CHAPTER 5

An Attempt to Identify Genetic Loci Responsible for Formation of Biofilm in *A. junii* BB1A

5.1. Introduction

5.1.1. Genetic Regulation of biofilm formation

Biofilm formation in microorganisms is a well-regulated developmental process that results in a complex population of cell types. The developmental process has been extensively studied in model organism such as *Pseudomonas aeruginosa* and can be divided into some basic key steps such as initiation, attachment, microcolony formation, biofilm maturation and dispersion (Sauer et al., 2002). Almost every step is governed and regulated by diverse types of genetic mechanism and it varies among different types of bacteria.

5.1.1.1. *Initiation of biofilm*

Genetic analysis of biofilm formation by many bacteria revealed multiple genetic pathways to initiate biofilm development. Initiation mechanism in many bacteria can be studied using mutant strains defective in biofilm formation. A simple screening method involving the growth and biofilm formation in 96-well plastic microtiter dishes has been extensively applied for the isolation of biofilm defective mutants among wide variety of organisms (Genevaux et al., 1996; Heilmann and Gotz, 1998; Mack et al., 1994; Nyvad and Kilian, 1990). These biofilm defective mutants have been used to identify genes required for development of biofilm. Similarily *Pseudomonas aeruginosa* mutants, defective in surface attachment, have also been described in a similar way (O'Toole and Kolter, 1998). *Vibrio cholerae* utilizes different pathways for initial attachment depending on the surface to which the organism attaches. It requires Tcp pilus for the attatchment to intestine (Herrington et al., 1988). While on abiotic surfaces, it is the pilus encoded by the *msh* locus that is required (Watnick et al., 1999). In *Acinetobacter baumanii*, the chaperone-usher secretion system leads to the formation of pili, which is further involved in the attachment and biofilm formation on to plastic surfaces (Tomaras et al., 2003).

5.1.1.2. *Quorum sensing*

An important characteristic of many biofilm forming bacteria is the density dependent expression of certain functions such as secretion of extracellular polymeric substances (EPS). It has been found that the secretion of EPS is under quorum-sensing control and has been widely studied in a number of bacterial model systems. In many species, including *P. aeruginosa*, the production of EPS occurs at high cell density (Davies et al., 1998; Sakuragi and Kolter, 2007). The study of bacterial strains with mutations in genes involved in the expression of signaling molecules and the analysis of differential gene expression in biofilms are contributing information on the molecular mechanisms of biofilm formation and the role of quorum sensing. In *A. baumanii* too the quorum sensing pathway senses the extracellular signal and directs the formation of biofilm. The strain *A. baumannii* M2 produces an acyl-homoserine lactone molecule, which is known to be a product of the *abaI* autoinducer synthase gene. Gaddy and Actis (2009) reviwed the biofilm regulation in *A. baumannii* and described that the inactivation of *abaI* gene results in a 30–40% biofilm reduction and hence the product of *abaI* gene is assumed to plays an important role in the biofilm formation in this bacterium. It was also observed by the investigators, that the addition of exogenous purified M2 acyl-homoserine lactone has restored biofilm maturation of the *abaI* mutant. Thus, quorum sensing pathway is important for the maturation of biofilm formed by *A. baumannii* as well as for other related human bacterial pathogens (Irie and Parsek, 2008). Similar mutational studies in *P. aeruginosa,* the mutant of which is unable to synthesize the quorum-sensing molecules acylhomoserine lactones (AHLs), revealed significantly altered biofilm architecture (Davies et al., 1998). This clearly demonstrated that these molecules regulate the formation of biofilm structures in these organisms.

The AHLs are the predominant signaling molecules in gram negative proteobacteria particulary belonging to α, β, and γ subdivisions. Structurally AHLs consist of a homoserine lactone ring covalently linked *via* an amide bond to an acyl side chain (ranging between 4 and 18 carbons) which may be saturated or unsaturated and with or without a hydroxy-, oxo- or no substituent on the carbon at the 3 position of the *N*-linked acyl chain (Atkinson et al., 2007). AHLs are usually synthesized by enzymes belonging to the LuxI family of AHL synthases. Much of the work on quorum sensing systems fall under the "two-gene" (an autoinducer synthase 'LuxI' together with a receptor molecule 'LuxR') model as defined in the Gram-negative Proteobacteria such as *Vibrio fischeri* system. Studies on the comparison between proteobacteria phylogeny as revealed by 16S ribosomal RNA sequences and phylogenies of LuxI and LuxR homologs shows a particularly high level of global similarity. In Gamma Proteobacteria the quorum sensing genes, which, although functionally similar to the *luxI/luxR* genes, have a markedly divergent sequence (Lerat and Moran, 2004). Studies of the crystal structures of EsaI

and LasI (the LuxI homologue) from *Pantoea stewartii* and *Pseudomonas aeruginosa* respectively have revealed that this large protein family belongs to the GCN5-related *N*-acetyltransferase protein family (Watson et al., 2002; Gould et al., 2004). The LuxI-like proteins are responsible for the biosynthesis of a specific acylated homoserine lactone signaling molecule (AHL) known as an autoinducer. The AHL concentration increases with increasing cell density. Once synthesized, AHLs gather extracellularly and diffuse into nearby bacterial cells where they usually interact with members of the LuxR family of transcriptional regulators (Fuqua et al., 1996; Swift et al., 2001; Cámara et al., 2002). The AHLs then bind to, and activate LuxR homologous proteins and the resulting LuxR protein/AHL complex activates or represses the related target structural gene(s) (Swift et al., 2001; Zhang et al., 2002; Cámara et al., 2002).

Table 5.1: Some examples of AHL dependent QS systems in gram negative gamma proteobacteria.

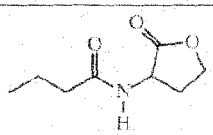
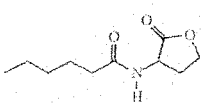
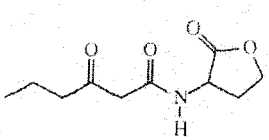
Signal molecule	Structure	Bacteria	Refrences
C4-HSL		*Aeromonas hydrophila, Aeromonas salmonicida, Pseudomonas aeruginosa, Serretia liquefaciens* MG1, *Serretia sp* ATCC39006,	Williams, 2007
C6-HSL		*Pseudomonas aureofaciens, Yersinia enterocolitica, Aeromonas hydrophila, Aeromonas salmonicida, Pseudomonas aeruginosa, Pseudomonas aureofaciens, Serretia sp* ATCC39006,	Williams, 2007
3-Oxo-C6-HSL		*E. carotovora, Vibrio fischeri, Yersinia enterocolitica Enterobacter agglomerans, Erwinia chrysanthemi, Pantoea stewartii, Yersinia pseudotuberculosis*	William, 2007; Swift et al., 1993; Nasser et al., 1998; Batchelor et al., 1997; Atkinson et al., 1999
3-oxo-C10-HSL		*Vibrio anguillarum*	Milton et al., 1997
3-oxo-C12-HSL		*Pseudomonas aeruginosa*	Gambello and Iglewski, 1991

Table 5.2: Bacteria with LuxI and LuxR homologue(s).

Organism	LuxR/LuxI homologues	Gene Bank Accsession no.
Aeromonas hydrophila	AhyR/AhyI	X89469
Aeromonas salmonicida	AsaR/AsaI	U65741
Agrobacterium tumefaciens	TraR/TraI	L17024,L22207
Burkholderia cepacia	CepR/CepI	AF330018,AF330012
Enterobacter agglomerans	EagR/EagI	X74300
Erwinia carotovora Subsp carotovora	CarR/ExpR ExpI /CarI	X74299,X80475,X72891
Erwinia chrysanthemi	ExpR/ExpI(EchR, EchI)	X96440
Escherichia coli	SdiA	AE005414
Pantoea stewartii	EsaR/EsaI	L32183,L32184
Pseudomonas aeruginosa	LasR/LasI	M59425
	RhlR/RhlI	L08962,U11811,U15644
Pseudomonas aureofaciens	PhzR/PhzI	L32729,L33724
Pseudomonas fluorescens	PhzR/PhzI	L48616
Pseudomonas syringae pv. tabaci	PsyR/PsyI	U39802
Ralstonia solanacearum	SolR/SolI	AF021840
Rhizobium etli	RaiR/RaiI	U92713
Rhizobium leguminosarum	RhiR	M98835
Rhodobacter sphaeroides	CerR/CerI	AF016298
Serratia liquefaciens	SwrR/SwrI	U22823
Vibrio anguillarum	VanR/VanI	U69677
Vibrio fischeri	LuxR/LuxI	M19039,M96844,M25752
Yersinia enterocolitica	YenR/YenI	X76082
Yersinia pestis	YpeR/YpeI	AF071401
Yersinia pseudotuberculosis	YpsR/YpsI	AF079973
	YtbR/YtbI	AF079136
Yersinia ruckeri	YukR/YukI	AF079135

5.1.1.3. Genetics regulation of EPS biosynthesis during biofilm formation

Extracellular polymeric substances (EPS) is the important constituent of biofilm which make up the intercellular space of microbial aggregates and form the structure and architecture of the biofilm matrix. The key functions of EPS involve the mediation of the initial attachment of cells to different substrata. Literature contains some evidences that the cell contact with the substratum stimulates the transcription of EPS genes (Dunne, 2002).

The regulatory mechanism of the EPS synthesis has not yet been well defined. EPS synthesis generally requires enzymes for the production of each nucleotide sugar precursor, separate transeferase for each monosaccharide in the subunit, one or more polymerase, and proteins involved in export of the polysaccharide (Coplin and Cook, 1990). Enzymes involved in the formation of EPS precursors, are known to be under separate control from mechanisms of gene expression associated with the EPS molecules biosynthesis (Sutherland, 2001). In number of bacterial species, the EPS synthesis is

controlled by a large cluster of EPS biosynthetic genes. Boucher et al. (1997) and Mejiaruiz et al. (1997) reported that in *Pseudomonas aeruginosa* and *Azotobacter vinelandii* cells, the gene cluster *algA, algC, algD, algE* and *algK* are involved in the regulation of both the processes of biosynthesis and the secretion of extracellular molecules. In *Staphylococcus epidermidis and S. aureus* biofim, four homologous proteins (IcaA, IcaD, IcaB, and IcaC) encoded by genes organized in a single operon (*icaADBC*) (Cramton et al., 1999; Heilmann et al., 1996) are responsible for the synthesis of exopolysaccharide known as β-1,6-linked *N*-acetylglucosamine, named PIA (polysaccharide intercellular adhesin) (Mack et al.,1996) and PNAG (poly-β-1,6-linked *N*-acetylglucosamine) respectively. Transposon mutagenesis in *E. coli* revealed that, a locus (*pgaABCD*) exhibiting significant similarity to the *icaADBC* operon of *S. aureus* is required for biofilm formation (Wang et al., 2004). The operon *pgaABCD* is present in diverse bacterial species and appears to be part of a horizontally transferred locus (Wang et al., 2004). The gene product of *pgaABCD* are involved in the synthesis of an unbranched β-1, 6-linked *N*-acetylglucosamine. Analysis of conserved protein domains predicted that PgaC is an N-glycosyltransferase homologous to *Ica*A; PgaB is a lipoprotein with putative polysaccharide *N*-deacetylase domains similar to those of IcaB while PgaA and PgaD have no functional homologies (Lasa, 2006). Choi et al. (2009) has shown the presence of *pgaABCD* locus in *Acinetobacter baumannii* that encodes proteins that synthesize cell-associated poly-β-(1-6)-*N*-acetylglucosamine (PNAG or PGA) and the locus is critical for biofilm formation. Deleting the *pga* locus resulted in an *A. baumannii* strain deficient in PNAG, and transcomplementation of the S1Δ*pga* strain with the *pgaABCD* genes fully restored the wild-type PNAG phenotype.

Table 5.3: Some biofilm forming bacteria having *pgaABCD* gene cluster.

Microorganisms	**Refrence**
E. coli	Wang et al., 2004
Actinobacillus pleuropneomoniae	Izano et al., 2007
Acinetobacter baumannii	Choi et al., 2009
Yersinia pestis	Izano et al., 2007
Yersinia enterocolitica	Izano et al., 2008
Bordetella parapertussis	Parise et al., 2007
Pectobacterium atrosepticum	Pérez-Mendoza et al., 2011

5.1.2. Methods to characterize genes involved in biofilm formation

One of the important objectives of molecular genetics is to identify and isolate genes regulating the expression of important traits. In the last few years, significant progress has been made in the area of detection and identification of genes involved in genetic traits of specific interest. Today, numbers of genome projects are adding tens of thousands of nucleotide sequences to the public databases each day. The main challenge is to translate these sequences into function. One approach commonly used is to search databases for well-characterized proteins that have similar amino acid sequences to the protein encoded by a new gene, and from there explore the gene's function further. Mutants that lack the gene or express an altered version of it can be used to study the function of genes in a cell or organism.

5.1.2.1. Induced Mutagenesis

Most genes were identified by the processes when the genes were mutated. By treating organisms with mutagens, large numbers of mutants can be obtained and then screened for a particular defect of interest. An alternative approach to chemical or radiation mutagenesis is called *insertional mutagenesis*. This method involves the random insertion of exogenous DNA (whose sequence is known) into the genome resulting in mutation if the inserted fragment interrupts a gene or its regulatory sequences. The inserted DNA then can be used as a molecular tag that assists in the subsequent identification and cloning of the disrupted gene.

5.1.2.2.Transposon mutagenesis

Transposon mutagenesis is an excellent strategy for defining the genetic basis of various bacterial traits such as biofilm formation. The two most exploited transposon systems are Tn5 (Goryshin et al. 2000). It is convenient to apply both *in vitro* and *in vivo* as it only requires the transposase catalysing the reaction and any DNA, including genetic markers, flanked by transposon end recognition sequences. Tn5 transposes via the cut and paste mechanism (Reznikoff, 2008) and is well understood and described (Reznikoff, 1993; 2003).

Tn5 is composed of terminal inverted repeats (IR) of the 1535 bp insertion element IS50, flanking a central region of 2748 bp with genes conferring resistance to antibiotics including kanamycin (kan), bleomycin (ble) and streptomycin (str) (De Bruijn and Lupski, 1984). The method of Rich and Willis (1990) uses restriction enzymes that do not cut within Tn5 and a primer extending outward from the ends of Tn5 inverted repeats.

5.1.3. Identifying the interrupted genes

Any phenotype obtained by transposon insertion is likely to occur because the insertion element has disrupted the corresponding gene, which means that the gene's sequences will flank the insertion site. Therefore, the method provides a short cut for isolating the target gene without going for linkage mapping and the assembly of large clone contigs. Theoretically, the simplest method of gene isolation is to generate a genomic library from the mutated organsims, and isolate clones containing the insertion element (and several kilobases of flanking DNA) by hybridization with the cloned transposon. However, there are methods to isolate flanking sequences in a matter of hours without the need for library preparation.

Flanked DNA is the DNA segments that are on either side of the transposon and is generated during transposon mutagenesis. Once a particular mutant affecting a specific trait has been isolated, the next challenge is to clone and sequence the mutated or flanked DNA.

In the past decades, many strategies have been developed to amplify unknown DNA flanking sequences. *Alu*-PCR, Inverse PCR (IPCR), and vectorette-PCR are PCR-based techniques used to amplify and clone unknown neighboring DNA junction sequences of the integrants.

5.1.3.1.IPCR

IPCR (Inverse PCR) is a modification of the PCR technique that is used to amplify cellular DNA adjacent to an integrated sequence after its intramolecular circularization (Collins et al., 1984; Ochman at al., 1988; Triglia et al., 1988; Silver and Keerikatte, 1989). The basic principle of IPCR is shown in figure 5.1. This method begins with the digestion of genomic DNA from the target tissue or cell line with a restriction enzyme which does not have a cutting site in the integrant. Subsequent intramolecular self-ligation of these DNA fragments generates a small monomeric circle. Within this circularized form of DNA, conventional PCR technique is applied to amplify the junction region by two primers in opposite directions on the known integrant sequences. Hence, in IPCR, the intramolecular circularization of template is a key step for amplification (Collins et al., 1984). To avoid intermolecular ligation, the concentration of input DNA has to be decreased and results in a larger volume for ligation.

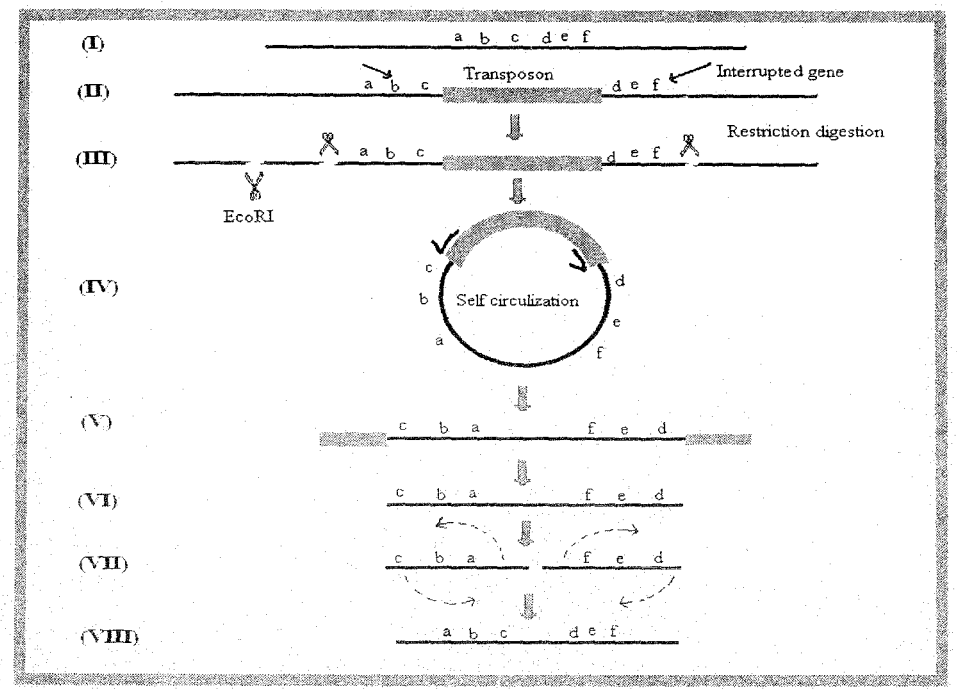


Fig. 5.1: Principle of inverse PCR (IPCR) technique. (I) The genomic DNA with genes a, b, c, d for an example (II) The insertion of transposon and disruption of gene (III) Genomic DNA is cleaved with restriction enzymes that do not cut inside the transposon, the transposon is liberated with the flanking genomic sequences. (IV) The cleave DNA is diluted and ligated to form circles. The circle containing the transposon insert is recognized by insert specific inverse primer (outward primer). (V) Amplification using this primer pair allows the flanking sequences to be cloned and sequenced subsequently (VII) to (VIII) the generated sequence represent inverted orientation and hence is reoriented in the original form.

5.1.4. PCR amplification with degenerate primer

Another approach to fish out the target gene is the PCR amplification using specific PCR primer, but this approach can only be used when some of the sequences of the target gene are known. In many cases one is interested to amplify a target gene whose sequences are not known in a particular organism. In such cases one has to design a primer from the homologous genes whose sequences are available in the data bank. Homology of a particular gene can be drawn using various bioinformatic tools more specifically sequence alignment software's such as CLUSTAL- W. This programme let the user to find out conserved sequences and can help them to make primers. Nucleotide sequence analyses with several alignment tools have revealed wide diversity with infrequent short stretches of similarity in *luxI* gene homologues.

Bacteria communicates with each other and can collectively form a group with properties not expressed when they are alone (Greenberg, 2000). Quorum sensing is the term used to define and quantify such behavior. Actually the individuals in a bacterial population senses the quanta of certain

chemical signals called autoinducer (AI) produced by themselves in course of growth (Shaw et al., 1997). N-acylated derivatives of L-homoserine lactone (acyl-HSLs) which are identified as AIs are the products of acyl-homoserine lactone synthase (Shaw et al., 1997). With few exceptions, these enzymes constitute an evolutionarily conserved family of homologues known as the *LuxI* family of autoinducer synthase (Gray and Garey, 2001). AI-based sensing mechanisms have been widely studied in gram negative bacteria (Greenberg 2000; Parsek and Greenberg, 2000; Swift et al.,1999) which in turn control several functions like exoenzyme synthesis, conjugation, antibiotic production, luminescence, metal tolerance and biofilm formation (Sarkar and Chakraborty, 2008).

These stretches are too small to design a universal primer for conducting PCR amplification. Identifying *luxI* homologues from species (whose complete genome sequences are not available) by DNA-DNA hybridization using a single probe would also not be possible. The most viable solution to the problem is to design degenerate primer set for the amplification of such sequences. A PCR primer is called degenerate if some of its position has several possible bases (Kwok et al., 1994.) For example in a degenerate primer like bchY fwd (5-CNCARANATGGYCCNGNTTYG-3) the R , Y and N are degenerate that means primer is a mixture in which one contain either A/G at position of R, similarly C/T at position Y and A/T/G/C at position N. Thus one in many primer mixtures will have the correct match with the target DNA so the probability of annealing of the correct primer decreases with the number of degeneracy introduced. The degeneracy can be defined as the product of number of bases available at the degenerate position in the primer for example in the above degenerate primer it is 2x2x4=16 therefore 1 in 16 primers have the correct match to the target DNA. This causes lower concentration of the specific primer in the mixtures of degenerate primer and will cause nonspecific PCR amplification as the Tm or annealing temperature of such primer is not clear. This problem can be solved by using Touchdown PCR. A Touchdown PCR (TD-PCR) involves an initial annealing temperature above the projected melting temperature (T_m) of the primers being used, then gradually transitions to a lower, more permissive annealing temperature over the course of successive cycles, thus will avoid amplifying nonspecific sequences (Don et al., 1991).

