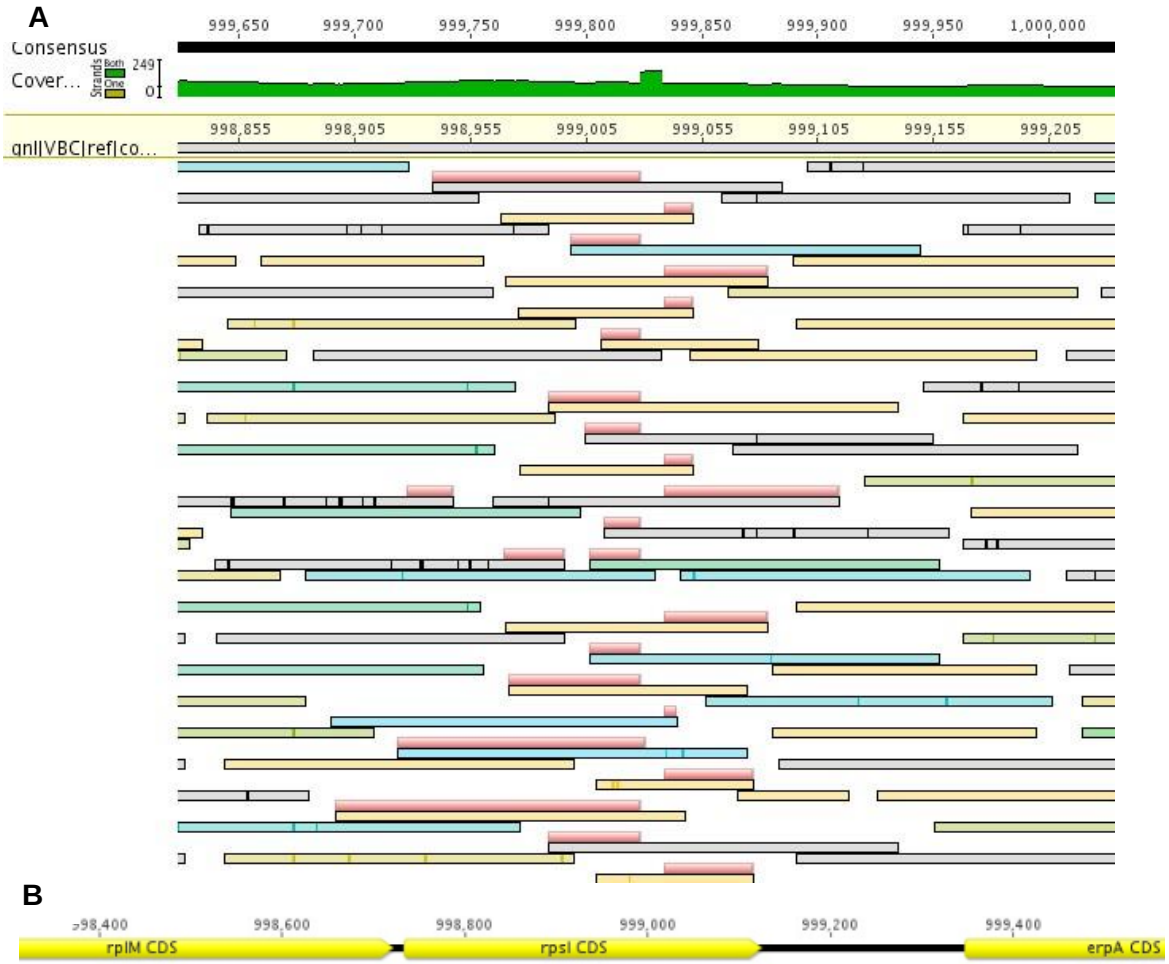

Fig. 16 (A) The alignment of reads from B74 mutant's genome against P0213-1 wild-type genome using Geneious v.6.1.8. In red are presented the reads unable to aligned with the reference genome. **(B) The unaligned reads localization.** The B74 mutant has the transposon inserted in *rpsI* gene.

Figure 16(A) presents a region of the alignment where some reads (red reads) did not align with the P0213-1 reference genome. The region shown in figure 16(A) is the only one that presents reads that cannot be aligned with the reference genome. To confirm if these red reads corresponded to the plasposon sequence, the sequence of kanamycin phosphotransferase gene was searched in the alignment and the red reads were the output. Through the localization of these reads it was possible to identify the gene disrupted by the plasposon. In the case shown in figure 16, the genome analyzed belongs to B74 mutant and the disrupted gene was *rpsI* encoding ribosomal protein S9 (Fig. 16(B)).

The second strategy to identify the gene disrupted by the plasposon was the restriction of each genome by a restriction endonuclease unable to cut the plasposon (such as EcoRI) followed by self-ligation of the fragments and electroporation of these ligations into *E. coli*. Following this procedure, it was possible to obtain colonies for each mutant genome ligations. Extraction of plasmid DNA from these colonies confirmed the presence of a plasmid with more than 5 kbp in all of them (data not shown). Sanger sequencing of the flanking regions was done at least from one colony. Disrupted genes were identified by aligning sequences of the flanking regions of each

mutant (Table S1) against the genome sequence of the wild-type strain and also against *B. multivorans* ATCC 17616.

Despite the success of both strategies, the second one was chosen to identify the remaining locations of plasposon insertion in the *B. multivorans* mutants' genome, as well as the identification of the disrupted genes (Table 4).

Table 4 Genes disrupted by the plasposon were identified in the *B. multivorans* selected mutants.

***B. multivorans* mutant**	**Chromosome**	**Start position**	**Locus tag**	**Gene name**	**Annotation**	**Homolog in *B. multivorans* ATCC 17616**
B63	2	1769684	PROKKA_05177	*xdhA_3*	Xanthine dehydrogenase molybdenum-binding subunit	Bmul_4199
B68	2	950289	PROKKA_04504	*fadR*	Fatty acid metabolism regulator protein	Bmul_3465
E28	2	2094940	PROKKA_05470	*pphA_2*	Phosphoenolpyruvate phosphomutase	Bmul_4471
E36	2	1789933	PROKKA_05198	*nemR_2*	HTH-type transcriptional repressor NemR	Bmul_4221
E50	2	1999993	PROKKA_05383	-	MMPL family protein	WK22_RS22775*
E70	1	2422886	PROKKA_02266	*ppsA*	Phosphoenolpyruvate synthase	Bmul_1274
F70	2	2351677	PROKKA_05712	-	Outer membrane porin protein precursor	Bmul_4600
G2	1	2877205	PROKKA_02681	*ttgR_3*	HTH-type transcriptional regulator TtgR	Bmul_1617
H21	1	777106 | 780959	PROKKA_00733 | PROKKA_00735	-	ImpA family type VI secretion-associated protein | Type VI secretion protein	Bmul_2923 | Bmul_2925
H42	1	1901049	PROKKA_01776	*ypeA*	Acetyltransferase YpeA	Bmul_0803
C1-28	1	2877205	PROKKA_02681	*ttgR_3*	HTH-type transcriptional regulator TtgR	Bmul_1617
C2-22	1	2877205	PROKKA_02681	*ttgR_3*	HTH-type transcriptional regulator TtgR	Bmul_1617
B74	1	998733	PROKKA_00932	*rpsI*	30S ribosomal protein S9	Bmul_2703
E89	2	1088239	PROKKA_04620	*cycA_3*	Cytochrome c-552 precursor	Bmul_3585
G38	1	702354	PROKKA_00659	*gltB_1*	Ferredoxin-dependent glutamate synthase 1	Bmul_0305

* Homolog found in *B. multivorans* AU1185, since no homologue was present in *B. multivorans* ATCC 17616.

Through the schematic representation present in figure 17, it is possible to observe that each disrupted gene identified is present in the chromosome 1 or in the chromosome 2 of clinical isolate *B. multivorans* P0213-1.

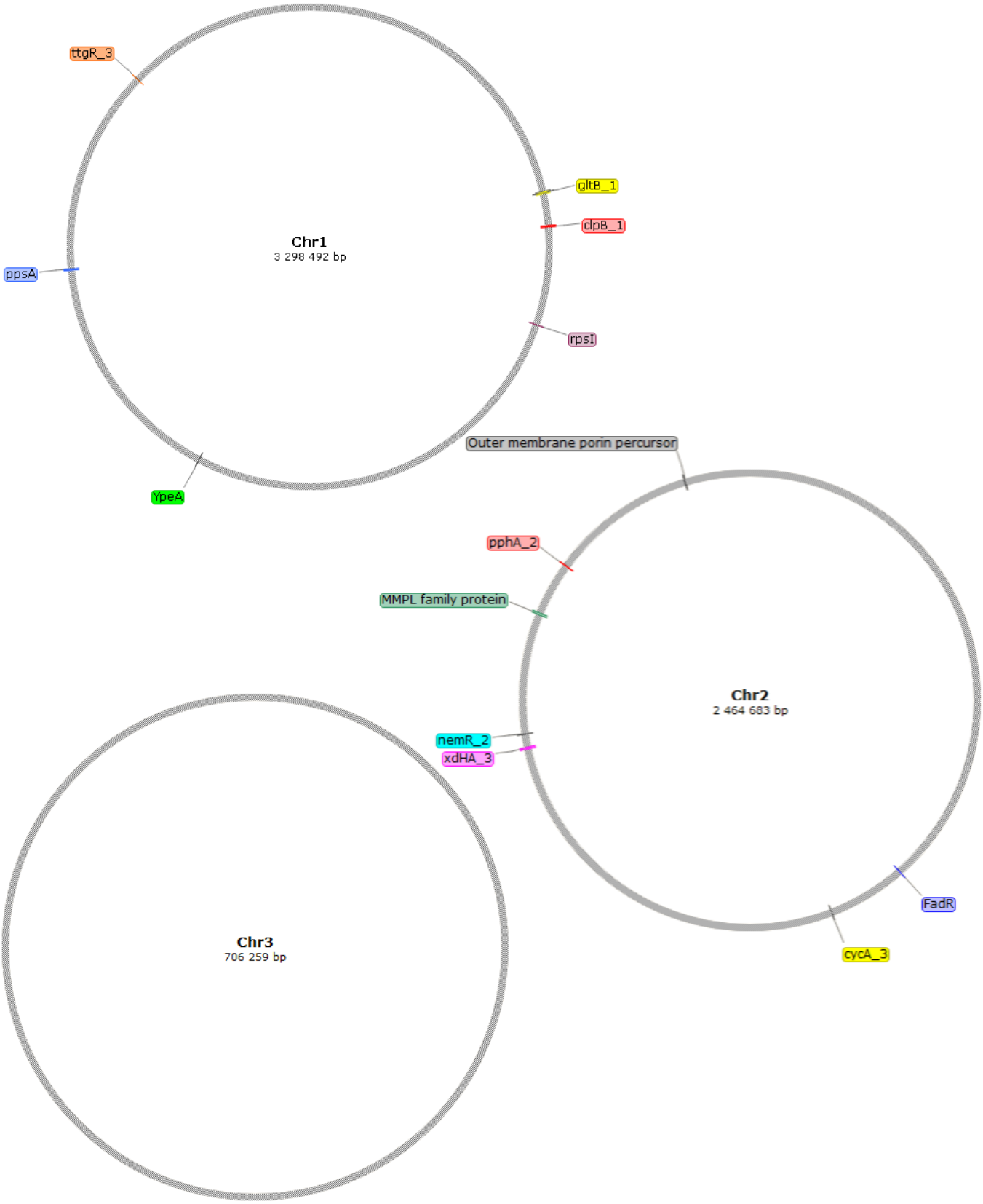

Fig. 17 Mapping of the disrupted genes identified among the 15 *B. multivorans* mutants in chromosomes 1 (Chr1), 2 (Chr2) and 3 (Chr3) (86)**. Chromosome representation is not at scale.**

3.4 Analysis of the mutated genes and their possible role in cellular aggregation

The five mutants producing large aggregates and very low amount of free cells have mutated genes included in the functional classes of carbon and energy production, lipid metabolism and nitrogen acquisition and assimilation. Figures 18 highlights the mutated genes and their gene neighbors. To easily obtain the flanking regions of the mutated gene we used the graphics for *B. multivorans* ATCC 17616 genome available at the *Burkholderia* genome database. Nevertheless, P0213-1 genome present exactly the same regions.

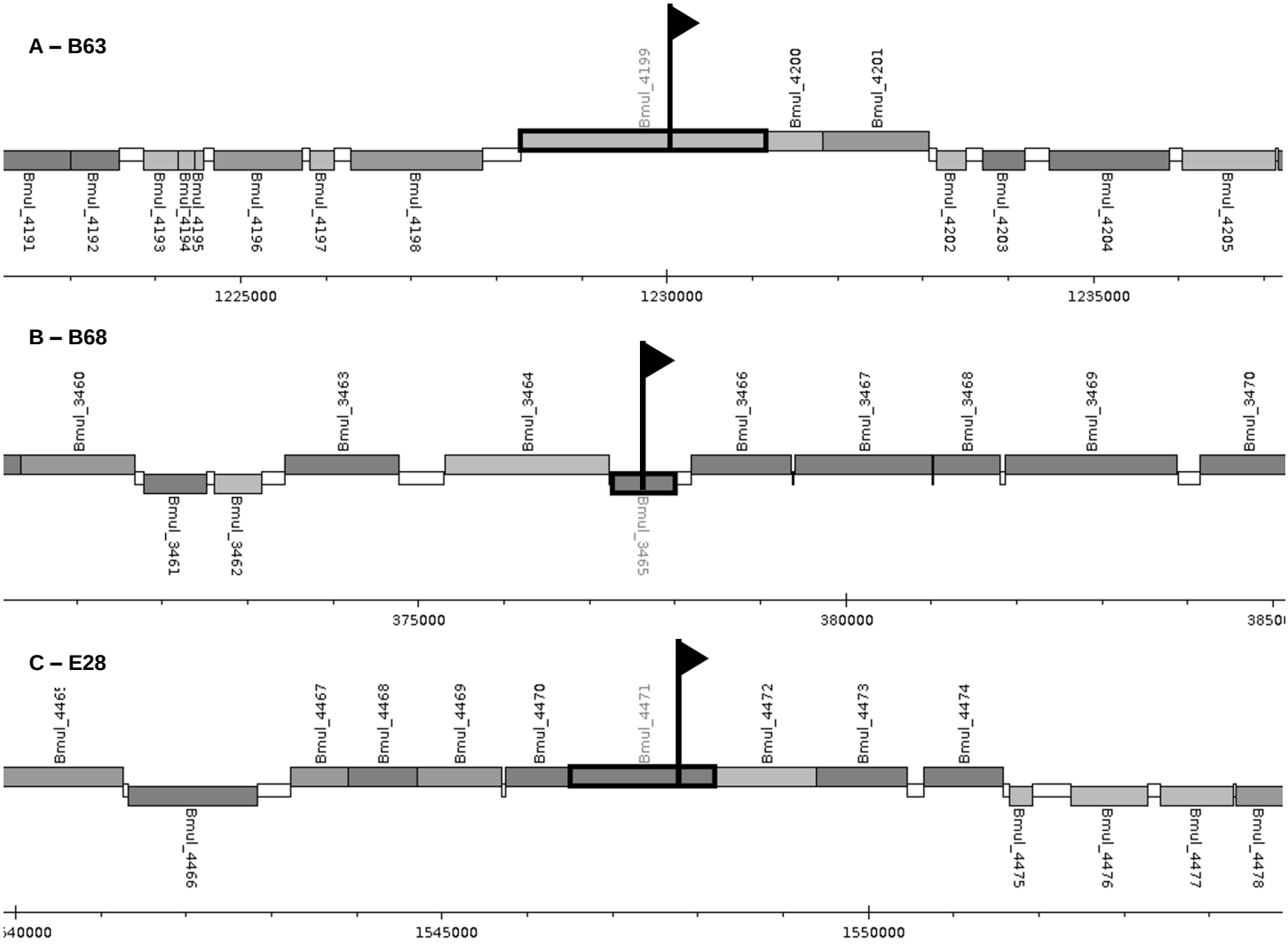

Fig. 18 Genes disrupted by insertion of the plasposon and their neighborhood in the *B. multivorans* genome. (A) In B63 mutant, the plasposon is inserted in *xdh*A_3 (Bmul_4199) gene (flagged). Together in the same operon is Bmul_4200 and Bmul_4201 genes encoding a 2Fe-2S iron-sulfur cluster binding domain-containing protein and a gluconate 2-dehydrogenase, respectively. **(B)** In B68 mutant, the plasposon is inserted in *fadR* (Bmul_3465) gene (flagged). Downstream of *fadR* are four genes belonging to an operon and upstream is a gene encoding a hypothetical protein. **(C)** In E28 mutant, the plasposon is inserted in *pphA*_2 (Bmul_4471) gene (flagged). This gene is within a putative 7 genes operon. **(D)** In E36 mutant, the plasposon is inserted in *nemR_2* (Bmul_4221) gene (flagged). Downstream of *nemR_2* is a gene encoding a hypothetical protein and upstream is a gene encoding an alcohol dehydrogenase. **(E)** In E50 mutant, the plasposon is inserted in WK22_RS22775 (considering the *B. multivorans* AU1185 genome) gene (flagged). This gene is within a possible operon with at least 18 genes.

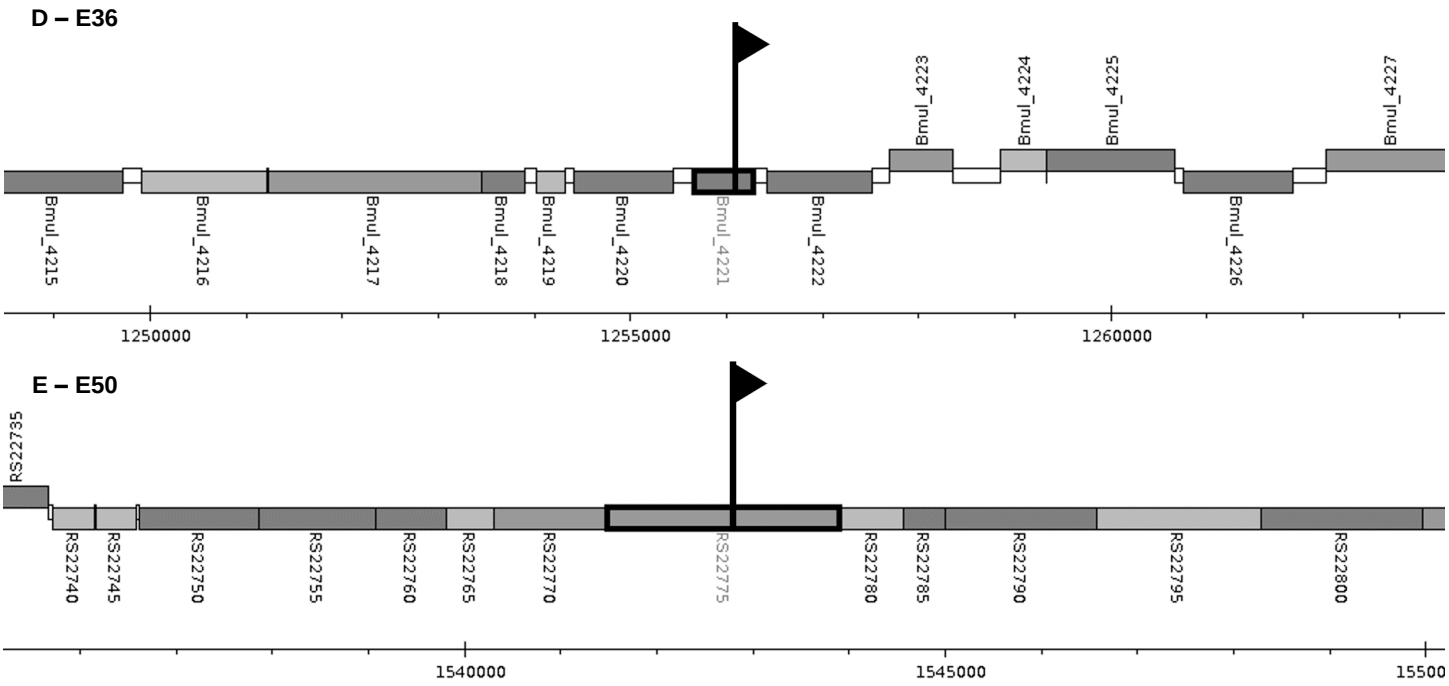

Fig. 18 Genes disrupted by insertion of the plasposon and their neighborhood in the *B. multivorans* genome (continuation). (A) In B63 mutant, the plasposon is inserted in *xdh*A_3 (Bmul_4199) gene (flagged). Together in the same operon is Bmul_4200 and Bmul_4201 genes encoding a 2Fe-2S iron-sulfur cluster binding domain-containing protein and a gluconate 2-dehydrogenase, respectively. **(B)** In B68 mutant, the plasposon is inserted in *fadR* (Bmul_3465) gene (flagged). Downstream of *fadR* are four genes belonging to an operon and upstream is a gene encoding a hypothetical protein. **(C)** In E28 mutant, the plasposon is inserted in *pphA*_2 (Bmul_4471) gene (flagged). This gene is within a putative 7 genes operon. **(D)** In E36 mutant, the plasposon is inserted in *nemR_2* (Bmul_4221) gene (flagged). Downstream of *nemR_2* is a gene encoding a hypothetical protein and upstream is a gene encoding an alcohol dehydrogenase. **(E)** In E50 mutant, the plasposon is inserted in WK22_RS22775 (considering the *B. multivorans* AU1185 genome) gene (flagged). This gene is within a possible operon with at least 18 genes.

B63 mutant has the gene *xdhA_3* disrupted by the plasposon. This gene is the first of a three members predicted operon and encodes the xanthine dehydrogenase molybdenum-binding subunit XdhA. This protein is required for the purine salvage pathways. Purine synthesis is essential, since their products provide the bases for DNA and RNA, cofactors, signaling molecules and carriers of energy. Purines can be synthesized by *de novo* pathways or recycled by salvage pathways. Therefore, the salvage pathway is a way of cell to preserve energy especially in starvation condition (87).

During starvation, both *E. coli* (88) and *Streptomyces coelicolor* (87) increase the production of (p)ppGpp (guanosine tetraphosphate and guanosine pentaphosphate), also called alarmones. As (p)ppGpp levels increase a significant reduction in GTP (guanosine triphosphate) concentrations takes place, thus the role of Xdh (xanthine dehydrogenase) in this metabolic pathway becomes critical in replenishing the GMP (guanosine monophosphate) / GTP pool in order to maintain the level of messengers like c-di-GMP (cyclic dimeric guanosine monophosphate) and (p)ppGpp (87) (Fig. 19).

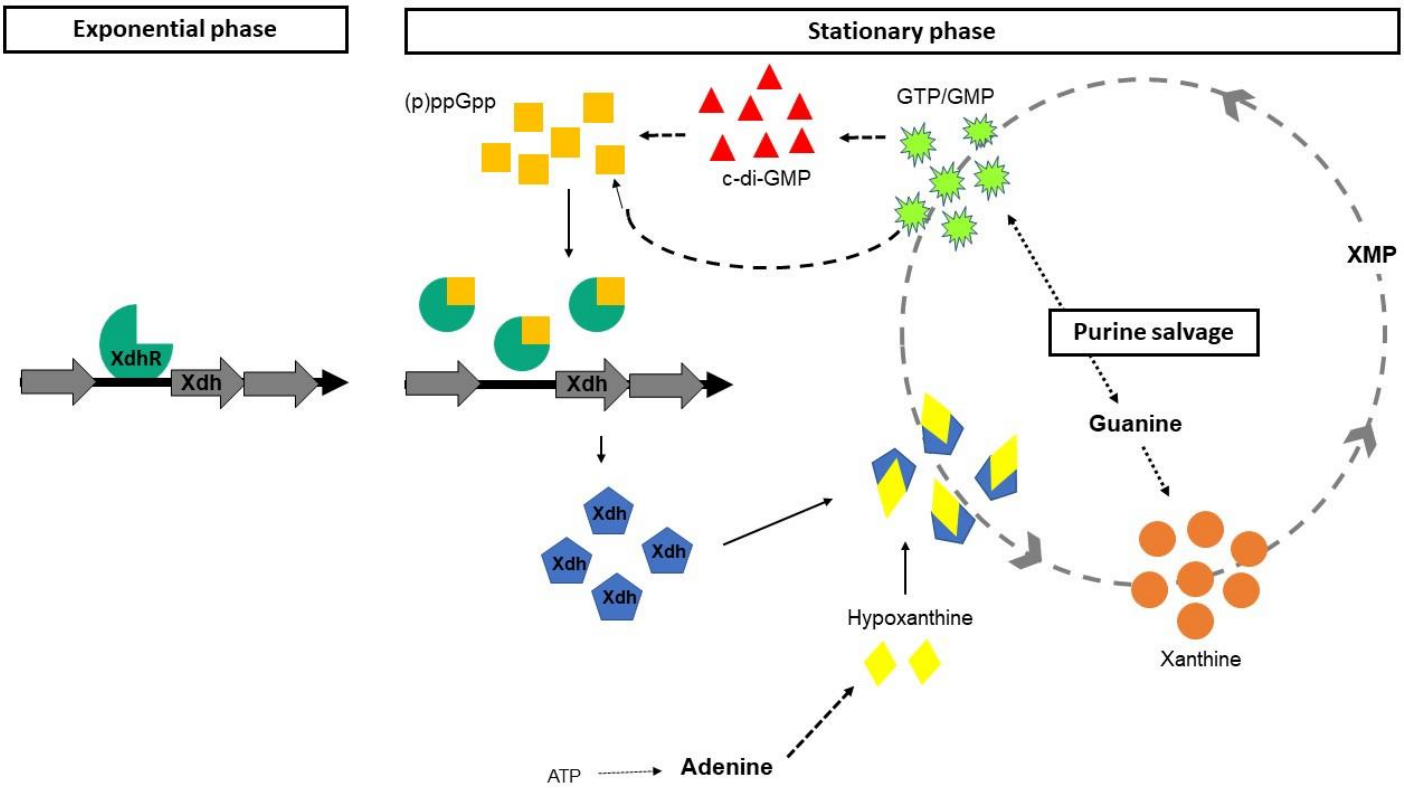

Fig. 19 Purine salvage pathway in *Streptomyces coelicolor*. XdhR (xanthine dehydrogenase regulator), in exponential phase, is bound to DNA, repressing the transcription of xanthine dehydrogenase (Xdh). When nutrients become limited, (p)ppGpp levels increase in response to starvation status and, xanthine dehydrogenase regulator, promotes transcription of *xdh* gene. Xdh protein, in its turn, catalyzes the reaction that produces xanthine from hypoxanthine. Xanthine is used as substrate and after several successive reactions the production of GMP/GTP occurs. GMP/GTP is necessary to produce c-di-GMP and alarmones. Thus, this pathway suffers autoregulation to maintain the GMP/GTP pool and, consequently, to maintain c-di-GMP/(p)ppGpp levels (adapted from (87)).

Since SM medium is poor in nitrogen, the fact that B63 mutant is unable to recycle purines, creates an even stronger nitrogen limitation, potentiating the formation of aggregates in detriment of free-living cells. Under these conditions, is likely that the levels of alarmone and c-di-GMP are altered, which will influence several phenotypic traits. Effectively several studies have shown that the intracellular messengers ((p)ppGpp and c-di-GMP) regulate biofilm formation, motility, biosynthesis of exopolysaccharides and virulence in several bacteria (89–91).

The mutated gene in B68 mutant encodes a transcriptional regulator of the TetR family. Usually, these regulators are located in the vicinity of the genes they regulate but are transcribed divergently. In here, it might regulate an operon of four genes (Fig. 18(B)) whose activities are: acyl-CoA dehydrogenase, propionyl-CoA carboxylase, enoyl-CoA hydrolase and carbamoyl-P-synthase L chain ATP-binding. These proteins are involved in the degradation of L-leucine (Fig. 20), in particular, in the conversion of 3-Methylbutanoyl-CoA into (S)-3-Hydroxy-3-methylglutanyl-CoA, this last one acts as precursor for terpenoid biosynthesis. Depending on whether this regulator is a repressor or an activator, the effect in the cell might be different. If it is a repressor, its mutation leads to constitutive expression of the four genes operon shifting metabolism into terpenoid biosynthesis. If it is an activator would shift metabolism into branched chain fatty-acid biosynthesis. In either case, it will induce additional metabolic stress (perhaps even affecting the

synthesis of signaling molecules such as cis-2-dodecenoic acid, BDSF and N-acyl-homoserine lactones) and it could cause the higher aggregation of cells observed.

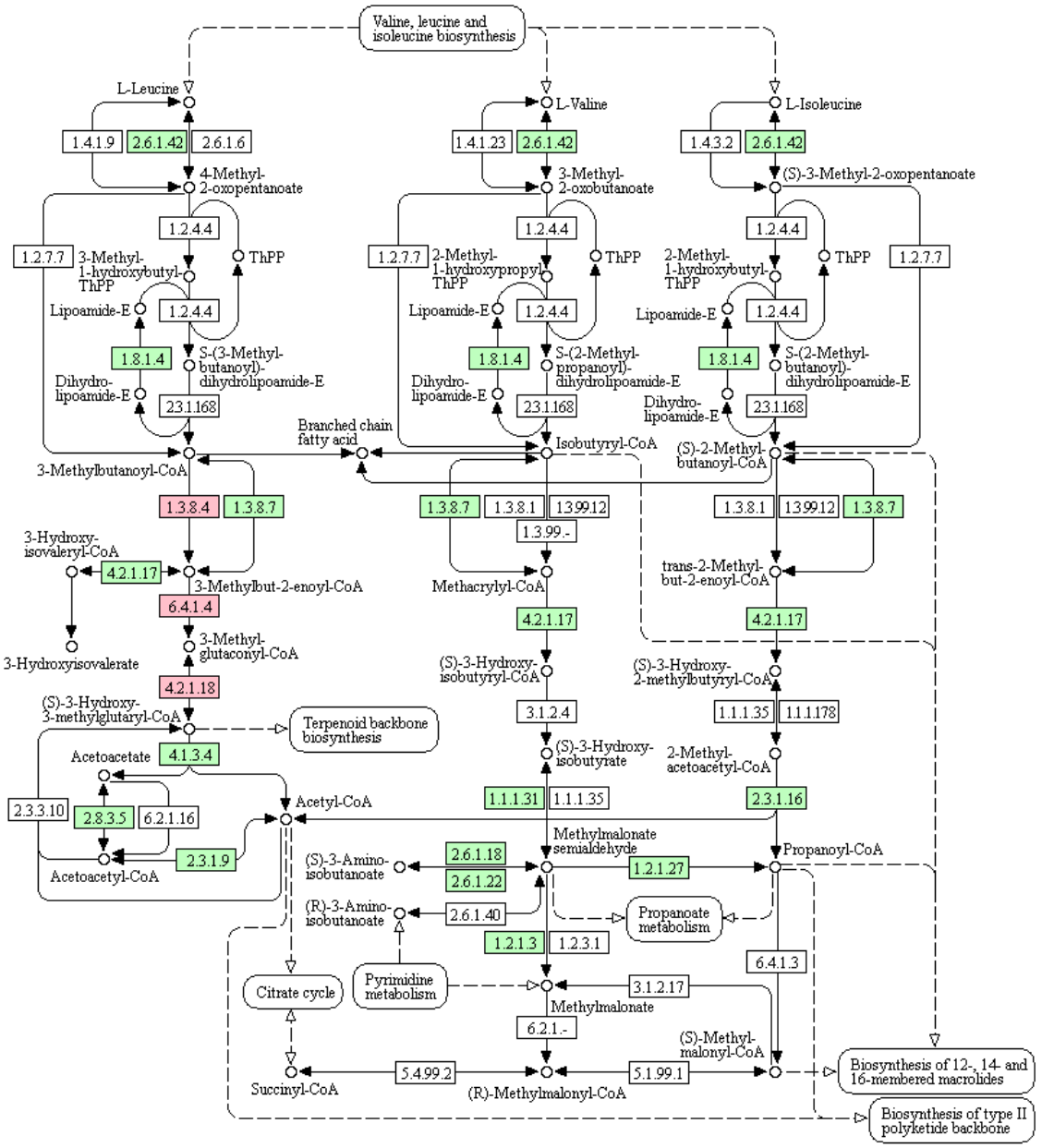

Fig. 20 Representation of valine, leucine and isoleucine degradation with genes highlighted in red: 1.3.8.4 – Bmul_3466; 6.4.1.4 – Bmul_3467 and Bmul_3469; 4.2.1.18 – Bmul_3468. These genes belong to an operon located downstream of *fadR* (gene disrupted in B68 mutant) (retrieved from (92)).

E28 mutant has the plasposon inserted into the fifth gene of a putative 7 genes operon. This mutated gene encodes a phosphoenolpyruvate phosphomutase, that together with the two last gene products of the operon are responsible for converting phosphoenolpyruvate into 2-aminoethylphosphonate (Fig. 21) in the phosphonate metabolism (93). This pathway is involved in the synthesis of some antibiotics, phosphonolipids, and surfactants. It could be that the absence of phosphonolipids in the membrane or the absence of a putative surfactant alters surface properties of cells, increasing cell aggregation.

