2015

Genetic and Transcriptomic Analyses of the Rice Pathogenic Bacterium, Burkholderia glumae, Reveal the Important Roles of the Regulatory Gene, tepR, for Bacterial Survival in Environmental Stresses

Jingyu Peng
Louisiana State University and Agricultural and Mechanical College, jingyu9009@gmail.com

GENETIC AND TRANSCRIPTOMIC ANALYSES OF THE RICE PATHOGENIC BACTERIUM, *BURKHOLDERIA GLUMAE*, REVEAL THE IMPORTANT ROLES OF THE REGULATORY GENE, *TEPR*, FOR BACTERIAL SURVIVAL IN ENVIRONMENTAL STRESSES

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agriculture and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Plant Pathology and Crop Physiology

by
Jingyu Peng
B.S., Shandong Agricultural University, 2013
December 2015

ACKNOWLEDGEMENTS

It has never been easy for me to gain such a great opportunity to study in Louisiana State University, one of the best public university in America. I have been so lucky to study Plant Pathology in such a harmonious and efficient department, the Department of Plant Pathology & Crop Physiology. It is my great honor to work with a group of graduate students and plant pathologists who are so passionate about research and science. I want to take this opportunity to give my greatest appreciation to all the person who have accompanied me, helped me, appraised and criticized me.

I want to give my greatest thanks to my parents, Bo Sun and Liqian Peng. They have never been hesitant to support me to study in America even at the cost that I could not see them for more than two years. Words can hardly explain how much they love me and I love them just as much.

I want to give my great appreciation to my advisor Dr. Jong Hyun Ham, who is always the brain of our lab. He is very enthusiastic about our research and has provided me with the best resource and environment for conducting researches. I feel so glad to work on the research project I am very interested in, and I am always so excited to discuss with him about our research progress and spark new ideas.

I want to thank Dr. Eric C. Achberger for his great help throughout the Microbial Genetics class in my first semester in America. He was so nice and patient to me, no matter what simple and silly questions I asked. His class opened me the door for molecular biology.

I want to thank Inderjit Barphagha, the research associate of our lab. She was the person who taught me all the laboratory techniques from the basic PCR to cloning techniques, hand by hand, step by step. She was never reserved in teaching me whatever she knew.

I want to thank my committee members Dr. Christopher A. Clark and Dr. J. Cameron Thrash. They both are the professors I respect most in LSU. I really appreciate them for their acceptance to advise my Master's research. It has always been great time to talk to them about my research progress and listen to their suggestions. I appreciate all the time they have taken, all the encouragement they have given, and all the advices they have provided!

Finally, I want to thank myself for never giving up! Thank myself for being the proudness of my parents! And one day, I wish to be the proudness of all the people who have helped me!

TABLE OF CONTENTS

ABSTRACT

Burkholderia glumae causes bacterial panicle blight in rice and leads to severe yield losses. *B. glumae* causes disease by producing multiple virulence factors including toxoflavin, lipase, flagella, type III secretion system, catalase, and exopolysaccharides (EPS). The production of toxoflavin, the most important virulence factor, is tightly regulated by TofI/TofR-mediated quorum sensing (QS) and the transcriptional activators ToxJ and ToxR. However, other virulence regulatory systems are poorly studied. Recently, *tepR*, encoding a sigma 54-dependent response regulator, was characterized to negatively regulate toxoflavin and extracellular protease production. In order to gain insights into the comprehensive biological functions of *tepR* in *B. glumae*, the genome-wide transcriptional profile dependent on *tepR* was investigated through comparative RNA sequencing (RNA-seq) analyses between the wild-type *B. glumae* strain 336gr-1 and its *tepR* knockout derivative, LSUPB401. The RNA-seq analyses yielded the identification of 238 differentially expressed genes (DEGs). Gene Ontology enrichment analysis of the DEGs indicated that genes involved in flagella assembly and stress response were up-regulated in LSUPB401. Kyoto Encyclopaedia of Genes and Genomes pathway enrichment further revealed that *tepR* positively regulated the gene cluster encoding a putative type VI secretion system (T6SS) and the genes responsible for a branched-chain amino acid ABC Transporter system but negatively regulated the genes involved in thiamine, cysteine and methionine metabolism. Phenotypic tests indicated that LSUPB401 was significantly more tolerant to heat stress but not stationary phase stress. LSUPB401 also largely lost its competition capacity against other bacteria tested, which was most likely due to the deficiency of the *tepR*-dependent

T6SS. Phylogenetic study indicated that TepR homologs are conserved in *Burkholderia* spp. and other plant and animal pathogenic bacteria. This indicated the modulation of bacterial behavior through the regulatory action of *tepR* is likely to be a widely occurring phenomenon in many bacteria. Finally, it was observed that knockout mutation of *qsmR* abolished the production of toxoflavin and extracellular protease in both wild type (336gr-1) and *ΔtepR* (LSUPB401) backgrounds, indicating that the positive regulatory function of *qsmR* on those bacterial behaviors is dominant over the negative function of *tepR* on them.