As the whole genome sequences of many organisms are available in public domain it is possible to design gene specific primers for PCR amplification. A number of computer programs like Amplify (Jarmon, 2004), simPCR (Rubin and Levy, 1996), PCRAna (Nishigaki et al., 2000), PUNS (Boutros and Okey, 2004), and Virtual PCR (Cao et al., 2005; Lexa et al., 2001) are available on-line which can predict the efficiency of PCR amplification with such designed primers. These programs however could not predict amplification with degenerate primer set (Cao et al., 2005).

5.2. Materials and Methods

5.2.1. Transposon mutagenesis

5.2.1.1. *Isolation of plasmid DNA*

The plasmid vector pSUP5011 (Simon et al., 1983) is a non-segregative, narrow host range plasmid carrying the transposon Tn5 with a broad host range mobilization sequence (mob) was isolated from the host strain *E. coli* S17-1 (Simon 1984) according to the method followed by kado and Leu (Kado and Leu, 1981).

Reagents used

Solution I: 50 mM glucose
25 mM Tris-Cl (pH-8.0)
10 mM EDTA (pH-8.0)

Sol. I was prepared from standard stocks in batches of a 100 ml; it was autoclaved for 15 min, at 15 Psi on liq-cycle and stored at 4 °C.

Solution II: 0.2 N NaOH (freshly diluted)
1% (w/v) SDS

It was prepared freshly and stored at room temperature.

Solution III: 5 M CH_3COOK 60.0 ml
CH_3COOH 11.5 ml
H_2O 28.5 ml

Sol. III was autoclaved for 15 min, at 15 Psi on liq-cycle and stored at 4 °C.

Isolation of plasmid DNA (Mini preparation): Plasmid (pSUP5011) bearing *E.coli* was grown overnight in Luria-Bertani (LB) medium at 37 °C.1.5 ml of this culture was taken in microcentrifuge tube and centrifuged at 5000 rpm for 5 min. The pellet was suspended in 100 µl of solution I, followed by 200 µl of solution II. The suspension was mixed by inverting the tubes and was incubated at room temperature for 5 min for complete lysis. Chilled solution III (150 µl) was added to the tubes and mixed gently. Tubes were chilled on ice for 10 min followed by centrifugation at 10, 000 rpm for 15- 20 min at 4 °C. The clear supernatant was transferred to the fresh tubes and double volume of isopropanol was added. Plasmid DNA was pelleted immediately by centrifugation at 10, 000 rpm for 20 min at 4 °C. Pellets

were rinsed with 70 % alcohol, air-dried and suspended in 30 µl of TE buffer (pH 8.0) each. The presence of DNA was checked by 0.8% agarose gel electrophoresis.

Agarose gel electrophoresis: 0.8% agarose (or less based on the increasing size of DNA fragment to be resolved) in 1 X TAE buffer supplemented with 1 µg/ml ethidium bromide was melted in a microwave or boiling water bath and then cooled to 50-60 °C before pouring in a casting tray fitted with a Teflon comb forming wells. DNA loading buffer (1 X) was mixed to the sample DNA prior to loading in the wells. Electrophoresis was performed in horizontal electrophoresis tank using 1X TAE buffer for 1-2 h. DNA could be directly viewed in a UV transilluminator.

5.2.1.2. *Transformation of suicidal vector pSUP5011*

Preparation of competent BB1A cells using $CaCl_2$: The suicidal vector pSUP5011 was used for transposon mutagenesis of *A. junni* BB1A. The vector was delivered by calcium chloride mediated artificial transformation. For this overnight grown BB1A cells were added in fresh 10ml LB medium containing 1.5 mM $NiCl_2$ and allowed to grow in shaker till the $O.D_{540}$ value reached at 0.04, the cells were then harvested at 6000 rpm for 15 min at 4 °C, washed with sterile ice-cold PBS buffer (pH 7.2) and finally suspended in 800 µl ice cold 0.1 M $CaCl_2$. The cell suspension was then stored at 4°C for 30 min. The cells were recovered by centrifugation at 4000 rpm for 10 min at 4°C. The fluid was decanted from the cell pellet, and was re-suspended in 400 µl ice cold 0.1M $CaCl_2$.

Transformation: 200 µl of the competent cells were transferred to a sterile microfuge tube. Plasmid DNA pSUP5011 (not more than 50 ng in a volume of 10 µl) was added to the tube. Gentle swirling of the microfuge tube was done to mix the contents. The tube was stored on ice for 30 min. It was then transferred to a water bath (42 °C) for exactly 90 sec followed by immediate transferring to an ice bath for chilling for 1-2 min. 800 µl of LB medium was added to the tube and incubated for 45 min in a water bath set at 37 °C to allow the bacteria to recover and to express antibiotic resistance marker encoded for the plasmid. Appropriate volume (upto 200 µl/ 9cm petriplates containing 15 ml medium each) of competent cells were transferred onto LB agar plate containing Kanamycin (50 µg/ml). A sterile bent glass rod was used to spread the transformed cells over the surface of the agar plate. The plates were left at room temperature until the liquid had been absorbed. Finally plates were inverted and incubated at 30 °C for 12-16 h for the appearance of colonies. From the Kan^r colonies the mutant was selected by replica plating on AB minimal agar plates plates containing different concentrations of $NiCl_2$. The mutant was selected on the basis of their higher nickel sensitivity compared to wild type.

Confirmation of the identity of the Kan^r mutant: To confirm that the mutant was derived from *A.junii* BB1A the genomic DNA of the mutant was isolated by boiling lysis and PCR amplification was done using primer (bbrc1 5′-TTCGATGGTGACCGCCGCT-3′ and bbrc2 5′-GCATCGCCATCGGCACG-3′) (Bhadra et al., 2006). To confirm the presence of transposon Tn5-*mob*, PCR amplification was done using forward primer F-5′-CCGACTGGGCTAAATCTGTG-3′ and reverse primer R-5′-CTCGTCCTGCAGTTCATTCA-3′. The primer was made from 5818 bp DNA of *Escherichia coli* transposon Tn5 (ACCESSION U00004 L19385, GI: 405822). For positive control 2 µl of plasmid pSUP5011 was taken and for negative control water was taken in place of template DNA. The reaction program was: 94 °C for 5 min and then 30 cycles each consisting of 30 sec at 94 °C, 30 sec at 52 °C, 1 min at 72 °C and finally at 72 °C for 7 min.

5.2.1.3. Determination of growth curve, biofilm formation and Minimum Inhibitory Concentration (MIC) of copper/nickel/ cadmium/ mercury/ zinc by *A. junii* BB1A and its metal sensitive mutant strain.

Growth curve of *A. junii* BB1A and its mutant was studied in BHI broth (the medium which supported maximum growth and biofilm formation: see chapter 1). Briefly 1% (v/v) of overnight grown culture was used to inoculate in 50 ml media and incubated at 30 °C with shaking (130 rpm). Optical densities of the cultures were determined in 3 h interval using spectrophotometer (Varian, Agilant technologies, Germany) at 600 nm. Biofilm formation by both strains was compared by using the standardized tissue culture plate method (TCP method) as described in chapter 1. MIC for copper or nickel or cadmium or zinc or mercury was studied by growing BB1A cells (wild or mutant) in AB minimal medium (metal non-complexing medium) containing different concentrations of copper or nickel or cadmium or zinc or mercury (as metal salts; $CuSO_4.5H_2O$, $NiCl_2.6H_2O$, $CdCl_2.H_2O$, $ZnSO_4.7H_2O$, and $HgCl_2$). Each experiment was repeated three times and MIC was determined as the lowest concentration of metal that completely inhibited growth after 48 h of incubation.

5.2.1.4. Isolation of genomic DNA and amplification of sequences flanking a Tn5 insertion in the chromosome of mutant cell

Isolation of total genomic DNA: Total cellular DNA preparation from mutant cells was made following modification of Marmur's procedure (Marmur, 1961) as described by Yates and Holmes (1987). Cells were grown in LB medium at 30 °C up to late log phase. Cells were harvested from 200 ml of the culture by centrifuging at 10,000 rpm for 10 min at 4 °C. The cell pellet was washed with distilled water and re-suspended in 2 ml of 0.15 M NaCl-0.1 M EDTA (pH 8.0) and the concentrated solution was

frozen at -20°C for at least 4 h. The frozen cells were rapidly thawed at 55 °C and lysozyme (1.5 mg/ml) was added. The mixture was then incubated at 37 °C for 15 min followed by the addition of SDS to the final concentration of 1% (w/v). The mixture was incubated at 55 °C for 1 h. Proteinase K was then added (5 mg/ml) and the incubation was continued at 55 °C until the solution became clear. The cell lysate was extracted with 2 ml of phenol: chloroform mixture (1:1) and then with chloroform at 4 °C. Finally the aqueous phase was collected in a test tube placed in ice. 1/10 volume of 3 M Na-acetate and double volume of chilled ethanol was then added. The DNA was spooled with a sterile bent glass rod. The DNA obtained was air dried and suspended in 1 X TE buffer (10:1). For quantification of DNA a UV-vis spectrophotometer (Thermospectronic) was used, and absorbance at 260 and 280 nm.

Inverse PCR (IPCR) strategy: The IPCR process used for amplification of flanking sequences of the Tn5 insertion is adapted from Huang et al. (2000). To amplify genomic DNA sequences flanking a Tn5 insertion in the chromosome of *A. junii* BB1A strain, circular substrates were generated by ligating *Eco*RI-digested genomic DNA. Tn5 was contained intact within one such circular molecule, as the transposon sequence is devoid of sites for cleavage by *Eco*RI. Genomic DNA was digested with *Eco*RI (5 U/µg of DNA) at 37 °C for the minimum time necessary to achieve complete digestion (2 h). Following phenol: chloroform extraction and ethanol precipitation, the digested DNA was selfligated, at a concentration of 0.3±0.5 µg/ml and in the presence of 3 U/ ml T4 DNA ligase (Promega) overnight at 15 °C. This low DNA concentration favours intramolecular circularization. The ligation mixture was extracted by phenol: chloroform (1:1), precipitated with ethanol, and resuspended in sterile distilled H_2O. The total DNA of *A. junii* BB1A without enzyme digestion was used as a control template for IPCR.

PCR amplification: A single oligonucleotide 5' GTTAGGAGGTCACATGG 3' [nucleotides 63 to 79 (complementary) and 5740 to 5756 of Tn5; Gene Bank accession no. U00004 L19385 (Berg, 1989; Mukhopadhyaya et al., 2000)] complementary to and extending outward from the ends of the inverted repeat of Tn5 was used to prime DNA synthesis in the polymerase chain reaction. The amplified product consisted of interrupted DNA sequences from *A.junii* BB1A, adjacent to both ends of the transposon insertion. PCR amplification was performed using 'PCR Amplification Kit' (GENEI, India), in 50 µl reaction volume. Each 50 µl PCR mix contained:-

3 µl of 10 mM dNTP mix,

5 µl of 10 X buffer containing 15 mM MgCl2,

12.5 pmol of primer (inverse primer single oligonucleotide),

10 ng of template DNA and

1U *Taq* DNA polymerase

Volume makeup with double distilled water

Taq DNA polymerase was added after adding all the ingredients of the mixture. The PCR was done in a GenAmp PCR system (Applied Biosystems). The reaction program was:-

94 °C for 5min and then

30 cycles each consisting of

94 °C for 30 sec,

55 °C for 30 sec , and

68 °C for 3 min

Finally at 68 °C for 7 min

Purity of DNA was tested with the help of UV spectrophotometer by determining the OD values at 260 and 280 nm.

Agarose gel electrophoresis: The amplicon was checked on 2% agarose.

5.2.1.5. Cloning and sequencing of the PCR product

Extraction of PCR product from gel: PCR product from the gel was extracted by freeze-thaw method (Tautz and Renz, 1983). Agarose gel slice containing the DNA fragment (PCR product) was cut out and transferred into 1.5ml tube. To this 50 µl of TE buffer was added and kept at -20 °C for 18 h. Centrifuged for 20min. at 10000 rpm and the supernatant was collected in a fresh tube. To the gel pellet, another 50 µl of TE buffer was added and freezing step was repeated, centrifuged. The supernatant was collected in 1.5 ml tube. The DNA from the pooled supernatant was precipitated with 2.5 volumes ethanol according to standard protocols (Sambrook et al., 1989).

Ligation of PCR product: The gel purified PCR products was cloned in pGEM®T-easy vector system II (Fig. 5.2) following company's protocol (Promega, USA). The vector (25 ng) and the insert DNA were taken in the molar ratio of 1:2 and were suspended in 4 µl sterile double distilled water. 5 µl of 2 X rapid

ligation buffer and 1 µl of T4 DNA ligase were added and thoroughly mixed by vortexing and centrifugation. The mixture was kept at 4 °C for at least 16 h, heated at 60 °C for 10 min and 5-7 µl was used to transform competent *E. coli* cells.

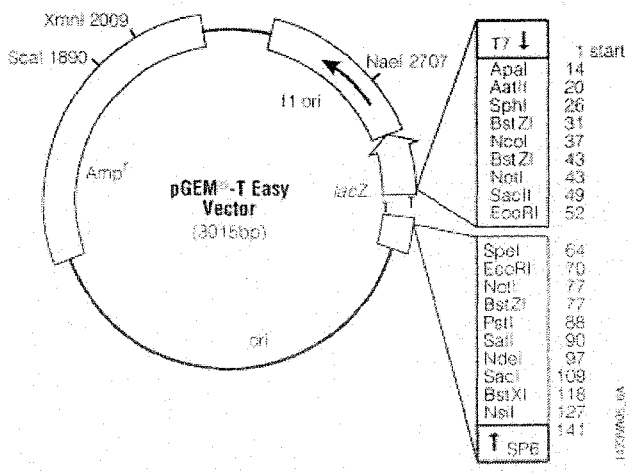


Fig. 5.2: pGEMT-easy vector system and its cloning site. (From www.promega.com)

Transformation of recombinant plasmids

Preparation of competent E. coli cells using $CaCl_2$: A single colony of *E. coli* JM109 from LB agar plate (freshly grown for 16-20 hours at 37°C) was taken and transferred into 10 ml of LB broth in a 100 ml flask. The culture was incubated for approximately 3 hours or more at 37°C with vigorous shaking to attain the viable cell number of 10^8 cells/ ml. The rest of the procedure is same as discussed in section 5.2.1.2.

Transformation: The procedure for transformation of the recombinant plasmid into the host *E. coli* JM109 was same as described for pSUP5011 in section 5.2.1.2. However the screening of successful clone was done by blue white screening on LB medium supplemented with X-gal and IPTG.

Identification of bacterial colonies that contain recombinant plasmids: The transformed cells were plated on LB agar plate containing 50 µg/ ml ampicillin. 0.5 mM isopropylthiogalactoside (IPTG) and 80 µg/ µl X-gal (5-bromo-4-chloro-3-indolyl-β-galactopyranoside) were added to ampicillin containing LB plates prior to spreading of transformed *E.coli* JM109 cells. The plate was incubated at 37 °C overnight. It was possible to recognize colonies that carried putative recombinant plasmids by bluewhite screening of the colonies. Insertion of foreign DNA into the polycloning site of plasmid vectors leads to the incapability of complementation. Bacteria carrying recombinant plasmid vector therefore produce

white colonies. However, bacteria containing re-circularized vectors (without any insert) could utilize chromogenic substrate X-Gal and thus form blue colonies. Recombinant plasmid from successful bacterial clones was isolated by alkaline lysis method described in section 5.2.1.1 and further checked by *Eco*RI restriction digestion followed by agarose gel electrophoresis.

DNA sequencing: The recombinant plasmid was directly used for sequencing of the inserts using primers for T7 and SP6 promoters. Nucleotide sequencing was performed with the ABI PRISM^{TM} Dye Terminator Cycle Sequencing Ready Reaction Kits (Perkin-Elmer) using specific primer and the reaction was analyzed in an 'ABI PRISM 377 DNA Sequencer'.

5.2.1.6. Sequence Analysis

Similarity search: For similarity searches, standard nucleotide BLAST (blastn suite) and standard protein BLAST (blastp suite) programs were used (Altschul et al., 1997) from the NCBI (National Center for Biotechnology Information) website http://www.ncbi.nlm.nih.

5.2.2. Application of Bioinformatics

5.2.2.1. Designing degenerate primer(s) for *luxI* homologues

Retrieval of luxI homologous sequences: Nucleotide sequences of *luxI* gene homologues in gamma-proteobacteria were retrieved from NCBI gene data bank (Table 5.4).

Table 5.4: List of gamma proteobacteria having *luxI* homologue sequences.

Gamma proteobacteria		NCBI Gene ID Coding for Acyl-homoserine lactone/ Autoinducer synthase	NCBI Gene ID Coding for 16s Ribosomal RNA
Aeromonadales	*Aeromonas salmonicida subsp. salmonicida* A449	4994878	4997900
	Aromonas hydrophila subsp hydrophila ATCC 7966	4488532	4490326
Pseudomonadales	*Pseudomonas syringae pv. tomato* DC3000	1185535	1182616
	Pseudomonas aeruginosa PAO1	881777	3240272
	Acinetobacter baumanni ATCC17978	4918338	5075456
Enterobacteriales	*Edwardsiella tarda* EIB 202	8608939	8609999
	Yersinia pestis CO92	1175290	1173916
	Yersinia pseudotuberculosis PB1/+	6257755	6261415
	Erwinia amylovora CFBP1430	8912769	8913707
	Erwinia pyrifoliae Ep1/96	8538907	8888002
	Erwinia tasmaneinsis Et 1/99	6298833	6299716
	Serratia proteamaculans 568	5602840	5603146
	Candidatus Hamiltonella defensa 5AT *(Acyrthosiphon pisum)*	7950473	7951350
	Pectobacterium atrosepticum SCRI1043	2884752	2883438
	Pectobacterium carotovorum subsp. Carotovorum PC1	8135146	8131134
	Pectobacterium wasabiae WPP163	8532872	8528550
	Sodalis glossinidius str. 'morsitans'	3868591	3866445
	Dickeya zeae Ech 1591	8117052	8118861
Vibrionales	*Vibrio fisheri* MJ11	6808033	6808213
	Photobacterium profundum SS9	3119949	3120506
Acidithiobacillales	*Acidithiobacillus ferrooxidans* ATCC 53993	6877646	6876510
Chromatiales	*Halothiobacillus neapolitanus c2*	8534107	8535301

Multiple Sequence alignment: Sequences from the selected genera were aligned using ClustalW multiple alignment program in SeaView 4 (http://pbil.univlyon1.fr/software/SeaView).

Designing degenerate primer: A phylogenetic tree of Acyl Homoserine lactone synthase gene sequences was prepared using distance method in SeaView 4 software. This was used as the guide tree to select genera of closely related sequences for designing usual primers or wherever necessary the degenerate primers. For degenerate primer, sequences of each cluster in the tree were separately aligned with

ClustalW in SeaView 4. Most similar stretches of nucleotides having minimum 10 A+T+C+G in the multiple alignments were tentatively chosen to test the properties close to an ideal primer for PCR reaction (Fig. 5.3 a and b).

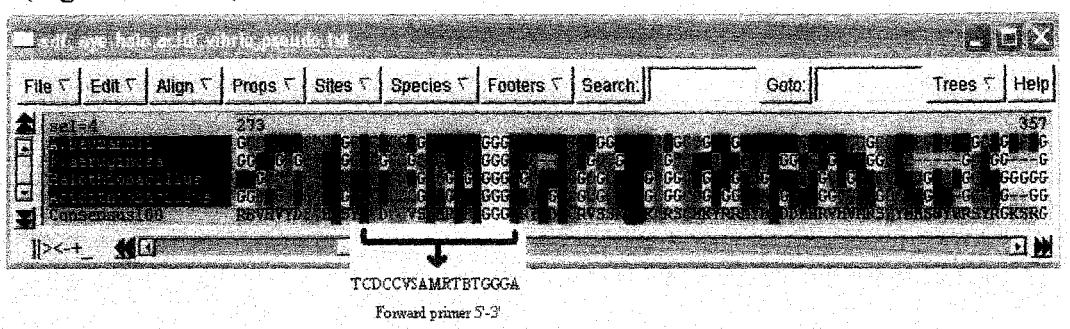


Fig. 5.3: a) Snapshot of SeaView window displaying ClustalW aligned *luxI* homologue sequences from gamma proteobacteria with degenerate letters just below the aligned sequences, which can be used to design degenerate primer (forward primer).