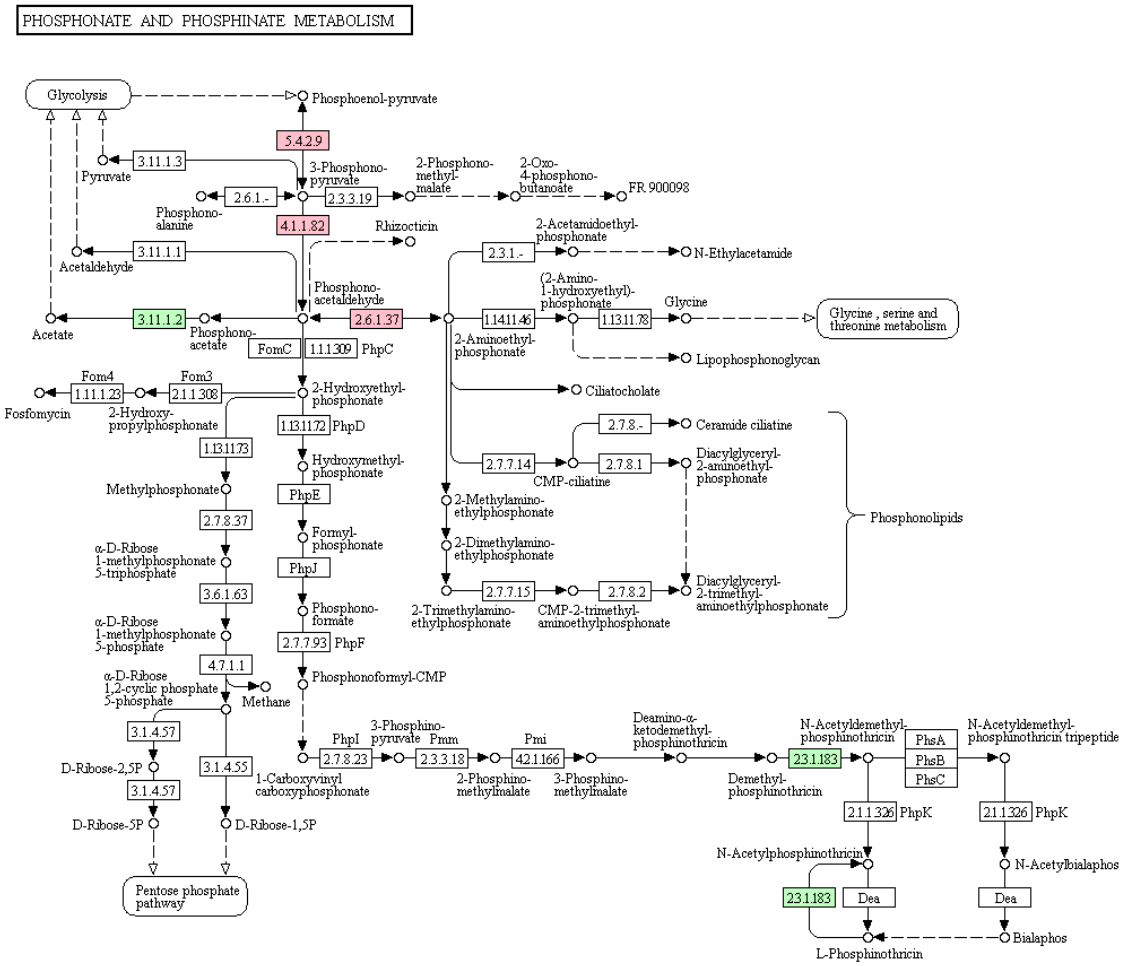

Fig. 21 Representation of phosphonate and phosphinate metabolism with genes highlighted in red: 5.4.2.9 – Bmul_4471 (gene disrupted in E28 mutant); 4.1.1.82 – Bmul_4472; 2.6.1.37 – Bmul_4473. These genes belong to a putative 7 genes operon (retrieved from (92)).

E36 mutant has a mutation in a gene encoding a TetR family transcriptional regulator. The genes located in its vicinity, a DNA-binding protein and an alcohol dehydrogenase, do not give clues into its possible role.

E50 mutant has the plasposon inserted into a gene whose product might be involved in transport functions at the membrane (94). This gene is within a possible operon with at least 18 genes and these might be involved in the synthesis of molecules with lipid backbones since some of the gene products are involved in fatty acid synthesis. This gene cluster is not present in microorganisms such as *B. multivorans* ATCC 17616 or *B. cenocepacia* J2315. Still, homologues are present in other *B. multivorans, B. vietnamiensis, B. territorii,* among others. Whatever is the function of this gene region, in its absence the formation of aggregates increases.

The next group of mutants is still able to form aggregates, but of smaller size and unstructured. In this category are included E70, F70, G2, C1-28, C2-22, H21 and H42 mutants. Their location in genome is shown in figure 22.

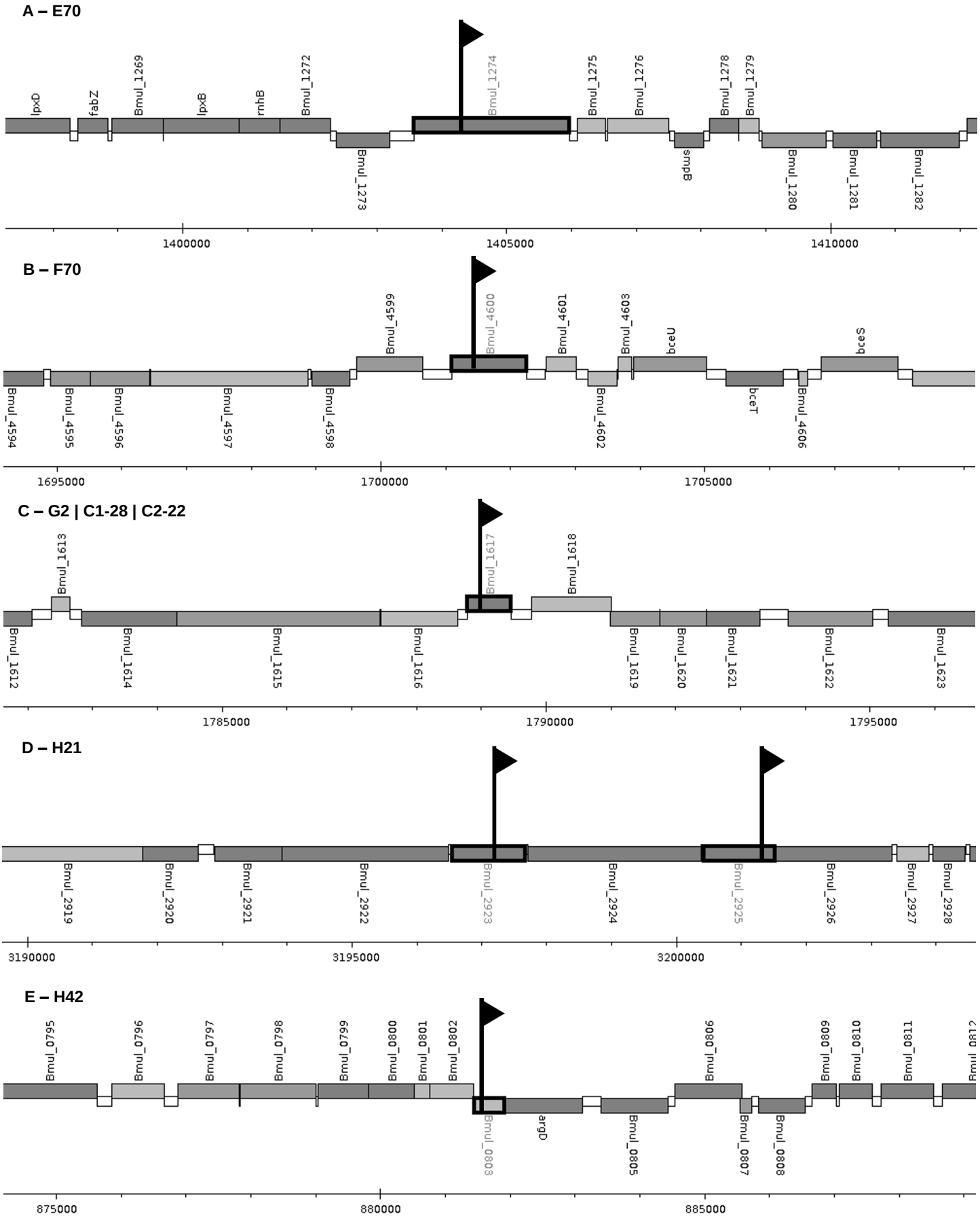

Fig. 22 Genes disrupted by insertion of the plasposon and their neighborhood in the *B. multivorans* genome. (A) In E70 mutant, the plasposon is inserted in *ppsA* (Bmul_1274) gene (flagged). Both genes upstream and downstream of *ppsA* gene are genes encoding hypothetical proteins. **(B)** In F70 mutant, the plasposon is inserted in Bmul_4600 gene (flagged). Downstream of Bmul_4600 gene is a gene encoding 3-

demethylubiquinone-9-3-methyltransferase and upstream is a gene encoding a ribonuclease BN. **(C)** In G2, C1-28 and C2-22 mutants, the plasposon is inserted in *ttgR_3* (Bmul_1617) gene (flagged). Downstream of *ttgR_3* are three genes belonging to an operon and upstream is a gene encoding a peptidase M23B. **(D)** In H21 mutant, the plasposon is inserted in Bmul_2923 and Bmul_2925 genes (flagged). These genes are within a possible operon**. (E)** In H42 mutant, the plasposon is inserted in *ypeA* (Bmul_0803) gene (flagged). Together in the same operon is *argD* gene which encodes acetylornithine transaminase protein.

E70 mutant is disrupted in gene *ppsA* encoding phosphoenolpyruvate synthase, an enzyme converting pyruvate into phosphoenolpyruvate. Utilization of carbon source of SM medium (mannitol) by the Entner-doudoroff pathway leads to glyceraldehyde-3P and pyruvate. The fact that the step of conversion of pyruvate into phosphoenolpyruvate is blocked leads probably to more pyruvate, more acetyl-CoA and perhaps an increase of lipid metabolic reactions. In a work using *B. cenocepacia*, Veselova and co-authors reported a mutant in phosphoenolpyruvate synthase which showed increased ability to produce N-acyl-homoserine lactones (AHL), the signal molecules of the quorum sensing system (95). Although we have no evidences that the same mechanism is acting in E70 mutant, we can also not exclude that quorum sensing might play some role in cellular aggregation. Further studies need to be done.

F70 mutant has the plasposon inserted into a gene encoding a putative porin. This type of proteins is involved in the transport of different substrate in and out of the cells. Possibly, it transports some compounds required for cells metabolism, which in the mutant are blocked.

G2 mutant, as well as C1-28 and C2-22, have a disruption in a gene encoding a transcriptional regulator of the TetR family. This gene is in the opposite direction of three genes encoding a putative RND efflux system. This efflux pump named Amr-OprM, as well as the repressor AmrR, are involved in aminoglycoside resistance as demonstrated in *B. vietnamiensis* (96). The constitutive expression in G2 mutant of the efflux pumps genes in the absence of antibiotics seems to have a negative effect in promoting the growth of cellular aggregates, but the mechanism behind that is unknown.

The recovery of the flanking regions from H21 mutant identified the plasposon in two adjacent location. One disrupts the gene encoding type VI secretion system protein TssA (homologue to Bmul_2923) and the other plasposon is inserted into *tssG* gene encoding another protein of the same secretion system. Another possibility is that the insertion of the plasposon caused the deletion of the region within the two detection sites. In both cases, since the disrupted genes belong to the same operon, we kept this mutant in further studies.

The type VI secretion system is a bacterial nanomachine used to deliver toxins directly into eukaryotic or prokaryotic target cells. As many secretion system, this one is an important virulence factor, being also implicated in biofilm formation (97). Although the exact function of *tssA* and *tssG* genes is unknown, *B. cenocepacia* K56F2 deletion mutants in these genes have shown their critical role for the type VI secretion system activity (98). H21 mutant has most likely an inactive type VI secretion system. The fact that it still forms aggregates, but of smaller size implicates this nanomachine in this process.

The last mutant of this group, H42, has the plasposon inserted into the *ypeA* gene encoding a putative acetyltransferase. This gene is the last of a 2-operon structure, being the first gene *argD*,

which is involved in the synthesis of N-acetylornitine (Fig. 22(E)). Whether ypeA acetyltransferase is also involved in this pathway and how it influences cellular aggregation is unknown.

The last group of mutants includes B74, E89 and G38 and are the most affected in cellular aggregates formation. Genes disrupted are shown in figure 23.

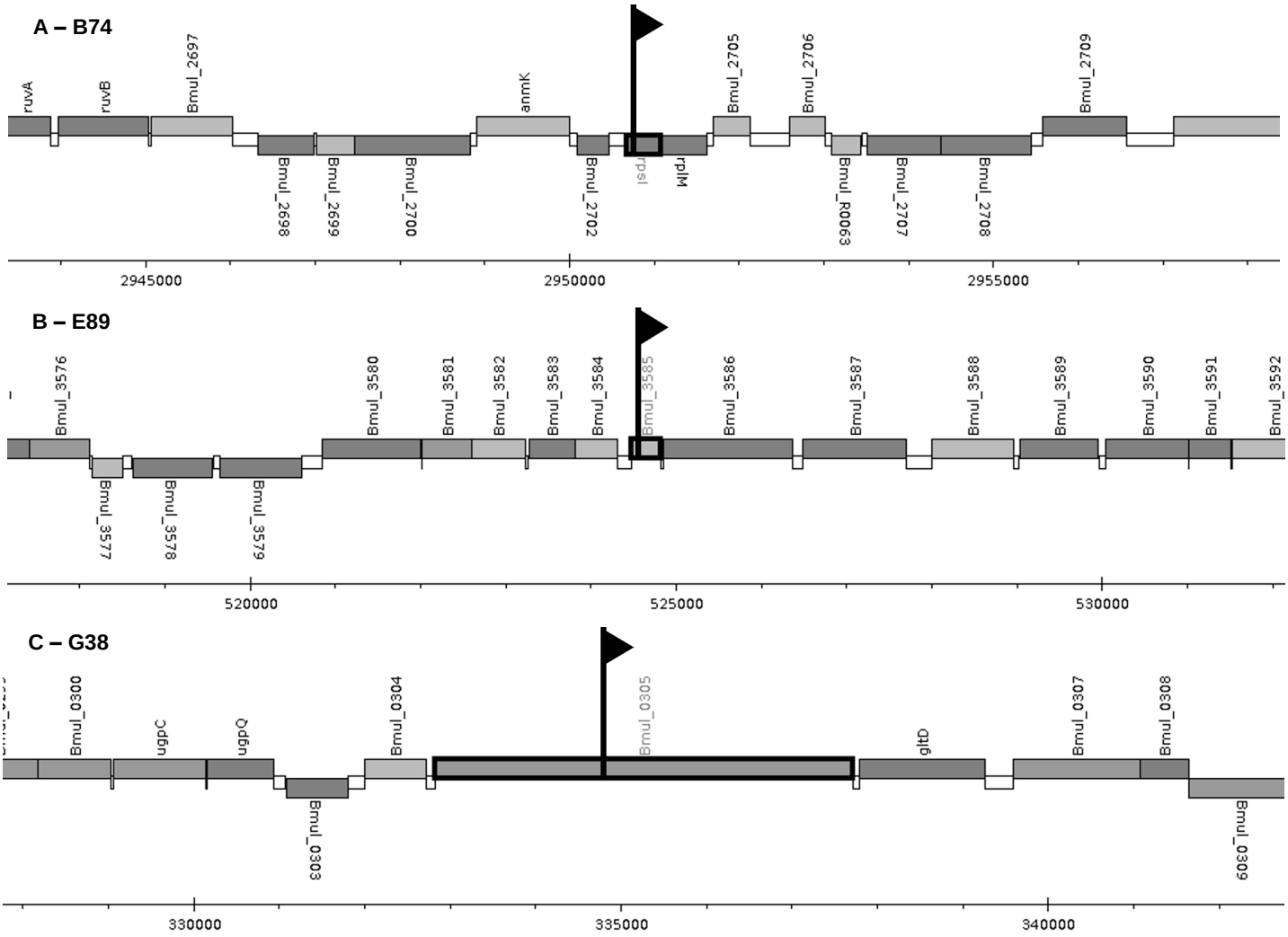

Fig. 23 Genes disrupted by insertion of the plasposon and their neighborhood in the B. multivorans genome. (A) In B74 mutant, the plasposon is inserted in *rpsI* gene (flagged). Together in the same operon is *rplM* gene encoding 50S ribosomal protein L13. Downstream of *rplM* is a gene encoding a hypothetical protein. Upstream is a gene encoding an iron-sulfur cluster insertion protein ErpA. **(B)** In E89 mutant, the plasposon is inserted in *cycA_3* (Bmul_3585) gene (flagged). Downstream of *cycA_3* gene is a gene encoding an aldehyde dehydrogenase and upstream is a gene encoding another cytochrome c. **(C)** In G38 mutant, the plasposon is inserted in *gltB_1* (Bmul_0305) gene (flagged). Downstream of *gltB_1* is a gene encoding a glutamate synthase subunit beta and upstream is a gene encoding a hypothetical protein.

In B74 mutant the disrupted gene, the *rpsI*. This gene encodes the 30S ribosomal protein S9 involved in translation. Ribosomal protein S9 has an important role in 30S subunit assembly reaction and, consequently, in the association of 30S and 50S subunits of ribosomes (99). Little is known about this protein. However, it was associated with the arise of *E. coli* kasugamycin-dependent mutants and also was reported that temperature can affect the protein functionality (100). Recently, this protein was suggested to be part of a protein-based vaccine against *Streptococcus pneumoniae* due to its immunogenicity. This result gives an indication that protein

S9 it might be a moonlight protein. Despite this characteristic, mice immunized with recombinant S9 protein and then challenged with a virulent strain of *S. pneumoniae* presented a significant reduction of bacteremia after 24 hours of infection, but died after four days of infection (101). The most likely hypothesis is that B74 mutant without a functional S9 protein, presents some defects in translation due to the possibility of uncorrected association of 30S and 50S subunits of ribosome and consequently the translation of some transcripts might be affected.

E89 mutant has the plasposon inserted in gene *cycA*_3 encoding a putative cytochrome c. The gene downstream, possibly from the same operon, encodes an enzyme metabolizing one of the steps of styrene degradation. Since cytochrome c-type proteins are involved in electron transfer, it is possible that one of the steps of this degradative pathway might involve this protein. The role of this pathway in cellular aggregation is unknown.

The last mutant of this group (G38 mutant) has *gltB* gene disrupted, which encodes glutamate synthase protein. As mentioned before, SM medium is poor in nitrogen and after 48 hours of growth cells probably face nitrogen starvation. Nitrogen is a major nutrient for cells and, for that reason, its metabolism is highly regulated by several mechanisms (102). For most bacteria, ammonium is the preferred source of nitrogen, however, when ammonium is not available at required concentrations, bacteria resort to other routes utilizing other sources of nitrogen (103).

Ammonium assimilation in *P. aeruginosa* (104) as well as in *Burkholderia cenocepacia* H111 (102) could be catalyzed by three different enzymes: an NADP-dependent glutamate dehydrogenase (NADP-GDH) and a combined action of glutamine synthetase (GS) and glutamate synthase (GOGAT). Lardi *et al.* (102), have determined which enzymes are involved in nitrogen metabolism in *B. cenocepacia* H111 under nitrogen limited conditions. The authors found that *gdh*A encoding for the glutamate dehydrogenase (GDH) was downregulated, suggesting that the via GS/GOGAT is used to assimilate ammonium in these conditions. Several enzymes required in nitrogen metabolism are regulated by a global nitrogen regulatory system (Ntr system) that senses the nitrogen status of cells (Fig. 24).

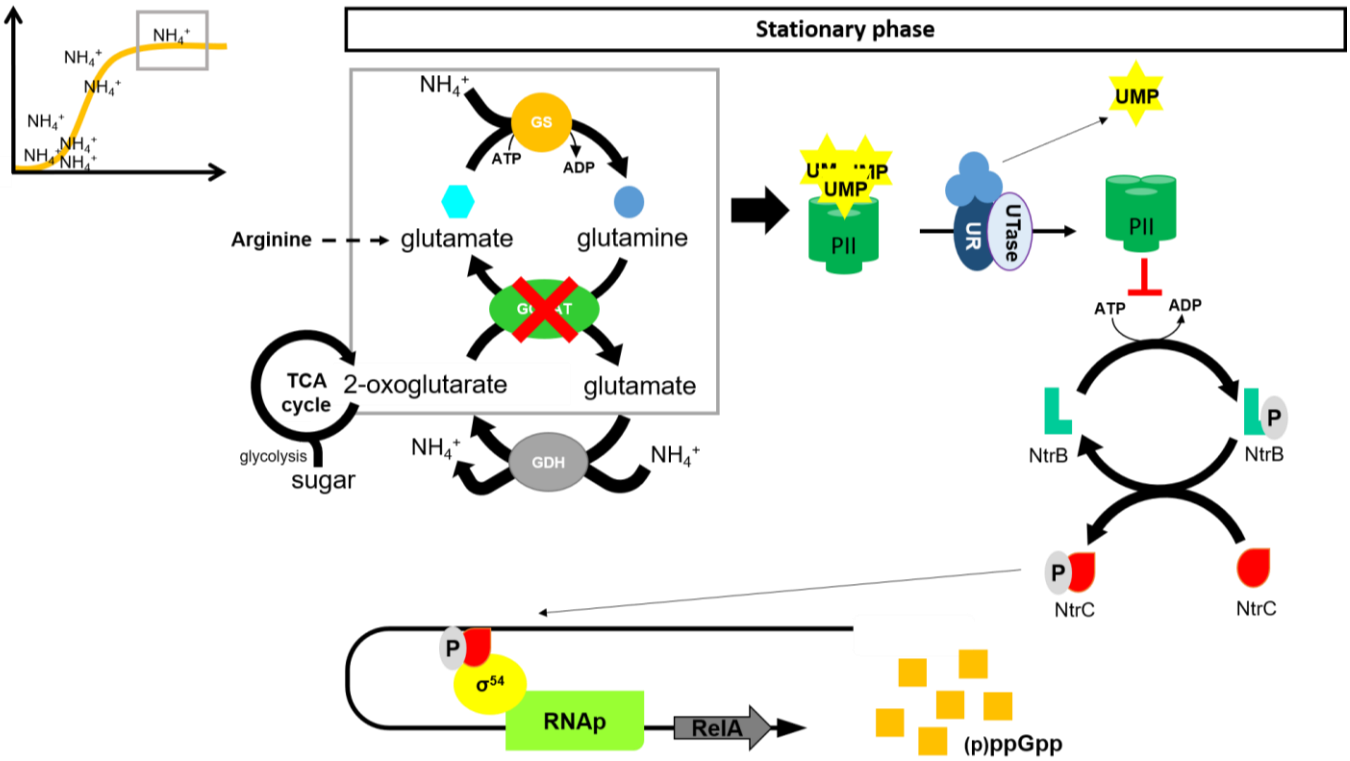

Fig. 24 Global nitrogen regulatory system (Ntr) and ammonium assimilation pathway. Under nitrogen limited conditions, the intracellular level of glutamine decreases, thereby activating the kinase activity of NtrB, which, in its turn, leads to phosphorylation of NtrC. Phosphorylated NtrC activates transcription at promoters dependent on the alternative sigma factor σ^{54} (RpoN), which binds to consensus regions of DNA (adapted from (105, 102, 106)).

In G38 mutant, one of the two enzymes involved in ammonium assimilation pathway is not functional, since *gltB* gene is disrupted. Consequently, the intracellular levels of glutamine might increase, once it cannot be used by non-functional GOGAT enzyme, but glutamate is still formed from arginine succinyltransferase (AST) pathway, allowing the continuous assimilation of NH_4^+ substrate (104). Therefore, the possible increase of glutamine levels and the decrease of 2-oxoglutarate levels might lead to the turned off of Ntr system. This event begins with the binding of glutamine to UR part of the bifunctional protein UR/UTase (uridyltransferase / uridylyl-removing enzyme) that remove UMP (uridine monophosphate) from the small trimeric signaling protein PII. In its turn, PII protein, responsible for the control of the NtrBC two-component regulatory system, is unable to activate the kinase activity of NtrB, that consequently would lead to phosphorylation of NtrC, an enhancer-binding protein (EBP). Thus, without NtrC phosphorylation, this protein together with the alternative sigma factor σ^{54} (or RpoN), cannot activate the transcription of several genes as, for example, the *rel*A gene responsible for alarmones (p)ppGpp synthesis (105, 102, 104).

The fact that this mutant still forms aggregates but of smaller size implicates somehow this Ntr system regulatory pathway in the aggregation process.

3.5 *In vitro* growth analysis of selected mutants

After having mapped the transposon location in the different mutants, it is important to estimate their growth kinetics to exclude possible effects on aggregation caused by growth impairment. Although aggregation was tested in SM medium, we have determined growth properties in LB. The reason is the lower amount of aggregates formed in LB, which allows a more reliable measurement of the optical density of the cultures. For that, *B. multivorans* P0213-1 and the plasposon mutants were grown in LB medium for 24 hours with 180 rpm orbital agitation as shown in figure 25.

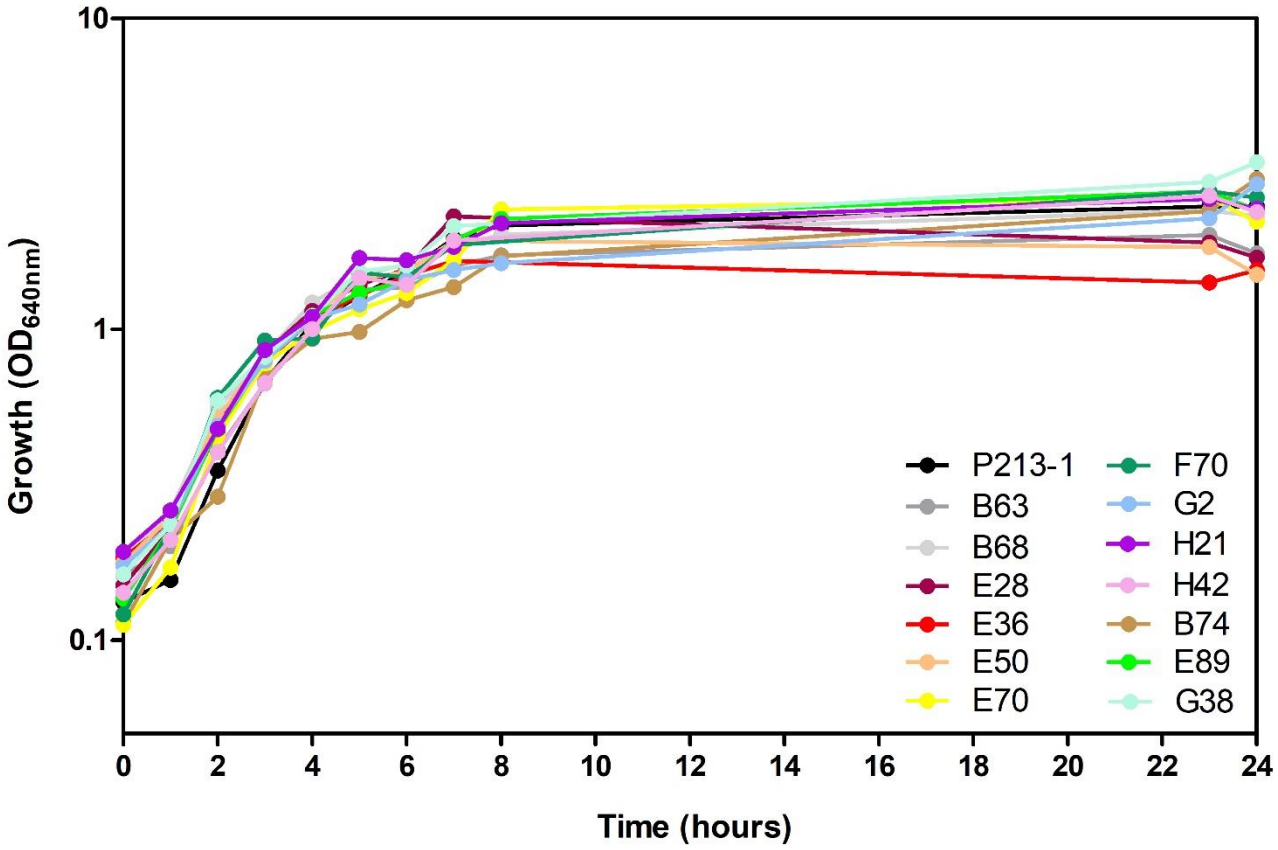

Fig. 25 Growth curves of the strains under study. Cultures were grown in LB medium at 37ºC, 180 rpm of orbital agitation, and OD_{640nm} was measured for 24 hours. Results are the means of data from three independent experiments. The standard deviation is below 5%.

When compared to the wild-type strain no significant differences were observed for the growth rate of the several mutants. Table 5 shows the estimated doubling time for each mutant and the wild-type strain.

Table 5 Doubling time of the strains under study. Cultures were grown in LB medium at 37ºC, 180 rpm of orbital agitation, and OD_{640nm} was measured for 24 hours. The doubling time was calculated from the growth rate of the exponential growth phase.

Bacterial strain	Doubling time (min ± SD*)	Bacterial strain	Doubling time (min ± SD*)
P0213-1	91.5 ± 5.3	F70	105.2 ± 4.3 (ns)
B63	98.5 ± 11.1 (ns)	G2	92.7 ± 8.3 (ns)
B68	93.9 ± 8.7 (ns)	H21	94.5 ± 13.9 (ns)
E28	90.3 ± 8.6 (ns)	H42	88.3 ± 2.4 (ns)
E36	104.3 ± 7.1 (ns)	B74	92.4 ± 7.2 (ns)
E50	94.2 ± 13.0 (ns)	E89	92.8 ± 2.7 (ns)
E70	102.8 ± 2.1 (ns)	G38	99.0 ± 6.4 (ns)

* Statistical difference in the doubling time of the mutant strains was tested in comparison with the wild-type (P0213-1). Significance level (one-way ANOVA followed by Dunnett's multiple comparisons test): ns, not statistically significant.

Based on table 5, E36, E70 and F70 mutants are the ones displaying a slower growth. Nevertheless, these differences were statistically not significant. Regarding the biomass formed after 24 hours, B63, E28, E36 and E50 seem to form less biomass.

Taking in account these data, it was concluded that, under the tested conditions, the disrupted genes interfere with the formation of planktonic cellular aggregates without interfering significantly with the growth kinetics.

3.6 Quantification of cellular aggregates and free cells produced by each selected mutant

Although it is possible to identify the differences between the size/number of aggregates of the wild-type and the mutants at naked eye, it has become necessary to quantify both aggregates and free cells, in order to determine, the percentage of planktonic cellular aggregates formation of each mutant. Following the procedure described in material and methods, it was estimated the dry-weight of free cells and aggregates for each mutant (Fig. 26).

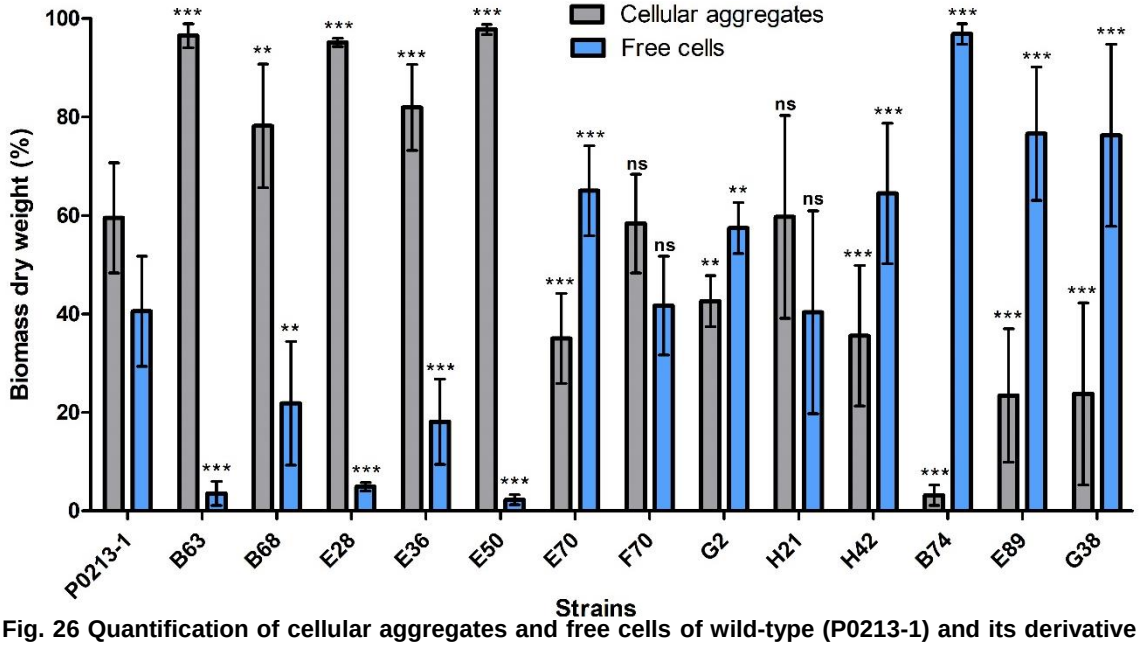

Fig. 26 Quantification of cellular aggregates and free cells of wild-type (P0213-1) and its derivative mutants. Error bars correspond to the standard deviations of the mean values of at least four independent experiments. Significance level (one-way ANOVA followed by Dunnett's multiple comparison test) between cellular aggregates and free cells of the wild-type (P0213-1) and cellular aggregates and free cells of the P0213-1 derivative mutants was determined: *, P<0.05; ***, P<0.001; ****, P<0.0001; ns, not statistically significant.