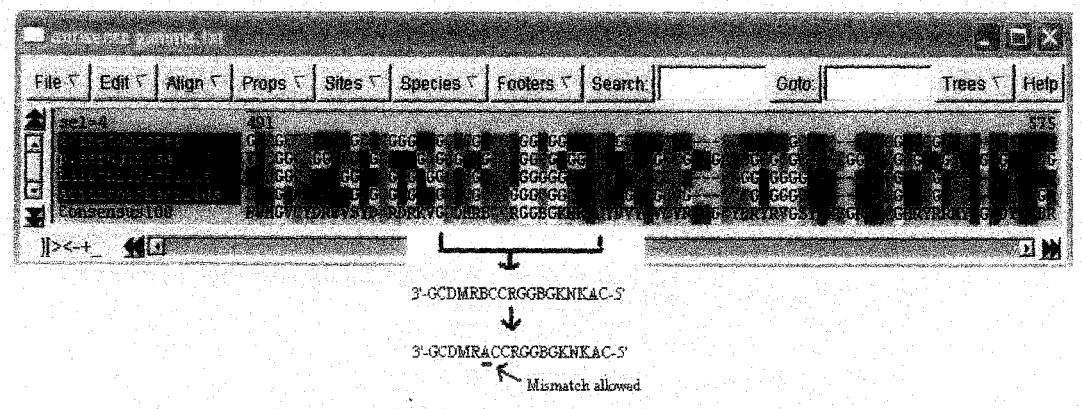


Fig. 5.3: b) Snapshot of SeaView window displaying ClustalW aligned *luxI* homologue sequences (antisense) from gamma proteobacteria with degenerate letters just below the aligned sequences, which can be used to design degenerate primer (reverse primer).

Fast PCR (http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm) was used to get the complementary sequence for making the reverse primer. The primers were basically analyzed for melting point, %GC, hairpin, self complementarities and primer dimer check using OligoCalc (www.basic.northwester.edu/biotools/oligocalc.html) and Oligo Analysis Tool (http://www.operon.com/technical/toolkit.aspx). Attention was also paid to have 3GC clamps at 3′ end for increasing the specificity of the primer. Using 100% consensus, degenerate sequences (letter codes were followed according to the recommendations of Nomenclature Committee of the International Union of Biochemistry and Molecular Biology) were constructed from the aligned sequence.

Insilico PCR Amplification: The designed degenerate primer pair was fed to the blank primer box followed by selection of bacterium (whose whole genome sequence is available in the database) in the interacting page of the online software, *In silico* PCR, for the output.

Wet lab study

Isolation of genomic DNA: As described in the section 5.2.1.4.

PCR amplification of DNA using degenerate primer (Touch down PCR): PCR amplification was performed using PCR Amplification Kit (GENEI, India), in 50 µl reaction volume, following instructions provided by the supplier. Each 50 µl PCR mix contained; 3 µl of 10mM dNTP mix, 5 µl of 10X buffer containing 15 mM $MgCl_2$, 12.5 pmol of each forward and reverse primer(degenerate primer), 10 ng of template DNA and 1U *Taq* DNA Polymerase. Taq DNA polymerase was added after adding all the ingredients of the mixture. The PCR was done in a GenAmp PCR system (Applied Biosystems). Since the primer was degenerate and has no single melting/ annealing temperatre, touch down PCR was performed. Touchdown PCR protocol consisted of two phases:

Phase 1:
94 °C for 8 min, followed by
30 cycles of
Denaturation at 94 °C for 1 min,
Annealing at variable temperatures for 1 min, and
Extension at 72 °C for 1 min

In the first cycle, the annealing temperature was set to 60 °C and, at each of the 29 subsequent cycles; the annealing temperature was decreased by 0.5 °C (i.e., it varied from 60 to 45 °C at 0.5 °C decrements along the 30 cycles).

Phase 2:
10 cycles of
94 °C for 1 min,
45 °C for 1 min, and
72 °C for 1 min
Finally at 72 °C for 7 min

Agarose gel electrophoresis: The amplicon was checked on 2% agarose.

Cloning, sequencing, and analysis of the PCR product (see section 5.2.1.5 and 5.2.1.6)

5.2.2.2. Amplification of *pgaC* gene from *A. junii* BB1A

Many species contains *pgaC* gene which encode glycosyltransferases. Literature shows that PgaC is a polysaccharide polymerase that uses UDP-GlcNAc as substrate (Wang et al., 2004) and the locus *pgaABCD* is critical for the formation of biofilm adhesin in *E.coli* (Wang et al., 2004) as well as in *A. baumannii* (Choi et al., 2009). For the PCR amplification of *pgaC*, primer was constructed from *A. junii* SH205 (nucleotide sequence 2185-3419, NCBI Accession no. ACPM01000018.1), showing 78% similarity with the 'biofilm PGA synthesis N-glycosyltransferase PgaC' of *A. baumanii* (NCBI gene ID: 7060465). The primer pair (Forward 5'-GAACGCCAAGACCCACCTTA-3' and Reverse 5'-TACCCGCTTGCAATCCCATT-3') was designed using NCBI Primer BLAST program. Genomic DNA from *A. junii* BB1A was isolated by the procedure described in section 5.2.1.4. PCR amplification was performed using 'PCR Amplification Kit' (GENEI, India), in 50 µl reaction volume. Each 50 µl PCR mix contained:-

3 µl of 10 mM dNTP mix,

5 µl of 10 X buffer containing 15 mM $MgCl_2$,

12.5 pmol of primer (inverse primer single oligonucleotide),

10 ng of template DNA and

1U *Taq* DNA polymerase

Volume makeup with double distilled water

Taq DNA polymerase was added after adding all the ingredients of the mixture. The PCR was done in a GenAmp PCR system (Applied Biosystems). The reaction program was:-

94 °C for 5min and then

30 cycles each consisting of

94 °C for 30 sec,

55 °C for 30 sec, and

72 °C for 1 min 30 sec

Finally at 72 °C for 7 min

Agarose gel electrophoresis: The amplicon was checked on 2% agarose. The procedure was similar to that described in section 5.2.1.1.

5.2.2.3. Structural and functional prediction of putatively homologous Pga proteins in *A. junii*

As the Pga proteins (PgaA, PgaB, PgaC, and PgaD) from diverse bacterial genera have been studied with respect to their role in the biofilm formation and regulation, it became necessary to look for the presence of similar proteins or genes in *A. junii.* Moreover one important molecule, N-acetyl-glucosamine probably encoded by the *pgaC* gene, has also been detected in the test strain BB1A (see chapter 2, section 2.3.16). This further supports the study of the genes or proteins responsible for the biosynthesis of N-acetyl-glucosamine. Since the present test strain BB1A has not been sequenced, the strain SH205, whose whole genome shotgun sequences is available on NCBI database, was used as the test/ equivalent model strain for A. junii.

Primary sequence analysis: Recently published *A. junii* SH205 genome on NCBI database was searched for gene loci similar to *pga* protein of *A. baumannii* (accession no. ACV90433.1, ACV90434.1, ACV90435.1, and ACV90436.1). ProtParam (Gasteiger et al., 2005) and SAPS (Brendel et al., 1992) was used for the primary sequence analysis. ProtParam estimates various physicochemical properties of protein like molecular weight, theoretical pI, and grand average of hydropathicity (GRAVY). SAPS (statistical analysis of protein sequences) evaluate a number of protein features like charge-clusters, hydrophobic regions, compositional domains etc.

Similarity search: For similarity searches, standard nucleotide BLAST (blastn suite) and standard protein BLAST (blastp suite) programs were used (Altschul et al., 1997) from the NCBI (National Center for Biotechnology Information) website http://www.ncbi.nlm.nih.

Sub-cellular localization: The Sub-cellular localization of proteins were predicted using ProtCompB (http://www.softberry.com), CELLO v.2.5 (Yu et al., 2004), and PSLpred hybrid module (Bhasin et al., 2005). ProtCompB uses combination of several methods such as linear discriminant function-based predictions, direct comparison with homologous proteins of known localization, prediction of functional peptide sequences etc., to identify the sub-cellular localization of bacterial proteins. CELLO v.2.5 is a subcellular localization prediction server (http://cello.life.nctu.edu.tw/) that identifies subcellular location of either DNA or protein sequences of gram negative, gram positive bacteria, or eukaryotes.

PSLpred is an SVM (support vector machine) based method to predict subcellular localization (cytoplasm, outer membrane, periplasm, inner membrane and extracellular) of prokaryotic protein (from gram negative bacteria).

Signal peptide prediction: Proteins that are meant for secretory pathways are usually synthesized with signal sequences (a short peptide present at the N-terminal), termed signal peptides (Blobel and Dobberstein, 1975), which direct the protein to a specific transporter complex in the membrane. The signal sequence is subsequently recognized and cleaved by signal peptidase during the export event. Signal peptide prediction was done using ProtCompB, SMART (Simple Modular Architecture Research Tool) (Letunic et al., 2012) and iPSORT (Bannai et al., 2002). iPSORT server (http://ipsort.hgc.jp/) predicts the signal peptides by applying the stored rules for various sequence features of known protein sorting signals.

Topology prediction: The Topology of proteins was predicted using TOPCONS server (http://topcons.net/), which combines an arbitrary number of topology predictions [OCTOPUS (Viklund and Elofsson, 2008), MEMSAT3 (Jones, 2007), SCAMPI (Bernsel et al., 2008), HMMTOP (Tusnady and Simon, 2001) and PRO- and PRODIV-TMHMM (Viklund and Elofsson, 2004)] into one consensus prediction and quantifies the reliability of the prediction based on the level of agreement between the underlying methods, both on the protein level and on the level of individual transmembrane regions.

Domains/patterns/motifs prediction: SMART (Simple Modular Architecture Research Tool) (Letunic et al., 2012) was used to recognize the presence of any domains in the proteins. PPSearch (http://www.ebi.ac.uk/Tools/ppsearch/), which uses PROSITE (Hulo et al., 2006) pattern database, was used to predict the protein motifs and functionally relevant patterns in the protein.

Protein protein interaction studies: In order to build a model showing the functional and physical interactions between the proteins, an online availabe program 'STRING' version 9.0 was used. STRING is a database of experimentally known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations. They are derived from four sources: (1) genomic context; (2) high-throughput experiments; (3) co-expression (conserved) and (4) known knowledge (Szklarczyk et al., 2011). This helps to determine the functional and physical interaction of the proteins involved herewith and also reveals the neighborhood display i.e., genes in the immediate neighborhood on the genome (within 300 bp on the same strand). As it has been known that groups of genes required for the same function have a propensity to show similar species coverage, are often

located in close proximity on the genome (in prokaryotes), and tend to be involved in gene-fusion events.

Generation of 3D structure of glycosyl transferase through homology modeling: Glycosyl transferase from *A. junii* SH205 showing significant similarity with the 'biofilm PGA synthesis N-glycosyltransferase PgaC' of *A. baumannii* was used for the homology modeling. Glycosyltransferases are the key enzymes that synthesize oligosaccharides, polysaccharides, and glycoconjugates by transferring the sugar moiety from an activated nucleotide-sugar donor to an acceptor molecule, which may be a growing oligosaccharide, a lipid, or a protein. HHpred (Homology detection & structure prediction by HMM-HMM comparison) (Söding et al., 2005) server (http://toolkit.tuebingen.mpg.de/hhpred) was used to find suitable template for building 3D model. It searches through the PDB database of proteins with solved 3D structure (Söding et al., 2005). The output of HHpred and HHsearch is a ranked list of database matches (including E-values and probabilities for a true relationship) and the pairwise query-database sequence alignments. The PDB ID was then used as a template in Swiss-Model (http://swissmodel.expasy.org/) (Schwede et al., 2003) to model the 3D structure for the protein of interest. Structure visualization was done in Swiss- PDB Viewer. Structure validation was carried out by using the program ProQ - Protein quality prediction server (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) (Wallner and Elofsson, 2003). The 3-D structure of PgaC homologue protein from *A. junii* SH205 was also compared with those of *E.coli* and *A. baumannii*.

Molecular docking: Molecular docking is a bioinformatics tool which uses the topographical features of protein and ligand and fits the two molecules in favorable configuration. The tool is primarily used for modeling protein-ligand interactions and has been used to study the structural basis of biological functions. The docking of glycosyl transferase (homologous to PgaC protein), from *A. junii* SH205, was performed against the ligand; UDP-GlcNAc (UDP-N-acetylglucosamine). The Graphical User Interface program 'AutoDock 4' (Morris et al., 1998), freely available for download at http://autodock.scripps.edu/downloads/, was used to prepare, run, and analyze the docking simulations. The ligand was drawn in ChemDraw Ultra 8.0 (ChemOffice package) assigned with proper 2D orientation. Energy of the molecules was minimized using UFF calculation in ArgusLab. The necessary hydrogen atoms, Kollman charges, and solvation parameters were then added to the modeled protein structure using the Autodock Tool. The energy minimized model was then read as input in the AutoDock, in order to carry out the docking simulation. Gasteiger charge was assigned and then non-polar hydrogens were merged. As the catalytic and binding site of the protein is not reported in the

literature, a grid box was created using the Autogrid program that covers the entire protein site and accommodates the ligand to move freely. The grid box size was set at 90, 90, and 90 Å for x, y and z respectively, and the grid center was set to 30.59, 15.822 and 3.497 Å for x, y and z respectively. Phe-215 (residues) was chosen for flexible docking given the minimum binding energy in comparison to other residues present in core of the protein. The Lamarckian Genetic Algorithm (LGA) was chosen to search for the best conformers. During the docking process, a maximum of 10 conformers was considered. Based on the free energy bonding data, out of 10-model result, one best model was picked up to analyze its interactions. The residues within 3Å of the ligand in complex structure were determined by using ArgusLab. The Ligplot of the interaction between UDP-GlcNAc and the proteins was generated by the PDBsum. This ligand-protein interaction, derived from *A. junii*, was also compared with that of *E. coli* and *A. baumannii*.

5.2.2.4. Phylogenetic distribution of Pga homologues.

Protein databases such as UniProt (http://www.ebi.ac.uk/uniprot/), PFAM (http://www.sanger.ac.uk/Sofware/Pfam/) and major sequence repositories such as National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), Ensembl (http://www.ensembl.org) and KEGG (http://www.genome.ad.jp/kegg/) provide direct access to many known, full-length *pga* gene sequences. Additional sequences were retrieved using bioinformatic search tools such as BLAST. The PgaA, PgaB, PgaC, and PgaD protein sequences spanning all major bacterial genera were assembled into a multiple sequence alignment using the ClustalW utility in Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.(Tamura et al., 2007). Phylogenetic tree construction was performed by MEGA4 neighbor-joining analysis of maximum likelihood distances for 500 bootstrap replicates.

5.3. Results

5.3.1. Transposon mutagenesis

Plasmid DNA (bearing Tn5 transposon having kanamycin resistane gene) was transformed into BB1A strain by $CaCl_2$ transformation method. Kanamycin resistant transformants were isolated on AB media supplemented with 50 µg/ ml Kanamycin (master plate). A single metal sensitive mutant (designated as BB1A:Tn5), failed to grow in presence of 1.5 mM $NiCl_2$, was isolated from master plate, by replica plate technique. Presence of Tn5 in the chromosomal DNA of BB1A:Tn5 was confirmed by the generation of a 1.6kb PCR product. Similar PCR product was obtained using plasmid pSUP5011 as positive control. No such product was found using genomic DNA of wild type BB1A or water as negative control.

Characterization of metal sensitive mutant

Minimum Inhibitory Concentration (MIC) to metal ions of the mutant BB1A:Tn5 in liquid AB medium was found to be 2.5 mM $NiCl_2$ or 2 mM $CuCl_2$ or 0.01 mM $CdCl_2$ or 0.001 mM $HgCl_2$ or 0.5 mM $ZnCl_2$ (Table 5.5). Comparison of growth curves of BB1A and its mutant BB1A:Tn5 in BHI medium enabled to identify the variations (Fig 5.4). Significant increase in cell yield was noted in mutant. The maximum OD_{600} recorded for wild type was 0.6 while it recorded 0.7 in case of mutant BB1A:Tn5. When biofilm formation in BHI grown cells was measured by TCP method, no biofilm formation was detected in the mutant strain (Fig 5.5).

Table 5.5: Comparison of wild type BB1A and its transposon mutant BB1A:Tn5

	MIC	
	BB1A	**BB1A:Tn5**
Cu^{2+}	4.0 mM	2.0 mM
Ni^{2+}	4.0 mM	2.5 mM
Cd^{2+}	0.5 mM	0.01 mM
Zn^{2+}	3.0 mM	0.5 mM
Hg^{2+}	0.010 mM	0.001 mM

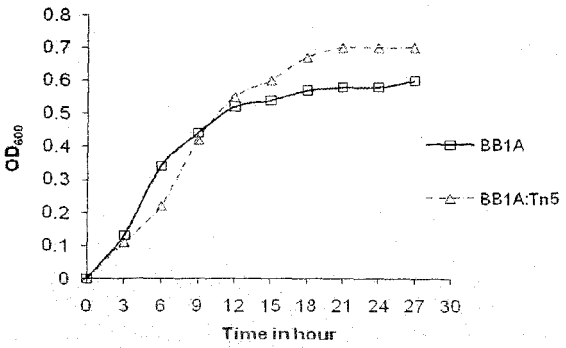


Fig. 5.4: Characteristic growth curves of wild type BB1A and its transpon mutant BB1A:Tn5 in BHI medium.

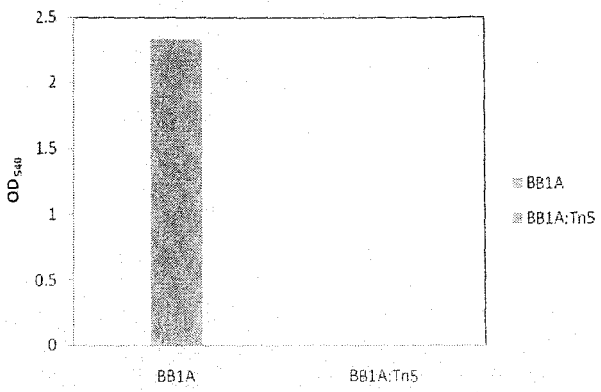


Fig. 5.5 : Quantification of biofilm in wild type BB1A and mutant BB1A:Tn5 in BHI medium.

Isolation and amplification of genomic DNA sequences flanking a Tn5 insertion in the chromosome of mutant cell

Inverse PCR was done to amplify the DNA sequences flanked at the ends of transposon. A single 550 bp amplicon (Fig 5.7) was generated by using the single primer (5' GTTAGGAGGTCACA 3').

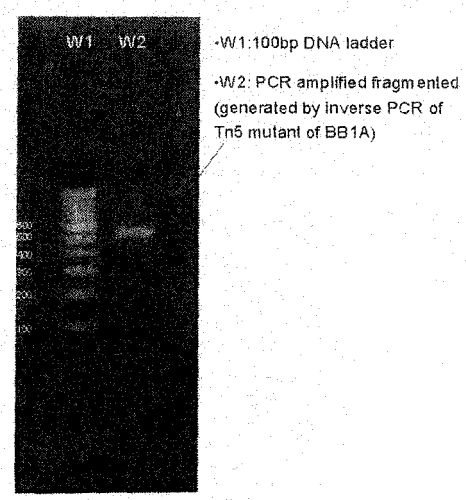


Fig.5.7: PCR amplified product obtained by Inverse PCR.

The PCR product was purified and cloned in pGEM T-easy vector prior to sequencing.The recombinant plasmids (pGEM T-easy vector carrying the insert) were transformed into the competent cells of *E.coli* JM109 and the clones containing recombinant plasmids were detected using blue-white screening method in X-gal containing plates. White colonies are picked up, plasmid isolated and digested with *E.co*RI to confirm the presences of insert. Sequencing of the insert by T7 primer revealed the following

sequence (Fig 5.6). The sequence obtained was cut at the *E.co*RI restriction site and is inverted and joined (Fig.5.7). The actual sequence is thus the invert of sequence generated by Inverse PCR. Similarity search by standard nucleotide BLAST revealed no significant result.

GTTAGGAGGTCACACATGGTCAGATGGTCATGAAGAGCAAACAGCAGTTAAAAGAAGATTTTGGTATCTCAGCAAAGCCT
ATTGCCATGTTTAAGTAAGTGGTTCATCCCCTTTTTGCCTCGGCTATGAAGGGGATAGCCAACAATGACCCCACGACTCAT
CTCAATTATCTCGCCATGACCTGGCTAAGTCCCAGTCCGATCTCTTGCCGTCCCGTTCACACCTGAATGACTTCTCCTGAAC
TTAAGTCTCAGACCTATTTGGCCGGTAATCCCTCTCGAATCTTGCTCTTTTTGCGCCCCGCCTTCCATCAACCTGTAACAGA
CCCAGCCATTTTCTGCGTAATTGTGATCTGAGTTGTAAAACGCAGATCATTATCTGTGTTTTACAGATTTAATGTTGACCTC
AAAAATCATGTGGGTAAGTATGATGGCCATCTAAGCAAAACTAAACACAGAGGATGACGCAATGAAAGCACTGGCTCGGT
TTGGCAAGGCCTTTGGCGGCTACAAGATGATTGATGTCCCACAACCCATGTGTGACCTCCTAAC

Fig. 5.6: The 551 bp amplicon (generated by inverse PCR) sequences. Underline portion is the *Eco*RI restriction site.

GCTCTCCCTAATGGCCGGTTTATCCAGACTCTGAATTCAAGTCCTCTTCAGTAAGTCCACACTTGCCCTGCCGTTCTCTAGC
CTGACCCTGAATCGGTCCAGTACCGCTCTATTAACTCTACTCAGCACCCCAGTAACAACCGATAGGGGAAGTATCGGCTCC
GTTTTTCCCCTACTTGGTGAATGAATTTGTACCGTTATCCGAAACGACTCTATGGTTTTAGAAGAAAATTGACGACAAACG
AGAAGTACTGGTAGACTGGTACCCAACACCCTGTAGTTAGTAGAACATCGGCGGTTTCCGGAACGGTTTGGCTCGGTCACG
AAAGTAACGCAGTAGGAGACACAAATCAAAACGAATCTACCGGTAGTATGAATGGGTGTACTAAAAACTCCAGTTGTAAT
TTAGACATTTTGTGTCTATTACTAGACGCAAAATGTTGAGTCTAGTGTTAATGCGTCTTTTACCGACCCAGACAATGTCCAA
CTACCTTCCGCCCCGCGTTTTTCTCGTTCTAA

Fig 5.7: Sequences from figure 5.8 has been cut at the *Eco*RI restriction site and rejoined in inverted position.