Results shown in figure 26 allow the subdivision of mutants into four groups as shown in table 6.

Table 6 The selected mutants were divided in four distinct groups. This division was based on the different percentage of planktonic cellular aggregates formed.

Percentage of planktonic cellular aggregates			
> 75%	**> 55 % and < 75%**	**> 25 % and < 55 %**	**<25 %**
B63	P0213-1	E70	B74
B68	F70	G2	E89
E28	H21	H42	G38
E36	-	-	-
E50	-	-	-

Based on data present in figure 26 and table 6, the percentage of planktonic cellular aggregates of the selected mutants its highly variable. Mutants producing more than 75% of biomass as aggregates include B63, B68, E28, E36 and E50 mutants (Fig. 26), confirming our preliminary visual inspection of figure 13. Searching the literature for similar phenotypes and their mutations allowed the identification of *Mycobacterium smegmatis* mutants with increased aggregation index (90). These two mutants had mutations in *rel*A gene responsible for (p)ppGpp synthesis and gene

*dsp*A involved in c-di-GMP synthesis, implicating these two messengers in cellular aggregation. Although we have no experimental evidence that B63 mutant has altered levels of such messengers that might affect the structure and hydrophobicity of the membranes as shown in the work of Gupta *et al.,* our B68, E28 and E50 mutants might have in common altered lipid metabolism which can affect synthesis of lipids either as signaling molecules or membrane composition. As discussed in the introduction, quorum sensing in *Rhodovulum sulfidophilum* was required for flocculation of this marine bacterium (30), being another example of the importance of signaling in cellular aggregation.

With aggregates ranging from 55-75% of biomass we found the wild-type strain P0213-1 and F70 and H21 mutants. Although these two mutants present the same amount of biomass distribution (between aggregates and free cells) as the wild-type strain, the average size of the aggregates and the microscopic structure was very different (Fig. 13).

The next group of mutants includes E70, G2 and H42, showing between 25 and 55% of the biomass in aggregates. The disrupted genes are involved in central metabolism (E70 and H42) or might de-repress an efflux pump. How these mutants affect aggregates formation still must be determined.

The last group of the mutants (B74, E89 and G38) present few macroscopic aggregates, while the majority of the biomass remaining as free cells. Regarding B74 mutant and the disruption of S9 ribosomal protein, no reports were found in the literature linking this mutation with the aggregation phenotype. However, is important to highlight a study implicating two other ribosomal proteins in biofilm formation by their interference in the regulation of bistability of *Bacillus subtilis* cells (107). It is then possible that *Burkholderia* S9 ribosomal protein is not only important in the assembly of the ribosome, but also for other mechanisms related to environmental adaption of bacteria, highlighting a possible character of a moonlight protein. A non-functional glutamate synthase in G38 mutant might cause increased levels of glutamine and the consequent inactivation of the NtrBC system. As reported before, NtrC is an enhancer-binding protein (EBP) that together with RpoN activates transcription of several genes (105). In line with that, it has been shown that *Aquincola tertiaricarbonis rpo*N gene was required for flocculation (5). Despite that, the genes involved in cells aggregation and controlled by RpoN remain to be determined. Although the role of the NtrBC system in *Burkholderia* aggregation in unknown, the *ntrC* and *rpoN* mutants of *B. cenocepacia* H111 showed downregulation of genes related to flagella and exopolysaccharide (105), two traits that might have some importance in the aggregation phenotype. Regarding E89 mutant, is difficult to predict the function of the disrupted cytochrome *c*-type encoding gene, but its location within genes putatively involved in the synthesis/degradation of some lipid intermediates, might suggest that this compound(s) is important for aggregation.

Together, mutant analysis seems to indicate that changes in lipid metabolism (and possibly signaling) have positive effect in aggregates formation, while the disruption of genes related with translation and regulation of nitrogen metabolism have a detrimental effect. Having more moderate effects on cellular aggregates we highlight an Omp porin and the type VI secretion system.

3.7 Genetic complementation of mutants

To confirm the importance of the mutated genes in cell aggregation, some of them were cloned in a replicative vector such as pBBR1MCS and complementation of the respective mutant was performed.

To complement the B68 mutant defective in a TetR-family regulator FadR putatively involved in terpene/fatty acid biosynthesis, gene PROKKA_04504 (Bmul_3465) together with the upstream region containing the putative promoter (Fig. 27(A)) were cloned into pBBR1MCS vector. The resulting plasmid PMF18-7, as well as the empty vector were introduced in the wild-type P0213-1 and B68 mutant. Cells were grown in SM medium at 37ºC with 180 rpm of orbital agitation for 48 hours and the percentage of biomass dry weight recovered from aggregates or free cells determined. As depicted in figure 27(B), the presence of the empty vector or the overexpression of *fad*R gene in the wild-type strain did not change the proportion of cells in the aggregates on free living cells. Complementation of B68 with *fad*R gene also did not alter significantly the proportion between free cells and aggregates (Fig. 27(B)) and we can conclude that complementation was not successful. Additionally, the introduction of the vector alone decreased aggregates biomass (P<0.001). Figure 27(C) shows microscopic images of aggregates obtained in all conditions, confirming their highly structured root-like forms, but no significant differences between them.

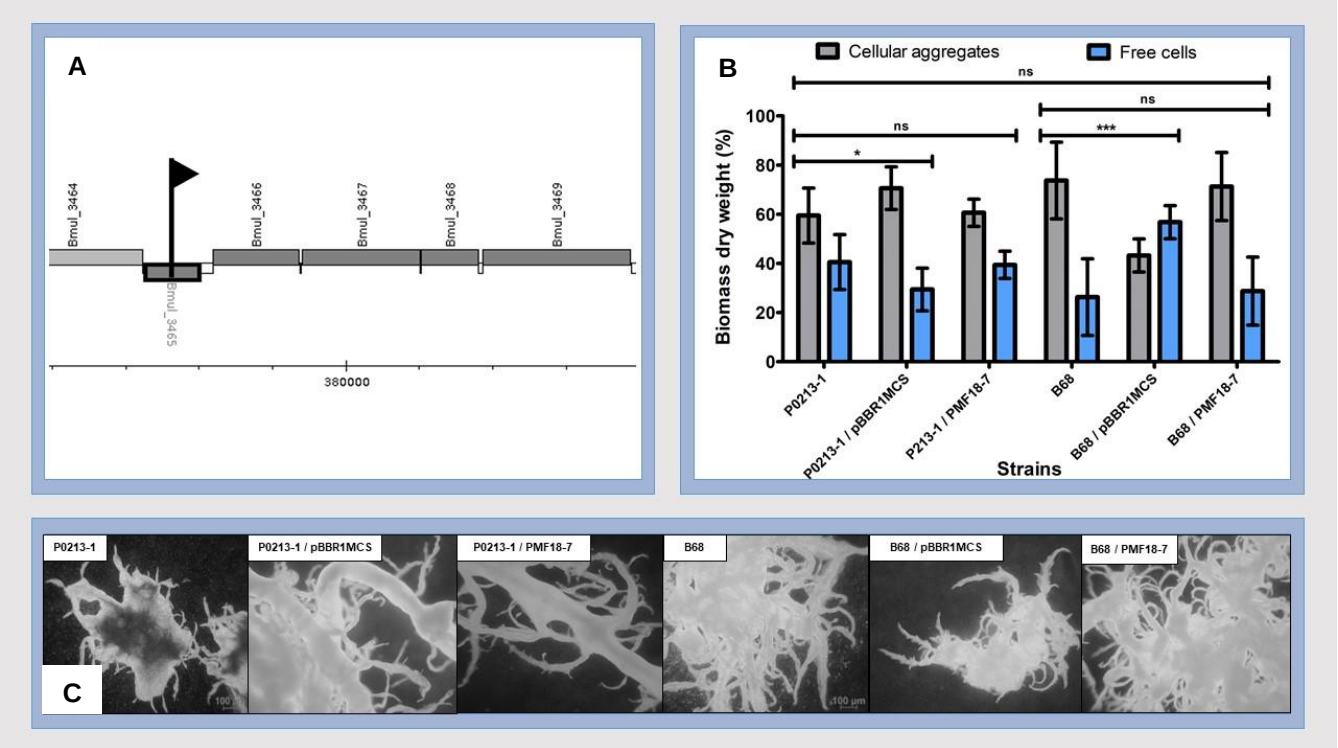

Fig. 27 Complementation of the B68 mutant. (A) Genomic location of the plasposon in *B. multivorans* B68 mutant showing a region from chromosome 2. In B68 mutant, the plasposon is inserted in *fadR* (Bmul_3465) gene (flagged). Downstream of *fadR* are four genes belonging to an operon and upstream is a gene encoding a hypothetical protein. **(B) Quantification of cellular aggregates and free cells of wild-type (P0213-1), mutant and complemented mutant.** Error bars correspond to the standard deviations of the mean values of at least four independent experiments. Significance level (one-way ANOVA followed by Turkey's multiple comparison test) between cellular aggregates and free cells of the wild-type (P0213-1) and

cellular aggregates and free cells of the P0213-1 derivative strains was determined: **, P<0.01; ***, P<0.001; ns, not statistically significant. **(C) Microscopic images of wild-type, mutant and complemented mutant.**

E36 mutant, similarly to B68, has the majority of the biomass in the form of aggregates and is mutated in another TetR-family transcriptional regulator encoded by PROKKA_05198 (Bmul_4221) (Fig. 28(A)). Both the upstream and gene coding region were cloned into pBBR1MCS giving rise to plasmid PMF18-6. Results of the complementation of E36 mutant shown in figure 28(B), showed that this approach was not successful since the percentage of aggregates of the mutant alone or with PMF18-6 did not change. A possible explanation is lack of gene expression if the upstream region does not contain a valid promoter. The microscopic images of the aggregates also do not show a significant difference.

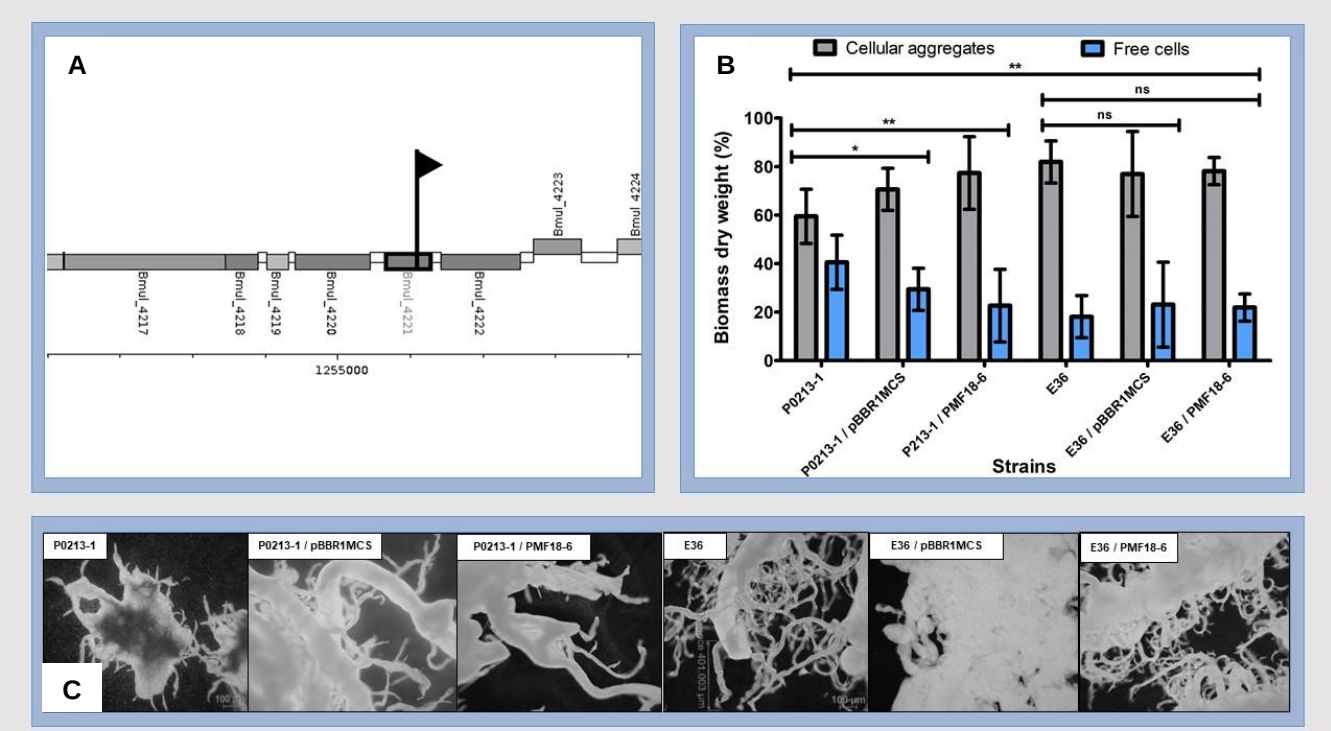

Fig. 28 Complementation of the E36 mutant. (A) Genomic location of the plasposon in *B. multivorans* E36 mutant showing a region from chromosome 2. In E36 mutant, the plasposon is inserted in *nemR_2* (Bmul_4221) gene (flagged). Downstream of *nemR_2* is a gene encoding a hypothetical protein and upstream is a gene encoding an alcohol dehydrogenase. **(B) Quantification of cellular aggregates and free cells of wild-type (P0213-1), mutant and complemented mutant.** Error bars correspond to the standard deviations of the mean values of at least four independent experiments. Significance level (one-way ANOVA followed by Turkey's multiple comparison test) between cellular aggregates and free cells of the wild-type (P0213-1) and cellular aggregates and free cells of the P0213-1 derivative strains was determined: *, P<0.5; ***, P<0.001; ns, not statistically significant. **(C) Microscopic images of wild-type, mutant and complemented mutant.**

G2 mutant, with the AmR repressor (PROKKA_02681, Bmul_1617) disrupted (Fig. 29(A)), produces more biomass as free cells than the wild-type strain. Cloning this gene with the respective upstream region into pBBR1MCS originated PMF18-8 plasmid. Introduction of PMF18-8 reverted the percentage of biomass in aggregates to values similar to the wild-type, but since the same effect was observed by the introduction of the vector alone, we cannot attribute this result to complementation with the gene (Fig. 29(B)). Microscopic images of figure 29(C) confirm the different structure of G2 mutant. Nevertheless, the G2/PMF18-8 aggregates seem to have

some of ramifications seen in the wild-type strain what is not visible in G2/pBBR1MCS. In conclusion, G2 mutant complemented with *amr*R gene did not restore the proportion between aggregates and free cells but might have an effect on the structure of the aggregates, which resemble more the wild-type ones.

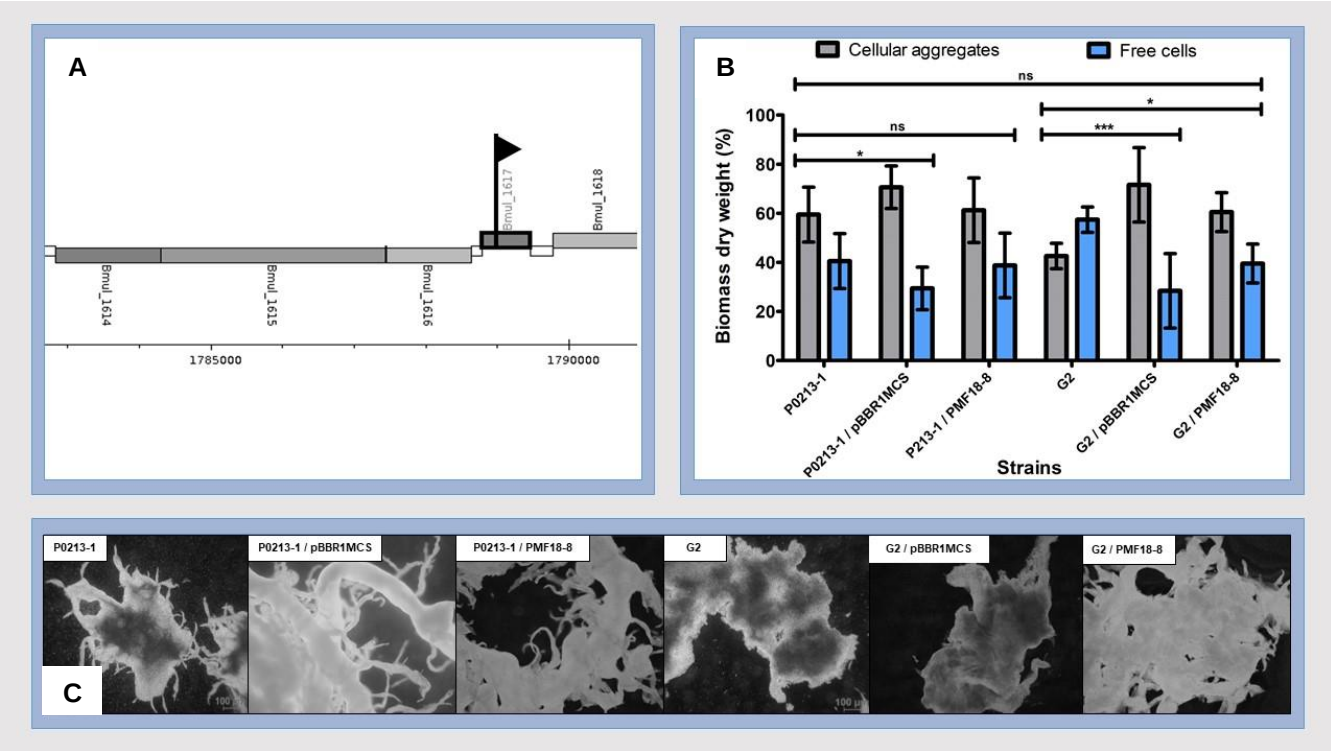

Fig. 29 Complementation of the G2 mutant. (A) Genomic location of the plasposon in *B. multivorans* G2 mutant showing a region from chromosome 1. In G2 mutant, the plasposon is inserted in *ttgR*_3 (Bmul_1617) gene (flagged). Downstream of *ttgR*_3 are three genes belonging to an operon and upstream is a gene encoding a peptidase M23B. **(B) Quantification of cellular aggregates and free cells of wild-type (P0213-1), mutant and complemented mutant.** Error bars correspond to the standard deviations of the mean values of at least four independent experiments. Significance level (one-way ANOVA followed by Turkey's multiple comparison test) between cellular aggregates and free cells of the wild-type (P0213-1) and cellular aggregates and free cells of the P0213-1 derivative strains was determined: *, P<0.5; ***, P<0.001; ns, not statistically significant. **(C) Microscopic images of wild-type, mutant and complemented mutant.**

The last complemented mutant was B74, mutated in *rps*I gene encoding ribosomal protein S9. Since this gene is in a putative operon together with *rpl*M gene, it was cloned the upstream region of *rpl*M gene and the *rps*I gene into pBBR1MCS. This constructed was named PMF18-4. Overexpression of this gene in the wild-type strain did not influence the distribution of biomass between aggregates and free cells (Fig. 30(B)). But, expression of the *rps*I gene in the mutant increased considerably the aggregates biomass (Fig. 30(B)) and size (Fig. 30(C)). Although, complementation did not reach wild-type levels, this experiment was successful.

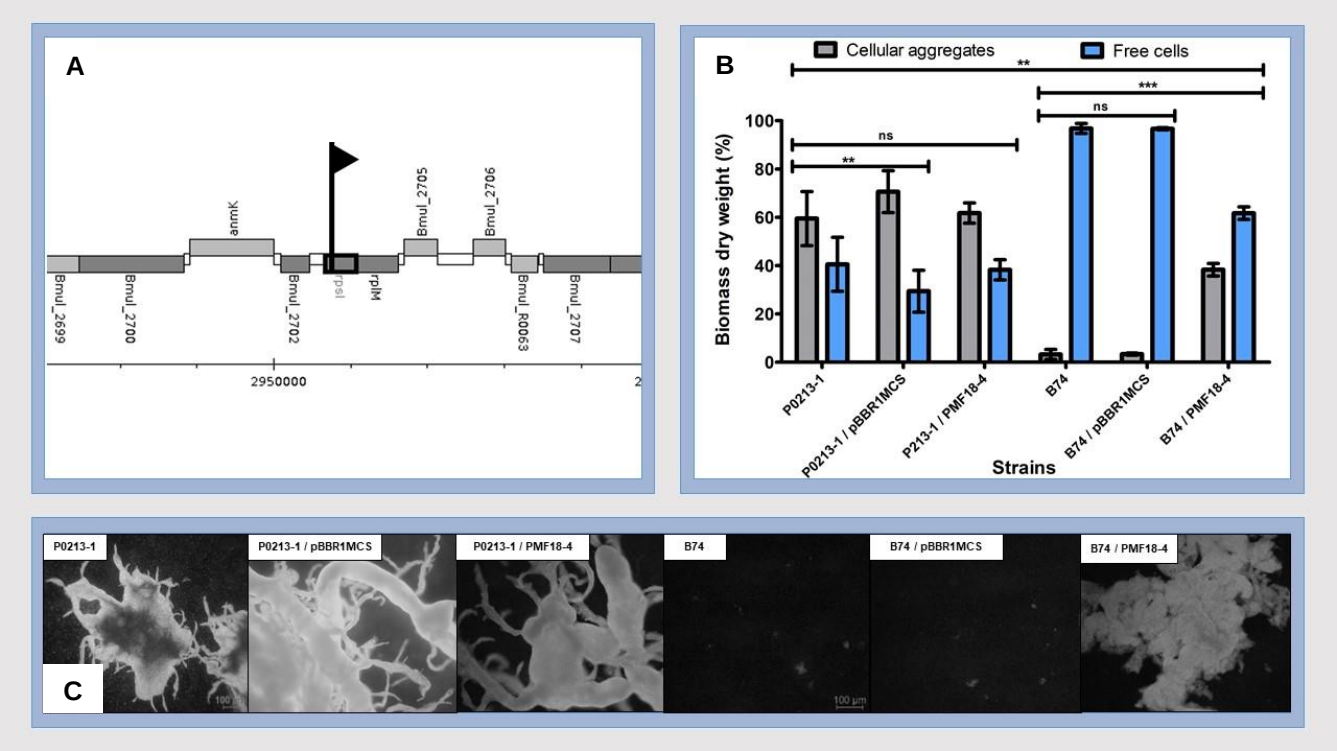

Fig. 30 Complementation of the B74 mutant. (A) Genomic location of the plasposon in *B. multivorans* B74 mutant showing a region from chromosome 1. Together in the same operon is *rplM* gene encoding 50S ribosomal protein L13. Downstream of *rplM* is a gene encoding a hypothetical protein. Upstream is a gene encoding an iron-sulfur cluster insertion protein ErpA. **(B) Quantification of cellular aggregates and free cells of wild-type (P0213-1), mutant and complemented mutant.** Error bars correspond to the standard deviations of the mean values of at least four independent experiments. Significance level (one-way ANOVA followed by Turkey's multiple comparison test) between cellular aggregates and free cells of the wild-type (P0213-1) and cellular aggregates and free cells of the P0213-1 derivative strains was determined: **, P<0.01; ***, P<0.001; ns, not statistically significant. **(C) Microscopic images of wild-type, mutant and complemented mutant.**

Complementation experiments of the other mutants of interest were started, but due to lack of time was not possible to finish them.

3.8 Surface-attached biofilm formation

In addition to the aggregates phenotypes already described, is our aim to identify if other phenotypes are altered as well in the selected mutants. Such traits include surface-attached biofilm formation, motility, antibiotic resistance, growth and virulence.

Surface-attached biofilm formation and planktonic cellular aggregates formation most likely share important steps of their development as described in the introductory chapter of this work. Thus, in here, we evaluated the ability of each mutant under study to produce surface-attached biofilms. For this, mutants were grown in microtiter plates for 48 hours and then, after staining with crystal violet, the absorbance at 590 nm (A_{590nm}) of the retained dye was measured (Fig. 31).

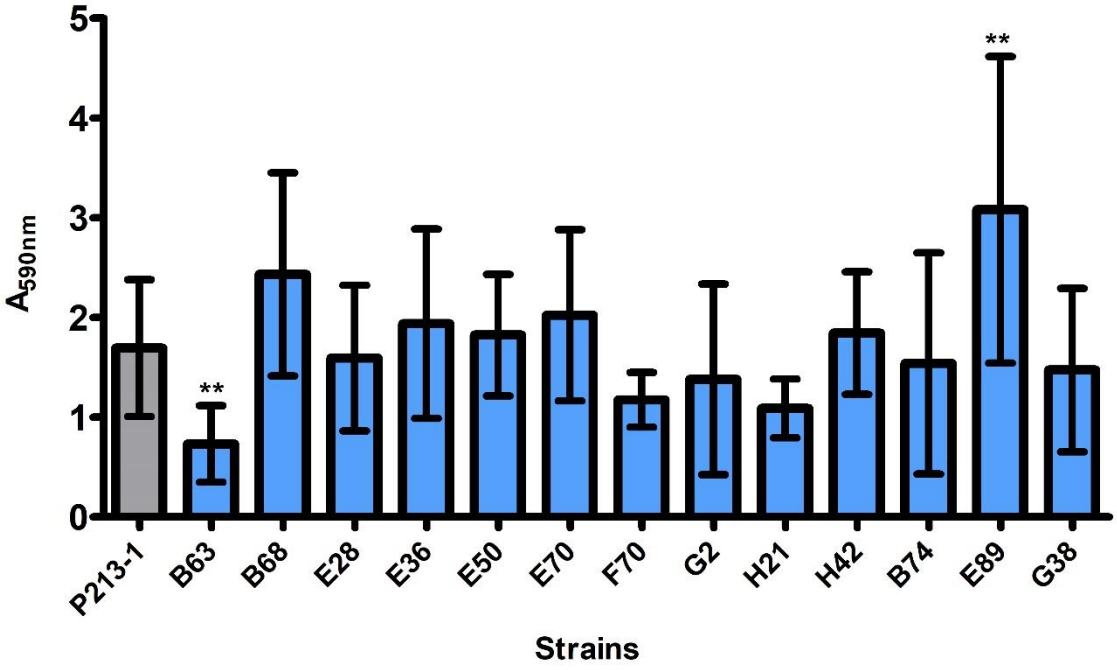

Fig. 31 Surface-attached biofilm formation of the P0213-1 and its derivative mutants was determined by absorbance measurement at 590 nm after growth for 48h at 37ºC in polystyrene microplates. Error bars correspond to the standard deviations of the mean values of at least seven independent experiments. Significance level (one-way ANOVA followed by Dunnett's multiple comparison test) between the wild-type (P0213-1) and the mutants was determined: **, P<0.01.

Of all tested mutants only two of them presented statistically significant differences in the surface-attached biofilm formation in comparison with the wild-type. Based on results shown in figure 31, the B63 mutant presented the lower absorbance and the E89 mutant has the highest surface-attached biofilm formation.

B63 has the xanthine dehydrogenase molybdenum-binding subunit encoding gene disrupted. As mentioned before this protein is necessary for purine salvage pathway to maintain the GTP levels and, consequently might affect signaling dependent on alarmones and c-di-GMP. With the aim of discovering the role of (p)ppGpp in *Pseudomonas aeruginosa*, Xu *et al.* (108) created a mutant (Δ*RS* mutant) unable to produce this intracellular molecule and tested its capacity of biofilm formation. Results have shown that biofilm formation by the Δ*RS* strain was significantly diminished compared to wild-type and complemented mutant.

In another study, a *Burkholderia cenocepacia* H111 mutant, that had a reduced intracellular c-di-GMP level, formed less biofilm when compared to a strain with a high c-di-GMP concentration (89). Taking this into account, perhaps the B63 mutant has defective signaling pathways, resulting in lower biofilm formation.

Regarding E89 mutant increased biofilm formation ability, it is difficult to grasp the role of the disrupted gene and therefore everything we might say is speculative.

Although steps in the formation of aggregates and surface-attached biofilms are described as similar, here we show that the same mutation can have distinct impact on both phenotypes. E89 mutant showed high formation of surface-attached biofilm and is among the group who presents less percentage of cellular aggregation. In case of the B63 mutant, the percentage of aggregates formation was high, while the absorbance measured to determine the capability of the surface-attached biofilm formation was the lowest one. These opposite results regarding these two phenotypes were previously reported in the literature further supporting the results obtain in this study (6, 90).

It is important to say, however, that the growth conditions of the different mutants in both experiments were different. To determine the aggregation capacity, the mutants were grown for 48 hours in SM medium at 37ºC with orbital agitation, while for the determination of surface-attached biofilm formation it was used LB medium, where mutants were grown for 48 hours at 37ºC without orbital agitation. These facts must be considered when discussing the differences between these two phenotypes. Thus, in this study, it should have been used the same medium for a better understanding of the mechanism for the development of these two types of structures formed by bacterial cells.

3.9 Swimming and swarming motilities

Motility might be one of the requirements for a successful cellular aggregation, since cellular aggregation depends on contact between bacterial cells through physical forces and biological mechanisms. Thus, motility allows prior approximation of bacterial cells for them to start aggregation (14).

To evaluate the motility of the *B. multivorans* strains, swimming and swarming agar plates were incubated for 24 hours and 48 hours, respectively, at 37ºC after inoculation. Then, motility zone diameter (cm) was measured for each of them and the results are present in figure 32 and 33.

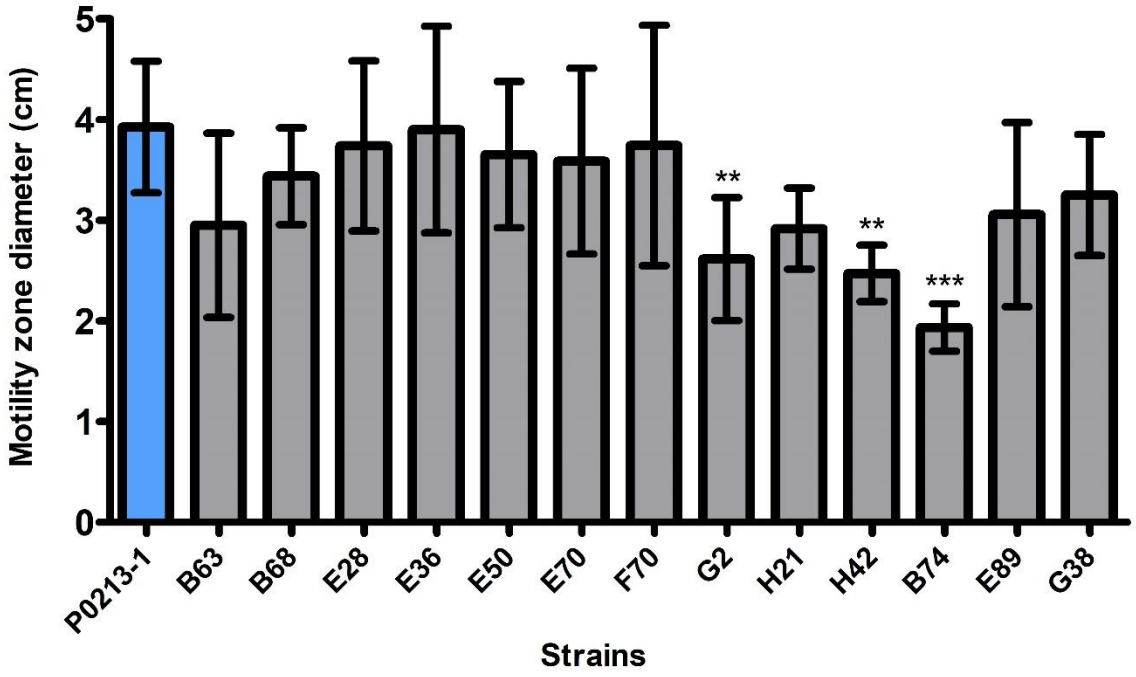

Fig. 32 Swimming motility of the P0213-1 and its derivative plasposon mutants was measured as the motility zone diameter after growth for 24h at 37ºC. Error bars correspond to the standard deviations of the mean values of at least two independent experiments. Significance level (one-way ANOVA followed by Dunnett's multiple comparison test) between the wild-type (P0213-1) and the mutants was determined: **, P<0.01; ***, P<0.001.

Based on figure 32, only 3 mutants (B74, G2 and H42) presented statistically significant decrease of swimming motility in comparison with the wild-type (P0213-1). Nevertheless, this observed decrease in motility is higher for some of the mutants producing less/smaller aggregates.

Swarming motility measurement has shown statistically significant decrease for B74, E28, E70 and G38 mutants (Fig. 33).

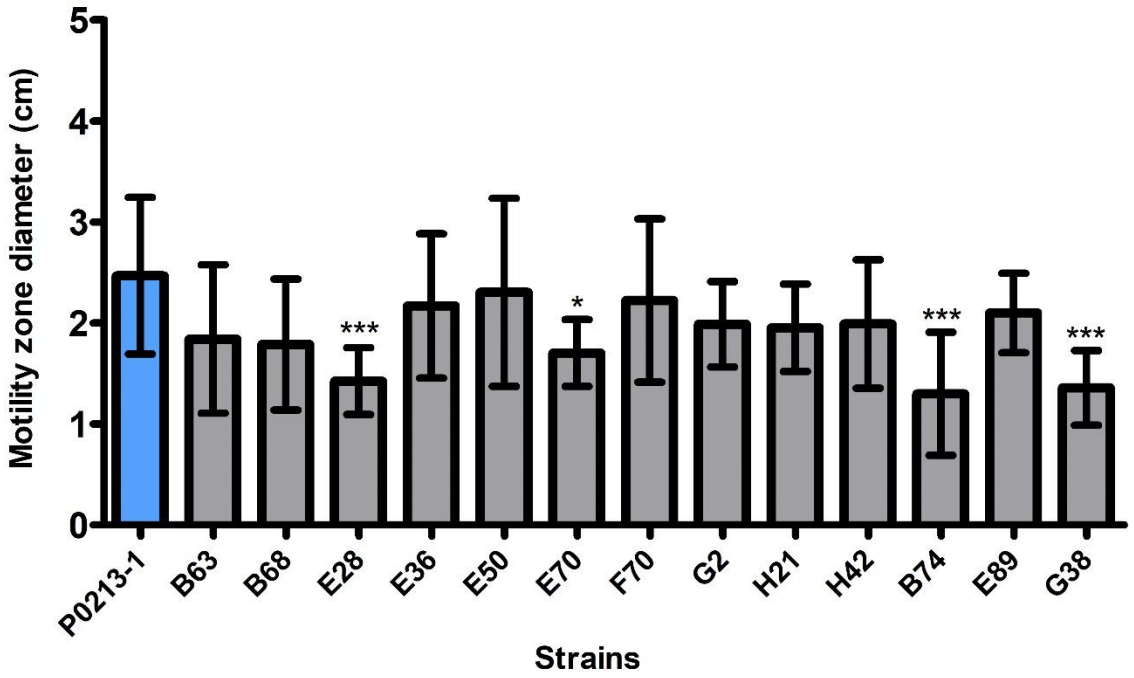

Fig. 33 Swarming motility of the P0213-1 and its derivative strains to antibiotics was measured as the motility zone diameter after growth for 24h at 37ºC. Error bars correspond to the standard deviations of the mean values of at least two independent experiments. Significance level (one-way ANOVA followed by Dunnett's multiple comparison test) between the wild-type (P0213-1) and the mutants was determined: *, P<0.05; ***, P<0.001.

Literature search correlating similar mutations as the ones here described and motility did not give useful results. Effectively, only in *Bacillus subtilis* was demonstrated that other two ribosomal proteins (S21 and S11) are important for the regulation of cell motility and biofilm formation (107). These ribosomal protein mutants had shown a decrease in motility, as observed in this study for the B74 mutant. These authors also determined that these two proteins interfere with two major operons that regulates the biofilm formation (107).

The other mutation from which some information is available is the one present in G38 mutant. As mentioned before, the G38 mutant presents a non-functional glutamate synthase, which might lead to an increase in the concentration of glutamine and consequently to the inactivation of the NtrBC system. Beyond the *bce*-I and *bce*-II clusters, which produce the EPS cepacian in *Burkholderia,* it was shown that the NtrBC system also regulates swarming motility. Indeed, the swarming ability of the *ntrC* mutant strain was lower than the ability of wild-type and complemented mutant, thus, suggesting that the swarm phenotype of the G38 mutant might be due to negative effects on the NtrBC system (105).

Motility is not always necessary for cells prior approximation, since cells can form the initial "seeding" aggregates by random collisions of motile cells or by Brownian motion. However, searching for ideal conditions by bacteria is more advantageous. So, despite the variation in aggregates formation here observed in some mutations none of them lost motility completely.

Bacteria swim in liquid and swarm over solid surface or in viscous environment. Thus, considering the experimental conditions used, the results of swimming motility are more important

to relate motility with the aggregation phenotype. By the results, it is possible to observed that the aggregation was not more than 50% for the strains that shown a statistically significant decrease in swimming motility when compared to the wild-type. However, this is not conserved through all strains tested since some of them present higher aggregation phenotype and less swimming motility than the wild-type.

In the other hand, the swarming motility could be more relevant to ensure surface-attached biofilm formation. But, as demonstrated in swimming motility, the results of swarming motility cannot be correlated to the formation of surface-attached biofilm formation. Together, these results suggest that motility was neither the more important feature for the aggregation phenotype or for surface-attached biofilm formation. It is also important to refer, that there are different mechanisms affecting motility in bacteria and this process is highly regulated.

3.10 Exopolysaccharide production

Cellular aggregation is mediated by bacterial products, more specifically, extracellular polymeric substances such as polysaccharides, extracellular DNA, and ions. In some bacterial species, exopolysaccharides promote the aggregation phenotype (10). Therefore, we also assessed exopolysaccharide production in our mutants.

To determine the production of exopolysaccharide, each mutant was inoculated onto YEM agar plates at 37ºC for 48 hours. Then, through visual inspection, it was possible to evaluate their mucoid phenotype (Fig. 34).

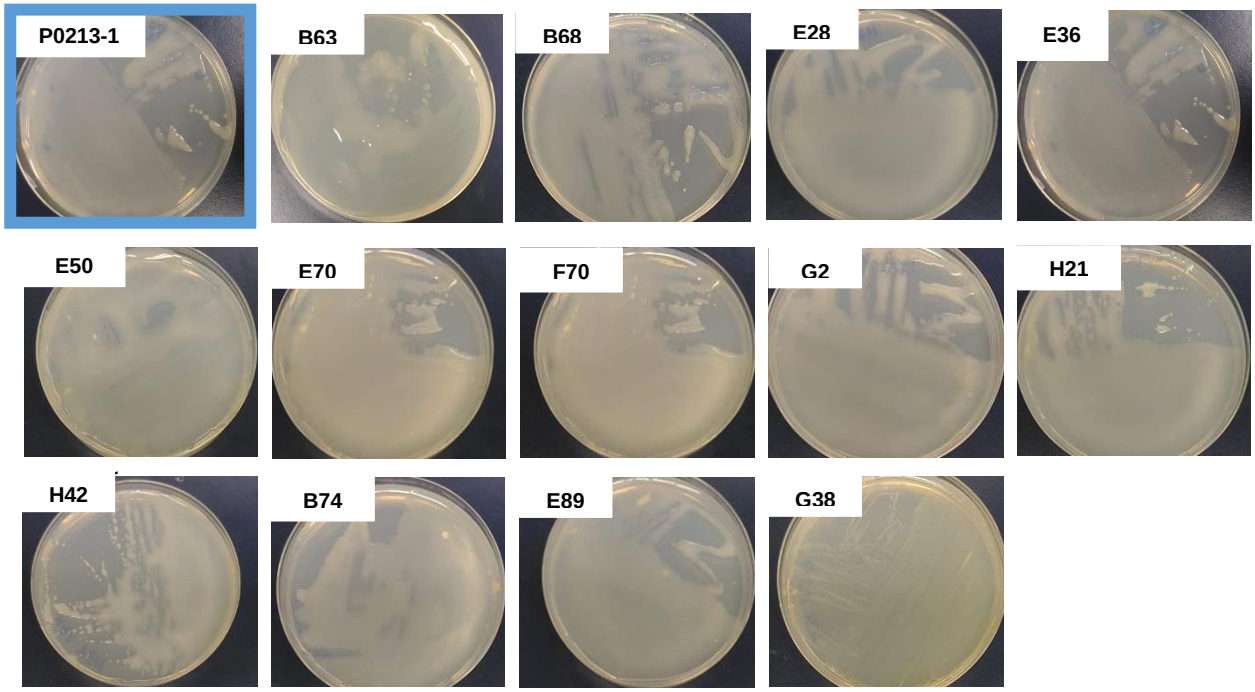

Fig. 34 Evaluation of mucoid phenotype in yeast extract mannitol medium (YEM) after 48 h of incubation at 37°C for wild-type (P0213-1) and its derivate mutants.

Results showed that, except for G38 mutant, all others are highly mucoid in YEM plates. As already explained before, the expression of *bce* genes directing the biosynthesis of cepacian are under control of NtrC and σ^{54} in *B. cenocepacia* (102, 105). Since our G38 mutant might have defects in NtrBC signaling, this is possibly the cause of the observed phenotype.

Although cellular aggregation has been shown to be mediated also by exopolysaccharides, there are some antagonistic reports (10, 66). This is also seen in this work, were two mutants (B74 and G38) produce fewer aggregates, but one is mucoid and the other nonmucoid. Another possibility is that cepacian might not have a fundamental role in this process as it has been reported by Silva *et al.* (66), but other polysaccharide produced by these strains might have.

3.11 Antimicrobial susceptibility

It has been reported that cells within planktonic cellular aggregates are more resistant against antimicrobials than free cells (6). Thus, it was important to determine if antimicrobial susceptibility is related to cellular aggregation.

Antimicrobial susceptibility of all mutants and wild-type was tested against ciprofloxacin, aztreonam, piperacillin plus tazobactam and tetracycline. Ciprofloxacin is a fluoroquinolone that works by inhibiting DNA replication; aztreonam is a β-lactam that inhibits cell wall biosynthesis; piperacillin is a β-lactam antibiotic that belongs to the class of penicillin and is widely used in combination with the β-lactamase inhibitor tazobactam, acting by inhibiting bacterial cell wall biosynthesis; and tetracycline is an aminoglycoside that inhibits protein synthesis.

The antimicrobial susceptibility was determined by measuring the diameter of the growth inhibition zone after 24 hours of incubation at 37°C (Fig.35).

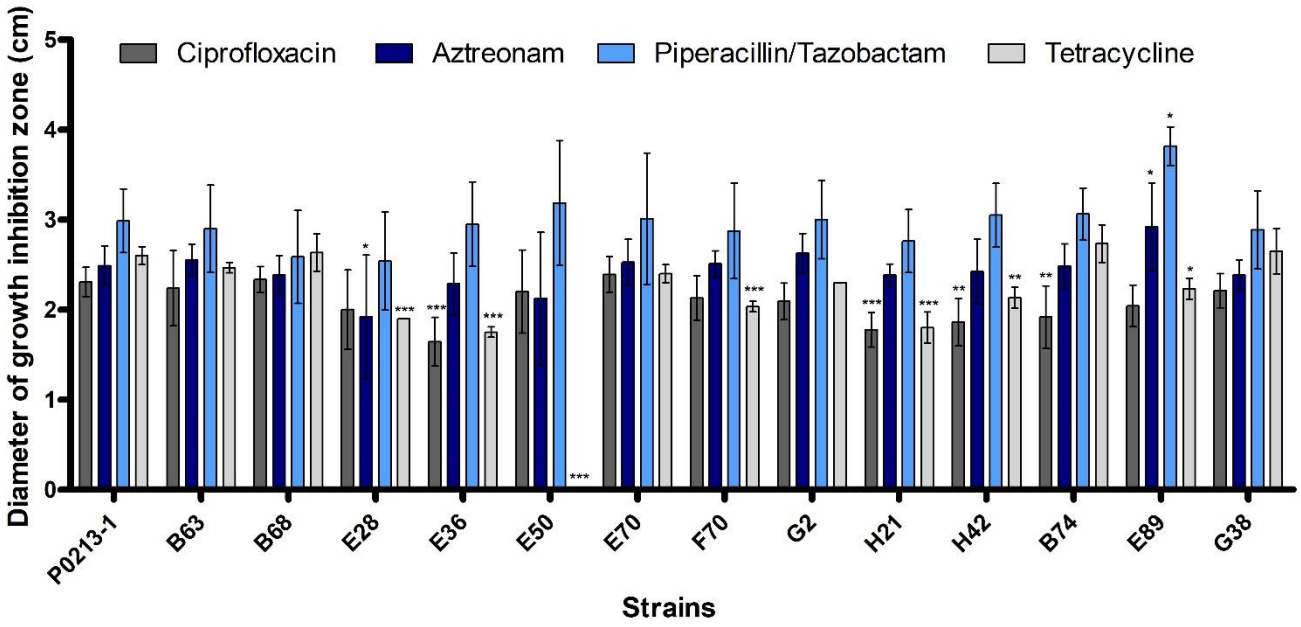

Fig. 35 Susceptibility of the P0213-1 and its derivative strains to antibiotics (ciprofloxacin, aztreonam piperacillin/tazobactam and tetracycline) was measured as the diameter of cell growth inhibition, after growth for 24h at 37ºC. Error bars correspond to the standard deviations of the mean values of at least two independent experiments. Significance level (one-way ANOVA followed by Dunnett's multiple comparison test) between the wild-type (P0213-1) and the mutants for each antibiotic tested was determined: *, P<0.5; **, P<0.01; ***, P<0.001

Regarding resistance against aztreonam and piperacillin/tazobactam, there were no differences between the wild-type and mutant strains, except for E89 which shows higher susceptibility. This mutant is one of three which produce lower amount of aggregates. For the other antibiotics, ciprofloxacin and tetracycline, we observed increased resistance for some of the mutants, while for others there was no difference. Since G2 mutant inactivates a repressor that controls the expression of an efflux pump, previously shown to be important for aminoglycoside resistance in *B. vietnamiensis*, we would expect this mutant to have increased resistance against tetracycline. Nevertheless, we obtained no difference between this mutant and the wild-type strain. Altogether, the ability to form aggregates did not alter significantly the antibiotic resistance profile of the strains under study, at least for the tested experimental conditions.

3.12 *In vitro* growth analysis of selected mutants in synthetic cystic fibrosis medium (SCFM)

The synthetic cystic fibrosis medium nutritionally mimics cystic fibrosis sputum (80). Since the biofilm formation model described by Worlitsch *et al*. (53) is very similar to the model of planktonic cellular aggregates, it was important to determine if the aggregation phenotype is related to the conditions of growth of the selected mutants. For this, the growth curves and corresponding growth rates in SCFM were determined for each selected mutant and compared with the wild-type strain. For this, all mutants were grown in SCFM medium for 24 hours with 180 rpm of orbital agitation at 37ºC (Fig. 36).

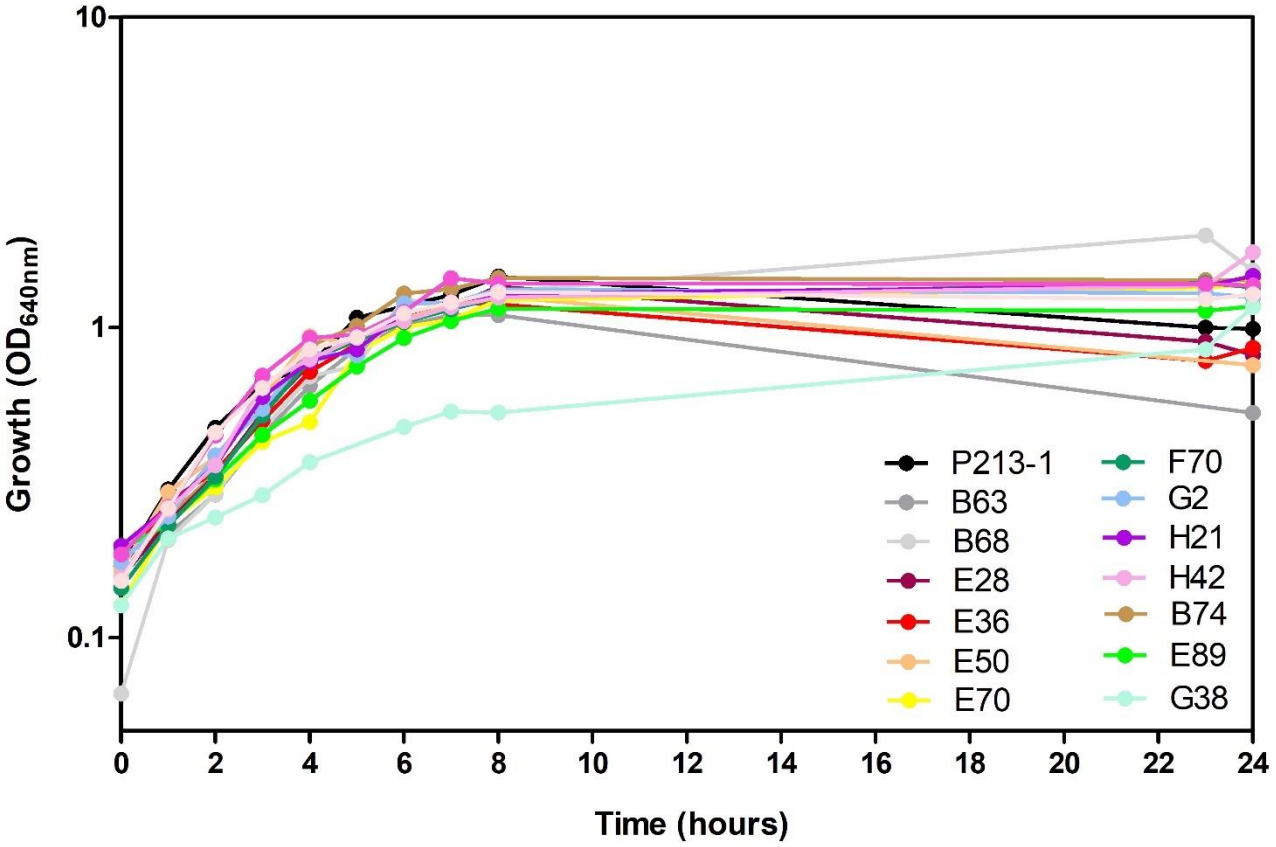

Fig. 36 Growth curves of wild-type P0213-1 and of the selected mutants. Cultures were grown in synthetic cystic fibrosis medium (SCFM) at 37ºC, 180 rpm of orbital agitation, and OD_{640nm} was measured for 24 hours. Results are the means of data from two independent experiments.

Growth curves shown in figure 36 confirm a very similar growth behavior, both in exponential and stationary phase of all strains, except for G38 mutant. Duplication time was also calculated for each mutant and wild-type strain as shown in table 7.

Table 7 Doubling time of the strains under study. Cultures were grown in synthetic cystic fibrosis medium (SCFM) at 37ºC, 180 rpm of orbital agitation, and OD_{640nm} was measured for 24 hours. The doubling time was calculated from the growth rate of the exponential growth phase.

Bacterial strain	Doubling time (min ± SD*)
P0213-1	138.1 ± 14.00
B63	123.5 ± 5.2 (ns)
B68	133.3 ± 0.2 (ns)
E28	129.6 ± 6.9 (ns)
E36	116.4 ± 15.4 (ns)
E50	144.6 ± 13.2 (ns)
E70	130.6 ± 0.2 (ns)

Bacterial strain	Doubling time (min ± SD*)
F70	102.7 ± 5.1**
G2	129.5 ± 7.5 (ns)
H21	133.9 ± 3.7 (ns)
H42	113.7 ± 2.4 (ns)
B74	120.5 ± 6.0 (ns)
E89	158.4 ± 14.7 (ns)
G38	244.1 ± 8.6***

* Statistical differences in the doubling time of the mutant strains was tested in comparison with the wild-type (P0213-1). Significance level (one-way ANOVA followed by Dunnett's multiple comparisons test): ns, not statistically significant; **, $p < 0.01$; ***, $p < 0.001$

According to this data, only two mutants present statistically significant differences in doubling time in comparison to the wild-type. The G38 mutant which presents an increase and F70 mutant which presents a decrease of the doubling time.

The synthetic cystic fibrosis medium nutritionally mimics cystic fibrosis sputum, that provides carbon, nitrogen and energy sources to support bacteria high density growth (80). Palmer *et al.* (80) analyzed the carbon utilization profiles in SCFM to determine the primary carbon sources consumed by *Pseudomonas aeruginosa* during growth in this medium. The authors determine that the first six carbon sources consumed were proline, alanine, arginine, lactate, glutamate and aspartate. After 3 hours of growth, the percentage of arginine in the medium was zero and after 3.5 hours the percentage of glutamate was approximately 5%. Extrapolating the results to *B. multivorans,* in the case of the G38 mutant, the glutamate synthase is non-functional and because of that this mutant is unable of produce glutamate through GS/GOGAT and through arginine, since the latter was totally consumed after 3 hours of growth. Taking into account the growth curve of the G38 mutant present in figure 36, it is possible to verify that after 2/3 hours growth begins to show a less pronounced slope compared to the other growth curves, reveling the higher doubling time, 244.1 ± 8.6 min, almost the double comparing with the wild-type P0213-1. Considering the disrupted gene of the G38 mutant, the higher doubling time presented by this mutant could be explained by the fact that bacterial cells cannot reset glutamate levels. Without this reset, bacterial cells have more difficulty to produce some components essentials for growth, like amino acids, co-factors, among others, which originate from glutamate (109).

The other mutant that presents a different doubling time, the F70 mutant, has a disrupted gene involved in the formation of an outer membrane protein. As is possible to see in table 7, the lack of a functional outer membrane protein, in SCFM, had a positive effect in growth of this mutant (doubling time of 102.7 ± 5.1 min).

Outer membranes porins were reported to be important in the uptake of nutrients, resistance to antibiotics, among others (110). However, the non-functional outer membrane porin of F70 mutant has not been characterized. Still, the STRING tool (111) identify some proteins related to the cell division that possibly interact with this protein.

3.13 Virulence determination in *Galleria mellonella*

To assess the virulence of the selected mutants and wild-type strain, *Galleria mellonella* was the chosen model of infection. This organism is widely used for studying bacterial human pathogens since it has an innate immune system similar to the innate immune systems of mammals (84). Ten larvae were injected per strain in a total of three experiments and survival was followed for three days post-infection (Fig. 37). Approximately $1x10^4$ bacterial cells were injected per larvae.

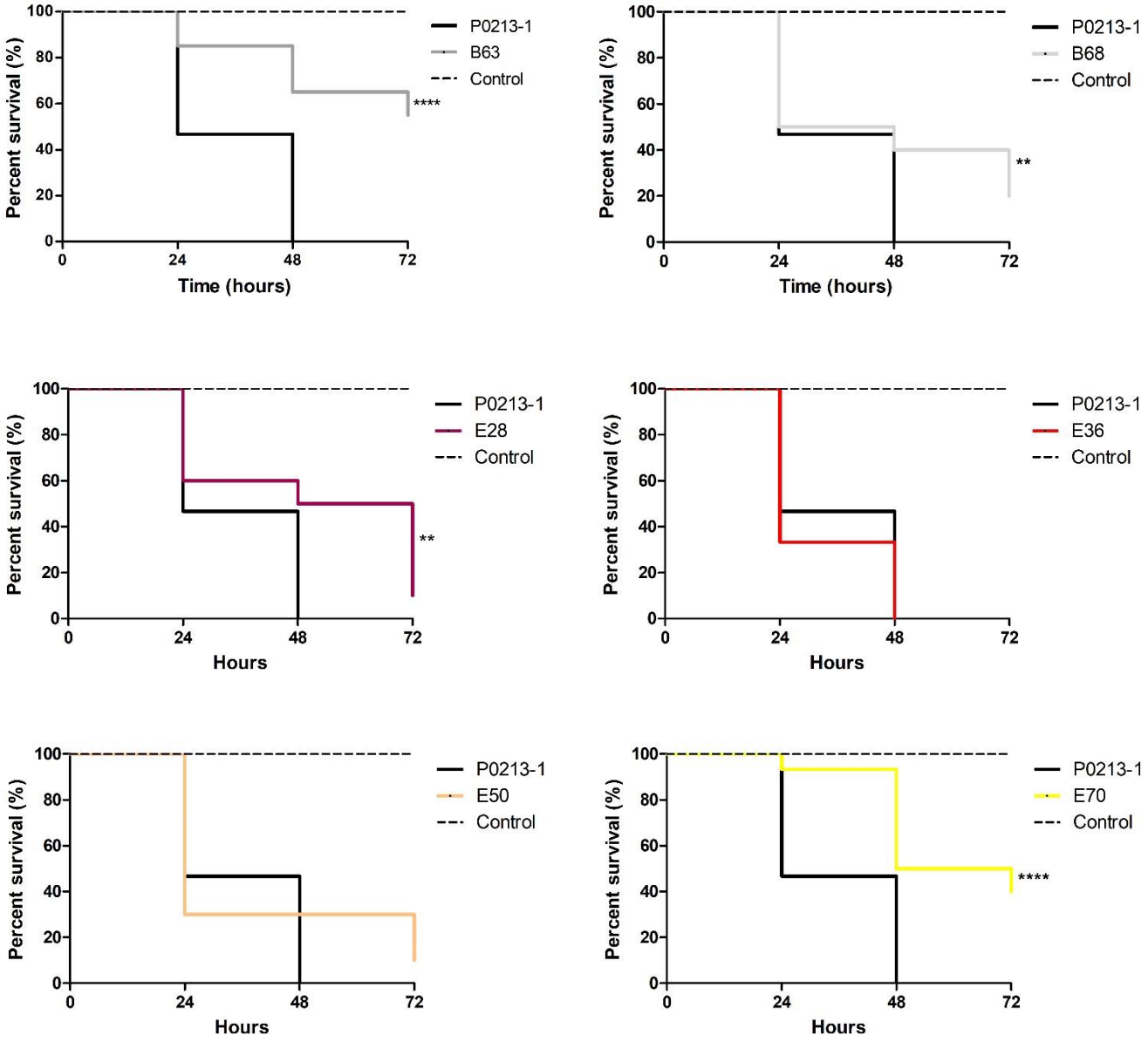

Fig. 37 Survival of *Galleria mellonella* larvae inoculated with *B. multivorans* (wild-type P0213-1 and the selected mutants). Triplicate groups of 10 larvae were inoculated with each isolate and survival was followed for three days post-infection. Larvae were injected with approximately $1x10^4$ bacterial cells. The control experiment without bacteria is also represented. Statistical significance of differences between the Kaplan-Meier curve of the first isolate and the subsequent ones was determined: **, P-value < 0.01; ****, P-value < 0.0001.

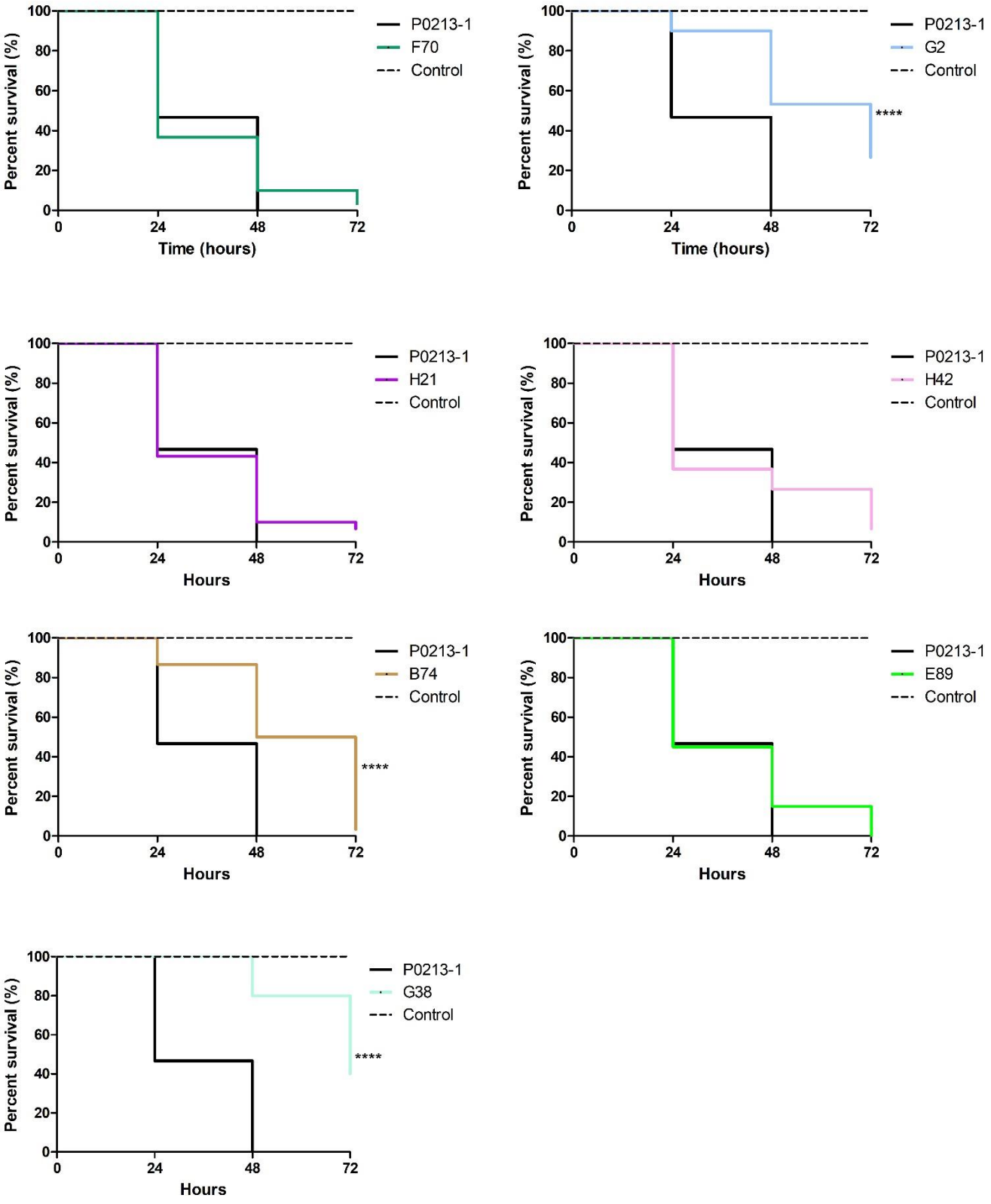

Fig.37 Survival of *Galleria mellonella* larvae inoculated with *B. multivorans* (wild-type P0213-1 and the selected mutants) (continuation). Triplicate groups of 10 larvae were inoculated with each isolate and survival was followed for three days post-infection. Larvae were injected with approximately $1x10^4$ bacterial cells. The control experiment without bacteria is also represented. Statistical significance of differences between the Kaplan-Meier curve of the first isolate and the subsequent ones was determined: **, P-value < 0.01; ****, P-value < 0.0001.

Considering the first group of mutants producing most of the biomass in the form of aggregates (B63, B68, E28, E36 and E50), all except E36 showed virulence attenuation when compared to the wild-type strain P0213-1. The one displaying higher attenuation is B63 mutant with xanthine dehydrogenase encoding gene disrupted. As already suggested before, this mutant might have defects in signaling pathways which would affect several cell traits, including virulence. For example, *P. aeruginosa* deficient in (p)ppGpp and c-di-GMP showed virulence attenuation in a mouse model of infection (108).

The second group of mutants producing smaller aggregates, but in higher number than the wild-type strain (G2, H21, H42, E70 and F70), only G2, H42 and E70 mutants show significant virulence attenuation (Fig. 37).

In the third group which includes B74, E89 and G38, we have seen virulence attenuation in B74 and G38, but not E89 (Fig. 37). These mutants are the ones where most of the cells are free living. The exact mechanisms of reduced virulence that are dependent an S9 ribosomal protein or on a protein involved in nitrogen metabolism are unknown.

From our analysis we did not see any correlation between the type/number of aggregates formation and virulence in the chosen infection model.

4 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Bacteria can proliferate as planktonic cells or as sessile communities. One example of these communities are biofilms, in which cells can be attached to a surface or can form non-attached planktonic cellular aggregates. The formation of planktonic cellular aggregates is of relevance in natural environments, but also during interaction with host cells resulting in pathogenicity. In patients with cystic fibrosis, microorganisms from the *Burkholderia cepacian* complex and *Pseudomonas aeruginosa* can form such cellular aggregates and often establish a chronic infection. For these reasons, it became relevant to understand the molecular mechanisms behind the formation of cellular aggregates in bacteria.

The aim of this work was to explore the unknown mechanisms, responsible for the formation of this type of aggregates in bacteria. For that, through plasposon mutagenesis, mutants of a clinical isolate, *B. multivorans* P0213-1, with altered aggregation phenotype were obtained. Then, 15 mutants were characterized regarding several phenotypes like motility, exopolysaccharide production, antimicrobial susceptibility, virulence in *G. mellonella* and growth kinetics. The main objective was trying to interconnect some of these phenotypes with the ability to form planktonic cellular aggregates and, consequently, to provide new possible mechanisms involved in the formation of cellular aggregates.

Although it was impossible to establish a linear relationship between each phenotype and aggregation ability, it is acceptable to assume that it was still possible to identify admissible players involved in the formation of aggregates by their role in the total set of tested phenotypes. Among that players are intracellular messengers (possibly (p)ppGpp and c-di-GMP and quorum sensing molecules), ribosomal proteins, outer membrane constituents and proteins involved in nitrogen and lipid metabolism.

Considering the results obtained, as well as the information from previous studies, it is possible to infer that the two intracellular messengers ((p)ppGpp and c-di-GMP) may be the key to one of the mechanisms that acts differentially in relation to the formation of the two types of biofilms considered in this study. Thus, it will be important to better understand the role of each intracellular messenger in aggregates formation, quantifying each of them over time, and infer which pathways might be influenced, for example, through transcriptomic analysis.

Based on the phenotypic results obtained with respect to the B74 mutant, it has been found that ribosomal proteins may be involved in other important processes in cells, in addition to their role in the ribosome assembly. Considering the results, it could be suggested a link between ribosomal proteins, cellular aggregation and motility. Point out further that this link was demonstrated before in *Bacillus subtillis*, but no kind of this association type was previously performed in gram-negative bacteria.

Also, quorum sensing system might be important (B68 mutant and E70 mutant) for the triggering of aggregation. This association had previously been established in relation to the formation of biofilms. Thus, this result highlights the importance of cell communication in phenotypes such as cellular aggregation in which there is a large cellular community.