5.3.2. Degenerate primer designing

5.3.2.1. *Degenerate primers for luxI homologue genes in γ-proteobacteria*

Phylogenetic analysis of the retrieved *luxI* homologue sequences from gamma proteobacteria has revealed two distinct clusters (Fig. 5.8a) which also correlates with the phylogenetic tree drawn from their 16S rRNA gene sequences (Fig. 5.8b). Two degenerate primer pairs were designed using SeaView 4 and were designated as Deg1F/Deg2R and Deg3F/Deg4R and their properties are shown in Table 5.6.

Using the *in silico* PCR tool (allowing one nucleotide mismatch but in one nucleotide at 3' end) it was found that the degenerate primer pair, Deg1F/Deg2R was capable of amplifying *luxI* homologue sequences from *Acinetobacter baumanii, Pseudomonas aeruginosa, Halothiobacillus neapolitanus,* and *Acidithiobacillus ferrooxidans* (Table 5.7a) and Deg3F/Deg4R enabled to amplify *luxI* homologue sequences from several bacterial species of the genera *Erwinia, Yersinia, Serratia, Pectobacterium, Edwardsiella* and *Dickeya* (Table 5.7b).

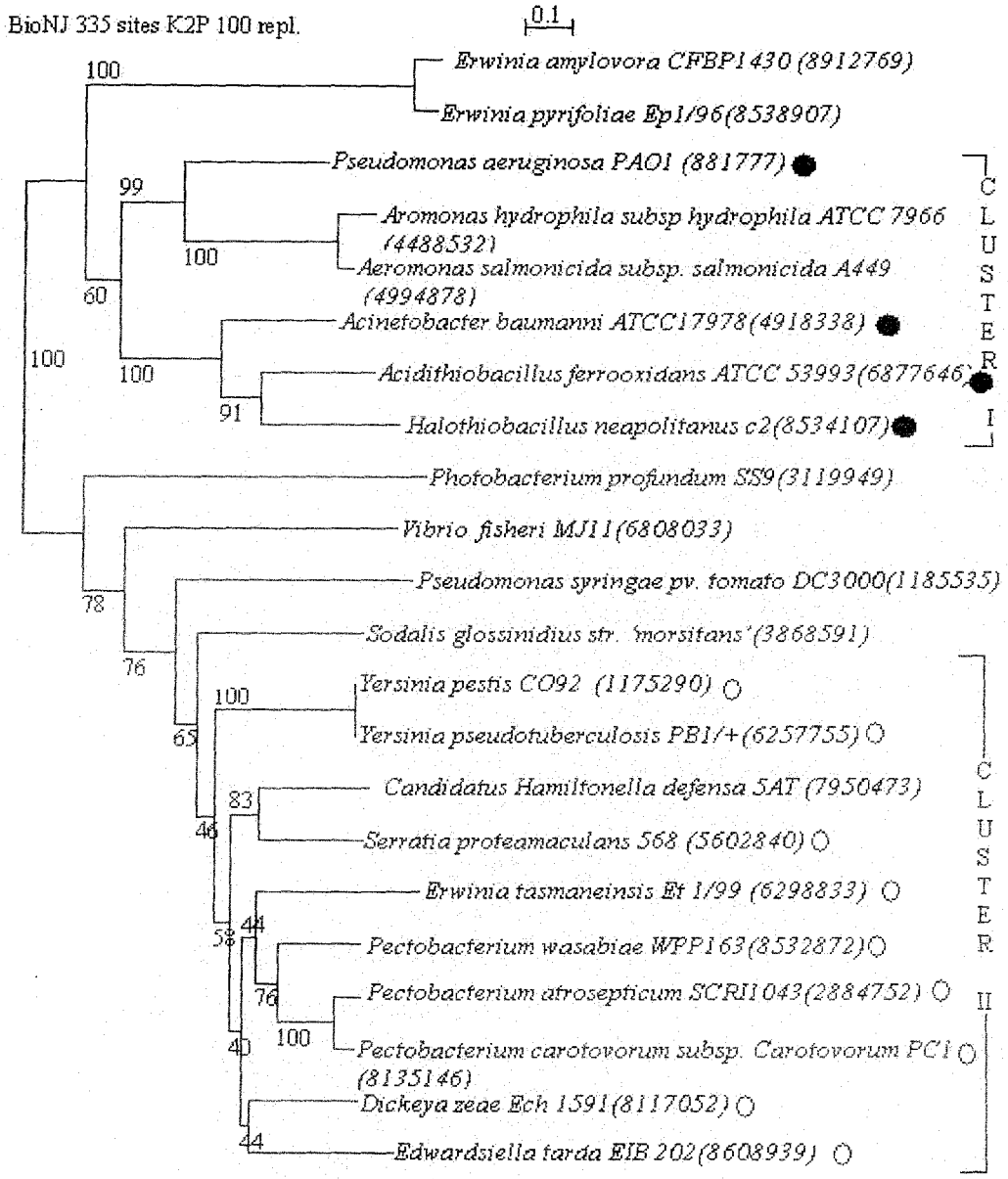


Fig. 5.8: a) Phylogenetic tree of *luxI* homologue sequences from gamma proteobacteria. Dendogram generated in SeaView version 4 using distance method and BioNJ algorithm (Gascuel 1997). NCBI gene IDs are given in parenthesis. Nucleotide-level distances are observed divergence with Kimura's two-parameter. Bootstrapped to 100 replicates. (•) *luxI* homologue sequences amplified with degenerate primer pair Deg1F/Deg2R, (○) *luxI* homologue sequences amplified with degenerate primer pair Deg3F/Deg4R.

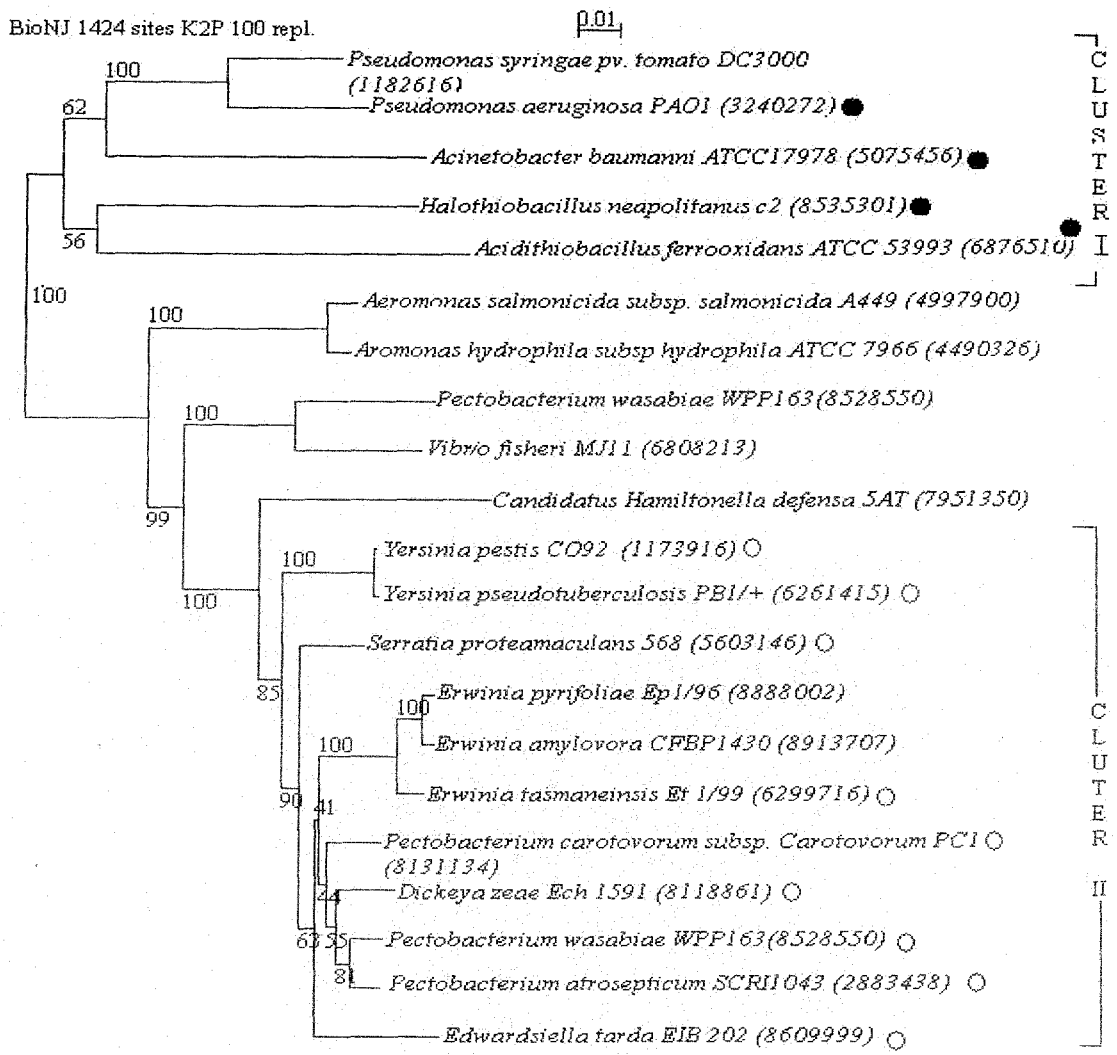


Fig. 5.8: b) Phylogenetic tree of 16S ribosomal RNA sequences from gamma proteobacteria. Dendogram generated in SeaView version 4 using distance method and BioNJ algorithm (Gascuel 1997). NCBI gene IDs are given in parenthesis. Nucleotide-level distances are observed divergence with Kimura's two-parameter. Bootstrapped to 100 replicates. (•) *luxI* homologue sequences amplified with degenerate primer pair Deg1F/Deg2R, (○) *luxI* homologue sequences amplified with degenerate primer pair Deg3F/Deg4R.

Table 5.6: Properties of designed degenerate primers.

Degenerate primers 5' to 3'	Length	degeneracy	TM in °C		% GC
			Oligoanalysis tool[a]	Oligocalc tool[b]	
Deg1F-TCDCCVSAMRTBTGGGA	17	216	58.8	42-54 (B),47-60 (SA)	58.8
Deg2R-CAKNKGBGGRCCARMDCG	18	1152	64.4	46-64(B), 51-70 (SA)	64.4
Deg3F-AARGAYMGDCTNCAHTGG	18	288	56.8	41-55 (B), 47-61 (SA)	56.8
Deg4R-AAYHYBCCADCCDGADC	17	972	57.1	40-57 (B), 45-62 (SA)	57.1

a=freely available at http://www.operon.com/technical/toolkit.aspx,

b=freely available at http://www.basic.northwestern.edu/biotools/oligocalc.html,

(B) = Basic, (SA) = Salt Adjusted

Table 5.7: a) *In silico* PCR amplification result using degenerate primer pair Deg1F/Deg2R.

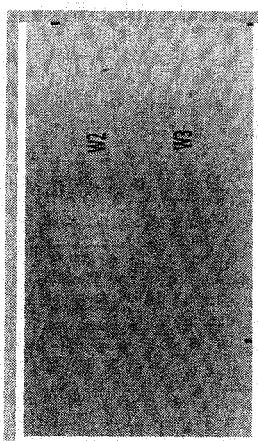
Organism	Amplicon	Sequence Name
Acinetobacter baumannii	238 bp, single band	Homoserine lactone synthase
Acidithiobacillus ferrooxidans	225 bp, single band	Autoindducer synthesis protein
Halothiobacillus neapolitanus c2	228 bp, single band	Acyl-homoserine-lactone synthase
Pseudomonas aeruginosa	213 bp, single band	Autoinducer synthesis protein *LasI*

Table 5.7: b) *In silico* PCR amplification result using degenerate primer pair Deg3F/ Deg4R.

Organism	Amplicon	Sequence Name
Pectobacterium atrosepticum SCRI1043	390 bp Single band	Acylhomoserine lactone synthase
Erwinia tasmaniensis	390 bp Single band	Acylhomoserine lactone synthase
Pectobacterium carotovorum subsp. carotovorum PC1	390 bp Single band	Acyl-homoserine-lactone synthase
Pectobacterium wasabiae WPP163	390 bp Single band	Acyl-homoserine-lactone synthase
Edwardsiella tarda	393 bp Single band	AHL synthase
Edwardsiella ictaluri	393 bp Single band	Autoinducer synthase putative
Dickeya zeae	390 bp Single band	Acylhomoserine lactone synthase
Dickeya dadanti	390 bp Single band	Acylhomoserine lactone synthase
Serratia preteamaculans 568	390 bp Single band	Autoinducer synthesis protein
Yersinia pestis	390 bp Single band	N-acylhomoserine lactone synthase
Yersinia entercolitica	390 bp Single band	N-acylhomoserine lactone synthase
Yersinia pseudotuberculosis	390 bp Single band	N-acylhomoserine lactone synthase

5.3.2.2. Wet lab study

PCR amplification of BB1A genome with the degenerate primer pair (Deg1F and Deg2R) resulted in a single amplicon (Fig 5.9a). Sequencing of the amplicon generated a 223 bp of nucleotide sequence (Fig. 5.9b).



a)
W2= 100bp ladder
W3=LuxI amplicon

b)
5'-
CAGATGAGGGCCAGCACGTGAAGAGTTAAAAGCGTTTATTTTAAAATTAAA
AAAATATGTACACGGCTAAACGTTTTTTGAGCTGCAATCTAGTCGATTTTCG
CTAGGATTGCAGCTGTTTTGTTTTTTGGGTATCAATATCCTACTTTATCATTA
TTGTCAGCTTTTTTATGGTTAGTGTGCTTGTTGATTATGGGATGGGCGCGCT
GGCCCTCTACTGA- 3'

Fig. 5.9: Nucleotide sequence (223bp) obtained from the amplicon generated with degenerate primer (Deg1F and Deg2R)

Table 5.8: Standard protein BLAST (blastn suite) of the amplicon showing similarity with the genus *Acinetobacter*.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
CP002080.1	Acinetobacter sp. DR1, complete genome	66.2	412	71%	1e-10	100%
CP002177.1	Acinetobacter calcoaceticus PHEA2, complete genome	60.8	339	75%	5e-09	100%
CP002522.1	Acinetobacter baumannii TCDC-AB0715, complete genome	31.9	124	36%	2.5	100%
CP001921.1	Acinetobacter baumannii 1656-2, complete genome	31.9	124	36%	2.5	100%
CP001172.1	Acinetobacter baumannii AB307-0294, complete genome	31.9	92.2	32%	2.5	100%
CP001182.1	Acinetobacter baumannii AB0057, complete genome	31.9	124	33%	2.5	100%
CP000863.1	Acinetobacter baumannii ACICU,complete genome	31.9	124	36%	2.5	100%
CU468230.2	Acinetobacter baumannii str. SDF, complete genome	31.9	152	34%	2.5	100%
CU459141.1	Acinetobacter baumannii str. AYE, complete genome	31.9	92.2	32%	2.5	100%
CP000521.1	Acinetobacter baumannii ATCC 17978, complete genome	31.9	216	34%	2.5	100%
CR543861.1	Acinetobacter sp. ADP1 complete genome	30.1	90.4	27%	8.7	100%
AB006902.2	Acinetobacter sp. NCIMB9871 cyclohexanol metabolic gene cluster (chnB, chnE, chnR, orf3, chnA, orf5, chnD, chnC), complete cds	30.1	30.1	9%	8.7	90%
AF282240.1	Acinetobacter sp. SE19 ChnZ (chnZ), pilin inverting protein (chnY), ChnX (chnX) genes, complete cds and cyclohexanol oxidation gene cluster, complete sequence	30.1	30.1	9%	8.7	90%

```
 57 atgtacacggctaaacgtatttttgagctgcaatctagtcgattt
      M  Y  T  A  K  R  I  F  E  L  Q  S  S  R  F
102 tcgctaggattgcagctgttttgtttttttgggtatcaatatccta
      S  L  G  L  Q  L  F  C  F  L  G  I  N  I  L
147 ctttatcaattattgtcagctttttttatggttagtgtgcttgttg
      L  Y  Q  L  L  S  A  F  L  W  L  V  C  L  L
192 attatgggatgggcgcgctggccctctactga 223
      I  M  G  W  A  R  W  P  S  T
```

Fig. 5.10: ORF obtained from the amplicon generated by LuxI degenerate primer. 55 amino acids long derived +3 frame. (MYTAKRIFELQSSRFSLGLQLFCFLGINILLYQLLSAFLWL /CLLIMGWARWPST)

Table 5.9: BLAST (protein-protein BLAST) analysis showing similarity of the amino acid sequences derived from the amplicon generated by LuxI degenerate primer, with that of other bacteria.

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
ZP_06066767.1	predicted protein [Acinetobacter junii SH205]	101	101	93%	3e-26	94%
ZP_03824368.1	conserved hypothetical protein [Acinetobacter sp. ATCC 27244]	43.5	43.5	86%	2e-04	48%
ZP_06728906.1	conserved hypothetical protein [Acinetobacter haemolyticus ATCC 19194]	43.5	43.5	86%	2e-04	48%
ZP_06057979.1	conserved hypothetical protein [Acinetobacter calcoaceticus RUH2202]	40.4	40.4	70%	0.002	51%
YP_004996779.1	unnamed protein product [Acinetobacter calcoaceticus PHEA-2]	38.1	38.1	70%	0.018	56%
ZP_06693154.1	conserved hypothetical protein [Acinetobacter sp. SH024]	37.7	37.7	70%	0.020	56%
ZP_04662514.1	hypothetical protein AbauAB_12918 [Acinetobacter baumannii AB900]	32.7	32.7	63%	2.2	46%
ZP_07238617.1	hypothetical protein AbauAB05_17406 [Acinetobacter baumannii AB058]	32.0	32.0	63%	2.2	46%
YP_001712417.1	hypothetical protein ABAYE0438 [Acinetobacter baumannii AYE]	32.3	32.3	63%	2.3	46%
ZP_05829274.1	conserved hypothetical protein [Acinetobacter baumannii ATCC 19606]	32.0	32.0	63%	3.8	43%
ABS90303.2	hypothetical protein A1S_3878 [Acinetobacter baumannii ATCC 17978]	32.0	32.0	63%	3.8	43%
YP_001847907.1	hypothetical protein ACICU_03248 [Acinetobacter baumannii ACICU] >gb|ACC58560.1| putative membrane protein [Acinetobacter baumannii ACICU]	31.6	31.6	63%	4.4	43%
EGT89049.1	hypothetical protein ABNIH1_17768 [Acinetobacter baumannii ABNIH1]	31.6	31.6	63%	4.6	43%
YP_308601.1	cysE gene product [Dehalococcoides sp. CBDB1]	32.3	32.3	89%	5.1	33%

Sequences producing significant alignments:						
Accession	**Description**	**Max score**	**Total score**	**Query coverage**	**E value**	**Max ident**
ZP_05824026.1	conserved hypothetical protein [Acinetobacter sp. RUH2624]	30.8	30.8	63%	9.4	43%

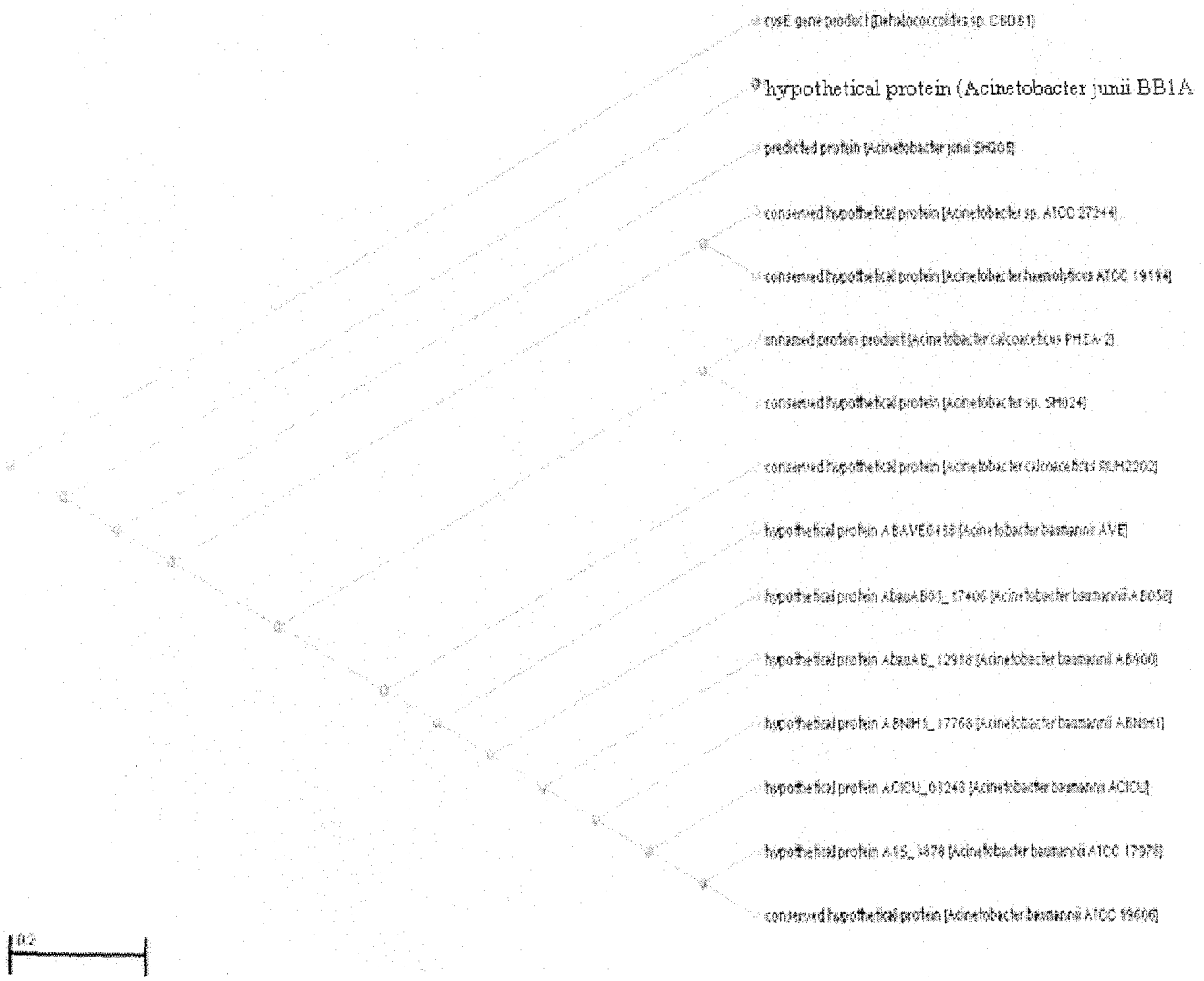


Fig. 5.10: Neighbour-joining phylogenetic tree based on amplicon sequences showing systematics. (Hypothetical protein product is from the amplicon sequence generated with degenerate primer designed for the amplification of *LuxI* homologue from *A. junii* BB1A)

5.3.3. Amlification of *pgaC* from *A. junii* BB1A

PCR amplification of BB1A genome with the primer pair (PgaC Forward and PgaC Reverse) resulted in a single amplicon (approximately 270 bp) (Fig. 5.11). The size of the amplicon was similar with that of *In silico* PCR amplification.