During this study many were the mutants that had disrupted genes related to proteins that are part of the cell membrane. One of the characteristics described as influencing the aggregate formation capacity is the cell surface properties, namely their charge. Under normal physiological conditions cell surfaces have a net negative charge, however, it is known that this charge is highly variable and changes depending on the environment conditions and on the components that make up their surface. Bacteria may contain different surface structures such as fimbriae, lipopolysaccharides, hydrophobic amino acids or extracellular polymeric substances which in turn may contain glycoproteins, polysaccharides or even attached acids. These surface structures in solution will be charged and therefore the presence and nature of these surface appendages will alter the bacterial cell charge. Thus, in these mutants it should have been interesting to determine in what ways the gene disruption may lead to change in the aggregation phenotype through its influence on cell surface properties. To that end, the zeta potential of the bacterium that gives the information about the cell surface charge could have been relevant for the discussion of the results obtained in these mutants.

Despite the several players identified in this study as being involved in the formation of planktonic cellular aggregates, there is much more work to do in order to understand these complex structures. Moreover, is important to highlight that the knowledge about cellular aggregates could be relevant to the development of potential therapeutics to improved treatments against *Bcc* bacteria.

5 REFERENCES

1. Vasudevan R. 2014. Biofilms: Microbial Cities of Scientific Significance. J Microbiol Exp 1:1–16.

2. Costerton JW, Lewandowski Z, Debeer D, Caldwell D, Korber D. 1994. Biofilms , the customized microniche. J Bacteriol 176:2137–2142.

3. Schleheck D, Barraud N, Klebensberger J, Webb JS, McDougald D, Rice SA, Kjelleberg S. 2009. *Pseudomonas aeruginosa* PAO1 preferentially grows as aggregates in liquid batch cultures and disperses upon starvation. PLoS One 4:e5513.

4. Percival SL, Malic S, Cruz H, Williams DW. 2011. Introduction to biofilms, p. 41–68. *In* Biofilms and Veterinary Medicine. Springer, Berlin, Heidelberg.

5. Yu D, Xia M, Zhang L, Song Y. 2017. RpoN is required for rloc formation but not for extracellular polysaccharide biosynthesis in a floc-forming *Aquincola tertiaricarbonis* strain. Appl Environm Microbiol 83:1–17.

6. Haaber J, Cohn MT, Frees D, Andersen TJ, Ingmer H. 2012. Planktonic aggregates of *Staphylococcus aureus* protect against common antibiotics. PLoS One 7:1–12.

7. Flemming HC, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S. 2016. Biofilms: An emergent form of bacterial life. Nat Rev Microbiol 14:563–575.

8. Kragh KN, Hutchison JB, Melaugh G, Rodesney C, Roberts AEL, Irie Y, Jensen P, Diggle SP, Allen RJ, Gordon V, Bjarnsholt T. 2016. Role of multicellular aggregates in biofilm formation. Appl Environm Microbiol 7:1–11.

9. Melaugh G, Hutchison J, Kasper N, Irie Y, Allen J. 2016. Shaping the growth behaviour of biofilms initiated from bacterial aggregates. PLoS One 11:1–18.

10. Bible AN, Khalsa-Moyers GK, Mukherjee T, Green CS, Mishra P, Purcell A, Aksenova A, Hurst GB, Alexandre G. 2015. Metabolic adaptations of *Azospirillum brasilense* to oxygen stress by cell-cell clumping and flocculation. Appl Environ Microbiol 81:8246–57.

11. Silva IN, Santos PM, Santos MR, Zlosnik JEA, Speert DP, Buskirk SW, Bruger EL, Waters CM, Cooper VS, Moreira M. 2016. Long-term evolution of *Burkholderia multivorans* during a chronic cystic fibrosis infection reveals shifting forces of selection. Am Soc Microbiol Journals 1:1–21.

12. Moreau-Marquis S, Stanton BAB, O'Toole GGA. 2008. *Pseudomonas aeruginosa* biofilm formation in the cystic fibrosis airway. Pulm Pharmacol Ther 21:595–599.

13. Lynch J. 2009. *Burkholderia cepacia* complex: impact on the cystic fibrosis lung lesion. Semin Respir Crit Care Med 30:569–610.

14. Dienerowitz M, Cowan L V., Gibson GM, Hay R, Padgett MJ, Phoenix VR. 2014. Optically

trapped bacteria pairs reveal discrete motile response to control aggregation upon cell-cell approach. Curr Microbiol 69:669–674.

15. Laganenka L, Colin R, Sourjik V. 2016. Chemotaxis towards autoinducer 2 mediates autoaggregation in *Escherichia coli*. Nat Commun 7:1–10.

16. Conrad JC. 2012. Physics of bacterial near-surface motility using flagella and type IV pili: Implications for biofilm formation. Res Microbiol 163:619–629.

17. Chaban B, Hughes HV, Beeby M. 2015. The flagellum in bacterial pathogens: For motility and a whole lot more. Semin Cell Dev Biol 46:91–103.

18. Alexandre G. 2015. Chemotaxis control of transient cell aggregation. J Bacteriol 197:3230–3237.

19. Stacy A, McNally L, Darch SE, Brown SP, Whiteley M. 2016. The biogeography of polymicrobial infection. Nat Rev Microbiol 14:93–105.

20. Jarrell KF, McBride MJ. 2008. The surprisingly diverse ways that prokaryotes move. Nat Rev Microbiol 6:466–476.

21. Hospenthal MK, Costa TRD, Waksman G. 2017. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. Nat Rev Microbiol 15:365–379.

22. Evans LDB, Hughes C, Fraser GM. 2014. Building a flagellum outside the bacterial cell. Trends Microbiol 22:566–572.

23. Pratt LA, Kolter R. 1998. Genetic analysis of *Escherichia coli* biofilm formation: Roles of flagella, motility, chemotaxis and type I pili. Mol Microbiol 30:285–293.

24. Foynes S, Dorrell N, Ward SJ, Stabler RA, McColm AA, Rycroft AN, Wren BW. 2000. *Helicobacter pylori* possesses two CheY response regulators and a histidine kinase sensor, CheA, which are essential for chemotaxis and colonization of the gastric mucosa. Infect Immun 68:2016–2023.

25. Guttenplan SB, Kearns DB. 2014. Regulation of flagellar motility during biofilm formation. FEMS Microbiol Rev 37:849–871.

26. Porter SL, Wadhams GH, Armitage JP. 2011. Signal processing in complex chemotaxis pathways. Nat Rev Microbiol 9:153–165.

27. Kirby JR. 2009. Chemotaxis-like regulatory systems: unique roles in diverse bacteria. Annu Rev Microbiol 63:45–59.

28. Kirn TJ, Lafferty MJ, Sandoe CMP, Taylor RK. 2000. Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by *Vibrio cholerae*. Mol Microbiol 35:896–910.

29. Tripathi L, Zhang Y, Lin Z. 2014. Bacterial sigma factors as targets for engineered or

synthetic transcriptional control. Front Bioeng Biotechnol 2:1–7.

30. Suzuki H, Daimon M, Awano T, Umekage S, Tanaka T, Kikuchi Y. 2009. Characterization of extracellular DNA production and flocculation of the marine photosynthetic bacterium *Rhodovulum sulfidophilum*. Appl Microbiol Biotechnol 84:349–356.

31. Das T, Sehar S, Manefield M. 2013. The roles of extracellular DNA in the structural integrity of extracellular polymeric substance and bacterial biofilm development. Environ Microbiol Rep 5:778–786.

32. Okshevsky M, Regina VR, Meyer RL. 2015. Extracellular DNA as a target for biofilm control. Curr Opin Biotechnol 33:73–80.

33. Qin Z, Ou Y, Yang L, Zhu Y, Tolker-Nielsen T, Molin S, Qu D. 2007. Role of autolysin-mediated DNA release in biofilm formation of *Staphylococcus epidermidis*. Microbiology 153:2083–2092.

34. LaSarre B, Federle MJ. 2013. Exploiting quorum sensing to confuse bacterial pathogens. Microbiol Mol Biol Rev 77:73–111.

35. Kaplan JB. 2010. Biofilm dispersal: Mechanisms, clinical implications, and potential therapeutic uses. J Dent Res 89:205–218.

36. Andersen DH. 1938. Cytic fibrosis of the pancreas and its relation to celiac disease. Am J Dis Child 56:344.

37. Cant N, Pollock N, Ford RC. 2014. CFTR structure and cystic fibrosis. Int J Biochem Cell Biol 52:15–25.

38. Bhagirath AY, Li Y, Somayajula D, Dadashi M, Badr S, Duan K. 2016. Cystic fibrosis lung environment and *Pseudomonas aeruginosa* infection. BMC Pulm Med 16:1–22.

39. Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, Bush A. 2015. Cystic fibrosis. Nat Rev Dis Prim 1:1–19.

40. Saint-Criq V, Gray MA. 2017. Role of CFTR in epithelial physiology. Cell Mol Life Sci 74:93–115.

41. Fanen P, Wohlhuter-Haddad A, Hinzpeter A. 2014. Genetics of cystic fibrosis: CFTR mutation classifications toward genotype-based CF therapies. Int J Biochem Cell Biol 52:94–102.

42. O'Sullivan BP, Freedman SD. 2009. Cystic fibrosis. Lancet 373:1891–1904.

43. European Cystic Fibrosis Society. 2013. ECFS Patient Registry Annual Data Report.

44. Marshall, B.; Elbert, A.; Petren, K.; Rizvi, S.; Fink, A.; Ostrenga, J.; Sewall, A.; Loeffler D. 2016. Patient registry - Annual Data Report 2015Cystic Fibrosis Foundation Patient Registry.

45. Silva A, Amorim A, Azevedo P, Lopes C, Gamboa F. 2016. Cystic fibrosis - characterization of the adult population in Portugal. Rev Port Pneumol 22:141–145.

46. Stoltz DA, Meyerholz DK, Welsh MJ. 2015. Origins of cystic fibrosis lung disease. N Engl J Med 372:351–362.

47. Seeley RR. 2003. Anatomy and physiology.

48. Berkebile AR, City I, City I. 2015. Effects of airway surface liquid pH on host defense in cystic fibrosis. Int J Biochem Cell Biol 52:124–129.

49. Engelhardt JF, Yankaskas JR, Ernst SA, Yang Y, Marino CR, Boucher RC, Cohn JA, Wilson JM. 1992. Submucosal glands are the predominant site of CFTR expression in the human bronchus. Nat Genet 2:240–248.

50. Boucher RC. 2007. Cystic fibrosis: a disease of vulnerability to airway surface dehydration. Trends Mol Med 13:231–240.

51. Elizur A, Cannon CL, Ferkol TW. 2008. Airway inflammation in cystic fibrosis. Chest 133:489–495.

52. Sriramulu DD, Lünsdorf H, Lam JS, Römling U. 2005. Microcolony formation: A novel biofilm model of *Pseudomonas aeruginosa* for the cystic fibrosis lung. J Med Microbiol 54:667–676.

53. Worlitzsch D, Tarran R, Ulrich M, Schwab U, Cekici A, Meyer KC, Birrer P, Bellon G, Berger J, Weiss T, Botzenhart K, Yankaskas JR, Randell S, Boucher RC, Döring G. 2002. Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. J Clin Invest 109:317–325.

54. Hassett DJ, Sutton MD, Schurr MJ, Herr AB, Caldwell CC, Matu JO. 2009. *Pseudomonas aeruginosa* hypoxic or anaerobic biofilm infections within cystic fibrosis airways. Trends Microbiol 17:130–138.

55. Hassett DJ, Cuppoletti J, Trapnell B, Lymar S V, Rowe JJ, Yoon S, Hilliard GM, Parvatiyar K, Kamani MC, Wozniak DJ, Hwang S-H, Mcdermott TR, Ochsner UA. 2002. Anaerobic metabolism and quorum sensing by *Pseudomonas aeruginosa* biofilms in chronically infected cystic fibrosis airways: rethinking antibiotic treatment strategies and drug targets. Adv Drug Deliv Rev 54:1425–1443.

56. Feldman M, Bryan R, Rajan S, Scheffler L, Brunnert S, Tang H, Prince A. 1998. Role of flagella in pathogenesis of *Pseudomonas aeruginosa* pulmonary infection. Infect Immun 66:43–51.

57. Caldara M, Friedlander RS, Kavanaugh NL, Aizenberg J, Foster KR, Ribbeck K. 2012. Mucin biopolymers prevent bacterial aggregation by retaining cells in the free-swimming state 22:2325–2330.

58. Arora SK, Ritchings BW, Almira EC, Lory S, Ramphal R. 1998. The *Pseudomonas aeruginosa* flagellar cap protein, FliD, is responsible for mucin adhesion. Infect Immun 66:1000–1007.

59. Cullen L, McClean S. 2015. Bacterial adaptation during chronic respiratory infections. Pathogens 4:66–89.

60. Eberl L, Tümmler B. 2004. *Pseudomonas aeruginosa* and *Burkholderia cepacia* in cystic fibrosis: genome evolution, interactions and adaptation. Int J Med Microbiol 294:123–131.

61. Hogardt M, Heesemann J. 2010. Adaptation of *Pseudomonas aeruginosa* during persistence in the cystic fibrosis lung. Int J Med Microbiol 300:557–562.

62. Hoboth C, Hoffmann R, Eichner A, Henke C, Schmoldt S, Imhof A, Heesemann J, Hogardt M. 2009. Dynamics of adaptive microevolution of hypermutable *Pseudomonas aeruginosa* during chronic pulmonary infection in patients with cystic fibrosis. J Infect Dis 200:118–130.

63. Petrova OE, Schurr JR, Schurr MJ, Sauer K. 2012. Microcolony formation by the opportunistic pathogen *Pseudomonas aeruginosa* requires pyruvate and pyruvate fermentation. Mol Microbiol 86:819–835.

64. McCarthy RR, Mooij MJ, Reen FJ, Lesouhaitier O, O'Gara F. 2014. A new regulator of pathogenicity (bvlR) is required for full virulence and tight microcolony formation in *Pseudomonas aeruginosa*. Microbiology 160:1488–1500.

65. Tomich M, Herfst CA, Golden JW, Mohr CD. 2002. Role of flagella in host cell invasion by *Burkholderia cepacia*. Infect Immun 70:1799–1806.

66. Silva IN, Ramires MJ, Azevedo LA, Guerreiro AR, Tavares AC, Becker JD, Moreira LM. 2017. The regulator LdhR of *Burkholderia multivorans* plays a role in carbon overflow and in planktonic cellular aggregates formation. Appllied Environ Microbiol 83:e01343-17.

67. Chandler JR, Duerkop BA, Hinz A, West TE, Herman JP, Churchill MEA, Skerrett SJ, Greenberg EP. 2009. Mutational analysis of *Burkholderia thailandensis* quorum sensing and self-aggregation. J Bacteriol 191:5901–5909.

68. Dennis JJ, Zylstra GJ, Dennis JJ. 1998. Plasposons : modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. Appl Environm Microbiol 64:2710–2715.

69. Kessler B, de Lorenzo V, Timmis KN. 1992. A general system to integrate lacZ fusions into the chromosomes of gram-negative eubacteria: regulation of the Pm promoter of the TOL plasmid studied with all controlling elements in monocopy. Mol Gen Genet 233:293–301.

70. Kovach ME, Phillips RW, Elzer PH, Roop RM, Peterson KM. 1994. pBBR1MCS: a broad-host-range cloning vector. Biotechniques 16:800–2.

71. Vieira J, Messing J. 1991. New pUC-derived cloning vectors with different selectable markers and DNA replication origins. Gene 100:189–94.

72. Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. J Bacteriol 149:114–122.

73. J S, Russel DW. 2001. Molecular cloning: a laboratory manual. New York.

74. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

75. Joshi NA, JN F. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33).

76. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, McLean J, Lasken R, Clingenpeel SR, Woyke T, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling Genomes and Mini-metagenomes from highly chimeric reads, p. 158–170. *In* . Springer, Berlin, Heidelberg.

77. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352.

78. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

79. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–9.

80. Palmer KL, Aye LM, Whiteley M. 2007. Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. J Bacteriol 189:8079–8087.

81. Zlosnik JEA, Hird TJ, Fraenkel MC, Moreira LM, Henry DA, Speert DP. 2008. Differential mucoid exopolysaccharide production by members of the *Burkholderia cepacia* complex. J Clin Microbiol 46:1470–1473.

82. Bauer AW, Kirby WM, Sherris JC, Turck M. 1966. Antibiotic susceptibility testing by a standardized single disk method. Am J Clin Pathol 45:493–6.

83. Ferreira AS, Leitao JH, Sousa SA, Cosme AM, Sa-Correia I, Moreira LM. 2007. Functional analysis of *Burkholderia cepacia* genes bceD and bceF, encoding a phosphotyrosine phosphatase and a tyrosine autokinase, respectively: role in exopolysaccharide biosynthesis and fiofilm formation. Appl Environ Microbiol 73:524–534.

84. Seed KD, Dennis JJ. 2008. Development of *Galleria mellonella* as an alternative infection

model for the *Burkholderia cepacia* complex. Infect Immun 76:1267–75.

85. Pessoa FD. 2017. Long-term evolution of *Burkholderia multivorans* bacteria during chronic respiratory infections of cystic fibrosis patients.

86. SnapGene software (from GSL Biotech; available at snapgene.com).

87. Sivapragasam S. 2015. Bacterial Xanthine Dehydrogenase Regulators.

88. Cashel M. 1975. Regulation of ppGpp and pppGpp. Annu Rev 301–318.

89. Schmid N, Suppiger A, Steiner E, Pessi G, Kaever V, Fazli M, Tolker-nielsen T, Jenal U, Eberl L. 2018. High intracellular c-di-GMP levels antagonize quorum sensing and virulence gene expression in *Burkholderia cenocepacia*. Microbiology 754–764.

90. Gupta KR, Kasetty S, Chatterji D. 2015. Novel functions of (p)ppGpp and cyclic di-GMP in mycobacterial physiology revealed by phenotype microarray analysis of wild-type and isogenic strains of *Mycobacterium smegmatis*. J Appl Environ Microbiol 81:2571–2578.

91. Boyd CD, O'Toole GA. 2012. Second messenger regulation of biofilm formation: Breakthroughs in understanding c-di-GMP effector systems. Annu Rev Cell Dev Biol 28:439–462.

92. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30.

93. Kulakova AN, Kulakov LA, Villarreal-Chiu JF, Gilbert JA, McGrath JW, Quinn JP. 2009. Expression of the phosphonoalanine-degradative gene cluster from *Variovorax* sp. Pal2 is induced by growth on phosphonoalanine and phosphonopyruvate. FEMS Microbiol Lett 292:100–106.

94. Tekaia F, Gordon S V, Garnier T, Brosch R, Barrell BG, Cole ST. 1999. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. Tuber Lung Dis 79:329–342.

95. Veselova MA, Lipasova VA, Zaitseva Y V., Koksharova OA, Chernukha MY, Romanova YM, Khmel' IA. 2012. Mutants of *Burkholderia cenocepacia* with a change in synthesis of N-acyl-homoserine lactones—Signal molecules of quorum sensing regulation. Russ J Genet 48:513–521.

96. Jassem AN, Forbes CM, Speert DP. 2014. Investigation of aminoglycoside resistance inducing conditions and a putative AmrAB-OprM efflux system in *Burkholderia vietnamiensis*. Ann Clin Microbiol Antimicrob 13:1–5.

97. Basler M. 2015. Type VI secretion system : secretion by a contractile nanomachine. Philos Trans B.

98. Aubert DF, Hu S, Valvano MA. 2015. Quantification of type VI secretion system activity in macrophages infected with *Burkholderia cenocepacia*. Microbiol (United Kingdom) 161:2161–2173.

99. Chen R. 1975. The primary structure of protein S9 from the 30S subunit of *Echerichia*. FEBS Lett 52:139–140.

100. Dabbs ER. 1978. Kasugamycin-dependent mutants of *Escherichia coli*. J Bacteriol 136:994–1001.

101. Araujo HF, Campos PC, Camargo DRA, Pereira FNR, Samuel ML, Oliveira MAA, Fortes-Dias CL, Leclercq SY. 2012. Immune response and protective efficacy of S9 ribosomal protein of *Streptococcus pneumoniae* in a model of sepsis. Can J Microbiol 58:1055–1062.

102. Lardi M, Aguilar C, Pedrioli A, Omasits U, Suppiger A, Cárcamo-Oyarce G, Schmid N, Ahrens CH, Eberl L, Pessi G. 2015. σ^{54}-Dependent response to nitrogen limitation and virulence in *Burkholderia cenocepacia* strain H111. Appl Environ Microbiol 81:4077–4089.

103. Eberl L, Ammendola A, Rothballer MH, Givskov M, Sternberg C, Kilstrup M, Schleifer KH, Molin S. 2000. Inactivation of gltB abolishes expression of the assimilatory nitrate reductase gene (nasB) in *Pseudomonas putida* KT2442. J Bacteriol 182:3368–3376.

104. Hashim S, Kwon DH, Abdelal A, Lu CD. 2004. The arginine regulatory protein mediates repression by arginine of the operons encoding glutamate synthase and anabolic glutamate dehydrogenase in *Pseudomonas aeruginosa*. J Bacteriol 186:3848–3854.

105. Liu Y, Lardi M, Pedrioli A, Eberl L, Pessi G. 2017. NtrC-dependent control of exopolysaccharide synthesis and motility in *Burkholderia cenocepacia* H111. PLoS One 12:1–22.

106. Brown DR, Barton G, Pan Z, Buck M, Wigneshweraraj S. 2014. Nitrogen stress response and stringent response are coupled in *Escherichia coli*. Nat Commun 5:7–10.

107. Takada H, Morita M, Shiwa Y, Sugimoto R, Suzuki S, Kawamura F, Yoshikawa H. 2014. Cell motility and biofilm formation in *Bacillus subtilis* are affected by the ribosomal proteins, S11 and S21. Biosci Biotechnol Biochem 78:898–907.

108. Xu X, Yu H, Zhang D, Xiong J, Qiu J, Xin R, He X, Sheng H, Cai W, Jiang L, Zhang K, Hu X. 2016. Role of ppGpp in *Pseudomonas aeruginosa* acute pulmonary infection and virulence regulation. Microbiol Res.

109. Huaman MA, Fiske CT, Jones TF, Warkentin J, Shepherd BE, Maruri F, Sterling TR. 2015. The many roles of glutamate in metabolism. J Ind Microb Biotechnol 143:951–959.

110. Arhin A, Boucher C. 2010. The outer membrane protein OprQ and adherence of *Pseudomonas aeruginosa* to human fibronectin. Microbiology 156:1415–1423.

111. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452.

6 SUPPLEMENTARY MATERIAL

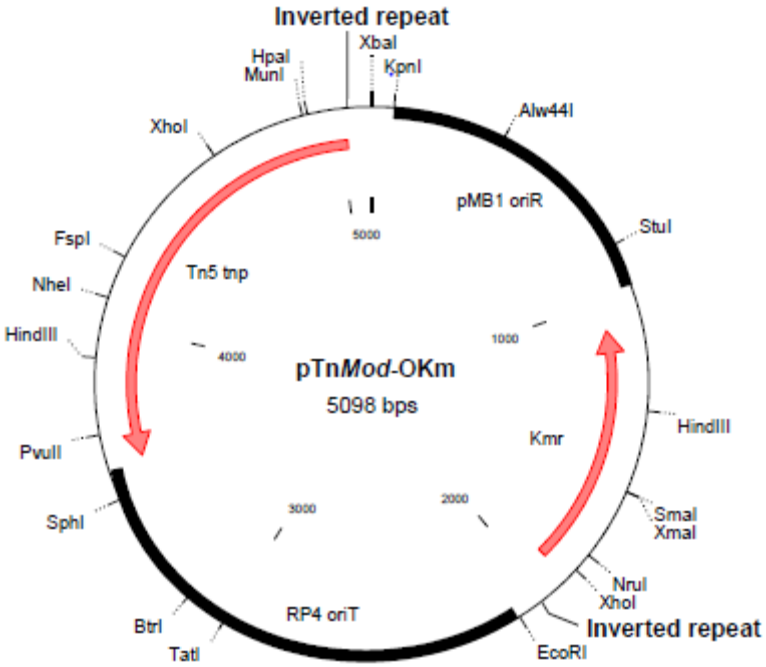

Fig. S1 Structure of transposon pTn*Mod*-ΩKm. Transposon is composed of genes that encode a transposase, and within the two inverted repeats sits an *E.coli* origin of replication and a selectable marker, in this case, a kanamycin marker (retrieved from (68)).

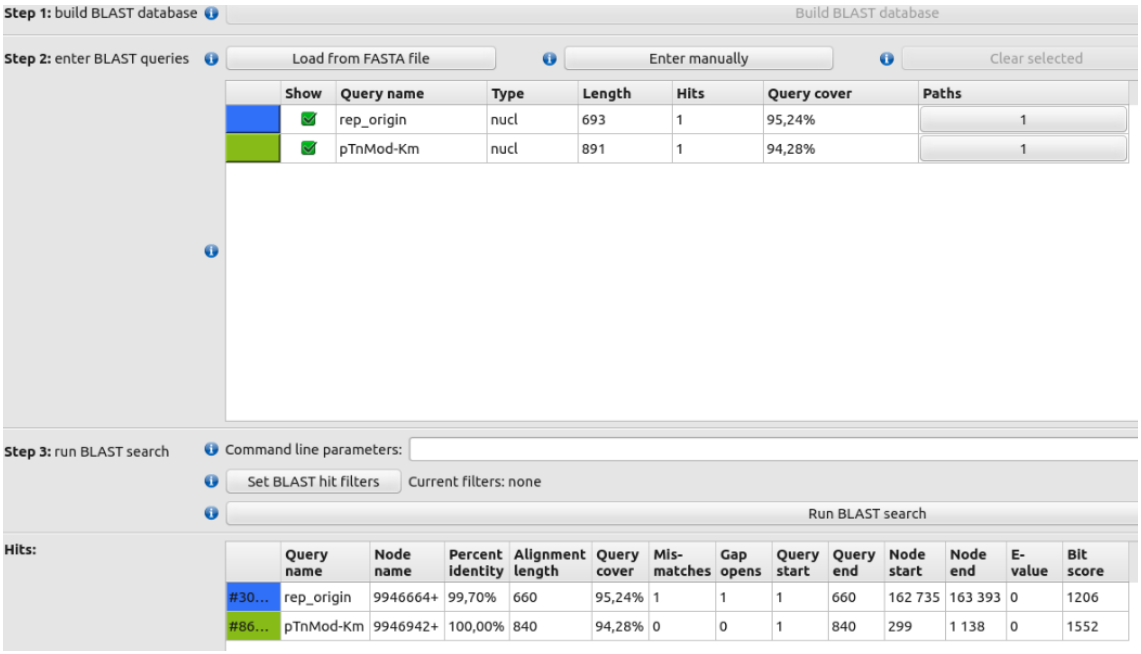

Fig. S2 Results from a BLAST search for the plasposon sequence using Bandage Tool. For sequences form different regions of plasposon, were tested and only one hit for each of them was found.

Table S1 Plasposon flanking sequences obtained from the plasmids recovered from the B63 and B68 mutants.

B63 mutant	
Seq A	TCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGGGCGAAGCCGACGCCGTAGCGCTTGC CGGGATTCGCGGCCTCGAATTCGACCTTCGCGAACGATGCTTCGGCGCCGGTGCCGAAGTCCTTCTGCACGCAGGCGAAG CCCTGTCTCTTGATCAGATCTGGCCGCCTAGGCCGGCCGCGATCGCCGGCGCGCCATGCATGTCGACGGTACCAGGAAAG AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC CCCTGACGAGGATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTC CCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCA CGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTAT CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG CCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG TAGCTCTTGATCCGGCAAACAAACCACCCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGC
Seq B	TGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAGATCTGATCAAGAGACAGGGCTTCGCCTGCGTGCAGAAGG ACTTCGGCACCGGCGCCGAAGCATCGTTCGCGAAGGTCGAATTCGAGGCCGCGAATCCCGGCAAGCGCTACGGCGTCGGC TTCGCCCTGACTCTTATACACAAGTGCGGCCGCACTAGTCTAGATTTAAATTAATTAAGAGCTCGGGGGGGGGGGGGAAA GCCACGTTGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCT TACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAA CATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATG GGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGA
Seq C	TGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCGATCTGATCAAGAGACAGGGCTTCGCCTGCGTGCAGAAGGACTTC GGCACCGGCGCCGAAGCATCGTTCGCGAAGGTCGAATTCGAGGCCGCGAATCCCGGCAAGCGCTACGGCGTCGGCTTCGC CCTGACTCTTATACACAAGTGCGGCCGCACTAGTCTAGATTTAAATTAATTAAGAGC
B68 mutant	
Seq A	CGTGGCTTTCCCCCCCCCCCCGGCTCTTAATTATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGGCC CCGTGCCGTTCACGAGCATCAGCCGCAGATGCGTACGCACGATGAACGCGAACTGCTGCTGCACGGTCGCCTCCTGCGGC ATCCCGTGCTCGAACGCCGCGATGATTTCGTCATACCAGTCGGCGATCACGCGCGCGCACAGCTCGCGCTTGCCGCGGAA ATAACTGAACACGGTCGCCTCCGACACGCCGACGCGTTGCGCGATCTCGGCGGCCGTCGCGTGCTCGTAACCCTTTTCGG CGAACACGTCGCGCCCGGCCCGCAGGATGTCCTGCACGCGCTGCTGGGATTTGCGCCCGGCGGGCGCGCGCGGCGCGTCG GCGCGCTCGGCTTTGGCGGTGGCGGTCATGGTGCGGGCGGTAGCGGCTGTCATCTCCGCATGATAACTGAGTGTCACTCA AAAAACTATTGACGCCCACCACGACGCCGCGCGAAACTGTGCAAAATTTCCGCAGTAATGGTCAATGAATCGGCGTATCG CCCGATTTGAGCAAAACTCAGAAATCGGCGCGTGCGACGCACTTTCTGGAGACGGAGGAGACATGAGCAACCTGCCCGGC GTGCAATTCATGCTCGGTGAAGACATCGAGATGCTGCGCGATGCCGTCGCGACATTCGCGGCGAAGGAAATCGCGCCGCG CGCGGCCGAGGTCGACCGTACCGACCAGTTTCCGATGGATCTCTGGAAGAAGTTCGGCGATCTCGGCGTGCTCGGGATGA CGGTGGCCGAGGAGTACGGCGGCGCGAACATGGGCTACACCGCGCACATGGTCGCGATGGAGGAGATCTCGCGCGCATCG GCGTCGATCGGCCTGTCGTACGGCGCGCACTCGAATCTCTGCGTGAACCAGATCCACCCGCAACGGCACCGAAGCGCAGA AGCGCAAGTATCTGCCGAAAGCTCGTCCTCGGGCGAACACAT
Seq B	CGTGGCTTTCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGGC CCCGTGCCGTTCACGAGCATCAGCCGCAGATGCGTACGCACGATGAACGCGAACTGCTGCTGCACGGTCGCCTCCTGCGG CATCCCGTGCTCGAACGCCGCGATGATTTCGTCATACCAGTCGGCGATCACGCGCGCGCACAGCTCGCGCTTGCCGCGGA AATAACTGAACACGGTCGCCTCCGACACGCCGACGCGTTGCGCGATCTCGGCGGCCGTCGCGTGCTCGTAACCCTTTTCG GCGAACACGTCGCGCCCGGCCCGCAGGATGTCCTGCACGCGCTGCTGGGATTTGCGCCCGGCGGGCGCGCGCGGCGCGTC GGCGCGCTCGGCTTTGGCGGTGGCGGTCATGGTGCGGGCGGTAGCGGCTGTCATCTCCGCATGATAACTGAGTGTCACTC AAAAAACTATTGACGCCCACCACGACGCCGCGCGAAACTGTGCAAAATTTCCGCAGTAATGGTCAATGAATCGGCGTATC GCCCGATTTGAGCAAAACTCAGAAATCGGCGCGTGCGACGCACTTTCTGGAGACGGAGGAGACATGAGCAACCTGCCCGG CGTGCAATTCATGCTCGGTGAAGACATCGAGATGCTGCGCGATGCCGTCGCGACATTCGCGGCGAAGGAAATCGCGCCGC GCGCGGCCGAGGTCGACCGTACCGACCAGTTTCCGATGGATCTCTGGAAGAAGTTCGGCGATCTCGGCGTGCTCGGGATG ACGGTGGCCGAGGAGTACAGCGGCGCGACATGGGCTACACCGCGCACATGGTCGCGATGGAGGAGATCTCGCGCGCATCG GCGTCGATCGGCCTGTCGTACGGGGCGCACTCGAATCTCTGCGTGA

Table S2 Plasposon flanking sequences obtained from the plasmids recovered from the B68 and B74 mutants.