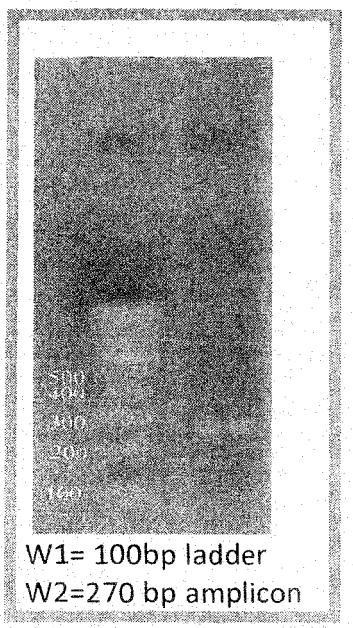


Fig. 5.11: Amplicon generated with the primer pair (PgaC Forward and PgaC Reverse).

5.3.4. Structural and functional prediction of Pga proteins from *A. junii*

BLAST (blastp suite) analysis at NCBI website, enabled the identification of four-gene locus in *A. junii* SH205 (the strain whose protein database is available), that shares similarity with various genetic loci encoding biofilm synthesis proteins in *A. baumannii* (Table 5.10).

Table. 5.10: Percentage maximum identity between proteins from *A. junii* and *A. baumanii*.

***A. baumanii* AB307-0294**	**% Identity**	***A. junii* SH205**	**% Identity**	***A. baumanii***
Biofilm PGA synthesis protein pgaA precursor (YP_002325205.1)	41 %	Biofilm synthesis protein (ZP_06065218.1)	41 %	PgaA (ACV90433.1)
Biofilm PGA synthesis lipoprotein PgaB (YP_002326570.1)	81 %	Polysaccharide deacetylase (ZP_06065265.1)	36 %	PgaB(ACV90434.1)
Biofilm PGA synthesis N-glycosyltransferase PgaC (YP_002326569.1)	89 %	Glycosyl transferase (ZP_06065266.1)	52 %	PgaC (ACV90435.1)
Biofilm PGA synthesis protein pgaD (YP_002325208.1)	43 %	Biofilm PGA synthesis protein pgaD (ZP_06065221.1)	43 %	PgaD (ACV90436.1)

In parenthesis: NCBI ref. seq. or accession number

Primary sequence analysis: Primanry sequence analysis of the protein sequences (homologues to PgaA, PgaB, PgaC, and PgaD) from *A. junii* SH205 revealed the molecular weight, theoretical pI and GRAVY (grand average of hydropathicity) index value.

Biofilm synthesis protein (ZP_06065218.1) is made of 811 amino acids with a predicted molecular weight of 95.2 KDa (SAPS tool) and an isoelectric point (pI) of 6.18 (ProtParam tool). The GRAVY (grand average of hydropathicity) index value-0.639 shows that it is probably a soluble protein (Kyte and Doolittle, 1982).

Polysaccharide deacetylase protein (ZP_06065265.1) is made of 665 amino acids with a predicted molecular weight of 76 KDa (SAPS tool) and an isoelectric point (pI) of 6.12 (ProtParam tool). The GRAVY index value -0.532 shows that it is probably a soluble protein (Kyte and Doolittle, 1982).

Glycosyl transferase (ZP_06065266.1) is made of 418 amino acids with a predicted molecular weight of 48.2 KDa (SAPS tool) and an isoelectric point (pI) of 8.57 (ProtParam tool). The GRAVY index value 0.172 shows that it is probably a hydrophobic protein (Kyte and Doolittle, 1982).

Biofilm PGA synthesis protein pgaD (ZP_06065221.1) is made of 154 amino acids with a predicted molecular weight of 18.6 KDa (SAPS tool) and an isoelectric point (pI) of 4.95 (ProtParam tool). The GRAVY index value 0.120 shows that it is probably a hydrophobic protein (Kyte and Doolittle, 1982).

Localization predictions: The cellular localization of all four proteins from *A. junii* SH205 was predicted using varios tools and is shown in table 5.11. All the tools gave similar result for a given protein. The result indicated that all the four proteins were putative membrane protein. This prompted to study the presence of signal peptide and potential cleavage sites in the protein (Table 5.12).

Table. 5.11: Subcellular localization of proteins from *A. junii*

Tools	**Protein sequence ZP_06065218.1 (Biofilm synthesis protein from *A. junii* SH205)**	**Protein sequence ZP_06065265.1 (Polysaccharide deacetylase protein from *A. junii* SH205)**	**Protein sequence ZP_06065266.1 (Glycosyl transeferase from *A. junii* SH205)**	**Protein sequence ZP_06065221.1 (Biofilm PGA synthesis protein PgaD from *A. junii* SH205)**
ProtCompB	Outer membrane protein (Integral Prediction of protein location: membrane bound Periplasmic)	Outer membrane (Integral Prediction of protein location: membrane bound Periplasmic)	Inner membrane (Integral Prediction of protein location: Inner Membrane)	Inner membrane (membrane bound Periplasmic)
CELLO v.2.5	Outer membrane protein	Periplasmic	Inner membrane	Inner membrane
PSLpred	Outer membrane protein	Periplasmic	Inner membrane	Inner membrane

Table.5.12: Signal peptide and cleavage position prediction of proteins from *A. junii*

Tools	**Protein sequence ZP_06065218.1 (Biofilm synthesis protein from *A. junii* SH205)**	**Protein sequence ZP_06065265.1 (Polysaccharide deacetylase protein from *A. junii* SH205)**	**Protein sequence ZP_06065266.1 (Glycosyl transeferase from *A. junii* SH205)**	**Protein sequence ZP_06065221.1 (Biofilm PGA synthesis protein PgaD from *A. junii* SH205)**
	Signal peptide (Cleavage position)	Signal peptide (Cleavage position)	Signal peptide (Cleavage position)	Signal peptide (Cleavage position)
ProtCompB	Present 1-15	Present1-28	Present 1-29	Present 1-47 (47)
iPSORT	Present 1-18 (no prediction)	Present (no prediction)	Present (no prediction)	Absent
SMART	Present 1-18 (18)	Present 1-28 (28)	Present 1-29	Absent

In parenthesis: predicted cleavage site

Topology predictions: The predicted topology of Pga homologue proteins from *A. junii* SH205 is shown in table 5.13. No transmembrane (TM) helix region was predicted in the biofilm synthesis protein (ZP_06065218.1) (table 5.13 a) from *A. junii* SH205. The consensus topology as predicted by TOPCONS revealed the presence of one TM helix region at position 6-26 in the polysaccharide

deacetylase protein (table 5.13 b). This region is predicted to be signal peptide. In glycosyl transferase protein (table 5.13 c), consensus prediction by TOPCONS revealed the presence of 5 TM regions, out of which the region 15-35 overlaps with the signal peptide region (1-28). The protein PgaD consists of two TM regions (Table 5.13 d), the N and the C-terminal of which is present inside the cytoplasm. The diagrammatic representation of topology of Pga homologue protein as predicted by the consensus topology tool TOPCONS and by their subcellular localization is shown in figure 5.12.

Table. 5.13 a): Predicted transmembrane (TM) helix positions of Protein sequence ZP_06065218.1 (Biofilm synthesis protein from *A. junii* SH205)

Tools	
SCAMPI-seq	No TM-regions predicted
SCAMPI-msa	No TM-regions predicted
PRODIV	No TM-regions predicted
PRO	No TM-regions predicted
OCTOPUS	No TM-regions predicted
TOPCONS	No TM-regions predicted

Table. 5.13 b): Predicted TM helix positions of Protein sequence ZP_06065265.1 (Polysaccharide deacetylase protein from *A. junii* SH205)

Tools	N-Terminal	C-terminal	TMS 1
SCAMPI-seq	IN	OUT	9-29
SCAMPI-msa	IN	OUT	6-26
PRODIV	OUT	IN	6-26
PRO	IN	OUT	No TM-regions predicted
OCTOPUS	IN	OUT	6-26
TOPCONS	IN	OUT	6-26 (in-->out)

Table. 5.13 c): Predicted TM helix positions of Protein sequence ZP_06065266.1 (Glycosyl transeferase from *A. junii* SH205)

Tools	N-Terminal	C-terminal	TMS 1	TMS 2	TMS 3	TMS 4	TMS 5
SCAMPI-seq	OUT	IN	15-35	294-314	316-336	339-359	377-397
SCAMPI-msa	OUT	IN	15-35	296-316	319-339	342-362	377-397
PRODIV	OUT	OUT	13-33	298-318	339-359	371-391	
PRO	IN	IN	7-27	307-327	339-359	380-400	
OCTOPUS	OUT	IN	15-35	296-316	319-339	342-362	377-397
TOPCONS	OUT	IN	15-35	296-316	319-339	341-361	377-397
			(out-->in)	(in-->out)	(out-->in)	(in-->out)	(out-->in)

Table. 5.13 d): Predicted TM helix positions of Protein sequence ZP_06065221.1 (Biofilm PGA synthesis protein PgaD from *A. junii* SH205)

Tools	N-Terminal	C-terminal	TMS 1	TMS 2
SCAMPI-seq	IN	IN	29-49	76-96
SCAMPI-msa	IN	IN	27-47	74-94
PRODIV	IN	IN	21-50	71-91
PRO	IN	IN	28-48	69-89
OCTOPUS	IN	IN	27-47	74-94
TOPCONS	IN	IN	28-48 (in-->out)	74-94 (out-->in)

Domains/patterns/motifs: The SMART result showed the presence of various structural domains or patterns or motifs in addition to the signal peptide prediction in the protein sequences studied. The sequence ZP_06065218.1 (Biofilm synthesis protein from *A. junii* SH205) was found to contain a coiled coil region (position: 364-389) and also predicted to contain a tetratricopeptide repeat (TPR) domain (position: 136-461) (d1fcha_, E-value: 2.00e-05) from the SCOP database. TPR repeats are also observed in the N-terminal domain of PgaA from *E.coli* (Itoh et al., 2008), where they are known to form elongated superhelical structure with a spiraling groove that mediates protein-protein interactions. Conserved domain analysis on the NCBI website revealed the presence of putative conserved domain. Members of this protein family are known to be poly-beta-1, 6 N-acetyl-D-glucosamine (PGA) export porin PgaA of gram negative bacteria which exports PGA across the outer membrane (Wang et al., 2004; Boehm et al., 2009).

The sequence ZP_06065265.1 (Polysaccharide deacetylase protein from *A. junii* SH205) was found to contain a polysaccharide deacetylase domain (position: 96-261) (Polysacc_deac_1, E-value: 2.20e-25) from the Pfam databse. Conserved domain search on NCBI website revealed that this domain is represented by poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase (PgaB), encoded by *Escherichia coli pgaB* gene from the pgaABCD operon, which affects biofilm development by promoting abiotic surface binding and intercellular adhesion.

The sequence ZP_06065266.1 (Glycosyl transeferase from *A. junii* SH205) was found to contain a Glycosyl transeferase domain (position: 56-223) (Glycos_transf_2, E-value= 8.50e-26:) from the Pfam databse and a Nucleotide-diphospho-sugar transferases (position:53-296) (d1qg8a_, E-value:8.00e-28). Conserved domain analysis of the protein sequence revealed that it belongs to Glycosyltransferase family A (GT-A) which includes diverse families of glycosyl transferases with a common GT-A type

structural fold. Conserved domain analysis also revealed the presence of residue 141-143 act as DxD motifs that, in all crystal structures are know to interact with the phosphate groups of nucleotide donor through the coordination of divalet cation, particularily Mn^{2+} (Breton et al., 2006). Glycosyltransferases are enzymes that synthesize oligosaccharides, polysaccharides, and glycoconjugates by transferring the sugar moiety from an activated nucleotide-sugar donor to an acceptor molecule, which may be a growing oligosaccharide, a lipid, or a protein. Litrature shows that PgaC is a polysaccharide polymerase that uses UDP-GlcNAc as substrate (Wang et al., 2004). The sequence ZP_06065266.1 also found to contain.

The sequence ZP_06065221.1 (Biofilm PGA synthesis protein PgaD from *A. junii* SH205) was found to contain S-adenosyl-L-methionine-dependent methyltransferases domain (position: 62-117) (d1jqea_, E-value: 7.00e+00) from SCOP database

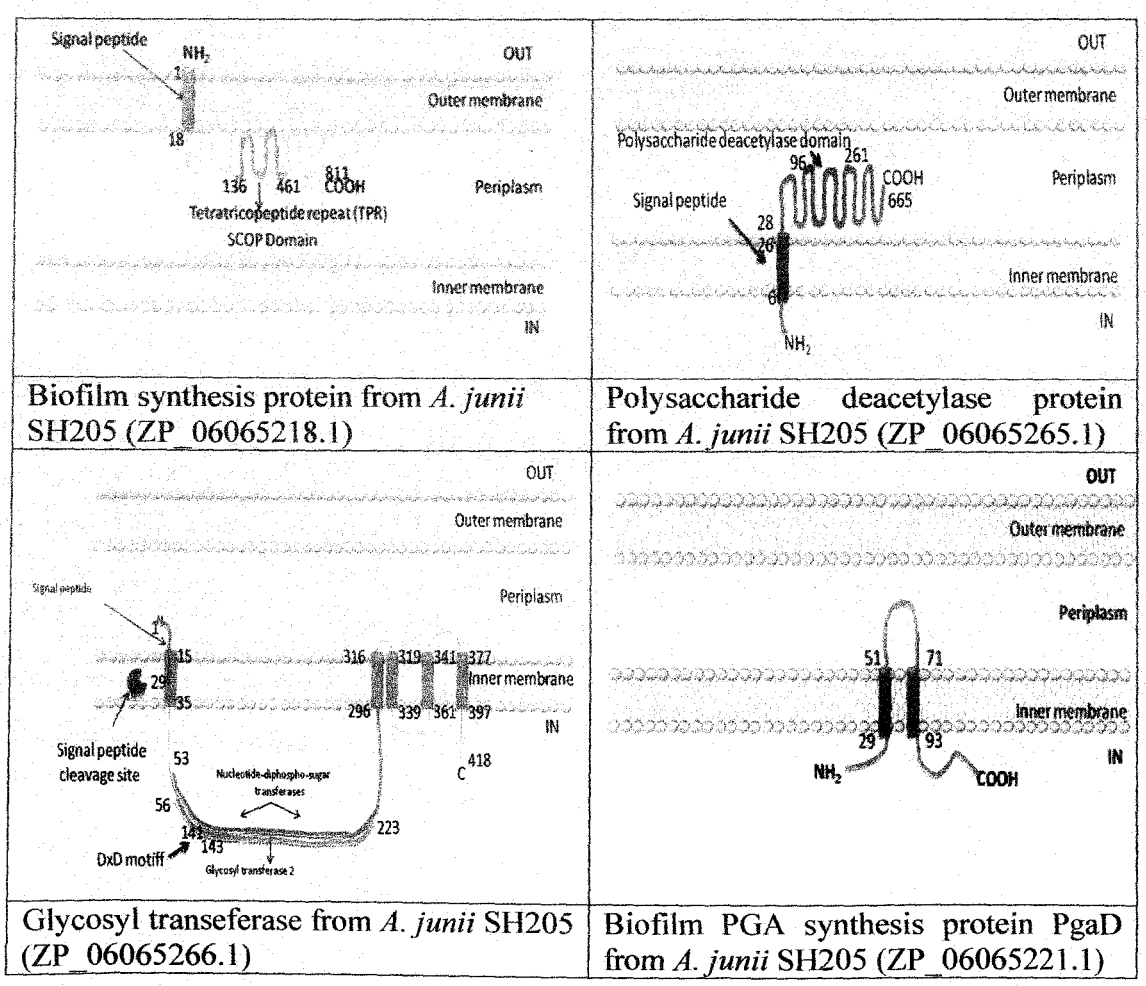


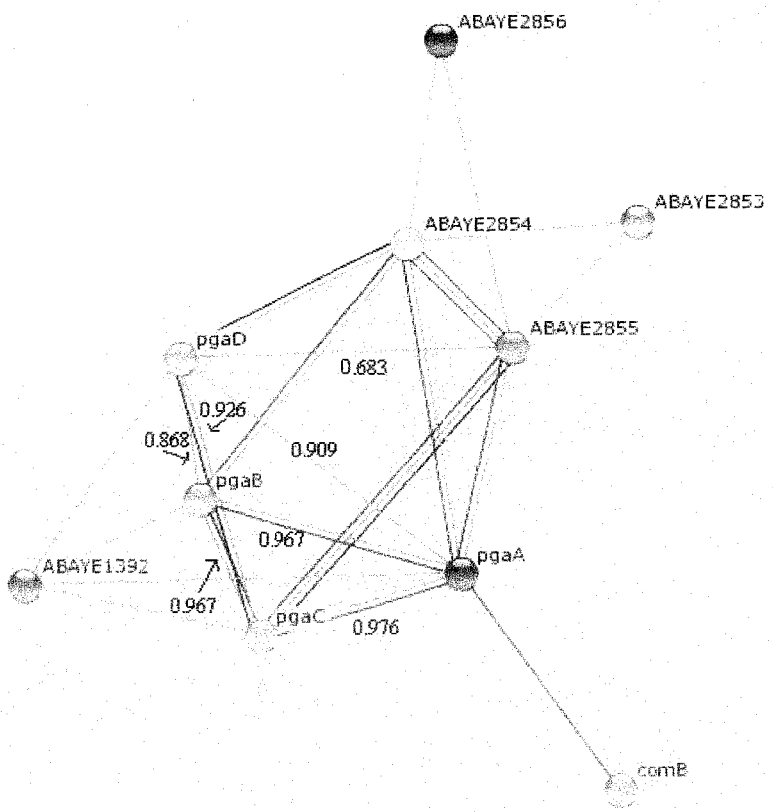
Fig. 5.12: Diagrammatic representation of Pga homologous proteins from *A. junii*

Pattern search in the protein sequence using PPsearch against PROSITE database revealed various putative phosphorylation sites as shown in the table 5.14. Glycosyl transferase protein revealed only Casein Kinase II (position: 105, 132, 364) and Protein Kinase C (position: 238) phosphorylation site. While, PgaD showed only Casein Kinase II (position: 6, 56,146) phosphorylation site.

Table. 5.14: Prosite patterns prediction using PPsearch tool.

	Protein sequence ZP_06065218.1 (Biofilm synthesis protein from *A. junii* SH205)	**Protein sequence ZP_06065265.1 (Polysaccharide deacetylase protein from *A. junii* SH205)**	**Protein sequence ZP_06065266.1 (Glycosyl transeferase from *A. junii* SH205)**	**Protein sequence ZP_06065221.1 (Biofilm PGA synthesis protein PgaD from *A. junii* SH205)**
Casein Kinase II	78, 99, 191, 220, 396, 397, 499, 509, 532, 633, 736	102, 132, 137, 156, 263, 353, 444, 456, 468, 470, 600	105, 132, 364	6, 56, 146
Tyrosine Kinase	98, 347, 518	-	-	-
Protein Kinase C	70, 99, 165, 269,406, 454, 509, 540, 593, 603, 662,736,754	243, 252, 438, 444, 533, 623	238	-

Protein protein interaction studies: The physical and function interaction between the four proteins from *A. junii* revealed that their interactions closely matches with the Pga proteins of *A. baumannii* AYE (whose protin protein interaction is already present in the STRING database) (Fig. 5.13). Figure shows the probabilistic confidence score between the proteins.The score is derived by separately benchmarking groups of associations against the manually curated functional classification scheme of the KEGG database (Kanehisa et al., 2010). The score represents a rough estimate of how likely a give association describes a function linkage between two proteins that is similar with the average pair of proteins annotated on the same 'map' or 'pathway' in KEGG (Szklarczyk et al., 2011). The neighbourhood display analysis (Fig. 5.14) shows the close association of the four genes in the immediate neighborhood on the genome (within 300 bp on the same strand) of *A. baumannii* AYE.



Predicted Functional Partners:

			Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
	ABAYE2855	putative lipoprotein (haemin storage system) (HmsF) (671 aa)	•		•	•			•		0.980
	pgaC	biofilm synthesis N-glycosyltransferase (PgaC-like) (415 aa)	•		•	•			•		0.976
	ABAYE2853	hypothetical protein (139 aa)	•								0.680
	comB	pilin like competence factor (333 aa)			•						0.454
	ABAYE2856	peptidyl-prolyl cis-trans isomerase (113 aa)	•								0.448
	ABAYE1392	hypothetical protein (278 aa)	•								0.404

Fig.13: Screenshot from the STRING website showing protein network and the interactions.

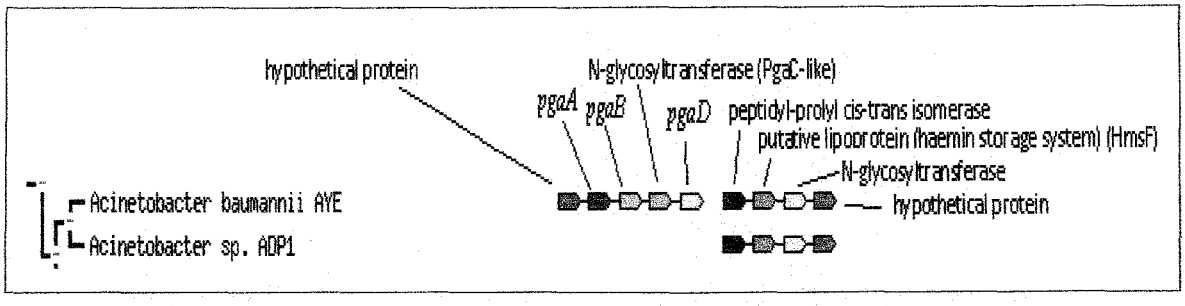


Fig. 5.14: Screenshot from the STRING website showing neighbourhood display of the genes encoding various proteins involved in the biosynthesis of poly-β-1, 6-linked *N*-acetylglucosamine.

Homology modeling of glycosyl transferase: The HHpred and HHsearch in protein data bank for the target sequence (ZP_06065266.1; Glycosyl transeferase from *A. junii* SH205) gave a best hit score with template 3bcv_A (crystal structure of a putative glycosyltransferase from *Bacteroides fragilis*) showing 100 % probability, E-value of 4E-32, 21 % identies with atomic resolution of its X-ray crystal structure obtained from diffraction studies being 2.35 Å. The sequence alignment of the query sequence (418 residues) and the template 3bcv_A (196 residues) is shown in figure 5.15a. The query sequence is modeled from 52-252 residues. A single 3D model was generated in swiss model and was validated with ProQ (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) and PROCHECK programs. Ribbon diagram of the modeled glycosyl transferase from *A. junii* SH205 is shown in figure 5.16b. Similarily homology modeling for the PgaC protein from *E. coli* (NCBI gene ID: 6967460) and *A. baumannii* (NCBI ref. seq. YP_002326569.1) was done and compared in the figure 5.16 c and d. The template search for them also resulted in best hit with 3bcv_A.

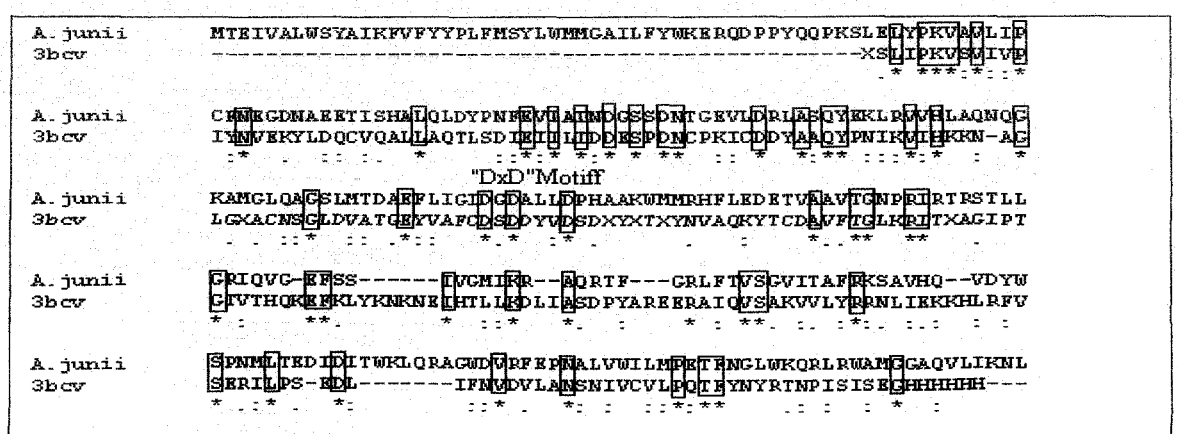


Fig.5.15 a): Alignment between target (glycosyl transferase of *A. junii* SH205) and template sequence (3bcv_A) obtained from ClustalW. The asterisk showed the identity of amino acids present in two protein sequences.

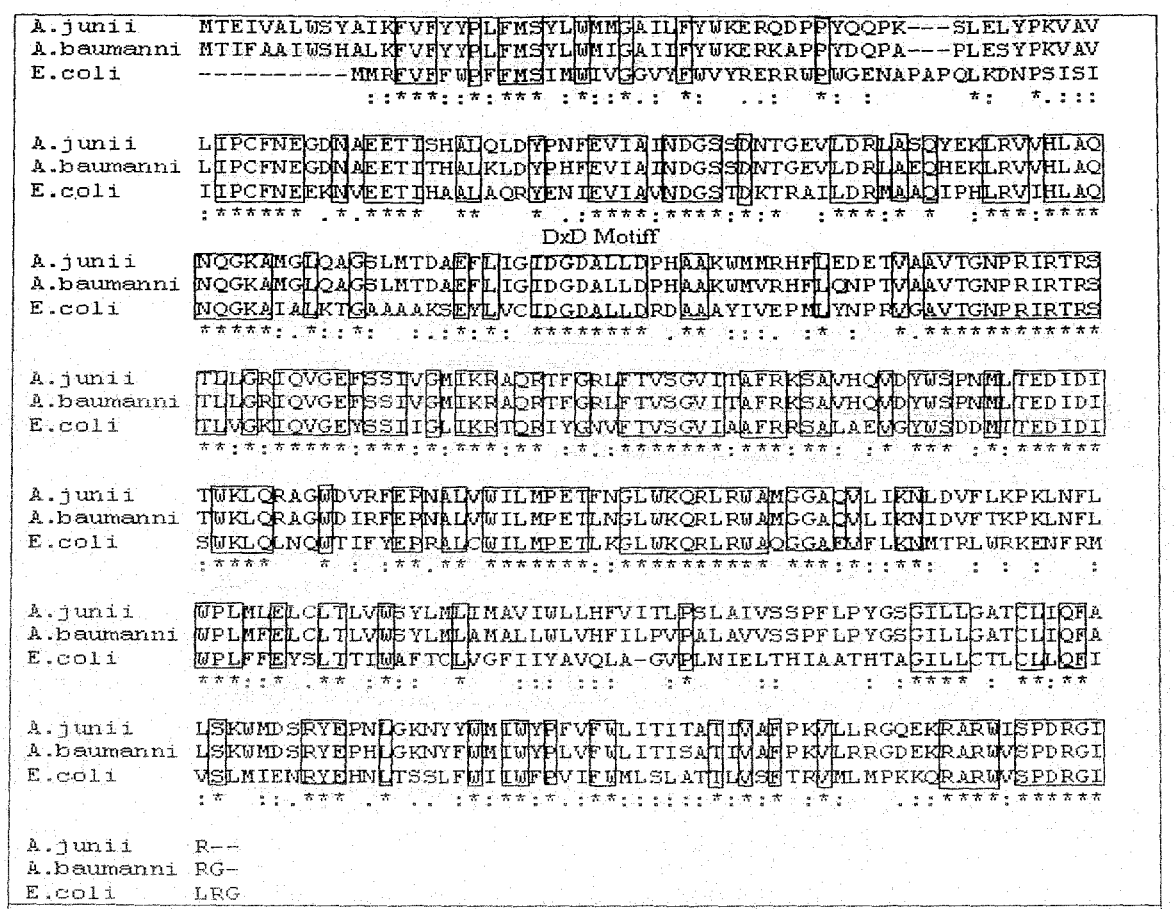


Fig. 5.15 b) Alignment between target glycosyl transferase of *A. junii*, *A. baumannii*, and *E. coli* obtained from ClustalW. The asterisk showed the identity of amino acids present in two protein sequences.

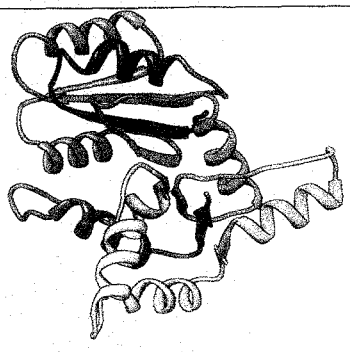
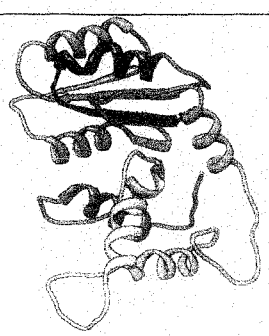
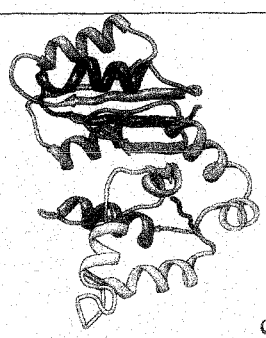
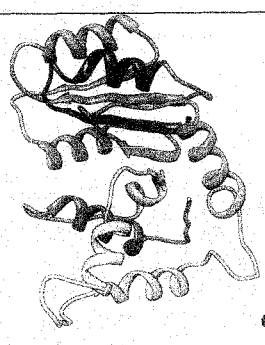


Fig. 5.16a): 3D model of the Template: 3bcv_A (crystal structure of a putative glycosyltransferase from *Bacteroides fragilis*)







A. junii	*A. baumannii*	*E. coli*
Template: 3bcv_A	Template: 3bcv_A	Template: 3bcv_A
Residue: 52 to 252	Residue: 52 to 252	Residue: 45 to 245
Predicted LGscore: 2.158	Predicted LGscore: 2.532	Predicted LGscore: 2.762
Predicted MaxSub:0.245	Predicted MaxSub:0.275	Predicted MaxSub:0.272
Different ranges of quality (ProQ web server*):	LGscore>1.5 fairly good model LGscore>2.5 very good model LGscore>4 extremly good model	MaxSub>0.1 fairly good model MaxSub>0.5 very good model MaxSub>0.8 extremly good model

Fig. 5.16: 3D homology models of PgaC/ glycosyl transferase. **b)** from *A. junii* SH205, **c)** from A. baumannii *A. baumanii* AB307-0294, **d)** from *E.coli* O157:H7
*http://www.sbc.su.se/~bjornw/ProQ/ProQ.html

Molecular docking: Further docking of the ligand UDP-GlucNAc with the modeled protein (glycosyl transferase from *A. junii*) was performed with AutoDock tool. Total 10 conformations were generated. Out of 10 docked conformations, the best docked conformations were chosen on the basis of lowest binding energy (table 5.15). These conformations were analyzed through Python Molecular Viewer for their interaction study shown in figure 5. 17a. In the predicted model, the neighbour residues within 3 Å was constituted by amino acid Phe 158, Gly 210, Val 211, Ile 212, Ala 214, Leu 241, and Phe 250. The residue 212-Ile and 250-Phe was common in *A. junii*, *A. baumanii* and in *E. coli* (table 5.16). Hydrogen bonds play an important role for the structure and function of biological molecules, especially for the enzyme catalysis (Crabtree R H 1998 Science 282, 5396). In binding mode, 6 hydrogen bonds of length

3.35 Å, 3.37 Å, 3.26 Å, 3.58 Å, 3.16 Å, 3.32 Å were detected between the ligand and protein residues at Phe 206, Glu 251, Ile 212, Ala 214, respectively.

Table. 5.15: Docking result of ligand (UDP-GlucNAc) with the glycosyl transferase/ PgaC

Organism	Docking energy (Kcal/mol)	No. of Hydrogen bond	Residues	Inhibition constant
***A. junii* SH205**	-10.68	6	Phe 206 (3.35 Å), Glu 251 (3.37 Å, 3.26 Å), Ile 212(3.58 Å), Ala 214 (3.16 Å, 3.32 Å)	14.59nM
A. baumannii	-11.24	4	Arg 249 (3.52 Å, 3.53 Å), Ser 209 (3.46 Å), Val 211 (3.09 Å)	6.83nM
E. coli	-9.93	7	Ala 214 (3.53 Å), Gln 242 (2.87 Å), Ile 212 (3.15 Å), Glu 251 (3.07 Å, 3.48 Å, 3.42 Å, 3.57 Å)	52.94nM

E. coli	*A. junii*	*A. baumannii*
153-Trp	158-Phe	206-Phe
157-His	210-Gly	207-Thr
212-Ile	211-Val	209-Ser
242-Gln	212-Ile	210-Gly
250-Phe	214-Ala	211-Val
251-Glu	241-Leu	212-Ile
	250-Phe	242-Gln
		250-Phe

Table 5.16: Comparison of neighbour amino acid residues within 3 Å of the predicted model (glycosyl transferase/ PgaC) from *A. junii, A. baumannii* and from *E. coli.* (shaded residues are common in one or other organism)

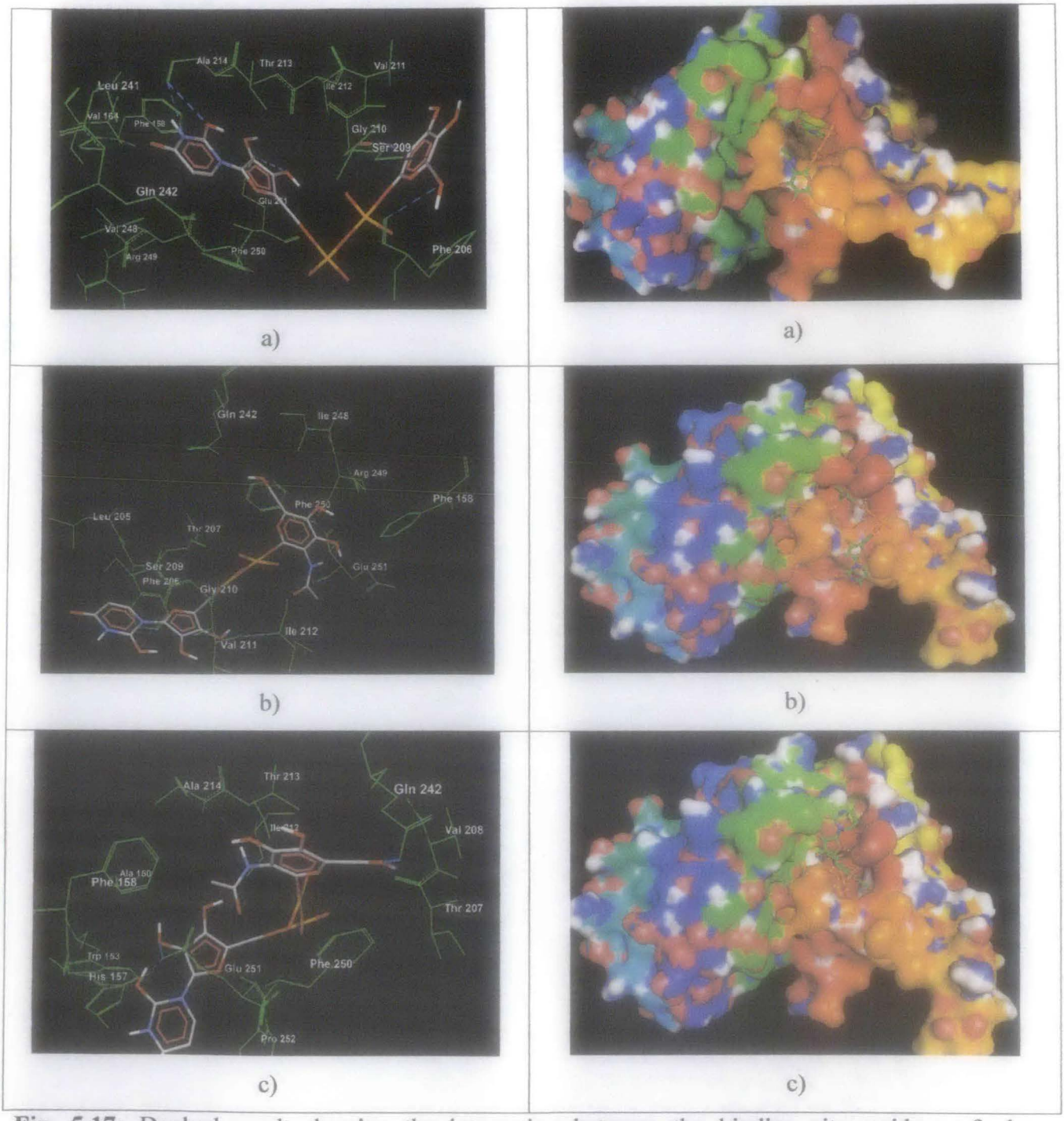


Fig. 5.17: Docked result showing the interaction between the binding site residues of glycosyl transferase and ligand UDP-GlcNAc. Figure **a), b),** and **c)** are the result obtained with the glycosyltransferase from *A. junii*, *A. baumannii,* and *E. coli* respectively. (blue colored dashed lines are the hydrogen bonds).

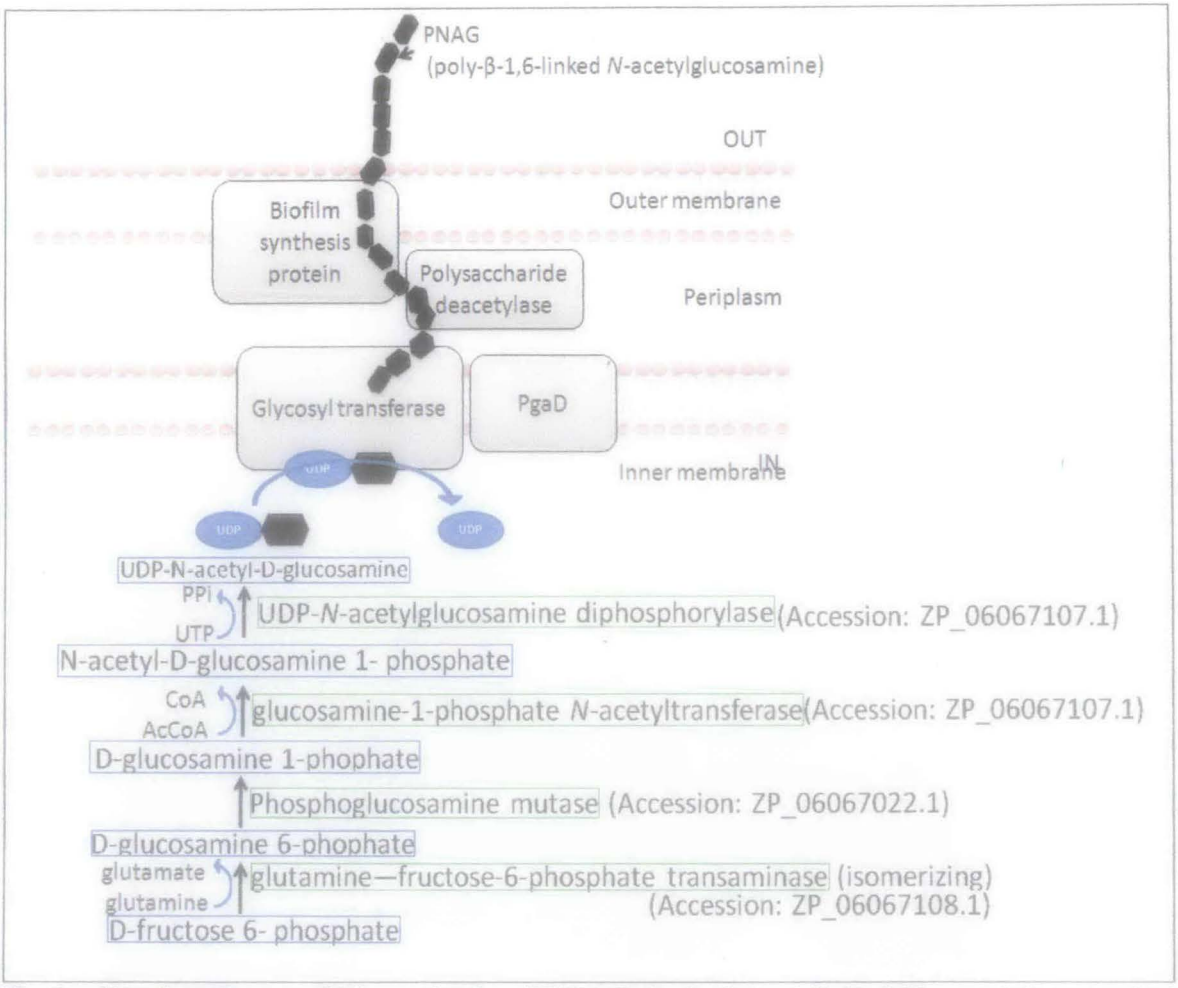


Fig. 5.18: Predicted pathway of biosynthesis of PNAG (poly-β-1, 6-linked *N*-acetylglucosamine) in *A. junii* as derived from the subcellular localization, topology prediction, and protein protein interaction of the four proteins showing homology with the Pga protein. The enzymatic pathway for the biosynthesis of UDP-N-acetyl-D-glucosamine was obtained from http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/polysacc/UDPGlcN.html; the accession no in the parenthesis represents the enzymes present in *A. junii* SH205

Phylogenetic distribution of *pga* homologues.