B68 mutant	
Seq C	CCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAGATCTGATCAAGAGACAGGCACGGGGCTCTG CGCACTGGTGCTGTCCGAAGGCCGCGCGAAGCAGCACGCGCTCAGCGACGAGCTCACCGCGCTGCAGCGCCGCTACACGG CGCCGCTGATGAGCGTGCTCGCGCGCGGCCAGGCGGCCGGGCAGGTGCGCCGCGACATGCCGCTGAGCCTGTTGCGCTCG ATGGTGTTCGGTCCGATCGAGCACGTGCTGTGGGACGCGATCCTCGGGCACCGCAAGCTCGATACGGAAACGACCGCGAC GCAGCTGATCGACATGCTCTGGGCGGCCGTGCAGCCGCCCGCGCCCGAGCAGGCCGCGCTCGTGCGGTTCAAAAACGAGG TCGCGGAAGCGATGAAGCGGCTGGAAGCGTCGCCGGAACGAGAATAGGTCGGCGCGTCACGCCGGGCGCGGACAATCGCC CGCGCCGTCCATCACGGCCGCCAGCGGTACACGACAAGACTCGAGCCGTTGTAATTGTCCTTCGTGATCACGTATTCGCC GGTCGAGCGCAAATACGCGCGGATGCCGTACATCGAGTCGACGTCGTTGCCGACATCCATCGCCGACGGGTTCGAGTTCG TCAGCGTCGTGACGAGGCCGCCCGTGTTCAGGTCGTACACGTCGATGTTCGGCACCGTGTGCACGTAGCCGACGAACAGA TAGTGGCCGGCTGCCGCGATCGACTTCGGGTTCGCGCTCGTCAGATCGATCGCGAGGCTCGGGGTGGTTTTGTTGCCGGC TTTCCAGCCGCGATAGATCTCGATATGGCCGTTCATCGCGGTCCAGTCCCAGCTGCCCGAAATGCCCTGCGCGAGGATCA TCGTGTCGCTGTCGGACTGATAGATGATGCGCGTGAGCGGCAACACCCGAACGCGGGAATCTGGATCTGCTCGGGCGCGC CCCACGAAAGG
B74 mutant	
Seq A	TTCTGCCTTTGCTAATGTTCTTTCTGGTACCGTCGAATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCGATC TGATCAAGAGACAGAACGCAGGCTTCGTCACGCGTGACGCGCGTGAAGTCGAGCGTAAGAAGGTCGGTCTGCACAAGGCA CGCCGCGCCAAGCAGTTCTCGAAGCGTTAATTCCGCTTCATGGCCGCGCCGCTTGCGGGCGGCGCCCGCCGGAAAAACCG CCAGCTTTCGCGCTGGCGGTTTTTTTATGCGCGCGTGCCGGACGGGCCGTCGCCGCCGCGCGCGCCAGGCCCGGCGCGGC GAAATCGACAAGAACCTGTTGATTTCATGGACTTCAGCGGGGGGCTGGGGGGGCGGCCCTACAATAGCGGCTAAAACTTT TTGGAGAGTTCGCATGAACGCTGTTACCGAATCCGCAGCAACGACCACCCCCACCGCCGGCTCCGTTCGTCTTCACCGAC GCCGCGGCCGACAAGGTCAAGCAATTGATCGACGAAGAGGGCAATCCCGACCTGAAGCTGCGCGTATTCGTGCAAGGCGG CGGCTGCTCGGGCTTCCAGTATGGCTTCACGTTCGACGAGGAAAGTCAACGAGGACGACACCGTGATGAACAAGAACGGG CGTCCCAGCTGGCTGATCGACTCGATGAGCTACCAGTATCTGGGTCGGGCGCCGATATCCGACTACCAGGACC
Seq B	GGGTCCCCCGATTCAGATGAACCTGGCTTTCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCA CTTGTGTATAAGAGTCAGGCCTGCGTTCGACAGGGCCGGCTTCAGCGTCGCGTCGTAGTCGATCAGCGCGCGCGTGATGC CGTGACGCACGGCGCCTGCGGTGGCCCGTTTCGCCGCCGCCCGACACGTTCACCTTGATGTCGAACGTTTGAGCGTGGTT CGTCAGTTCCAGCGGCTGACGCACGATCATCAGCGACGTTTCACGTGCGAAGTAGTCAGCGATGGGCTTGCCGTTGACGA TGATGTCGCCCTTGCCAGCCTTGATGAAGACACGAGCGACTGCGCTCTTGCGGCGGCGGGGACGGGGGTTCCAGTTACCG ATCATGTGGGCTCCCCTTAGATCTCGAGCGCCTTCGGCTGTTGAGCCGAGTGCGGATGCGTCGCCCCCCGCGTAGACCTT CAGCTTCTTGATCATCGCGTAGCCGAGCGGGCCCCTTTCGGCAGCATGCCCTTGACCGCCTTCTCGAGCGCACGGCCCGG GAAGCGTTCCCTGCATCTTGCCGAACGTCGTTTCCTAGATACCGCCCGGGTAGCCCGAGTGACTGTAG
Seq C	TTCTGCTTTGCTAATGTTCTTTCCTGGTACCGTCGAATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCGATC TGATCAGAGACAGAACGCAGGCTTCGTCACGCGTGACGCGCGTGAAGTCGAGCGTAAGAAGGTCGGTCTGCACAAGGCAC GCCGCGCCAAGCAGTTCTCGAAGCGTTAATTCCGCTTCATGGCCGCGCCGCTTGCGGGCGGCGCCCGCCGGAAAAACCGC CAGCTTTCGCGCTGGCGGTTTTTTTATGCGCCCGTGCCGGACGGGCCGTCGCCGCCG
Seq D	GGTCAGCGTCAGCAATGACCGTGGCTTTCCCCCCCCCCGACTCTTAATAATTTAAATCTAGACTAGTGCGGCCGCACTTG TGTATAAGAGTCAGGCCTGCGTTCGACAGGGCCGGCTTCAGCGTCGCGTCGTAGTCGATCAGCGCGCGCGTGATGCCGTG ACGCACTGCGCCTGCCTGACCCGTTTCGCCGCCGCCCGAACGTTCACCTTGATGTCGAACGTTTGAGCGTGGTTCGTCAG TTCCAGCGGCTGACGCACGATCATCAGCGACGTTTCCCGTGC

Table S3 Plasposon flanking sequences obtained from the plasmids recovered from the E28 mutant.

E28 mutant	
Seq A	TTAGGACCGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGCCTAGTGCGGCCGCACTTGTGTATAAG AGTCAGAGCACCTTCTGGCCGAACAGCCCGCGGTCGTCGGCGGCCGAGCAGAACGCGAAGTCGCGCACGCTCTGGTTGGT CGGCTCGGTCAGCGACGAATCGACGACGACGCTGAACTCGGCCTCGCTCTCGGCGAGATCGCGCAGGATGTAGCTGCGGA ACAGCAGGTCGCCGTACGAGATCACGGTGTCGCCGGTCAGGCGCTCGGCCGCGCAGGCGAGCGATGCGAGTTCGCCCGTT TCGGCGTGGCGCTCGTTGACGACGAGCTTGATGCCCGACGTGTCGATCGCATCGGCGCGATAGCCGCCGACGACGGTGAT GTCGTTCACGCCGTGCTTCTTGAAGCCGTCGACGAGCCAGCGCAGCAGCGGCTTGCCGGCAACGGGCAGCATCACCTTCG GCTTGTCCTCGGTGACGGCCTCGAGCCCCTTGCCGCGGCTCGCGGCCAGCACGACTGCCGCGTTCGACGCGCGCGACGAC GACGACAGGTAGATACGCTCGGCCGCCGAGTATTCGTCGGCGTCCTGCAGACGGAAGATCTCGTTGACCGACGCGACGCG ATCCTCGACGTTGATCAGCGTTTCGCTTTCGTGGATCTCGCGGGCGACGGCCTGCATCGCCGATGCCGACGCGCGGATCA GATGGTTCGCCCAGATCACGGTGCTGATGCCGGCCTGGCGGAACACGTCGGTGGGCGTGCTGTAGTACTTGGTCGGCACG ATCACGAGCGGCGCCTTGCCGCTCCATTCGCGCGCGAACTGCAGGATTTTCGTCCGGGGCGCGACAGCTTGCTGTGGATC AGGATCGCGTCGGCGCCGGCTTCCGCATAGGCGTTCGCGCGGCGCAGTGCCTCGTCCATGCCCCAGCCCGCGATCAGCGC TTCGACGCGCGCGACGATCGAGAAGTCGGGGTCGCTCTGCGAAATCCCTTGCCGGCCTTGGATCTTGCCCGCAAGAATTC GCTTCGCGATGGCAACAGGTCCTGTGGCAGCAGGCTGCGGCGCCACGCGTCGGTGAGCGGTGACCGGGCGGATCAGCCAG CAGCAATCATGGCTCTTCACGGTGGCCATCAAGTATCGAGCCGGGCGCCGCGATGAG
Seq B	ATCCGTGCCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGC CAGATCTGATCAAGAGACAGGAAGGTGCTGCTCCAGCGCGTGTCGAGCGATGCCGCCGCCGGCACGCCGCACGGCCGCTG GATCGGGCTGCTGAACGTGCGCGGCGCCGGCGTCGAGCGTCTGAAGGCGATGCTCGCGACGCTGCAGGCACGCGCCGATT TCGACACGCTGGACATCCCCGCGCTGCTCAACGCGCTGATCGACGCCGGCGAGAAGATCGAGGTGCAGTACGTGCACGGC CACTGGCGCGGCGTCAACGATCTCGAAGACTTCCGCCGCGCGGGCGACTTCGCGCACGGACAAACGCCCCTGTCGGAACC GGGCGCCGGCAACGGAGGCGCGCAATGATCGAAGCGGCCCAGTTCGTCGAGGCCGCACGCGCACGCGGCTTCGACTGGTA CGCGGGCGTGCCCTGCTCGTACCTGACGCCGTTCATCAACTACGTGCTGCAGGATCCGACGCTGCACTACCTGTCGGCCG CCAACGAAGGCGACGCGGTCGCGCTGATCGCCGGCGCGACGCTCGGCGGCAAGCGCGGGATCGCGATGATGCAGAACTCG GGGCTCGGCAACGCCGTCAGCCCGCTCACGTCGCTCACCTGGACCTTCCGCCTGCCGCAGCTGCTGATCGTCACCTGGCG CGGTCAGCCGGGCGTCGCCGACGAGCCGCAGCACGCACTGATGGGGCCGATCACGCCCGCGATGCTCGACACGATGGAGA TCCCGTGGGAGACGTTCCCGACCGACCCCCGAACAGGTCGGCCCGGCACTCGACCGCGCGATCGCGCACATGGACGCGAC GGGTCGCCCGTACGCGCTCGTGATGCAGAAAGGCAGCGTCGCGCCGTATGAGCTGAAGGCGAATCCGGCTGCGCCGCGCG CGCATGTCGCCCGCGCGGTCGTCGTCGCGTGCGGCGTCGCCCCGATGCATGGCCCGACGCGTCACGATGCGGCTGCAGCG CGTGATCGCGCATACGCCCGGTTCGACTCGGACCGTTCGTGGGCTCGCCGTTCCACCCGGCCTTTCTGGCGGGGCCG
Seq C	TTAGTATCCGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAA GAGTCAGAGCACCTTCTGGCCGAACAGCCCGCGGTCGTCGGCGGCCGAGCAGAACGCGAAGTCGCGCACGCTCTGGTTGG TCGGCTCGGTCAGCGACGAATCGACGACGACGCTGAACTCGGCCTCGCTCTCGGCGAGATCGCGCAGGATGTAGCTGCGG AACAGCAGGTCGCCGTACGAGATCACGGTGTCGCCGGTCAGGCGCTCGGCCGCGCAGGCGAGCGATGCGAGTTCGCCCGT TTCGGCGTGGCGCTCGTTGACGACGAGCTTGATGCCCGACGTGTCGATCGCATCGGCGCGATAGCCGCCGACGACGGTGA TGTCGTTCACGCCGTGCTTCTTGAAGCCGTCGACGAGCCAGCGCAGCAGCGGCTTGCCGGCAACGGGCAGCATCACCTTC GGCTTGTCCTCGGTGACGGCCTCGAGCCCCTTGCCGCGGCTCGCGGCCAGCACGACTGCCGCGTTCGACGCGCGCGACGA CGACGACAGGTAGATACGCTCGGCCGCCGAGTATTCGTCGGCGTCCTGCAGACGGAAGATCTCGTTGACCGACGCGACGC GATCCTCGACGTTGATCAGCGTTTCGCTTTCGTGGATCTCGCGGGCGACGGCCTGCATCGCCGATGCCGACGCGCGGATC AGATGGTTCGCCCAGATCACGGTGCTGATGCCGGCCTGGCGGAACACGTCGGTGGGCGTGCTGTAGTACTTGGTCGGCAC GATCACGAGCGGCGCCTTGCCGCTCCATTCGCGCGCGAACTGCAGGATTTCGTCGGGGCGCCGACAGCTTGCTGTGGATC AGGATCGCGTCGGCGCCCGGCTTCCGCATAGGGCGTTCGCGCGGCGCAGTGCCTCGTCCATGCCCCAGCCCGCCGATCAG CGCTTCGACGCGCGCGACGATCCGAGAAGTCGGGGTCCGCTCCTGCGAATCCTTGCCCGGCCCTTGATCCTTGCCGCAGA AATTCGCTTTCCGCGATGGCCACAGGTCGTGCCAGCAAGGCTGCGGCGCACGCCGTTCGGTGGAGCGTGACCCGGGCCGG GATTCCAGCAGCCAGCCATCCAATGCTCCTTCGAGGCGCGTGTGGGGCTCGAATCTAGAGACCACGGACAGCCCGATATG TGTCGCGCCCAATTCAA
Seq D	ATTCGTGCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCC AGATCTGATCAAGAGACAGGAAGGTGCTGCTCCAGCGCGTGTCGAGCGATGCCGCCGCCGGCACGCCGCACGGCCGCTGG ATCGGGCTGCTGAACGTGCGCGGCGCCGGCGTCGAGCGTCTGAAGGCGATGCTCGCGACGCTGCAGGCACGCGCCGATTT CGACACGCTGGACATCCCCGCGCTGCTCAACGCGCTGATCGACGCCGGCGAGAAGATCGAGGTGCAGTACGTGCACGGCC ACTGGCGCGGCGTCAACGATCTCGAAGACTTCCGCCGCGCGGGCGACTTCGCGCACGGACAAACGCCCCTGTCGGAACCG GGCGCCGGCAACGGAGGCGCGCAATGATCGAAGCGGCCCAGTTCGTCGAGGCCGCACGCGCACGCGGCTTCGACTGGTAC GCGGGCGTGCCCTGCTCGTACCTGACGCCGTTCATCAACTACGTGCTGCAGGATCCGACGCTGCACTACCTGTCGGCCGC CAACGAAGGCGACGCGGTCGCGCTGATCGCCGGCGCGACGCTCGGCGGCAAGCGCGGGATCGCGATGATGCAGAACTCGG GGCTCGGCAACGCCGTCAGCCCGCTCACGTCGCTCACCTGGACCTTCCGCCTGCCGCAGCTGCTGATCGTCACCTGGCGC GGTCAGCCGGGCGTCGCCGACGAGCCGCAGCACGCACTGATGGGGCCGATCACGCCCGCGATGCTCGACACGATGGAGAT CCCGTGGGAGACGTTCCCGACCGACCCCGAACAGGTCGGCCCGGCACTCGACCGCGCGATCGCGCACATGGACGCGACGG GTCGCCCGTACGCGCTCGTGATGCAGAAAGGCAGCGTCGCGCCGTATGAGCTGAAAGGCGAATCCCGGCTGCGCCGCGCG CCCATGTCGCCGCGCGGTCGTCGTCGCGTTGCGCGTCGCCCGATGCATGGCCGACGCGTCACGATGCGCTGCAGCGCGTG ATCGCGCAATACCGCCGGTCGACCTCGAACCGTCGTTGCTCGCGTCCACCCGGCTTCTGCGGCCGGGAACTTGTATCGCC GCCTCCGACGAATTCGAA

Table S4 Plasposon flanking sequences obtained from the plasmids recovered from the E36 mutant.

E36 mutant	
Seq A	TTCAGCCCGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAG AGTCAGCGCTCGAACAGGTTGTCGAGATGCTCGAGGTAGCCTTCGAAATACGACTCGAGCAGCGCTTCGCCGAACGCCTC CTTCGAGCCGAAGTAATGGTAGAACGACCCCTTCGGGATGCCCGCCGCAGCGAGAATCTCGTTCAGCCCGACCGCCGAAA ACCCCTTGTGGAGCATGATCGGCTTCGCGATGTTGAGGATGTGCTGGCGGACTTCCGCGCCCGAGGATGCATTCATGATG GCGTATGGTAGCAGCGATTAGACCGGTCGTCTAACGGCGTCGGGGCGACGTTTCGCGCGTTGCAGTTTGGTCGCTTTGCA GGTTTGCAGGTCGCGGTGTACGTTGACCTGCTCGGGCTTGTTTAGCCTCATGCCGGTTTGGGCGCGCGGCGCCGCGCGAT CGGTTTTGTGGGGCGTACGGCCGACACCGGCCGGATCGGCCCGCGACGCGCCGTCACACGCGGCGCCGACGCCCGCTCGG CACGCAGCGCGTCGAGCTCGGCCGCGCTTGCCGCCTGTGCCGAGCGTGCGCCCTCGAGCTGCCCTTCCAGGACGCCGCAG CGATGCTGGCTGTCGGCAAGCTGCGCCTGCAGCGCGCCGATTTCGCGGCGAAGTTGTGCGTCGCGTTCGTCGGCGCGCGC GTTCGTCGCGTCGAGTTCCTTTTGCAGTCGTGCGGCAGCCTGCCGCTCGCGATCGATTTCCTGCAGCGCGCGGCGCTCCG ACGCGCGCAGACGCTCTTCCGCCCGGCTCGCATCGTCGCGCAGCCGATCGAGCTGCGTGCCGAAGTCGGCGCGCGCTTGT GCGAGCGCGCGAGCGCCTTCCGCGTTATCGGCCTTCAGCGTTCGATTTCGGCTTGCTGCGCGCGGCGCGTGGCGTCGTCG GCCGCCCTCGCCTGCTCCAGTTTCCTGAATCCTCACCCTGGTGCCGCGAGCAACGCGGCTGTGCGCTGCTCGAGCCGCGG TTCGGTGGCGGGCCAGCTCGACTGCAAAGGGCGGCGACCTCGCCTTGGCGCTGCCGGTTCCGGCTCGACTCCGGCGCGCC AGTGGGGGTCAAGGCCGGCGGCCGCATTCGGCCTGGACGATCTGTCCAACGTCGCGGACCCAATTCTCCGG
Seq B	ATCGTGCCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCC AGATCTGATCAAGAGACAGGTTCGAGCGCCAGCCCGGCACGGGCGCCGAGCGGATGATGACGTACTGGAGCAATTGGCTG CACACGCAGTGCGCGGACGATCCCGAGGGCAAGTGCCTCGCGGTGAAGCTGGGCGCGGAAGTGTCCGACCTGTCCGAAGC GATGCGCGCGGTACTGCGGCGCGGCACGAGCCAGATCGTCGAACGGCTGGCGGCATGCATCGAAGCCGGCCTCGCGGACG GTTCGCTGCGCGGCATCGAGGATCCGTCGCACACCGCTGCGCTGCTGTACGAGCTGTGGCTCGGTGCGACGCTGCTGGAG AAGATCCACCGCAACCGCAAGCCGCTCGAAACCGCGATGGCCGCGACGCGGCAGCTGCTGAACCTGCCGCCGTCCGCGCC CGACGAGGTGCGGACATAAGCGGGCGCGCTACGGCGCGCCTTGTTTTTACCTATCACACTAGACGACTGGTCTACTCATG ACGATCGCGATTCCTTCCCCGTTTTTGCCCTCGCAGCATCCGCGCCCGCAAGGGCAACGGGGCGATCGGCCGCGCATCGC GGATCACGCGTAGTCCACTCGTCCTTTCCTGATTCGATCGTGTGCGTGCACACGCATCCTTGTTGACGGAGGCTCACATG TCGCAAAGCAAGACCGCAAACCGCCGTATCGTCCTCAATTCGCGCCCGGTCGGCGCGCCCACCGCCAACGACTTCCGCCT CGAAACCGGCGACGTGCCGACGCCCGGCGCCGGGCAGGTGCTGCTGCGCACGGTATGGCTGTCGCTCGATCCGTACATGC GCGGCCGGATGAGCGACGCGCCGTCGTATGCGCCGCCCGTGCAACTCGGCGACGTGATGGTCGGCCGGCACGATCAGCCG CGTCGTATCGTCGAACCCTGCCGGCGTTTCCGCGGAGGGGCGACCTCGTCGTCGCGACGGGCGCTGGCAGGATTATGCGC TGTTCGACGCAGCGACCTGATTCGGCTTCGCCCGGCGATTCCCGCATCCGTCGCGCCGCGCCTCGGCGTGGCTCGGCATG CCGGATTCACGGTTAACCGCCTGCTACGATTCGCCGAGCCGAAGGGCCGTGTG
Seq C	TTCAGCCCGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGCACTAGTGCGGCCGCACTTGTGTATAA GAGTCAGCGCTCGAACAGGTTGTCGAGATGCTCGAGGTAGCCTTCGAAATACGACTCGAGCAGCGCTTCGCCGAACGCCT CCTTCGAGCCGAAGTAATGGTAGAACGACCCCTTCGGGATGCCCGCCGCAGCGAGAATCTCGTTCAGCCCGACCGCCGAA AACCCCTTGTGGAGCATGATCGGCTTCGCGATGTTGAGGATGTGCTGGCGGACTTCCGCGCCCGAGGATGCATTCATGAT GGCGTATGGTAGCAGCGATTAGACCGGTCGTCTAACGGCGTCGGGGCGACGTTTCGCGCGTTGCAGTTTGGTCGCTTTGC AGGTTTGCAGGTCGCGGTGTACGTTGACCTGCTCGGGCTTGTTTAGCCTCATGCCGGTTTGGGCGCGCGGCGCCGCGCGA TCGGTTTTGTGGGGCGTACGGCCGACACCGGCCGGATCGGCCCGCGACGCGCCGTCACACGCGGCGCCGACGCCCGCTCG GCACGCAGCGCGTCGAGCTCGGCCGCGCTTGCCGCCTGTGCCGAGCGTGCGCCCTCGAGCTGCCCTTCCAGGACGCCGCA GCGATGCTGGCTGTCGGCAAGCTGCGCCTGCAGCGCGCCGATTTCGCGGCGAAGTTGTGCGTCGCGTTCGTCGGCGCGCG CGTTCGTCGCGTCGAGTTCCTTTTGCAGTCGTGCGGCAGCCTGCCGCTCGCGATCGATTTCCTGCAGCGCGCGGCGCTCC GACGCGCGCAGACGCTCTTCCGCCCGGCTCGCATCGTCGCGCAGCCGATCGAGCTGCGTGCCGAAGTCGGCGCGCGCTTG TGCGAGCGCGCGAGCGCCTTCCGCGTTATCGGCCTTCAGGCGTTCGATTTCGGCTTGCTGCGCGCGGCGCGTGGCGTCGT CGGCCGCCCTCGCCTGCTCCAGTTTCCTGAATCCTCACCTGTGCCGCGGAGCAACGCGGCTGTGCGCTGCTCGAGCGCGG TTTTCGGTGCGGGGCCAGCTCGACTGCAAGGCGGCGACTTCGGGCCTGACCGCTGCCCGTCGGCTCGACTCGCCGCGCAA TTGCGTCAGCCGCGGCGGCCGCAATCGCCTGGACGGATCGGGGTTCCACCATAATGGT
Seq D	TTCTGCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAG ATCTGATCAAGAGACAGGTTCGAGCGCCAGCCCGGCACGGGCGCCGAGCGGATGATGACGTACTGGAGCAATTGGCTGCA CACGCAGTGCGCGGACGATCCCGAGGGCAAGTGCCTCGCGGTGAAGCTGGGCGCGGAAGTGTCCGACCTGTCCGAAGCGA TGCGCGCGGTACTGCGGCGCGGCACGAGCCAGATCGTCGAACGGCTGGCGGCATGCATCGAAGCCGGCCTCGCGGACGGT TCGCTGCGCGGCATCGAGGATCCGTCGCACACCGCTGCGCTGCTGTACGAGCTGTGGCTCGGTGCGACGCTGCTGGAGAA GATCCACCGCAACCGCAAGCCGCTCGAAACCGCGATGGCCGCGACGCGGCAGCTGCTGAACCTGCCGCCGTCCGCGCCCG ACGAGGTGCGGACATAAGCGGGCGCGCTACGGCGCGCCTTGTTTTTACCTATCACACTAGACGACTGGTCTACTCATGAC GATCGCGATTCCTTCCCCGTTTTTGCCCTCGCAGCATCCGCGCCCGCAAGGGCAACGGGGCGATCGGCCGCGCATCGCGG ATCACGCGTAGTCCACTCGTCCTTTCCTGATTCGATCGTGTGCGTGCACACGCATCCTTGTTGACGGAGGCTCACATGTC GCAAAGCAAGACCGCAAACCGCCGTATCGTCCTCAATTCGCGCCCGGTCGGCGCGCCCACCGCCAACGACTTCCGCCTCG AAACCGGCGACGTGCCGACGCCCGGCGCCGGGCAGGTGCTGCTGCGCACGGTATGGCTGTCGCTCGATCCGTACATGCGC GGCCGGATGAGCGACGCGCCGTCGTATGCGCCGCCCGTGCAACTCGGCGACGTGATGGTCGGCGGCACGATCAGCCGCGT CGTATCGTCGAACCTGCCGGCGTTCCGCGAGGGCGACCTCGTCGTCGCGACGGGCGCTGGCAGGATTATGCGCTGTCCGA CGGCAGCGACCTGATTCCGCTCGGCCGCGATTCCCGCATTCGTCGCGCGGCGCTCGGCGTGCCTCGCCATGCCGATCACC CGGCGTACACGCTGGCTTACCGGATTCGGCCAGGCGAGCCAGGTGAACGGT

Table S5 Plasposon flanking sequences obtained from the plasmids recovered from the E50 mutant.

E50 mutant	
Seq A	TTAGTCCCGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGCACTAGTGCGGCCGCACTTGTGTATAA GAGTCAGGCACCGCGGCCAGCGTGACGAGCAGCCAGGCGGCGCACAGCAGGCGCCGCCGCGCGCGTTCGCGCGCCGTGCC GGCAACGTCGCGCGTCACGTCCATTACCGCGCCTCCAGCCACGCGCGCTCGGCGGCGGCGGGCGCGGCGACCGGCGTCGC GCGCTCGAATTCCGACGATCGCGAAGCACGCGATCTGCTTGAGCGCGGGGAACGGCACCCACGCGAGAATCGCGTAACCG AGCAGGCTCGTCGCGAGCGCCACCGCGAGCGCCGGCCGCACCGCGGCCGCACCGCGCCGCGCGTCCCATTGCCGCCCTGC CGACAAATACGCGACGAAGTACTGGATCGAATAGTCGACGGCTTCGCCGATCAGGCTCACGCCGAACACCAGCGTCAGCA GATGCAGCTTGCCGAAGATGAACAACGTCGCGGCCAGCGCGCAGACGATGCCGAAGGCGGTCGATGCAAACGCGAGCATC AGCAGCCGCGGCGAACGGAACACGCTCATCATCAGCAGCGCGATGCCGACGAGCGACACCGCGCCGATCACGTGCACGTC GCGCTCGGCACTCGCACGCGCGGCTTCCGCATAGAACACGGCGCCCGCACGGTCGACGCGCACGCCGCGAAATTCGACCG CGAGCGCGTGCTCCGCATCGGCCGTCGCCACCCGTACCGCGTGCTGAATCTTCGATTCGTACGCCGAGCCCGGCAGCGAC GCCATCACGAGGACGCCCGTCGCGCCGCCGCGATGCGACACCAGCATGCCGTCCTCGAGCGACAGGTCGGATGCGGCGAG CGGCAACTTGCCGAGCCAGTGCTGCAGCCAGCCGAACGGATCGTCGGCCACCGCGGTCTGCAAGCCGTCGCCGGGCGGCG CGTAAAGCCGTCGCGCGAGCTGCTCGCCGAGCGTCGCATCGCGCTGCTCGAGCGATCGGCGATCCGCGGTCGGTCAACAA AACCGGAAGCGGGGCGCAGTTAGGTGGTCCGCGAATCTGCGACCAGTCGAACGCCGCACCTCGACGGTACCGGAACCGTA ACGCGGCCCGTTGCCGGGCGTTCGCCGAAACGGTTCCGTGGCGGTCCTATTCGACGCAGCACCGAGAC
Seq B	CTTTCGTGCCTTTTGCTCATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGG CCAGATCTGATCAAGAGACAGCCGCGGTGCTGTCCATCGTGTGGCGCTTTGTCAGCGGCCCGGTACCGCTGCAGACGAAT CTGCTCGCGCTGCTGCCGGCCACCGAAGCGAATCCGGTCGCGGAGCGCGCCGTCGACATGCTCGGCGAGGCGCTCGGCAA TCGCACCGTCTACCTCGTGTCCGACGCGGACGCCGCGCGCGCGAAGGCGGCCGCGAAGCGGTTCGGCGAGCGGCTCGGCG CCAGCGGCGCGTTCCGTTCGGTGACGGTCGAGGTGCCGCCGTTCGACCTGTCGCAGATCGCGGCCACCTACCTGCCGGCC CGCTTCGGTTTGTTGACCGACGCGGATCGCCGATCGCTCGAGCAGCGCGATGCGACGCTCGGCGAGCAGCTCGCGCGACG GCTTTACGCGCCGCCCGGCGACGGCTTGCAGACCGCGGTGGCCGACGATCCGTTCGGCTGGCTGCAGCACTGGCTCGGCA AGTTGCCGCTCGCCGCATCCGACCTGTCGCTCGAGGACGGCATGCTGGTGTCGCATCGCGGCGGCGCGACGGGCGTCCTC GTGATGGCGTCGCTGCCGGGCTCGGCGTACGAATCGAAGATTCAGCACGCGGTACGGGTGGCGACGGCCGATGCGGAGCA CGCGCTCGCGGTCGAATTTCGCGGCGTGCGCGTCGACCGTGCGGGCGCCGTGTTCTATGCGGAAGCCGCGCGTGCGAGTG CCGAGCGCGACGTGCACGTGATCGGCGCGGTGTCGCTCGTCGGCATCGCGCTGCTGATGATGAGCGTGTTCCGTTCGCCG CGGCTGCTGATGCTCGCGTTTTGCATCGACCGCCTTCGGCATCGTCTGCGCGCTGGCCCGCGACGTTGTTCATCTTCGGC AAGCTGCATCTGCTGACGCTTGGTGTTCGGCGTGAGCCCTGATCGGCGAAGCCGTCGACTATTTCGATCCAGTACTTTCG TCGCGGTATTTGTCCGGCAGGGCCGCCAATTGGGAACGCGGCGCCGCCGGTGCGCCCGCGTTGCGCCGGGCGCTCCGCGG ATGGCGCCTCGCGACCAGTCCTGGCTCGGTTACGGATTCTCGCGGTGGGGTGCCGGTCCCCGGGCGCTCCAAGCCCAGAT ATCTGCG
Seq C	TTAGGAACGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAG AGTCAGGCACCGCGGCCAGCGTGACGAGCAGCCAGGCGGCGCACAGCAGGCGCCGCCGCGCGCGTTCGCGCGCCGTGCCG GCAACGTCGCGCGTCACGTCCATTACCGCGCCTCCAGCCACGCGCGCTCGGCGGCGGCGGGCGCGGCGACCGGCGTCGCG CGCTCGAATTCCGACGATCGCGAAGCACGCGATCTGCTTGAGCGCGGGGAACGGCACCCACGCGAGAATCGCGTAACCGA GCAGGCTCGTCGCGAGCGCCACCGCGAGCGCCGGCCGCACCGCGGCCGCACCGCGCCGCGCGTCCCATTGCCGCCCTGCC GACAAATACGCGACGAAGTACTGGATCGAATAGTCGACGGCTTCGCCGATCAGGCTCACGCCGAACACCAGCGTCAGCAG ATGCAGCTTGCCGAAGATGAACAACGTCGCGGCCAGCGCGCAGACGATGCCGAAGGCGGTCGATGCAAACGCGAGCATCA GCAGCCGCGGCGAACGGAACACGCTCATCATCAGCAGCGCGATGCCGACGAGCGACACCGCGCCGATCACGTGCACGTCG CGCTCGGCACTCGCACGCGCGGCTTCCGCATAGAACACGGCGCCCGCACGGTCGACGCGCACGCCGCGAAATTCGACCGC GAGCGCGTGCTCCGCATCGGCCGTCGCCACCCGTACCGCGTGCTGAATCTTCGATTCGTACGCCGAGCCCGGCAGCGACG CCATCACGAGGACGCCCGTCGCGCCGCCGCGATGCGACACCAGCATGCCGTCCTCGAGCGACAGGTCGGATGCGGCGAGC GGCAACTTGCCGAGCCAGTGCTGCAGCCAGCCGAACGGATCGTCGGCCACCCGCGGTCTGCAAGCCGTCGCCCGGGCGGC GCGTAAAGCCGTCGCGCGAGCCTGCTCGCCGAGCGTCGCATCGCGCTGCTCGAGCGGATCCGCGATCCGCGTCGGTCAAA CAACCGAAGCCGGCCGGCAGTTAGGTTGGTCGCGATCTGCGGAACAGGTCGGAACCGGGGCACCTGGACCGTACCGACGG AACGCGCCGGTGGCGAAGCCCGGTGCCCGGAACCGGGTTGCGGGGGCCGGCCATTTAC
Seq D	ATGTCTGCTTTGCTAATGTTCTTTCCTGGTACCGTCGACAGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAG ATCTGATCAAGAGACAGCCGCGGTGCTGTCCATCGTGTGGCGCTTTGTCAGCGGCCCGGTACCGCTGCAGACGAATCTGC TCGCGCTGCTGCCGGCCACCGAAGCGAATCCGGTCGCGGAGCGCGCCGTCGACATGCTCGGCGAGGCGCTCGGCAATCGC ACCGTCTACCTCGTGTCCGACGCGGACGCCGCGCGCGCGAAGGCGGCCGCGAAGCGGTTCGGCGAGCGGCTCGGCGCCAG CGGCGCGTTCCGTTCGGTGACGGTCGAGGTGCCGCCGTTCGACCTGTCGCAGATCGCGGCCACCTACCTGCCGGCCCGCT TCGGTTTGTTGACCGACGCGGATCGCCGATCGCTCGAGCAGCGCGATGCGACGCTCGGCGAGCAGCTCGCGCGACGGCTT TACGCGCCGCCCGGCGACGGCTTGCAGACCGCGGTGGCCGACGATCCGTTCGGCTGGCTGCAGCACTGGCTCGGCAAGTT GCCGCTCGCCGCATCCGACCTGTCGCTCGAGGACGGCATGCTGGTGTCGCATCGCGGCGGCGCGACGGGCGTCCTCGTGA TGGCGTCGCTGCCGGGCTCGGCGTACGAATCGAAGATTCAGCACGCGGTACGGGTGGCGACGGCCGATGCGGAGCACGCG CTCGCGGTCGAATTTCGCGGCGTGCGCGTCGACCGTGCGGGCGCCGTGTTCTATGCGGAAGCCGCGCGTGCGAGTGCCGA GCGCGACGTGCACGTGATCGGCGCGGTGTCGCTCGTCGGCATCGCGCTGCTGATGATGAGCGTGTTCCGTTCGCCGCGGC TGCTGATGCTCGCGTTTGCATCGACCGCCTTCGGCATCGTCTGCGCGCTGGCCGCGACGTTGTTCATCTTCGGCAAGCTG CATCTGCTGACGCTGGTGTTCGGCGTGAGCCTGATCGGCGAAGCCGTCGACTATTCGATCCAGTACTTCGTCGCGTATTT TGTCGGCAGGGCGGCAATTGGGACCCCGCCGGCTCCTGTGCGCCCGCGGTGGCGGCACGACGCTCCGCGGTTGGCCGCTC TCCGGACAGACCTTGCCTCGGTTTAATG

Table S6 Plasposon flanking sequences obtained from the plasmids recovered from the E70 mutant.