The operon *pgaABCD* is present in many bacterial species and come outs to be part of a horizontally transferred locus (Wang et al., 2004). BLAST analysis at the NCBI website with the *pga* gene products as query sequences revealed homologous loci in eubacteria. Table 5.17 shows a list of bacterial species having *pgaA*, *pgaB*, *pgaC*, and *pgaD* gene loci. The phylogentic tree of the PgaA, PgaB, PgaC, and PgaD proteins/or homologous proteins as obtained by the pairwise alignment were shown in figure 5.19 a, b, c and d.

Table 5.17: List of bacterial species with *pga* genes.

Organism	NCBI-Gene ID			
	PgaA	***pgaB***	***pgaC***	***pgaD***
Acinetobacter baumannii	7057332	7059442	7060465	7058980
Acinetobacter sp. ADP1	-----	2879399	2879400	2879401
Acinetobacter oleivorans	9381695	9383358	9383357	9383356
Acinetobacter calcoaceticus	11639831	11639830	11639829	11639828
Actinobacillus pleuropneumoniae	4848631	6396756	4848633	5851322
Aeromonas salmonicida	4996544	-----	4997906	-----
Enterobacter aerogenes	10791962	10791961	10791960	10791959
Enterobacter cloacae	9126786	9126787	9126788	9126789
Escherichia coli O157:H7	209397134	959161	6967460	959159
Kangiella koreensis	8369967	-----	8369969	8369970
Klebsiella pneumoniae	5340165	5340164	6938667	12543472
Pseudomonas fluorescens	3480755	7821075	3480757	3480758
Pectobacterium wasabiae	8533056	8533057	8533058	8533059
Pectobacterium atrosepticum	2881305	2881304	2881306	2881884
Pectobacterium carotovorum	8135288	-----	-----	-----
Pelobacter carbinolicus	3725497	3725496	3725495	3725494
Photobacterium profundum	-----	3120157	3120156	-----
Staphylococcus aureus	-----	12329884 Hypothetical	12425465	-----
Stenotrophomonas maltophilia	12743697	6477209	12743699	12743700
Yersinia enterocolitica	10302168	12599348	10302166	10302165
Yersinia pseudotuberculosis	6259356	2955805	2955807	2955808
Xanthomonas oryzae	3266184	3266064	3266094	3266205
Hahella chejuensis	-----	3839789	3839790	-----
Aggregatibacter actinomycetemcomitans	8535885	8535886	8535887	-----
Aggregatibacter aphrophilus	8122288	8122287	8122286	-----
Desulfobacterium autotrophicum	-----	-----	7501999	-----
Shigella flexneri	-----	-----	4209018	4209017
Thioalkalivibrio sulfidophilus	7316143	7316142	-----	7316140
Pusillimonas sp.	10603728	10603109	10603932	10604974
Ralstonia solanacearum	9412771	9412772	9412773	9414310
Corallococcus coralloides	-----	11986827	11981365	-----

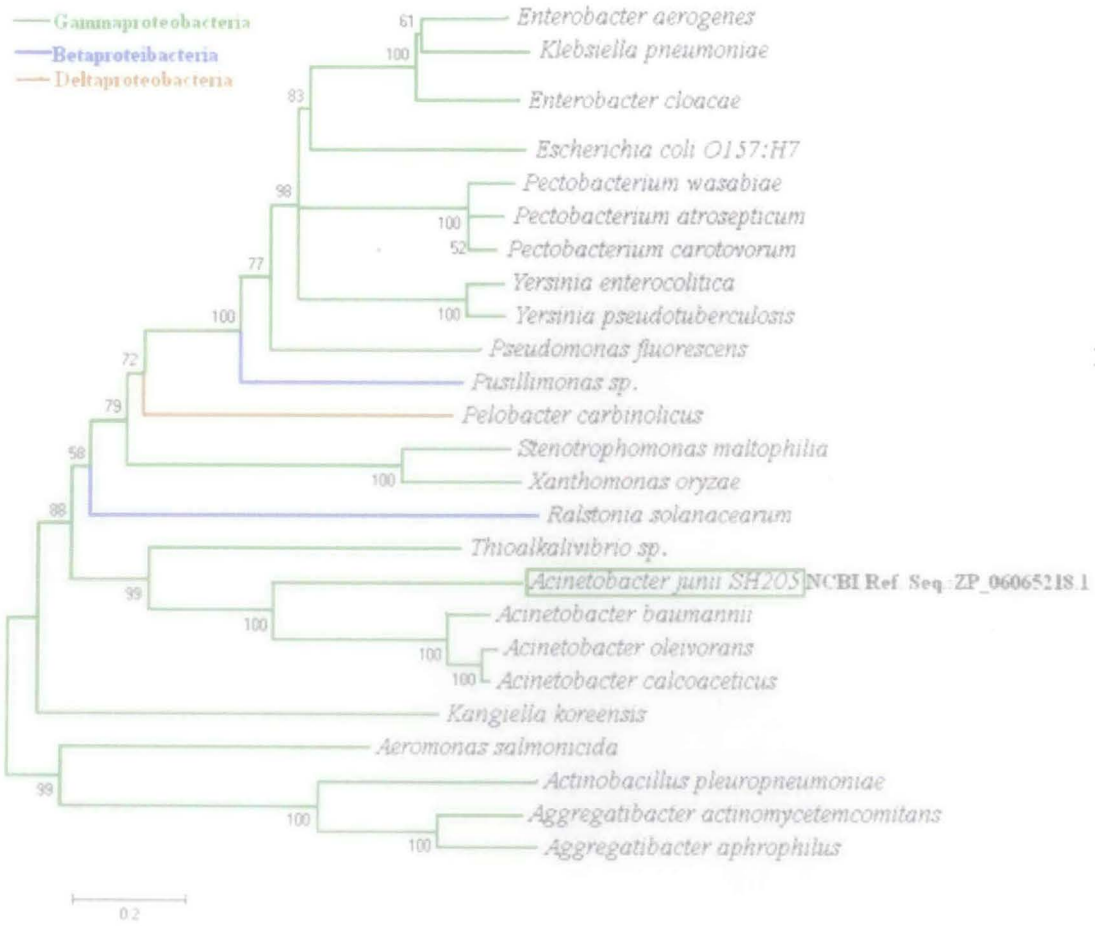


Fig. 5.19: a) Evolutionary relationships of PgaA protein from 25 taxa

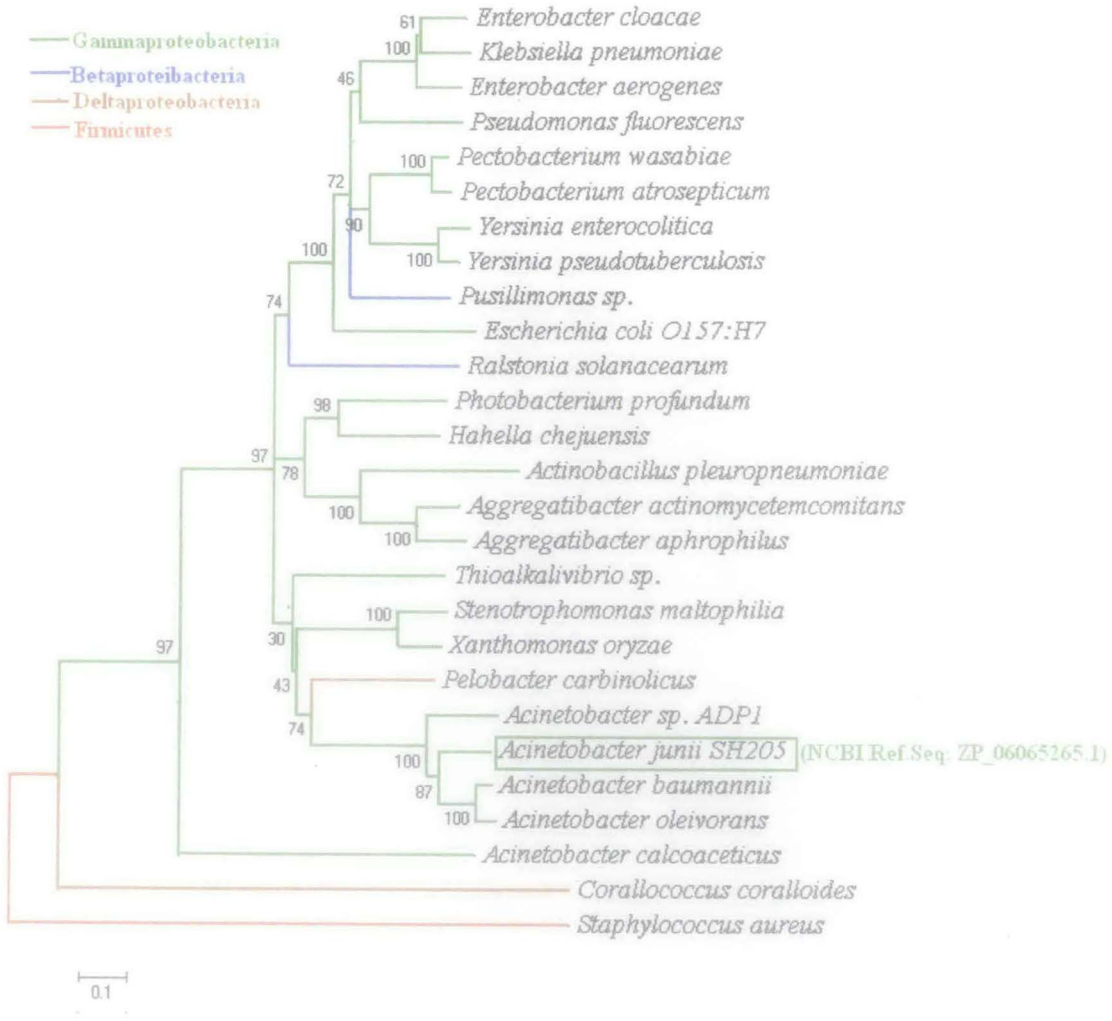


Fig. 5.19: b) Evolutionary relationships of PgaB protein from 27 taxa

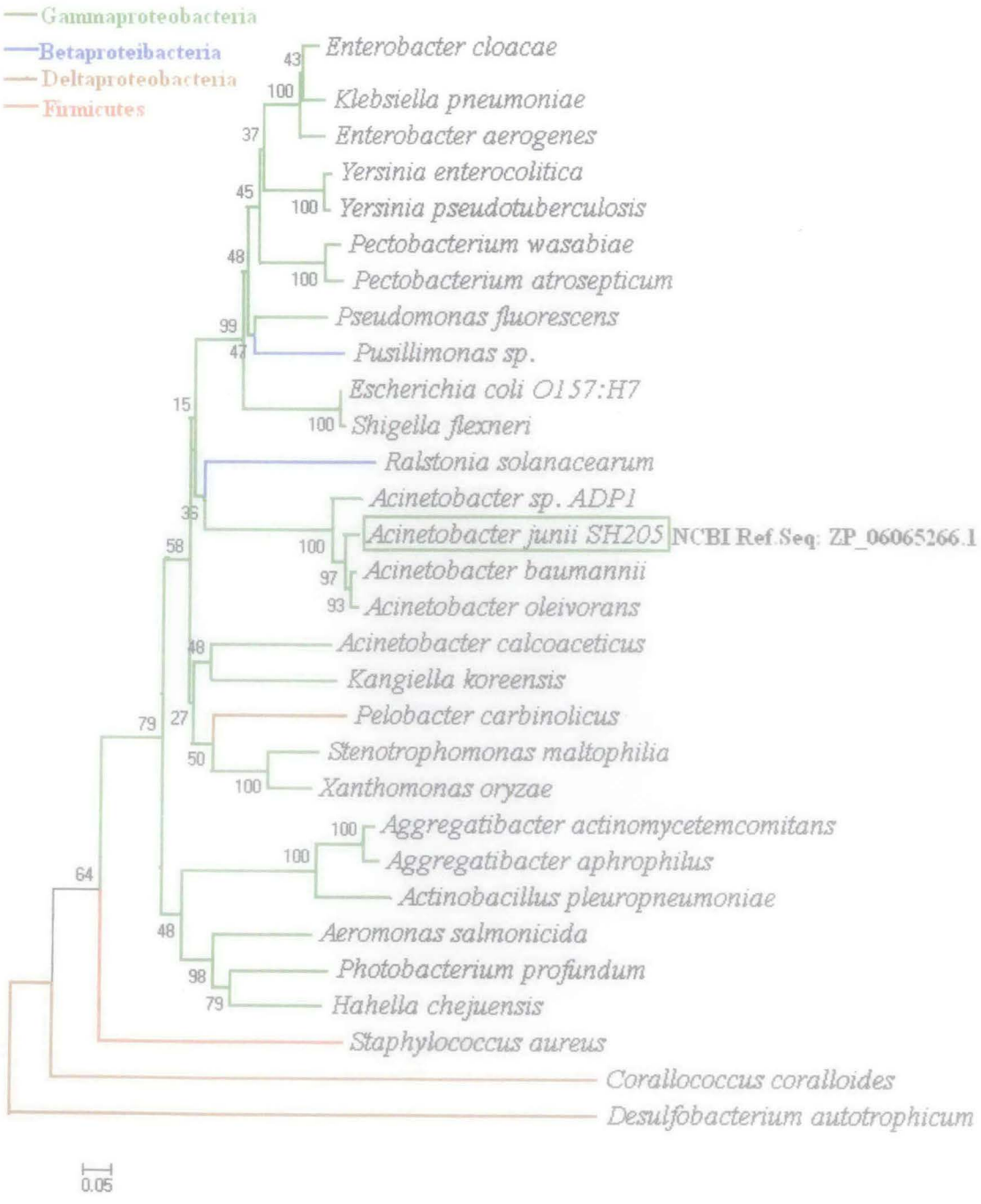


Fig. 5.19: c) Evolutionary relationships of PgaC protein from 30 taxa

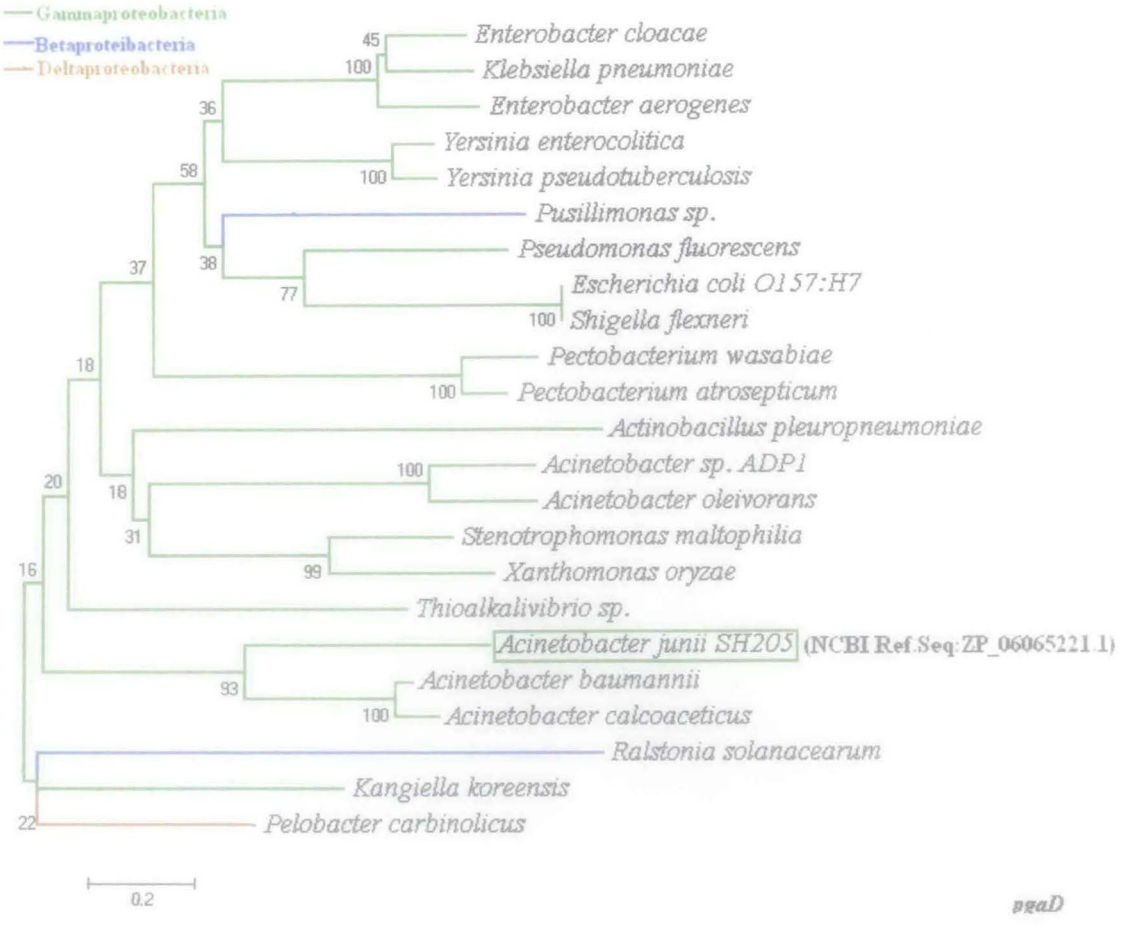


Fig. 5.19: d) Evolutionary relationships of PgaD protein from 23 taxa

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 500 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the genes analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 71 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

5.4. Discussion

Transposon mutagenesis: In the present study, transposon mutagenesis with *E.coli* transposon (Tn5 carried by plasmid pSUP5011) was done in order to isolate Tn5 insertion mutant of biofilm producing *A.junii* BB1A. In any transposon mutagenesis experiment, it is desirable to show linkage of the transposon insertions with the observed phenotype. Insertion of transposon (Tn5) led to the isolation of a single BB1A mutant strain (BB1A:Tn5) having increased sensitivity towards metal/metalloid ions and with the inability to produce biofilm. Comparison of growth curve in BHI medium indicated that the growth of planktonic cell is not cell number dependent in mutant as it can continue its growth and surpasses the maximum yield shown by the wild type cells under similar condition (Fig. 5.4). The BLAST analysis of the 223 bp DNA sequence flanked by the transposon revealed no significant result.

Degenerate Primer: The degenerate primer was manually designed using SeaView 4 since it allowed display of degenerate letters just below the aligned sequence (Fig 5.3a and b) and has an easy to use graphical interface (Gouy et al., 2010). This has enabled us to select short nucleotide sequence containing of at least 10 A+T+G+C which is the prime requisite for in-silico PCR amplification. The *in silico* PCR software program linked with NCBI genome data bank was found to be important software tool which allowed amplification using degenerate primer. The output of this program showed the nucleotide positions of the genome region which have undergone amplification, length of the amplicon and a simulation of the electrophoretic mobility on agarose gel. Further information of each amplicon may be obtained following the corresponding links: DNA sequences, list of ORFs that are included in the amplicon, and a link to the NCBI site, which displays a map of the chromosome around the amplicon (Bikandi et al., 2004).

Functional ability of the designed primers to amplify the target gene(s) could therefore be worked out in silico by accessing whole genome sequence (including plasmids when available) in the database. In all cases it would be mandatory to consider primers have been designed correctly, so that they do not form dimers, hairpin or any other aberrant structures preventing amplification.

Identification of two clusters in the phylogenetic tree constructed with *luxI* homologues has helped us to select the bacterial species for preparing the degenerate primers since the task has become easier because of lesser degeneracy among sequences of individual clusters. The applicability of the degenerate primers could be further enhanced if more mismatches are allowed under real wet lab conditions. To our observation in-silico PCR amplification with Deg1F/Deg2R failed with *Aeromonas salmonicida* genome template as the present version of the software, *In silico* PCR, did not allow more than two mismatches.Hence an exercise to design degenerate primers (when necessary in situations like

luxI sequences) for amplification of gene homologue of interest with the help of a combination of SeaView 4 and *In Silico* PCR programs would be very useful for the researchers in real wet lab situations.

Wet lab amplification of *A.junii* BB1A genome with the degenerate primer Deg1F/Deg2R initially resulted in non specific amplification (a common problem with degenerate primers) which was rectified by performing Touch Down PCR (TD-PCR). In TD-PCR the annealing temperature was decreased in increments for every subsequent set of cycles. The primer annealed at the highest temperature which is least-permissive of nonspecific binding that it is able to tolerate. Thus, the first sequence amplified is the one between the regions of greatest primer specificity; it is most likely that this is the sequence of interest. These fragments will be further amplified during subsequent rounds at lower temperatures, and thus out-competed the nonspecific sequences to which the primers may bind at those lower temperatures. If the primer initially (during the higher-temperature phases) binds to the sequence of interest, subsequent rounds of polymerase chain reaction can be performed upon the product to further amplify those fragments. The analysis of the amplicon generated by degenerate primer (primer Deg1F/Deg2R, designed for *luxI* homologue), revealed no similarity with the *luxI* homologue sequences; instead it showed 94% identity with the predicted protein of *A.junii* SH205. However the genome database of the available *A.junii* SH205 did show the presence of homoserine/ homoserine lactone efflux protein (NCBI Reference Sequence: ZP_06067918.1) but not the LuxI hmologue. This efflux protein is known to conduct the efflux of homoserine and homoserine lactone. Many bacteria use N-acyl homoserine lactone (AHL) as the key signalling molecules in the quorum sensing pathway. AHL based quorum sensing phenomena in BB1A strain was detected by using the reporter strain *Agrobacterium tumefaciens* NTL4 (pZLR4) (Chapter 4). It now seems that in the BB1A strain, the AHL based quorum sensing phenomena was controlled by genes dissimilar to *luxI*.

To survive under adverse conditions such as metal stress, *A. junii* BB1A cells must have mechanisms to avoid contact with toxic metals. The strain BB1A forms biofilm, the EPS (Extracellular polymeric substances) component of which was shown to bind metal ions and have higher tolerance to toxic metals compared to the same cells which were temprorily deficient in EPS (Chapter 4). The biofilm formation in many bacteria has been reported to be under the regulation of *pgaABCD* operon. In *A. baumannii, pgaABCD* locus is known to encode proteins involved in the production of Poly-β-1-6-*N*-Acetylglucosamine, which was found to be critical for biofilm formation (Choi et al., 2009). In the present work, PCR amplification of *A. junii* BB1A genome with the primer derived from the *pgaC* gene of *A. junii* SH205 generated a single amplicon of expected size, revealing the probability of similar *pga*ABCD locus in the test strain. BLASTp search in the protein data bank of *A.junii* SH205 revealed

the presences of a cluster of four genes, similar to the *pga*ABCD locus of *A. baumannii*. The key enzyme, glycosyl transferase, involved in the biosynthesis of Poly-β-1-6-*N*-Acetylglucosamine, from *A. junii* SH205 is predicted to be an inner membrane protein with 5 transmembrane helix regions. In the in silico work, 3D homology model of glycosyl transferase was constructed, using the SwissModel online server, and a refined model after energy minimization was obtained.The model was further used for Molecular Docking of UDP-GlcNAc. Docking of the modeled protein with ligand showed the best binding site. In the predicted model, the neighbour residues within 3 Å was constituted by amino acid Phe 158, Gly 210, Val 211, Ile 212, Ala 214, Leu 241, and Phe 250.

The phylogenetic analysis of known *pga* genes puts a broad overview of this gene family and lays a foundation to advance studies on their distribution, structure and functional diversity in other species. The fundamental cellular roles of these genes in bacteria and their evolutionary relationship thus can be elucidated through the studies of homologue genes.

5.5. References

Atkinson, S., Cámara, M. and Williams, P. 2007. N-Acylhomoserine Lactones, Quorum Sensing, and Biofilm Development in Gram-negative Bacteria. In: Kjelleberg, S and Givskov, M. (eds), The Biofilm Mode of Life: Mechanisms and Adaptations. Horizon Scientific Press, pp 95-122.

Atkinson, S., Throup, J.P., Stewart, G.S.A.B. and Williams, P. 1999. A hierarchical quorum-sensing system in *Yersinia pseudotuberculosis* is involved in the regulation of motility and clumping. Mol. Microbiol. **33**:1267-1277.

Atschul, S.F., Madden, T.L., Schaffer A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST; a new generation of protein database search programs. Nucleic. Acids. Res. **25**: 3389-3402.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S. 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics, **18**:298-305.

Batchelor, S.E., Cooper, M., Chhabra, S.R., Glover, L.A., Stewart, G.S.A.B., Williams, P. and Prosser, J.I. 1997. Cell density-regulated recovery of starved biofilm populations of ammonia-oxidising bacteria. Appl. Environ. Microbiol. **63**:2281-2286.

Berg, D.E. 1989. Transposon Tn5. In: Berg, D.E. and Howe, M.M. (eds), Mobile DNA. Washington, D.C.: American Society for Microbiology. pp 185-210.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. and Elofsson, A. Prediction of membrane-protein topology from first principles. Proc. Natl Acad. Sci. USA. **105**:7177–7181.

Bhadra, B., Nanda, A.K. and Chakraborty, R. 2006. Inducible nickel resistance in a river isolate of India phylogenetically ascertained as a novel strain of *Acinetobacter junii*. World J. Microbiol. Biotechnol. **22**:225-232.

Bhasin, M., Garg, A. and Raghava, G.P.S. 2005. PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics, **21**:2522-2524.

Bikandi, J., San Millán, R., Rementeria, A. and Garaizar, J. 2004. *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR, and endonuclease restriction. Bioinformatics **20**:798-799.

Blobel, G. and Dobberstein, B.1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Bio. **67**: 835-851.

Boehm, A., Steiner, S., Zaehringer, F., Casanova, A., Hamburger, F., Ritz, D., Keck, W., Ackermann, M., Schirmer, T. and Jenal, U. 2009.Second messenger signalling governs *Escherichia coli* biofilm induction upon ribosomal stress. Mol. Microbiol. 72: 1500-1516.

Boucher, J.C., Schurr, M.J., Yu, H., Rowen, D.W., Deretic, V. 1997. *Pseudomonas aeruginosa* in cystic fibrosis: Role of mucC in the regulation of alginate production and stress sensitivity. Microbiol. **143**:3473-3480.

Boutros, P.C. and Okey, A.B. 2004. PUNS: transcriptomic- and genomic-in silico PCR for enhanced primer design. Bioinformatics **20**: 2399-2400.

Brendel, V., Bucher, P., Nourbakhsh, I.R., Blaisdell, B.E. and Karlin, S. 1992. Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA, **89**:2002-2006.

Breton, C., Šnajdrová, L., Jeanneau, C., Koca, J. and Imberty, A. 2006. Structures and mechanisms of glycosyltransferases. Glycobiology. 16: 29R–37R.

Cámara, M., Williams, P. and Hardman, A. 2002. Controlling infection by tuning in and turning down the volume of bacterial small-talk. Lancet Infect. Dis. **2**:667-676.

Cao, Y., Wang, L., Xu, K., Kou, C., Zhang, Y., Wei, G., He, J., Wang, Y. and Zhao, L. 2005. Information theory-based algorithm for *in silico* prediction of PCR products with whole genomic sequences as templates. BMC Bioinformatics **6**: 190.

Choi, A.H.K., Slamti, L., Avci, F.Y., Pier, G.B. and Maira-Litrán, T. 2009. The *pga*ABCD locus of *Acinetobacter baumannii* encodes the production of poly-β-1-6-n-acetylglucosamine, which is critical for biofilm formation. J. Bacteriol. **191**:5953-5963.

Collins, F. S. and Weissman, S. M. 1984. Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. Proc. Natl. Acad. Sci. USA **81**: 6812-6816

Coplin, D.L. and Cook, D. 1990. Molecular genetics of extracellular polysaccharide biosynthesis in vascular phytopathogenic bacteria. Mol. Plant-Microbe Interactions. **3**:271-279.

Cramton, S.E., Gerke, C., Schnell, N.F., Nichols, W.W. and Gotz, F. 1999. The intercellular adhesion (ica) locus is present in *Staphylococcus aureus* and is required for biofilm formation. Infect. Immun. **67**:5427-5433.

Davies, D.G., Parsek, M.R., Pearson, J.P., Iglewski, B.H., Costerton JW, Greenberg, E.P. 1998. The involvement of cell-to-cell signals in the development of a bacterial biofilm. Science **280**:295-298.

De Bruijn, F. J. and Lupski, J. R. 1984. The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - a review. Gene **27**:131-149.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. **19**:4008.

Dunne, W.M. 2002. Bacterial adhesion: seen any good biofilms lately? Clin. Microbiol. Rev. **15**:155-166.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evol. **39**:783-791.

Fuqua, C., Winans, S.C., and Greenberg, E.P. 1996. Census and consensus in bacterial ecosystems: The LuxR-LuxI family of quorum-sensing transcriptional regulators. Annu. Rev. Microbiol. **50**:727-751.

Gaddy, J.A. and Actis, L.A. 2009.Regulation of *Acinetobacter baumannii* biofilm formation. Future Microbiol. **4**:273-278.

Gambello, M.J. and Iglewski, B.H. 1991. Cloning and characterisation of the *Pseudomonas aeruginosalas R* gene, a transcriptional activator of elastase expression. J. Bacteriol. **173**:3000-3009.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook. Walker, J.M(ed), The Proteomics Protocols Handbook, Humana Pres. pp. 571-607.

Genevaux, P., Muller, S. and Bauda, P. 1996. A rapid screening procedure to identify mini-Tn10 insertion mutants of *Escherichia coli* K-12 with altered adhesion properties. FEMS Microbiol. Lett. **142**:27-30.

Goryshin, I.Y., Jendrisak, J., Hoffman, L.M., Meis, R. and Reznikoff, W.S. 2000. Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat. Biotechnol. **18**:97-100.

Gould, T.A., Schweizer, H.P. and Churchill, M.E.A. 2004. Structure of the *Pseudomonas aeruginosa* acylhomoserine lactone synthase LasI. Mol. Microbiol. **53**:1135-1146.

Gouy, M., Guindon, S. and Gascue, O. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. Mol. Biol. Evol. **27**:221-224.

Gray, K.M. and Garey, J.R. 2001. The evolution of bacterial LuxI and LuxR quorum sensing regulators. Microbiol. **147**: 2379-2387.

Greenberg, E.P. 2000. Acyl-homoserine lactone quorum sensing in bacteria. J. Microbiol. **38**: 117-121.

Heilmann, C. and Gotz, C. 1998. Further characterization of *Staphylococcus epidermidis* transposon mutants deficient in primary attachment or intercellular adhesion. Zentvalbl. Bakteriol. **287**:69-83.

Heilmann, C., Schweitzer, O., Gerke, C., Vanittanakom, N., Mack, D. and Götz, F. 1996. Molecular basis of intercellular adhesion in the biofilm-forming Staphylococcus epidermidis. Mol. Microbiol. **20**:1083-1091.

Herrington, D.A., Hall, R.H., Losonsky, G., Mekalanos, J.J., Taylor, R.K. and Levine, M.M. 1988. Toxin, toxin-coregulated pilus, and the toxR regulon are essential for *Vibrio cholerae* pathogenesis in humans. J. Exp. Med. **168**:1487-1492.

Huang, G., Zhang, L. and Birch, R.G. 2000. Rapid amplification and cloning of Tn5 flanking fragments by inverse PCR. Lett. Appl. Microbiol. **31**:149-153.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M. and Sigrist, C.J. 2006. The PROSITE database. Nucleic Acids Res. **34**:D227-230.

Irie, Y. and Parsek, M.R. 2008. Quorum sensing and microbial biofilms. Curr. Top. Microbiol. Immunol. **322**:67-84.

Itoh, Y., Rice, J.D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., Beveridge, T.J., Preston, F. and Romeo, T. 2008.Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. J. Bacteriol. 190: 3670-3680.

Izano, E.A., Sadovskaya, I., Vinogradov, E., Mulks, M.H., Velliyagounder, K., Ragunath, C., Kher, W.B., Ramasubbu, N., Jabbouri, S., Perry, M.B. and Kaplan, J.B. 2007. Poly-N-acetylglucosamine mediates biofilm formation and antibiotic resistance in *Actinobacillus pleuropneumoniae*. Microb. Pathog. **43**:1-9.

Izano, E.A., Sadovskaya, I., Wang, H., Vinogradov, E., Ragunath, C., Ramasubbu, N., Jabbouri, S., Perry, M.B. and Kaplan, J.B. 2008. Poly-N-acetylglucosamine mediates biofilm formation and detergent resistance in *Aggregatibacter actinomycetemcomitans*. Microb. Pathog. **44**:52-60.

Jarman, S.N. 2004. Amplicon: software for designing PCR primers on aligned DNA sequences. Bioinfomatics Applications Notes **20**: 1644–1645.

Jones DT. 2007. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics. **23**:538–544.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. **38**: D355-D360.

Kwok, S., Chang, S.Y., Sninsky, J.J. and Wang, A.A. 1994. Guide to the design and use of mismatch and degenerate primers; Cold Spring Harbor Laboratory ISSN 1054-9805/94, PCR methods and Appl. pp 39-47.

Kyte. J. and Doolittle, R.F.1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105-32.

Lasa, I. 2006. Towards the identification of the common features of bacterial biofilm development. Int. Microbiol. **9**:21-28.

Lerat, E. and Moran, N.A. 2004.Evolutionary history of quorum-sensing systems in bacteria. Mol. Biol. Evol. **21**:903-913.

Letunic, I., Doerks, T. and Bork, P. 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. **40**: D302-305.

Lexa, M., Horak, J. and Brzobohaty, B. 2001. Virtual PCR. Bioinformatics. **17**:192-193.

Mack, D., Nedelmann, M., Krokotsch, A., Schwarzkopf, A., Heesemann, J. and Laufs, R. 1994. Characterization of transposon mutants of biofilm-producing *Staphylococcus epidermidis* impaired in

the accumulative phase of biofilm production: genetic identification of a hexosamine-containing polysaccharide intracellular adhesin. Infect. Immun. **62**:3244-3253.

Mack, D., W. Fischer, A. Krokotsch, K. Leopold, R. Hartmann, H. Egge, and R. Laufs. 1996. The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear β-1, 6-linked glucosaminoglycan: purification and structural analysis. J. Bacteriol. **178**:175-183.

Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208-218.

Mejiaruiz, H., Moreno, S., Guzman, J., Najera, R., Leon, R., Soberonchavez, G. and Espin, G. 1997. Isolation and characterization of an *Azotobacter vinelandii* algk mutant. FEMS Microbiol. Lett. **156**:101-106.

Milton, D.L., Hardman, A., Camara, M., Chhabra, S.R., Bycroft, B.W., Stewart, G.S.A.B. and Williams, P. 1997. Quorum sensing in *Vibrio anguillarum*: Characterization of the *vanI/R* locus and identification of the autoinducer *N*-(3-oxodecanoyl)-L-homoserine lactone. J. Bacteriol. **179**:3004-3012.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. 1998. Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function, J. Computational Chemistry. **19**: 1639-1662.

Mukhopadhyaya, P.N., Deb, C., Lahiri, C. and Roy, P. 2000. A soxA gene, encoding a diheme cytochrome c, and a sox locus, essential for sulfur oxidation in a new sulfur lithotrophic bacterium. J. Bacteriol. **182**:4278-4287.

Nasser, W., Bouillant, M.L., Salmond, G. and Reverchon, S. 1998. Characterisation of the *expI-expR* locus directing the synthesis of two *N*-acyl-homoserine lactone signal molecules. Mol. Microbiol. **29**:1391-1405.

Nishigaki, K., Saito, A., Takashi, H. and Naimuddin, M. 2000. Whole genome sequence- enabled prediction of sequences performed for random PCR products of *Escherichia coli*. Nucleic Acids Res. **28**: 1879-1884.

Niu, C., Clemmer, K.M., Bonomo, R.A. and Rather, P.N. 2008. Isolation and characterization of an autoinducer synthase from *Acinetobacter baumannii*. J. Bacteriol. **190**:3386-3392.

Nyvad, B. and Kilian, M. 1990. Comparison of the initial *Streptococcal* microflora on dental enamel in caries-active and in caries-inactive individuals. Caries Res. **24**:267-272.

Ochman, H., Gerber, A.S. and Hartl, D.L. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics **120**: 621-623

O'Toole, G.A. and Kolter, R. 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. Mol. Microbiol. **30**:295-304.

Parise, G., Mishra, M., Itoh, Y., Romeo, T. and Deora, R. 2007. Role of a putative polysaccharide locus in *Bordetella* biofilm development. J. Bacteriol. **189**: 750-760.

Parsek, M.R. and Greenberg, E.P. 2000. Acyl-homoserine lactone quorum sensing in Gram-negative bacteria: a signaling mechanism involved in associations with higher organisms. Proc. Natl. Acad. Sci. **97**: 8789-8793.

Pérez-Mendoza, D., Coulthurst, S.J., Sanjuán, J. and Salmond, G.P.C. 2011. N-Acetylglucosamine-dependent biofilm formation inPectobacterium atrosepticum is cryptic and activated by elevated c-di-GMP levels. Microbiol. **157**:3340-3348.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configuration. J. Mol. Biol. 7: 95-99.

Reznikoff, W.S. 1993. The Tn5 transposon. Annu. Rev. Microbiol. **47**:945-963.

Reznikoff, W.S. 2003. Tn5 as a model for understanding DNA transposition. Genome Res. **47**:1199-1206.

Reznikoff, W.S. 2008. Transposon Tn5. Annu. Rev. Genet. **42**: 269-286.

Rubin, E. and Levy, A.A. 1996. A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res. **24**: 3538-3545.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406-425.

Sakuragi, Y. and Kolter, R. 2007. Quorum-sensing regulation of the biofilm matrix genes (pel) of *Pseudomonas aeruginosa*. J. Bacteriol. **189**:5383-5386.

Sambrook, J., Fritsch, E. and Maniatis, T. 1989. Molecular Cloning. Cold Spring Harbor laboratory press, Cold Spring Harbor, NY.

Sarkar, S. and Chakraborty, R. 2008. Quorum sensing in metal tolerance of *Acinetobacter junii* BB1A is associated with biofilm production. FEMS Microbiol. Lett. **282**: 160-165.

Sauer, K., Camper, A.K., Ehrlich, G.D., Costerton, J.W. and Davies, D.G. 2002. *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. J. Bacteriol. **184**:1140-1154.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. **31**:3381-3385.

Shaw, P.D., Ping, G., Daly, S.L., Cha, C., Cronan, J.E., Rinehart, K.L. and Farrand, S.K. 1997. Detecting and characterizing N-acyl-homoserine lactone signal molecules by thin layer chromatography. Proc. Natl. Acad. Sci. **94**: 6036-6041.

Silver J. and Keerikatte V. 1989. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. **63**: 1924-1928.

Simon, R. 1984. High frequency mobilization of gram-negative bacteria replicons by the *in vitro* constructed Tn*5-mob* transposon. Mol. Gen. Genet.**196**:413-420.

Sutherland, I.W. 2001. Microbial polysaccharides from gram-negative bacteria. Int. Dairy J. **11**:663.

Swift, S., Downie, J.A., Whitehead, N.A., Barnard, A.M.L., Salmond, G.P.C. and Williams, P. 2001. Quorum sensing as a population-density-dependent determinant of bacterial physiology. Adv. Microb. Physiol. **45**:199-270.

Swift, S., Williams, P. and Stewart, G.S.A.B. 1999. *N*-Acyl homoserine lactones and quorum sensing in proteobacteria. In: Dunny,G.M. and Winans,S.C. (eds), Cell-Cell Signaling in Bacteria; American Society for Microbiology Press, Washington, DC. pp. 291–313.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J. and von Mering, C., 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. **39**: D561-D568.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596-1599.

Tautz, D. and Renz, M. 1983. An optimized freeze-squeeze method for recovery of DNA fragments from agarose gels. Anal. Biochem.**132**:14-19.

Tomaras, A.P., Dorsey, C.W., Edelmann, R.E. and Actis, L.A. 2003. Attachment to and biofilm formation on abiotic surfaces by *Acinetobacter baumannii*: involvement of a novel chaperone-usher pili assembly system. Microbiol. **149**:3473-3484.

Triglia, T., Peterson, M.G. and Kemp, D.J. 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequence. Nucleic Acids Res. **16**: 8186

Tusnady, G.E and Simon, I. 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics **17**:849-850.

Viklund H, Elofsson A. 2008. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics **24**:1662-1668.

Viklund, H. and Elofsson, A. 2004. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. Protein Sci. **13**:1908-1917.

Wang, X., Preston, F. and Romeo, T. 2004.The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. J. Bacteriol. **186**: 2724-2734.

Watnick, P.I., Fullner, K.J. and Kolter, R. 1999. A role for the mannose-sensitive hemagglutinin in biofilm formation by *Vibrio cholerae* El Tor. J. Bacteriol. **181**:3606-3609.

Watson, W.T., Minogue, T.D., Val, D.L., Beck von Bodman, S. and Churchill, M.E.A. 2002. Structural basis and specificity of acyl homoserine lactone signal production in bacterial quorum sensing. Mol. Cell. **9**:685-694.

Williams, P. 2007.Quorum sensing, communication and cross kingdom signaling in the bacterial world. Microbiol. **153**:3923-3928.

Yu, C.S., Lin, C.J. and Hwang, J.K. 2005. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. **13**:1402-1406.

Zhang, R.G., Pappas, T., Brace, J.L., Miller, P.C., Oulmassov, T., Molyneauz, J.M., Anderson, J.C., Bashkin, J.K., Winans, S.C. and Joachimiak, A. 2002.Structure of a bacterial quorum sensing transcription factor complexed with pheromone and DNA. Nature **417**:971-974.

Zuckerkandl, E. and Pauling, L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson, V. and Vogel, H.J. (eds), Evolving Genes and Proteins, Academic Press, New York. pp. 97-166.