E70 mutant	
Seq A	TTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGACTTCGAACTGCGCGCGGATTTCCTGCTCGAGGCGC GGCTGCAGCGGCGCATCGACGATCCACTTGCGAATTCCATCTTGATCAGCTTCGAGCCGATCGAGCGGCGGATGATCGGG TACTTGTCCTGCGCGAGCGTGGTCTTGAACACGTAGAACTCGTCCGGGTTCACGGCGCCCTGCACGACCGTTTCGCCGAG ACCGTAGCTCGACGTGATGAACACGGCGTCCTTGAAGCCCGACTCGGTGTCGATCGTGAACATCACGCCGGCCGCACCGA CGTCCGAACGGACCATGCGCTGCACGCCGGCCGACAGCGCGACTTCGGCATGCGTGAAGCCCTTGTGCACGCGATACGAG ATCGCGCGGTCGTTGTACAGCGACGCGAACACGTGCTTCATGCGGTCGAGCACGTCCTCGATGCCGACGACGTTCAGATA CGACTCCTGCTGACCGGCGAACGAGGCGTCGGGCAGATCTTCGGCCGTGGCGGACGAACGCACGGCGAACGACAGCTCGC CCGGCGAGCCGCTTTGCAGCACTTCGAACCTGTCTCTTGATCAGATCTGGCCGCCTAGGCCGGCCGCGATCGCCGGCGCG CCATGCATGTCGACGGTACCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGA CAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCG
Seq B	CCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAGATCTGATCAAGAGACAGGTTCGAAGTGCTG CAAAGCGGCTCGCCGGGCGAGCTGTCGTTCGCCGTGCGTTCGTCCGCCACGGCCGAAGATCTGCCCGACGCCTCGTTCGC CGGTCAGCAGGAGTCGTATCTGAACGTCGTCGGCATCGAGGACGTGCTCGACCGCATGAAGCACGTGTTCGCGTCGCTGT ACAACGACCGCGCGATCTCGTATCGCGTGCACAAGGGCTTCACGCATGCCGAAGTCGCGCTGTCGGCCGGCGTGCAGCGC ATGGTCCGTTCGGACGTCGGTGCGGCCGGCGTGATGTTCACGATCGACACCGAGTCGGGCTTCAAGGACGCCGTGTTCAT CACGTCGAGCTACGGTCTCGGCGAAACGGTCGTGCAGGGCGCCGTGAACCCGGACGAGTTCTACGTGTTCAAGACCACGC TCGCGCAGGACAAGTACCCGATCATCCGCCGCTCGATCGGCTCGAAGCTGATCAAGATGGAATTCGCAAGTGGATCGTCG ATGCGCCGCTGCAGCCGCGCCTCGAGCAGGAAATCCGCGCGCAGTTCGAAGTCTGACTCTTATACACAAGTGCGGCCGCA CTAGTCTAGATTTAAATTAATTAAGAGCTCGGGGGGGGGGGGGAAAGCCACGTTGTGTCTCAAAATCTCTGATGTTACAT TGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCC ATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTTCCAACATGGATGCTGATTTATATGGGTATAAATGGGC TC
Seq C	GTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGAC TTCGAACTGCGCGCGGATTTCCTGCTCGAGGCGCGGCTGCAGCGGCGCATCGACGATCCACTTGCGAATTCCATCTTGAT CAGCTTCGAGCCGATCGAGCGGCGGATGATCGGGTACTTGTCCTGCGCGAGCGTGGTCTTGAACACGTAGAACTCGTCCG GGTTCACGGCGCCCTGCACGACCGTTTCGCCGAGACCGTAGCTCGACGTGATGAACACGGCGTCCTTGAAGCCCGACTCG GTGTCGATCGTGAACATCACGCCGGCCGCACCGACGTCCGAACGGACCATGCGCTGCACGCCGGCCGACAGCGCGACTTC GGCATGCGTGAAGCCCTTGTGCACGCGATACGAGATCGCGCGGTCGTTGTACAGCGACGCGAACACGTGCTTCATGCGGT CGAGCACGTCCTCGATGCCGACGACGTTCAGATACGACTCCTGCTGACCGGCGAACGAGGCGTCGGGCAGATCTTCGGCC GTGGCGGACGAACGCACGGCGAACGACAGCTCGCCCGGCGAGCCGCTTTGCAGCACTTCGAACCTGTCTCTTGATCAGAT CTGGCCGCCTAGGCCGGCCGCGATCGCCGGCGCGCCATGCATGTCGACGGTACCAGGAAAGAACATGTGAGCAAAAGGCC AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAGCGTGGCGCTTTCTCATAG CTCA
Seq D	ATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAGATCTGATCAAGAGACAGGTTCGAAGTGCTGCAAAGCGG CTCGCCGGGCGAGCTGTCGTTCGCCGTGCGTTCGTCCGCCACGGCCGAAGATCTGCCCGACGCCTCGTTCGCCGGTCAGC AGGAGTCGTATCTGAACGTCGTCGGCATCGAGGACGTGCTCGACCGCATGAAGCACGTGTTCGCGTCGCTGTACAACGAC CGCGCGATCTCGTATCGCGTGCACAAGGGCTTCACGCATGCCGAAGTCGCGCTGTCGGCCGGCGTGCAGCGCATGGTCCG TTCGGACGTCGGTGCGGCCGGCGTGATGTTCACGATCGACACCGAGTCGGGCTTCAAGGACGCCGTGTTCATCACGTCGA GCTACGGTCTCGGCGAAACGGTCGTGCAGGGCGCCGTGAACCCGGACGAGTTCTACGTGTTCAAGACCACGCTCGCGCAG GACAAGTACCCGATCATCCGCCGCTCGATCGGCTCGAAGCTGATCAAGATGGAATTCGCAAGTGGATCGTCGATGCGCCG CTGCAGCCGCGCCTCGAGCAGGAAATCCGCGCGCAGTTCGAAGTCTGACTCTTATACACAAGTGCGGCCGCACTAGTCTA GATTTAAATTAATTAAGAGCTCGGGGGGGGGGGGGGAAAGCCACGTTGTGTCTCAAAATCTCTGATGTTACATTGCACAAG ATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTTATGAGCCATATTCA ACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATA ATGTCGGGCAATCAGGTGCGACAATCTATCCGATTTGTATGGGAAGCCCGATGCGCCCAGAGTTTGTTTCTTGAACCATG GCAAAGGTAGGCGTTGCCCAATG

Table S7 Plasposon flanking sequences obtained from the plasmids recovered from the E89 mutant.

E89 mutant	
Seq A	GGTCCCGAGAAAACATGTACCTGGCTTTCCCCCCCCCCCCGGCTCTTAATAATTTAATCTAGACTAGTGCGGCCGCACTT GTGTATAAGAGTCAGAATTCGATGCGCCGGGCACGCGCATCACGTACTCGCGGCCGGCCGGCACATGCGTGAAATAGCCG AGCGAATTCGGTCCTGTCTCTTGATCAGATCTGGCCGCCTAGGCCGGCCGCGATCGCCGGCGCGCCATGCATGTCGACGG TACCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCAT AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATA CCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTA CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAATCGATGAG CTCGGGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGTGTTGCTGACTCATACCAGCCTGATCGCCCATCATCCAGC CAGAAGGTGAGGGAGGCACGTTTGATGGAGAGCCTTGTTGTAGTGACCAGTGGTGAATTGGACCTTTTGCCTTGCAGGAC GTCGCTGCGAGATGCTGATCTGATCCTTTCACTCCTCAGACAAAGAGGTTTTCG
Seq B	TTCGTGCCCTTTTGTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCC AGATCTGATCAAGAGACAGGACCGAATTCGCTCGGCTATTTCACGCATGTGCCGGCCGGCCGCGAGTACGTGATGCGCGT GCCCGGCGCATCGAATTCTGACTCTTATACACAAGTGCGGCCGCACTAGTCTAGATTTAAATTAATTAAGAGCTCGGGGG GGGGGGGGAAAGCCACGTTGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAA AACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGA TTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTA TCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATG AGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCA TGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGT TGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTC GTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGGCTGGCC TGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGGTGATTTCTCA CTTGATAACCCTTATTTTTGACGAGGGGGAAATTAATAGGTTGTATTGATGTTGGGACGAGTCGGAAATCGCAGAGCGAT ACCAGGATCTGCCATCCTAATGGGAACTGCCTCGGTGGAGTTTTCTCTTTCATTACAGAAACGGGCTTTTTCAAAAATAT GGTATTGGATAAATTCCTGAATTGATTAAATTGGCAGTTCATTTGGATGTTCGATGAGTTTTCTATTCAGATTGGTTAAT TTGGGTGTGAAAACCTGGGGCACGAGAC

Table S8 Plasposon flanking sequences obtained from the plasmids recovered from the F70 mutant.

F70 mutant	
Seq A	CTAGAGGCATGCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGA CGAATGCCTGGCGGTTGAAGATCGTGTTTGCTGCGCTCAGCTTGCCCGTGCCCGGGTTGAAGCCGTTTTCGATCTGGAAG ATTGCCTTCAGGCCGCCGCCCAGGTCTTCAGCACCCTTCAGGCCCCAGCGGCTGCCTTGCAGGTTGCCCGACAGCATTTG CCACATGTTGTTGGCTTGGCCGTCATTGCCATGAACATAGGCAATCGACGTATCGATCACGCCGTACAGCGTCACGCTGG ATTGAGCCTGTGCAGCGCCGGCGGCGCCCAGCAGGGCGAGCGAGAGGGCCGAGAGAGCAAATTTCTTCATCGTTGAGTAT CTCCACGCAAGTGATTCGTGATTCGTTATTTGTTGCGGATTGGAGAATAGCTCAGTGCCATACAGGCAAGAAAGACTAAA AAATAAAGTGTCTCGAAAAGATCACACCAACGAAAAGTCCATATATTTCAAGCACATCGCGAACTGTTCCTGTTTTTGAA ATAGTTGACATTTGCTGCGCTGCGATTGTTGCGGAGGATGGCGCCCACCACAACATGACGCGTGTTTTGGCCGGGCCGCG CACTGAAATCGGGGTGTGCAAACGTTTGCGATTGCAGCGTTGCGCATGCAAGTGCCGCGCCAGCGCGACGTGCCTATCGG CTGAACAAAAGGGATCGCGGAAAAGGGAAGTTCGACGTGAGGAGCGCGACCGCTTGCACGACCGCGCGTCGCCGGCGTAA TCGCGTGCCGGCCGCTACGCCGGAGCCCCGGTCAATCGTCGCGGGTGCGCTCGCCCGGCGGGCGGGCGGCTCGCGTATCG TGCGCCGTTGCATCGGGCGCCGCCGCGTCGCGCGCCGCCGCATCCGTGCGCCGATATGCCTCGCCGCGCGCGGCCGAGAA TTCGACCATCCGCGTGCCCGCTGGTACCTTGCACTGCGGTGACGCGCACGACGCGCGAGTCGGGGAAAATCGATGCGTAG AACGCCGCGGCTTCTTCGCTTCGGTCGAATACCAGAAGGAACGGCGTGATCTCTGAATCGTCATCGCGAAATCCTCCCTT TGTCAAGTTGGGGTCGACGCAAAACCCGCCGGCCGCCCTCCATCCGACGGACGTCCGGCCGAACCCGTTCCGAATTCCGC ACATTGCCCGCGCAACGCGGAGAGAGCCACAACTCCCT
Seq B	ATAAAGGGGAAGGGCCGCCAAGGGGGTGCTGTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGAT CGCGGCCGGCCTAGGCGGCCAGATCTGATCAAGAGACAGGGCATTCGTCGGTCTGCAAAGCAATCAGTACGGTACGCTGA CGCTCGGCCGCCAGTACGACCCGGTCGTCGACCTTGTCCAGGCAGTCACGGCTGACAACTACTTCGGCAGCTTCTTCGCG ACCCCGGGTGACGTCGACAACAACGACAACAGCCTGCGCGTCAGCAACGCGATCAAGTACACGTCGCCCGTCTACGCCGG CTTCCAGTTCGAAGGCATGTACGGCCTGAGCGGCATCGCAGGCAAGCCGGGCCAGGGTCAGACGTGGTCGGCTGCTGCGG CCTACAACAACGGCCCGATCGGCATCGCCGCCGGCTACTTCTACGCGAACAACCCGTCGCCGACGGCCAGCGCGGTCCGT GCCGGCTGGGGCTCGACGACGTCGGACAACATCGTCGACGGCCCGATCAACGCAGGCTACGTGACGGCGAAGTCGATCGG CATCGCGCAAGTCGCAGCTCAGTACGCGATCGGCCCGGTGACGTTCGGCCTCGGCTACAGCAACGCGCAATACAAGCCGG ACGCTTACTCGGCGTTCTCGTCGACCGAGAAGTACAACACGGGCCGCGGCTTCGTGACGTACCAGGTCACGGCGCCGCTG CTGCTGGGTCTCGGCTACTCGTACACGAAGGCAAGCGGCAACACGGACGCGAAGTACCACCAGGTTTCGCTGGGCGCAGA CTACTCGCTGTCGAAGCGCACGGACGTCTACCTGGTCGGCGCGTACCAGCACGCCAGCGGCACGCAGCTCAACGCGGACG GCACGACGTCGGCAGCGCAAGCATCGATCGGCTCGTACGGCTATGCCGGCACGAAGTCGCAGAAATGGTTGCCCTCGGCC TGGGCCCAGTTCTAAAACGACGTATCCGGCGTACAGCGCGTCTCGACGCGCTGCCGCCGACTGAAAGCAGTCACCGGTTT CGGCCGCGACTGCTTTTTTTCGTCGTCGGCGCTCGGGCGTCGATCCGCCGCCGGGTGCGCTGACGACGTCGGGGGTCGAT CGCACTGTCCGCGCGATCATGGGAGATTGCTCGGTGGCGCATTTCGATTTGACGGTCCGGCGACCGTGCGTTCCGATGGA GGCGCCGCGATC

Table S9 Plasposon flanking sequences obtained from the plasmids recovered from the G2 mutant.

G2 mutant	
Seq A	CTAGTGCGGCCGCACTTGTGTATAAGAGTCAGCCTCCATCTTGTTGCGGTAGTGGCCGTACACGGCGCCGCGCGACATGC CGGCGGCCTCGGCGAGGTCCGCCATCGCGGTTTGCGCGACGCCCTTGTCGAGCAGCACGAGCTCGGCGGCGTCGAGGATC CGGTGCTTGATGGCGAGCGATTCTTCCCGGGTCTTGCGAGCCATGTATCGAGTTTCCTTACAGTACGCTGTCGCTTTATT GAGACAGTCTGACGATTGGATGGAGGATCGATGTATTTAATCAGTCGTGACTGATTATAATCGGCGACCCTGAGCCGCCG CATTGTATCGGCGGCGCACGATCCGAGGTCCCAGATGAATGACAAACGCTCGCTGTCGCGCCGGCTGACGCCGCTCGCGC TGGCCGTCGCCGTGGCCGCGGCCGGCTGCGGAAAGGGCAAACCGGATGCCGAGCCGGCGGCCGTGCAGAAGGCGACCGTC GTCACGGTGCGTCCGGTCGCGGTGCCGATGACCGTCGAGCTGCCGGGCCGGCTCGATGCCTATCGACGCGCCGAAGTGCG CGCGCGCGTCGCCGGCATCGTCACGGCACGCACGTACGACGAAGGCCAGGAAGTGAAGCAGGGCGCCGTGCTGTTTCGCA TCGATCCCGCGCCGTTGAAAGCGGCACGCGATGCGGCGCAGGGCGCGCTCGCCAAGGCGCAGGCCGCGGCGGCCGCGGCC GCCGACAAGCGGCGGCGCTATGCCGATCTCGTGCGCGATCACGCCGTCAGCGAGCGCGACTACACGGGAGGCCGTCGCGG CCGACACGCAGGCGAAAGCGGAAGTCGCGTCGGCGAAAGCCGAACTCGCGCGTGCGCAGCTGCAGCTCGACTACGCGAAC GTGACCGCGCCGATTTCCGGGCGGGCGCGGCGCGCGCTCGTGACCGAAGG
Seq B	GCCGGCGATCGCGGCCGGCCTAGGCGGCCGATCTGATCAAGCAGACAGAGATGGAGGTGTGCCTCGCGATGTGCGACCGC GCGTTCGCGCGCACGTCGGAGGGCTTCGAGGCCGGCGACGGGCTGCCGGCGTTCGCGACGCTGCGGCACGCCGCGTCGCA CTATCTGCGCCAGTGCGGCGAGCCCGGTTCGATGCAGCGCGTGCTCGTGATCCTCTACACGAAGTGCGAGCAAAGCGACG AGAACGCGGCGCTGCTGCGTCGCCGGATGCTGCTCGAACTGCAGACGCTGCGGATCACGAAGGCGCTGCTGCGCCGCGCG ATCGAGGCGGGCGAACTCGCGGCCGACCTCGACGTGCATCTGGCCGCCGTCTATCTCGTCTCGCTGCTCGAAGGCGTGTT CGCGTCGATGATCTGGACGAACCGGCTGCGCGGCAATCTGTGGAACGACGCGGAAGCGATGCTCGATGCGGGCTTCGACG CAGTGCGGACCTCCGCCGCGCTGCGGCGGCCCGCACACAAATTGCCGAAGCCCGCGAAACGATCGCAATAAATGCCGATT TAACCCGATTTACCGCACTGCGAAACGAGTGAGTCGACCCTCTTTAAAATTTTTAACATCTTCCGGAACATCGGTAGACT CGCGCTGTCAGCTCCTCTTTAGCGTCTTCGTGATAACCGGGTTCGACCCGGCATGCACGCCGCCGTTCGAAGCGGGTTTG CCCGCGCGTTGGGTCGAACGGAAACGACACAGGCCCCCTGGCGGGGCCTGTTTTGTTGCACGCCCCGCCGCCCGCCCCGT CGCCGTTTCGACGGATGAGCCGCCGGCGGTCCCGCCCCACACCCCCGTGAAACCGCTTTTTCTGGGATGTGGGACCCTTG GTCAAGTTAGCC
Seq C	ACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGCCTCCATCTTGTTGCGGTAGTGGCCGTACACGGCGCCGCGCGACATG CCGGCGGCCTCGGCGAGGTCCGCCATCGCGGTTTGCGCGACGCCCTTGTCGAGCAGCACGAGCTCGGCGGCGTCGAGGAT CCGGTGCTTGATGGCGAGCGATTCTTCCCGGGTCTTGCGAGCCATGTATCGAGTTTCCTTACAGTACGCTGTCGCTTTAT TGAGACAGTCTGACGATTGGATGGAGGATCGATGTATTTAATCAGTCGTGACTGATTATAATCGGCGACCCTGAGCCGCC GCATTGTATCGGCGGCGCACGATCCGAGGTCCCAGATGAATGACAAACGCTCGCTGTCGCGCCGGCTGACGCCGCTCGCG CTGGCCGTCGCCGTGGCCGCGGCCGGCTGCGGAAAGGGCAAACCGGATGCCGAGCCGGCGGCCGTGCAGAAGGCGACCGT CGTCACGGTGCGTCCGGTCGCGGTGCCGATGACCGTCGAGCTGCCGGGCCGGCTCGATGCCTATCGACGCGCCGAAGTGC GCGCGCGCGTCGCCGGCATCGTCACGGCACGCACGTACGACGAAGGCCAGGAAGTGAAGCAGGGCGCCGTGCTGTTTCGC ATCGATCCCGCGCCGTTGAAAGCGGCACGCGATGCGGCGCAGGCGCGCTCGCCAAGGCGCAGGCCGCGGCGGCCGCGGCC GCCGACAAGCGGCGGCGCTATGCCGATCTCGTGCGCGATCACGCCGTCAGCGAGCGCGACTACACGGAGGCCGTCGCGGC CCGACACGCAGGCGAAGGCGGAGGTCGCGTCGGCGAAAGGCCGAACTCGCGCGTGCGCAGCTGCAGCCTCGACTACGCGA ACGTGACCGCGCCCGATTTCC
Seq D	ATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCAGATCTGATCAAGAGACAGAGATGGAGGTGTGCCTCGCGATGTG CGACCGCGCGTTCGCGCGCACGTCGGAGGGCTTCGAGGCCGGCGACGGGCTGCCGGCGTTCGCGACGCTGCGGCACGCCG CGTCGCACTATCTGCGCCAGTGCGGCGAGCCCGGTTCGATGCAGCGCGTGCTCGTGATCCTCTACACGAAGTGCGAGCAA AGCGACGAGAACGCGGCGCTGCTGCGTCGCCGGATGCTGCTCGAACTGCAGACGCTGCGGATCACGAAGGCGCTGCTGCG CCGCGCGATCGAGGCGGGCGAACTCGCGGCCGACCTCGACGTGCATCTGGCCGCCGTCTATCTCGTCTCGCTGCTCGAAG GCGTGTTCGCGTCGATGATCTGGACGAACCGGCTGCGCGGCAATCTGTGGAACGACGCGGAAGCGATGCTCGATGCGGGC TTCGACGCAGTGCGGACCTCCGCCGCGCTGCGGCGGCCCGCACACAAATTGCCGAAGCCCGCGAAACGATCGCAATAAAT GCCGATTTAACCCGATTTACCGCACTGCGAAACGAGTGAGTCGACCCTCTTTAAAATTTTTAACATCTTCCGGAACATCG GTAGACTCGCGCTGTCAGCTCTCTTTAGCGTCTTCGTGATAACCGGGTTCGACCCGGCATGCACGCCGCCGTTCGAAGCG GGTTTGCCCGCGCGTTGGGTCGAACGGGAAACGACACAGGCCCCCTGGCGGGGCCTGTTTTGTTGCACGCCCGCCGCCCG CCCCGTCGCCG

Table S10 Plasposon flanking sequences obtained from the plasmids recovered from the G38 mutant.

G38 mutant	
Seq A	TTGACGCCGTGGCTTTCCCCCCCCCCCCGGCTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGA GTCAGGCGCGAAGTGATGCGTTTCGCGCGCCGAGCCCGTCTCGACGACGAGGCCCGTGCTCGTGCGCAGCCCTTGCTGGA CGAGGTGCGTGTGGATCGCCGAGGTAGCGAGCAGCGCCGGAATCGCGACATGCTCGGCGTCGGTCTTGCGGTCCGACACG ATCAGGATGTTGTAGCCGGACTTCACGGCGTCGACGGCTTCCGCGCACAGCGACGCGAGACGCGCCTCGATGCCTTCCTT GCCCCACGCGACCGGGTAGCAGATGTTCAGTTCATACGCGCTGAACTTGCCGCCCGTGTATTGATCGATCGCGCGGATCT TCGCGATGTCCTTGAAGTCGAGCACCGGCTGCGACACTTCGAGACGCATCGGCGGGTTGATGTTGTTGGTGTCGAGCAGG TTCGGCTTCGGGCCGATGAACGACACGAGCGACATCACCATGTTCTCGCGGATCGGGTCGATCGGCGGGTTCGTGACCTG CGCGAACAGCTGCTTGAAGTAGTGATAGAGCGTCTTGTTCTTGTTCGACATCACCGCGAGCGGCGAGTCGTTGCCCATCG AGCCGACCGCTTCCTCGCCCTGCTGCGCCATCGGCGCCATCAGAAACTTCAGGTCTTCCTGCGTGTAGCCGAACGCCTGC TGGCGGTCGAGCAGCGACGCGCCCTGCGCGCGGCCGGCCGCGACTTCCTCGGCCTTCGGCTCGATCTCGTCGAGCTTGAT GCGCACGGCGTCGATCCAGCTCTTGTACGGCTTCGCGTTCGCGAGGTTGTCCTTGAGCTCCTTGTCGTCGATGATGCGGG CCGTGTTCCATGTCGATCAGGAACATCTTGCCCGGGCTGCAGACGCCACTTTCTTGACGATCTTCGATTCCGGGAATCGG CAGCGTGCCGGCTTCCGGACGCCCATGATCACGAAGGTCCTTCGTCGGTGAC
Seq B	TTCTGCCTTTTGCTAATGTTCTTTCCTGGTACCGTCGTCATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCA GATCTGATCAAGAGACAGACTTCGCGCTGCTCGCCGGCTACGGCGCGGAAGCCGTGCATCCGTACCTCGCGATGGAGACG CTTGCGAAGATGGCCGAAGGGCTGTCGGGCGACCTGTCGCCGGAGAAGGCCATCTACAACTTCACGAAGGCGATCGGCAA GGGCCTGCAGAAAGTGATGTCGAAGATGGGCATCTCGACGTACATGTCGTACACGGGCGCGCAGATCTTCGAAGCGCTCG GCCTGTCGAGCGACCTCGTCGAGAAGTACTTCAAGGGCACCGCATCGAAGGTCGGCGGGATCGGCCTGTTCGAAGTCGCG GAAGAGGCGATCCGCCTGCACCGCGACGCGTTCGGCGACAACCCGGTCCTGCGCGACATGCTCGACGCGGGCGGCGAGTA CGCGTACCGCGTGCGCGGCGAAGACCACATGTGGACGCCCGACTCGATCGCGAAGCTGCAGCACGCGACGCGCAGCAACT CGTACCAGACGTACAAGGAATACGCGCATCTGATCAACGATCAGACCAAGCGTCACATGACGTTCCGCGGGCTGTTCGAG TTCAAGGTCCAGCCGACCAAGGCGATCCCGATCGACGACGTCGAACCGGCGAAGGAAATCGTCAAGCGCTTCGCGACGGG CGCGATGTCGCTCGGCTCGATCAGCACCGAAGCGCACGCGACGCTCGCGATCGCGATGAACCGGATCGGCGGCAAGTCGA ACACCGGCGAAAGCGGCGAAGACGAGAACCGCTATCGCAACGAGCTGCGCGGCATTCCGATCAAGGCCGGCGAGACGCTG AAGTCGGTGATCGGCGACGAAGTCGTCTCCGACATTCCGCTGAAAGGAAGGCGATTCGCTGCGCTCGAAGATCAAGCAGG TCGCGTCGGGCCCGCTTCGGCGTGACGGCCGAGTATCTCGCGTCGGCCGACCAGATCCAGATCAAGATGGCGCAGGGCGC GGAAGCCCGGGCGAAAGGCGGCCAGCTGCCCGGGCCACAAGGTGTCCCGAGTACATCGGCAAGGCTGGCGTACTCGGTGC CCGGGTGGTCCGGGCCCTGATCCTCGCGGCCCGGCCGGCAAC
Seq C	TTGAACCGTGGCTTTCCCCCCCCCCCCGGCTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAG TCAGGCGCGAAGTGATGCGTTTCGCGCGCCGAGCCCGTCTCGACGACGAGGCCCGTGCTCGTGCGCAGCCCTTGCTGGAC GAGGTGCGTGTGGATCGCCGAGGTAGCGAGCAGCGCCGGAATCGCGACATGCTCGGCGTCGGTCTTGCGGTCCGACACGA TCAGGATGTTGTAGCCGGACTTCACGGCGTCGACGGCTTCCGCGCACAGCGACGCGAGACGCGCCTCGATGCCTTCCTTG CCCCACGCGACCGGGTAGCAGATGTTCAGTTCATACGCGCTGAACTTGCCGCCCGTGTATTGATCGATCGCGCGGATCTT CGCGATGTCCTTGAAGTCGAGCACCGGCTGCGACACTTCGAGACGCATCGGCGGGTTGATGTTGTTGGTGTCGAGCAGGT TCGGCTTCGGGCCGATGAACGACACGAGCGACATCACCATGTTCTCGCGGATCGGGTCGATCGGCGGGTTCGTGACCTGC GCGAACAGCTGCTTGAAGTAGTGATAGAGCGTCTTGTTCTTGTTCGACATCACCGCGAGCGGCGAGTCGTTGCCCATCGA GCCGACCGCTTCCTCGCCCTGCTGCGCCATCGGCGCCATCAGAAACTTCAGGTCTTCCTGCGTGTAGCCGAACGCCTGCT GGCGGTCGAGCAGCGACGCGCCCTGCGCGCGGCCGGCCGCGACTTCCTCGGCCTTCGGCTCGATCTCGTCGAGCTTGATG CGCACGGCGTCGATCCAGCTCTTGTACGGCTTCGCGTTCGCGAGGTTGTCCTTGAGCTCCTTGTCGTCGATGATGCGGCC GTGTTCCATGTCGATCAGGAACATCTTGCCCGGCTGCAGACGCCACTTCTTGACGATCTTTCGATTCCCGGAATCGGCAG CGTGCCGGGCTTTCCGACGCCATGATCACGAGTCGTCGTCGTGACGATGTAGCGCGCCCGGACGCAGCCCGTTTACGTCG AGCGTCGCGCCCGATCTGGCGACCGTCGTGAACGCGATCGCGCCGGCGTTCCACGCTTCATCATCGCGGCCGTGTATTTC GTAGACCGCCGCCGGGCGGGACTC
Seq D	TTCTGCCTTTGCTAATGTTCTTTCCTGGTACCGTCGAATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCGAT CTGATCAAGAGACAGACTTCGCGCTGCTCGCCGGCTACGGCGCGGAAGCCGTGCATCCGTACCTCGCGATGGAGACGCTT GCGAAGATGGCCGAAGGGCTGTCGGGCGACCTGTCGCCGGAGAAGGCCATCTACAACTTCACGAAGGCGATCGGCAAGGG CCTGCAGAAAGTGATGTCGAAGATGGGCATCTCGACGTACATGTCGTACACGGGCGCGCAGATCTTCGAAGCGCTCGGCC TGTCGAGCGACCTCGTCGAGAAGTACTTCAAGGGCACCGCATCGAAGGTCGGCGGGATCGGCCTGTTCGAAGTCGCGGAA GAGGCGATCCGCCTGCACCGCGACGCGTTCGGCGACAACCCGGTCCTGCGCGACATGCTCGACGCGGGCGGCGAGTACGC GTACCGCGTGCGCGGCGAAGACCACATGTGGACGCCCGACTCGATCGCGAAGCTGCAGCACGCGACGCGCAGCAACTCGT ACCAGACGTACAAGGAATACGCGCATCTGATCAACGATCAGACCAAGCGTCACATGACGTTCCGCGGGCTGTTCGAGTTC AAGGTCCAGCCGACCAAGGCGATCCCGATCGACGACGTCGAACCGGCGAAGGAAATCGTCAAGCGCTTCGCGACGGGCGC GATGTCGCTCGGCTCGATCAGCACCGAAGCGCACGCGACGCTCGCGATCGCGATGAACCGGATCGGCGGCAAGTCGAACA CCGGCGAAGGCGGCGAAGACGAGAACCGCTATCGCAACGAGCTGCGCGGCATTCCGATCAAGGCCGGCGAGACGCTGAAG TCGGTGATCGGCGACGAAGTCGTCTCCGACATTCCGCTGAAGGAAGGCGATTCGCTGCGCTCGAAGATCAAGCAGGTCGC GTCGGGCCGCTTCGGCGTGACGGCCGGAGTATCTCGCGGTCGGCCGGACCAGATCCAGATCAGATGGGCGCAAGGGCGCG GAAAGCCCGGGCGAAAGGCCGGCCAGCTGCCGGGGGCCACAAGGGTGGTCCGAGGTACATTCGGCAAGGCCTGCGTTACT CCGTGCCGGGGTGTTCCGGCCTG

Table S11 Plasposon flanking sequences obtained from the plasmids recovered from the H21 mutant.

H21 mutant	
Seq A	ATCGTGCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCCA GATCTGATCAAGAGACAGGTTCTACGCGGCGTGGAAGAAATACCGGCTGCCGTTCCACTACGAGCTCGATCGCGACGAGC GCTATCTGCCGCTGCTGCTAGCGATCGCGGGCGTGCCGAGCGACGACGTGCGCGACAGTCTCGCGGCAGGCGCGGGCGGC GTGCTCGACGAAGCCGTCGCCGGCTATGCGCTCGCCGCGCGGCACCGGCCGATGTCGGCCGCCTATCTGCAGCGTACGCT GTCCGATTATTTCCGCGTGCCGGTGAAGATCGATCAGTTCGTCGGCAAGTGGTACGACGTGCCGCCCGATCAGCTGAGCG TGCTCGGCCAGGTCAACGCGGTGCTCGGCGCGACGGCGCTGGTCGGCGAACGCGTGTGGCAGCGCGACATGCGCGCCCGG ATCGTGATCGGCCCGCTGTCCAAGCGCGACTACGAGGCGTTCCTGCCGGGCGGCGCGCAGGCCGTTGCGCTCGAGCGGAT GTTGACGCTGCTCGCCGGCGTCACGCTCGAATACGAAGTGAAGCTCGTGCTGAAGCGGACCGAAGTCGGCGCGAGCGTAC TCGGCGCCGGCTCGCGGCTCGGCTGGGACGCGTTCCTCTGCACGCGCGACGCCGACGAAGACCGGTCCGATGCGCGCTAC GAGCTGCACGTGATTCACTGATACCGAACGACAACAAACACGCAAACGAACCTGAGATCGACGCCATGAGCACGCCTCTG AAGACCCTCATCACGAAACTGAATCCGCTGTGCCGGCATGCGGCCGAGCGCGCGGCGAGCGCGTGTCTTGCGCGCGGCCA TTACGAGGTCGATCTGGAGCATCTGTTTCTCGCGCTGCTCGAGGAGCCGGCGGGCGACCTGTCGCTCGCGCTGCGCGCAA GCCGCATCGACGCGCATGCGTTGCGCGCCCGATCTCGAGCGCGAGCTCACGCGGCCCTGAAGGACCCGGCAACACGCGCA ACGCCCGTAATTCTCGGTGGCAATCTGATCGCGCTTGTTCGAAGCAAGGCCCTGGCCTGGATCGCGTCGCTCGAACTTCG CAGCTCGGACCCGCAATCGCTTCGGGGCCAACTGGCTGGCTCGCGCTTGCTTAACCGCGCCGGAACCTGCACAAGTCGGC GCGACAGCCGGGAATAGGATGTCGGGGGGGCGCCT
Seq B	ATCAGCGCCGTGGCTTTCCCCCCCCCCCCGACTCTTAATTTATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAG AGTCAGAATTCGTTCGAGAACAGCAGGTCGTCGCCGGTGGGCGACGCGTCGCTGATCGGCGTCAGCAGCTCGGGCAGATT GATCGGCATGGTTCAGTAAATCCGTTGATGCATCGAAAGGGTCGCGCTCATTCGACCGTGTATGCGAACTCGCCGGCGTC GTCCGCGCGCACCGCGATGCGCGCAATCGACGCGCCGTCGGCGATCCGGCCCAGCACGTGGCCCGCGATCTCCGGCAGCA GCGTGCCGTTCAGGATGTGGTCGACGTTGCGTGCGCCCGAGTCGACTTCGGTGCAGCGTGCGAGGACCGCTTCGACGAGC GATTCGTCCCATTCGAACGCGGCCTTGTGGTTCGCCTCGATGCGGCTGCGAATCCGTTCGAGCTTCAGCTCGATGATCTC GGCGAGCACGTCGTCGGAGATCGGGTAGTACGGCACCACCTTCATCCGGCCGAGGAACGCGGGCTTGAACGCCTTGTACA GCTGCGGGCGCAGCGCTTCCGCGAGCGCATCGGGATCGGGCAGTTCCTCGGCCGGCTTGTTCAGGCACGCCTGCATCACC GCGGCCGAACCGACGTTCGACGTCAGGATGATCAGCGTGTTGCGGAAGTCGATCTCGCGCCCTTGGGCGTCGTCCATCAT GCCCTTGTCGAAAACCTGGGAAGAACATCTCGAGGGACGTCCCGGATGCGCTTTCTCGACCTCGTCGAGCAGCACGACCG AATACGGATTTGGGCCGCACGGCCTCGGTCAG
Seq C	TAATCCCGGTGGCCAACCGTGGCTTTCCCCCCCCCCCCCGACTCTTAATGAAATAAAATCTAGATAGTGCGGCCGCACTT GTGTATAATAGTAGGCGTAGAACAGCGCGGTCGCGCGGTTCGAGAACACGTCGAAGAATGCACGCGCCGCATGGTCGCGC TTCAGGTGCTCGCGCGCCGCGATCTGTTCGGTGTAGTGCAGCGGCAGCGCGCCCTGCGCGCCGAGCAGCCCGAAGAACGC CGGCGTCAGTTCGACGCGCCCGAGCTCGCCGCCCGCGACCGCCGCTTCGCGCTGCTCGGCCGAGTCGAGCGGCAGGCCGT CGTCGTCGAACGAGCGGGCGCCTTCGATCTCGCTCGGCGGAAAGCCGAGCGACAGCGTATTGCGGAACGCGATCCGGTGC GCGACGACGTCGCCTTGCCGCCACGCGCCGGGCGCATCGGCGGCCTGCCGCGCGAACAGCCCTTCGAGCACGCGTACCGC CTGGAAGAATTCTACTGTTTGGGTGTATGAGCTATTTGCAAAGGAAATTGGTGATGACAAAGCTCGGCGCTATTTGAAGA AAATCGACTATGGCAACGCCGATCCTTTTTGTCCGGTGTTGGGTTGAAGGTGAAGCCGGTCGGGGCCGCAGCGGGGGCCG GCTTTTCAGCCTTGCCCCCCTGCTTCGGCCGCCGTGGCTCCGGCGTGTTGGGTGCCGGCGCGGGTTCCGCAGCCTTGGCC TGCGGTGCGGGCACATCGGCGGGCTTGGCCTTGATGTGGCCGGCTGGCGTGCGGAGCGGAACGGTCTCGTAGGAGAACTT GACCTTTCCCGTTTTCCGCATGTGCTCCCAAAAGGGGGACGAAGCGCATAAGCGGACGCTTAAAGGCCGGCTT
Seq D	TGCAGTGGACACGTGGCTTTCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTAT AAGAGTCAGAATTCGTTCGAGAACAGCAGGTCGTCGCCGGTGGGCGACGCGTCGCTGATCGGCGTCAGCAGCTCGGGCAG ATTGATCGGCATGGTTCAGTAAATCCGTTGATGCATCGAAAGGGTCGCGCTCATTCGACCGTGTATGCGAACTCGCCGGC GTCGTCCGCGCGCACCGCGATGCGCGCAATCGACGCGCCGTCGGCGATCCGGCCCAGCACGTGGCCCGCGATCTCCGGCA GCAGCGTGCCGTTCAGGATGTGGTCGACGTTGCGTGCGCCCGAGTCGACTTCGGTGCAGCGTGCGAGGACCGCTTCGACG AGCGATTCGTCCCATTCGAACGCGGCCTTGTGGTTCGCCTCGATGCGGCTGCGAATCCGTTCGAGCTTCAGCTCGATGAT CTCGGCGAGCACGTCGTCGGAGATCGGGTAGTACGGCACCACCTTCATCCGGCCGAGGAACCCGGGCTTGAACGCCTTGT ACAGCTGCGGGCGCAGCGCTTCCGCGAGCGCATCGGGATCGGGCAGTTCCTCGGCCGGCTTGTTCAGGCACGCCTGCATC ACCGCGGCCGAACCGACGTTCGACGTCAGGATGATCAGCGTGTTGCGGAAGTCGATCTCGCGCCCAGGGTCGGCGTCCAT CATGCCCTTGTCGAAAACCTGGAAGAACATCTCGAGGACGTCCGGATGCGCTTTCTCGACCTCGTCGAGCAGCACGACCG AATACGGATTGCGCCCGCCGGGCCTCGGTCAGCACGCCGCCTTCGCCGTAACCGACGTAGCCCGGCGGCGAACCCTTCAG CCCGGACACGCTGTGCGCCTCCTGATACTCGCTCATGTTGATCGTGATCATCTTGCGCTCGCCGCCGTAGAGGATGTCCG CGAGCGCCAGCGCCCGTCTCGTCTTGCCGACGCCCGACGGCCCGACGAACATGAACACCGCGCGCGCTTGTCGGATCCTC GAGTTCGCGACGCGGTGCCCACCGGCGTGGCGCGATTCGCTCCGAGCGCATTGTCCTTGCGGATCCCGCGCGCGGCGAGC ACCGGCTCGAATTGGACCGTGCCGATTCTCGTCCTTCACCCAATTCG

Table S12 Plasposon flanking sequences obtained from the plasmids recovered from the H21 and H42 mutants.

H21 mutant	
Seq E	ATTGTGCCCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGC CAGATCTGATCAAGAGACAGGTTCTACGCGGCGTGGAAGAAATACCGGCTGCCGTTCCACTACGAGCTCGATCGCGACGA GCGCTATCTGCCGCTGCTGCTAGCGATCGCGGGCGTGCCGAGCGACGACGTGCGCGACAGTCTCGCGGCAGGCGCGGGCG GCGTGCTCGACGAAGCCGTCGCCGGCTATGCGCTCGCCGCGCGGCACCGGCCGATGTCGGCCGCCTATCTGCAGCGTACG CTGTCCGATTATTTCCGCGTGCCGGTGAAGATCGATCAGTTCGTCGGCAAGTGGTACGACGTGCCGCCCGATCAGCTGAG CGTGCTCGGCCAGGTCAACGCGGTGCTCGGCGCGACGGCGCTGGTCGGCGAACGCGTGTGGCAGCGCGACATGCGCGCCC GGATCGTGATCGGCCCGCTGTCCAAGCGCGACTACGAGGCGTTCCTGCCGGGCGGCGCGCAGGCCGTTGCGCTCGAGCGG ATGTTGACGCTGCTCGCCGGCGTCACGCTCGAATACGAAGTGAAGCTCGTGCTGAAGCGGACCGAAGTCGGCGCGAGCGT ACTCGGCGCCGGCTCGCGGCTCGGCTGGGACGCGTTCCTCTGCACGCGCGACGCCGACGAAGACCGGTCCGATGCGCGCT ACGAGCTGCACGTGATTCACTGATACCGAACGACAACAAACACGCAAACGAACCTGAGATCGACGCCATGAGCACGCCTC TGAAGACCCTCATCACGAAACTGAATCCGCTGTGCCGGCATGCGGCCGAGCGCGCGGCGAGCGCGTGTCTTGCGCGCGGC CATTACGAGGTCGATCTGGAGCATCTGTTTCTCGCGCTGCTCGAGGAGCCCGGCGGGCGACCTGTCGCTCGCGCTGCGCG CAAGCCCGCATCGACGCGCATGCGTTGCGCGCCGATCTCGAGCGCGGAGCTCACGCGCCTGAAGGACCGGCCAACACGGC GGCACGCCCCGTATTTCTCGGTGCATTCTGATCCGCGCGCCTTGTTCCGAGCAGCCTGCTGGATCCGCGGTCGCTCGAAC CTCGCAGCTCGCGCATTCCCGTTCCGGCCACTGCTGGCTCGCGGCTGCCTACCGGGGCCGGCAACCTTCGCACAGAG
H42 mutant	
Seq A	TTAGACACCGTGGCTTTCCCCCCCCCCCCCGGCTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAA GAGTCAGGCCGCACGTCGTTGGCGACGCCGAACGAATAGAGCCAGCCGCGATGGCCGTCGAAGCCGGCCATCAGCGTGCC GACGACGCGCGTGCCGCAGACCGCGACGAAGAACAGCTCCGGCTGCGTCGCGAGCTTCAGCTCGATCGAGCGCAGCGGAT CGCGGTGCGGCGGCGCGCCGGCCGCATCGTAATGCGGAAACGCGTCGCGCCACACCGCGAATACGGCATCGGTGTCGGCG CGTTCGAACGGGCGGATAACGACGGCATCGGCGGCGGCATCCATCGGCGGGTTCTCCTCAGAGCGTGTCGAGCACCGACC GCAGCATCGCCATCATCTGGTCGATTTCCTCGGTCGTCACGTTCAGCGCGGGCATGAAGCGCAGCAGGTTCGGGCGCGCG GCGTTCAGCAGCAGCCCGTCGGGCTGCATGTCGCGCGCCTTCTCGACGATCTGCGGGCCGATGTCCTTGCCGAGCAGCAG CGCGCGCAGCAGCCCTTCGCCGCGTTCGCCTTCGAAGCCGCGCTCCTCCGACAGCTCGAGCAGCTTGCGCTTCAGGTATT CGCTGCGCGCACGCACGCCCTCGAGGAAGCCCGGCGCGACCAGTTGCGAGATCACTGAATGGCCGACGGCCGTCATCAGC GGATTGCCGTTGTACGTGCCGCCCTGATCGCCGGCTTCGAATACCGCGACGTCGGACTTCGACAGCAGCGCGCCGAGCGG CACGCCGCCGCCGATGCCCTTGGCGAGCGTCATGATGTCCGGCTCGATGCCCGACAGCTCGTATGCGAACAGCGTGCCCG CGCGGCCGCACCCGCTTTGCACTTCGTCGACGATCAGCAGCAGGTTGTGCTGGCGCGTCAGCTCGCGCAGCGCGCGCATA AACTCGCGCGTCGCCGGAATCACGCCGCCTTCGCCCTGGATCGGCTCGAGCATCACCGCGACCGTCTTCTCGTTGATCAG CTTCTCGACCGACTCGATGTTGTTCAGCTCGGCCCTTTCGGAAGCCGCCACCTGCGGCGCGTAAGGCGTGGTCCAGCTTG CTGCCCCTCCGCCGACTCGTCCAGCTGCGCCGTGAAGCTGTGAATCCCGAAAACGTGTATGAAT
Seq B	GGACTGAATACGTGGCTTTCCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTAT AAGAGTCAGGCCGCACGTCGTTGGCGACGCCGAACGAATAGAGCCAGCCGCGATGGCCGTCGAAGCCGGCCATCAGCGTG CCGACGACGCGCGTGCCGCAGACCGCGACGAAGAACAGCTCCGGCTGCGTCGCGAGCTTCAGCTCGATCGAGCGCAGCGG ATCGCGGTGCGGCGGCGCGCCGGCCGCATCGTAATGCGGAAACGCGTCGCGCCACACCGCGAATACGGCATCGGTGTCGG CGCGTTCGAACGGGCGGATAACGACGGCATCGGCGGCGGCATCCATCGGCGGGTTCTCCTCAGAGCGTGTCGAGCACCGA CCGCAGCATCGCCATCATCTGGTCGATTTCCTCGGTCGTCACGTTCAGCGCGGGCATGAAGCGCAGCAGGTTCGGGCGCG CGGCGTTCAGCAGCAGCCCGTCGGGCTGCATGTCGCGCGCCTTCTCGACGATCTGCGGGCCGATGTCCTTGCCGAGCAGC AGCGCGCGCAGCAGCCCTTCGCCGCGTTCGCCTTCGAAGCCGCGCTCCTCCGACAGCTCGAGCAGCTTGCGCTTCAGGTA TTCGCTGCGCGCACGCACGCCCTCGAGGAAGCCCGGCGCGACCAGTTGCGAGATCACTGAATGGCCGACGGCCGTCATCA GCGGATTGCCGTTGTACGTGCCGCCCTGATCGCCGGCTTCGAATACCGCGACGTCGGACTTCGACAGCAGCGCGCCGAGC GGCACGCCGCCGCCGATGCCCTTGGCGAGCGTCATGATGTCCCGGCTCGATGCCCGACAGCTCGTAGCGAACAGCGTGCC CGCGCGGCCGCACCCGCTTTGCACTTCGTCGACGATCAGCAGCAGGTTGTGCTGGCGCGTCAGCTCGCGCAGCGCGCGCA TAAACTCGCGCGTCGCCGGAATCACGCCCCCCTTTCGCCCTGGATCGGCTCGAGCATTCACCCGCGACCGTCTTTCTCGT TTGATCAGGCTTCTCCGGACCGACTCGATTGGTTGGTTCAGGCTCGGCTTTCGGAGGCCCGGGCACCCTGGCGGGGCGCG GTAGAGGCGTGTTCCAGGCTTGCCTGCGCCTCCGCCGGAATCGTCAGCTGCGCCGTGAGCCTTGTTGATTCGAAACGTGT GGATGTGGT

Table S13 Plasposon flanking sequences obtained from the plasmids recovered from the H42 and C1-28 mutants.

H42 mutant	
Seq C	TAGAGACCCTGCCTTTTGCTAATGTTCTTTCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGC GGCCAGATCTGATCAAGAGACAGACGTGCGGCGGCTCGGAGTCGGCCGCGCGCTGATCGCGCATGCCGAGCGTGCGCTCG CCGCGCTCGGCTGCCCGAAGATCAACCTGCAGGTGCTGCCCGGCAACGACGACGCATGCCGCTTCTACGCGGCGCTCGGC TATCGCATCGAAGAGCGCATCTCGTTCGGCAAGACGCTGCCGGCCGCCTGACGCGACGGCCGGCAGCGTCGCGTCAGCGC GTGATCGGCCCGGCGTCCGGCAGCGGGTTGCCGAGCGCGTCGGTCGGCACCGCCTGCTTTTCGAAGGCGGCGAAATAGGC GGCCCACTGCGTATCGCTCACGCGCACGATCAGCGCGCCGTCCTGCCAGACGTCGTTGTGATCGTTATGGTCGTTGCCTG CGCGGTGCAGGTAGTTCGGGCCCGTCGAGCCCTGGTTCATGTGCGTGTCGTGTATCCCGTTGCCCTGCGCATACGTGCGG CCGAACACGACCACGTCGAACTGCCCCGCCTGCGCGGCGTTCACGAGCCGCAGCAGCGACGGAATCGGCTCCGGATGCTC GGTGCCGTCCATCACCGCGCTTGCGCGCCACGTGCCCGTTTCGTTCAGGATGTCGCTGCGCAGATAGTCGAGCGCCGGCA GCGCGCTCTGCCCGGTGAGATCGGTATAGCCTTCCGGCGCGGCCGCGAGCGTCTGCGTGACCGGATGATGGAAATCGTAG ACGAGCTTGTAATACAGCAAATCGTCCGCGTCGTTGGTGCCGACGTTGATCGCGACATCCCAGTCGCCGCCGGGCACGGC CAGCGTCAGGTGGACGTGATACTGCGTCTCGCCGCCGTGCGGTGAACCTTTCAGTCTTGCGACGGCCTTCACCTTCGCCT TCACGAAACCGTAATTCGAGACCTCATGCGTGCTCTCCATGCGCCGAGGGGCTCATCATGCTACGCGCTGCGCGTGTCGC TTGCCCGACAGCCGCGCCGCCGCCTGCCGTGCCGCATGGCGCCGGCACGATCGTCCGAATGGAGTCGAATCGCGGCGCCA CAGTTCGGATGCCCGGCGCTCGCAATGAAAAGGCCGTCACGAACTTCGCGTGGACAGCTTCGAAAGGCCCCCCTGGTAATA
Seq D	TGCAGTGGACACGTGGCTTTCCCCCCCCCCCCGACTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTAT AAGAGTCAGAATTCGTTCGAGAACAGCAGGTCGTCGCCGGTGGGCGACGCGTCGCTGATCGGCGTCAGCAGCTCGGGCAG ATTGATCGGCATGGTTCAGTAAATCCGTTGATGCATCGAAAGGGTCGCGCTCATTCGACCGTGTATGCGAACTCGCCGGC GTCGTCCGCGCGCACCGCGATGCGCGCAATCGACGCGCCGTCGGCGATCCGGCCCAGCACGTGGCCCGCGATCTCCGGCA GCAGCGTGCCGTTCAGGATGTGGTCGACGTTGCGTGCGCCCGAGTCGACTTCGGTGCAGCGTGCGAGGACCGCTTCGACG AGCGATTCGTCCCATTCGAACGCGGCCTTGTGGTTCGCCTCGATGCGGCTGCGAATCCGTTCGAGCTTCAGCTCGATGAT CTCGGCGAGCACGTCGTCGGAGATCGGGTAGTACGGCACCACCTTCATCCGGCCGAGGAACCCGGGCTTGAACGCCTTGT ACAGCTGCGGGCGCAGCGCTTCCGCGAGCGCATCGGGATCGGGCAGTTCCTCGGCCGGCTTGTTCAGGCACGCCTGCATC ACCGCGGCCGAACCGACGTTCGACGTCAGGATGATCAGCGTGTTGCGGAAGTCGATCTCGCGCCCAGGGTCGGCGTCCAT CATGCCCTTGTCGAAAACCTGGAAGAACATCTCGAGGACGTCCGGATGCGCTTTCTCGACCTCGTCGAGCAGCACGACCG AATACGGATTGCGCCCGCCGGGCCTCGGTCAGCACGCCGCCTTCGCCGTAACCGACGTAGCCCGGCGGCGAACCCTTCAG CCCGGACACGCTGTGCGCCTCCTGATACTCGCTCATGTTGATCGTGATCATCTTGCGCTCGCCGCCGTAGAGGATGTCCG CGAGCGCCAGCGCCCGTCTCGTCTTGCCGACGCCCGACGGCCCGACGAACATGAACACCGCGCGCGCTTGTCGGATCCTC GAGTTCGCGACGCGGTGCCCACCGGCGTGGCGCGATTCGCTCCGAGCGCATTGTCCTTGCGGATCCCGCGCGCGGCGAGC ACCGGCTCGAATTGGACCGTGCCGATTCTCGTCCTTCACCCAATTCG
C1-28 mutant	
Seq A	TTAGCAGCGTGGCTTTCCCCCCCCCCCCCGGCTCTTAATTAATTTAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAG AGTCAGCCTTCGAGCAGCGAGACGAGATAGACGGCGGCCAGATGCACGTCGAGGTCGGCCGCGAGTTCGCCCGCCTCGAT CGCGCGGCGCAGCAGCGCCTTCGTGATCCGCAGCGTCTGCAGTTCGAGCAGCATCCGGCGACGCAGCAGCGCCGCGTTCT CGTCGCTTTGCTCGCACTTCGTGTAGAGGATCACGAGCACGCGCTGCATCGAACCGGGCTCGCCGCACTGGCGCAGATAG TGCGACGCGGCGTGCCGCAGCGTCGCGAACGCCGGCAGCCCGTCGCCGGCCTCGAAGCCCTCCGACGTGCGCGCGAACGC GCGGTCGCACATCGCGAGGCACACCTCCATCTTGTTGCGGTAGTGGCCGTACACGGCGCCGCGCGACATGCCGGCGGCCT CGGCGAGGTCCGCCATCGCGGTTTGCGCGACGCCCTTGTCGAGCAGCACGAGCTCGGCGGCGTCGAGGATCCGGTGCTTG ATGGCGAGCGATTCTTCCCGGGTCTTGCGAGCCATGTATCGAGTTTCCTTACAGTACGCTGTCGCTTTATTGAGACAGTC TGACGATTGGATGGAGGATCGATGTATTTAATCAGTCGTGACTGATTATAATCGGCGACCCTGAGCCGCCGCATTGTATC GGCGGCGCACGATCCGAGGTCCCAGATGAATGACAAACGCTCGCTGTCGCGCCGGCTGACGCCGCTCGCGCTGGCCGTCG CCGTGGCCGCGGCCGGCTGCGGAAAGGGCAAACCGGATGCCGAGCCGGCGGCCGTGCAGAAAGGCGACCGTCGTCACGGT GCGTCCCGGTCGCGGTGCCGATGACCGTCGAGCTGCCGGGCCGGCTCGATGCCTTTCGACGCGCCCGAAGTGCGCGCGCG CGTCGCCGGCATCGTCACGGCACGCACGTACGACGAAAGCCAGGAGTGAAGCAGGGCCGCCGTGCTGTTTCGCATCGAAT CCGCGCCGTTGGAAAGCGCACCGCGATGCGGCGCAGGCGCGGCTTCGCAGGCCGCAAGCCGGGGCGGCCGCGTCCGCGAC AGCGGCGGCGCTATGCGATCTCGTTGCGCGATTCACGCCGTCAGGCCAAGCCGCGCACTATACACTCACGCGAAG

Table S14 Plasposon flanking sequences obtained from the plasmids recovered from the C1-28 and C2-22 mutants.

C1-28 mutant	
Seq B	GAAGGGCCGGCTGCCCTTTGCTAATGTTCTTTCCTGGTACCGTGCGAGGAGCATGGCGCGCCGGCGATCGCGGCCGGCCT AGGCGGCCAGAATCTGATCAAGCCCACAGGCTCGAAGGCGTGTTCGCGTCGATGATCTGGACGAACCGGCTGCGCGGCAA TCTGTGGAACGACGCGGAAGCGATGCTCGATGCGGGCTTCGACGCAGTGCGGACCTCCGCCGCGCTGCGGCGGCCCGCAC ACAAATTGCCGAAGCCCGCGAAACGATCGCAATAAATGCCGCTTTAACCCGATTTACCGCACTGCGAAACGAGTGAGTCG ACCCTCTTTAAAATTTTTAACATCTTCCGGAACATCGGTAGACTCGCGCTGTCAGCTCTCTTTAGCGTCTTCGTGATAAC CGGGTTCGACCCGGCATGCACGCCGCCGTTCGAAGCGGGTTTGCCCGCGCGTTGGGTCGAACGGAAACGACACAGGCCCC CTGGCGGGGCCTGTTTTGTTGCACGCCCGCCGCCCGCCCCGTCGCCGTTCGACGGATGAGCCGCCGGCGGTCCCGCCCAC ACCCCCGTGAACCGCTTTTCTGGATGTGGACCCTTGGTCAAGTTAGCCGCAACTCGTGATGCTTCCGCGCCGGGCCGCGC GCAGCGCTTCGTGCGCGCGCTGGTGCCCGCCGCGGTGCTCGGCGCGCTCGCCGCGCTCGGGCTCGCCGCCCGCGATGCCG GCCCGTCTCGCACGGCCCGGCCCCCGTCGACGCCGACACGGTCGTGCTGCCGCAGCGCGTGCTGTCCGACACGCCCGTTT TCCCACCCGCACAACATTTCGTGACGAGCGAACTCGCCCGCACGATCGGGAGCGCAACGCGCAGCCTGGCCTGTTTCGGA GCGGTCAGCGCGCACCGCAACGACATGCTGGGCTACGGCGGAACCTGTTTACGCTCGCGCCGCCCGAGCCGCTGCACGGC ATGCGCGCGGGGCACGATCGAAGCTCTCGGCTCGCCGACACCGCTGAACCGGCATCAGCGTGCCCGCCGGAATTTCCGGA ATT
C2-22 mutant	
Seq A	TAATTCCCCATAGATTTAAAGGATTTTTTTTTTTTTTTTAAGTGAAAACTGGCTTTCCCCCCCCCCCCGACTCTTAATAA ATAAATCTAGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGCCTTCGCAGCAGCGAGACGAGATAGACGGCGGCCAGA TGCACGTCGAGGTCGGCCGCGAGTTCGCCCGCCTCGATCGCGCGGCGCAGCAGCGCCTTCGTGATCCGCAGCGTCTGCAG TTCGAGCAGCATCCGGCGACGCAGCAGCGCCGCGTTCTCGTCGCTTTGCTCGCACTTCGTGTAGAGGATCACGAGCACGC GCTGCATCGAACCGGGCTCGCCGCACTGGCGCAGATAGTGCGACGCGGCGTGCCGCAGCGTCGCGAACGCCGGCAGCCCG TCGCCGGCCTCGAAGCCCTCCGACGTGCGCGCGAACGCGCGGTCGCACATCGCGAGGCACACCTCCATCTTGTTGCGGTA GTGGCCGTACACGGCGCCGCGCGACATGCCGGCGGCCTCGGCGAGGTCCGCCATCGCGGTTTGCGCGACGCCCTTGTCGA GCAGCACGAGCTCGGCGGCGTCGAGGATCCGGTGCTTGATGGCGAGCGATTCTTCCCGGGTCTTGCGAGCCATGTATCGA GTTTCCTTACAGTACGCTGTCGCTTTATTGAGACAGTCTGACGATTGGATGGAGGATCGATGTATTTAATCAGTCGTGAC TGATTATAATCGGCGACCCTGAGCCGCCGCATTGTATCGGCGGCGCACGATCCGAGGTCCCAGATGAATGACAAACGCTC GCTGTCGCGCCGGCTGACGCCGCTCGCGCTGGCCGTCGCCGTGGGCCGCGGCCGGCTGCGGAAAGGGCAAACCGGATGCC GAGCCGGCGGCCGTGCAGAAAGGCGACCGTCCTCACGGTGCGTCCGGTCGCGGTGCCCGATGACCGTCGAGCTGCCGGGC CGGGCTCGATGCCTATCGACCGCGCCGGAGTTGCGCGCGCGCGTCGCCGGCATCGTACGGGCACGCAACGTACGACGAAG GCCAGAAGTGAAAGCAAGGGCGCCGGTGCCTGTTCCGCATCGAATTCCCGCGCCGGTTGAAGGCGCGCACCCGATGCCGG CGCAAGGGCGCCGCCTCGGCCAAAGGCGCAATGCTCGGGCCGTCCGCTGTTCCGGCTCGACATAATGCTGTTTGCCCTAT ATGTGGCCGCCGTGATTATCTTGC
Seq B	ATCGTGCCTTTTGCTAATGTTCTTTCCTGGTACCGTCGACATGCATGGCGCGCCGGCGATCGCGGCCGGCCTAGGCGGCC AGATCTGATCAAGAGACAGTCGGAGGGCTTCGAGGCCGGCGACGGGCTGCCGGCACGTCACGACGCTGCTAAACGCCGCG TGCCACTATCTGCTCCAATTCGGCTAGCCAAGTTAAATCCAGCACATGATCGTAAACCTCTAGACGAAGTACTAAAAA
Seq C	TTGGGGGGGGGGAAAAAGAAAAAACAAAAAAATTTGTGGTTTTAAAAGAAGGAAAAAGATTTTTTTTGTTTTGAAGGGTG GTTTGAGAAAAGCAATATAAAAGCAAAACAAGACAAGTGACCCGGCTTTCCCCCCCCCCGGCTCTTATTAATTTAAATCT AGACTAGTGCGGCCGCACTTGTGTATAAGAGTCAGGCCCTCCGACGTGCGCGCGAACGCGCGGTCGCACATCGCGAGGCA CACCTCCATCTTGTTGCGGTAGTGGCCGTACACGGCGCCGCGCGACATGCCGGCGGCCTCGGCGAGGTCCGCCATCGCGG TTTGCGCGACGCCCTTGTCGAGCAGCACGAGCTCGGCGGCGTCGAGGATCCGGTGCTTGATGGCGAGCGATTCTTCCCGG GTCTTGCGAGCCATGTATCGAGTTTCCTTACAGTACGCTGTCGCTTTATTGAGACAGTCTGACGATTGGATGGAGGATCG ATGTATTTAATCAGTCGTGACTGATTATAATCGGCGACCCTGAGCCGCCGCATTGTATCGGCGGCGCACGATCCGAGGTC CCAGATGAATGACAAACGCTCGCTGTCGCGCCGGCTGACGCCGCTCGCGCTGGCCGTCGCCGTGGCCGCGGCCGGCTGCG GAAAGGGCAAACCGGATGCCGAGCCGGCGGCCGTGCAGAAGGCGACCGTCGTCACGGTGCGTCCGGTCGCGGTGCCGATG ACCGTCGAGCTGCCGGGCCGGCTCGATGCCTATCGACGCGCCGAAGTGCGCGCGCGCGTCGCCGGCATCGTCACGGCACG CACGTACGACGAAGGCCAGGAAGTGAAGCAGGGCGCCGTGCTGTTTCGCATCGATCCCGCGCCGTTGAAAGCGGCACGCG ATGCGGGCGCAGGGGCGCGCTCGCCAAGGCGGCAGGCCGCGGCGGCCGCGGGCCGCCGACAAGCGGGCGGGCGGCTATGC CGATCTCGTGCGGCGATCACGCCCGTCAGGCGGAGGCGGCGACTACACGGGAGGCCGGTTCGCGGGCCCGACACCGCCAG GCGGAAGGGCGGGGAGGGTTCGGCGTTCGGCCGAAAAGGCCGGAAACTTCCGCGGCGTTGGCGCCAGGCTGGCAGGCTCG GAACTTACCGCCGGAACGGTTGGAACCGGCGCCCCGGAATTTTTCCGGGACGGAGCGGCGGCGACCGGCCGCCTTCGGTG TGGTACCCCCTGATATG