CHAPTER 1: INTRODUCTION

Bacterial Panicle Blight (BPB) of Rice caused by *Burkholderia glumae*

Introduction to BPB. Panicle blighting has been a serious problem for rice production for many years in the rice growing regions in the southern United States. However, panicle blighting was considered to be caused by physiological disorder until 1996-1997 (Shahjahan et al., 2000). Later, *B. glumae*, together with *B. gladioli*, were isolated and identified to be the major cause of BPB (Nandakumar et al., 2005, Nandakumar et al., 2009). Typical symptoms of BPB involve panicle blight, sheath rot and seedling blight (Sayler et al., 2007, Tsushima, 1996). Usually, stripe symptoms appear first on leaves. The sheath of rice turns brown and develops necrotic lesions. Affected panicles gradually develop straw-colored florets with seedling blighting (Shahjahan, 2000). In severe cases, panicles become upright due to the failure of grain filling (Figure 1.1) (Nandakumar et al., 2009).

Figure 1.1. BPB symptoms on the rice cultivar CL111 which was spray-inoculated with a wild-type *B. glumae* strain 336gr-1 in Crowley, LA, USA.

Besides BPB on rice, *B. glumae* was also reported to cause bacterial wilt in other crops, such as potato, tomato and eggplant, posing great agricultural and economic losses (Jeong et al., 2003).

Infection by *B. glumae* can be confused with other bacterial infections, especially *Pseudomonas* spp., which can cause similar symptoms. For example, the bacterial stripe disease, which also causes grain discoloration, was reported to be caused by *Acidovorax citrulli* (previously named *Pseudomonas avenae*) (Kadota and Ohuchi, 1983). *Pseudomonas fuscovaginae* can cause extended water-soaking and necrosis symptoms, and grain discoloration before the emergence of panicles (Zeigler and Alvarez, 1987). Also, *B. glumae* was detected from seeds that did not show any symptoms, and the isolated strains could still infect rice and cause the BPB disease (Luo et al., 2007). These factors make it hard to correctly identify the pathogen and disease based solely on the symptoms of the infected rice.

Characteristics of *B. glumae*. *B. glumae* is a gram negative bacterium. It is rod-shaped, aerobic and nonfluorescent, and has 1-3 polar flagella. It can produce a yellow-green, water-soluble pigment on various media (Schaad et al., 2001). Because of the pigment, the colony looks grayish white or yellow. The arginine dehydrolase reaction, oxidase reaction and nitrate reduction reaction for *B. glumae* are negative (Cottyn et al., 1996). The total genome size of a species of *Burkholderia* is usually greater than 8Mbp (Holden et al., 2004). The extensively studied *B. glumae* strain BGR1 has two chromosomes and four plasmids, with 3,906,529 base pairs in chromosome 1 and 2,827,355 base pairs in chromosome 2, respectively (Lim et al., 2009). A total of 6305

genes have been found in the genome with an average GC content of 68.2% (Lim et al., 2009).

Epidemiology of BPB. Bacterial pathogens can survive in various environmental niches, such as water, plant, and soil. *B. glumae* has been considered to multiply on the phylloplane of rice plants during the growing stage (Matsuda and Sato, 1987; Tsushima et al., 1996) and survives in both inside and surface of rice seeds, rice tissue or weeds in the field overwinter (Tsushimi et al., 1989). The disease can emerge successfully next year under favorable environmental conditions. The prevalence of the disease can be affected by multiple factors, such as temperature, humidity, disease severity of previous year and host susceptibility (Tsushima and Naito, 1991; Tsushima et al., 1996). Especially, BPB tends to outbreak in seasons with high temperature and long-lasting rains (Zeigler and Alvarez, 1990; Mew, 1992). In Louisiana, the disease broke out in 1995, when high temperature lasted during the growing season. High humidity was also reported to favor the disease, especially during the flowering stage (Tsushima et al., 1995). The severity of the disease is also influenced by the inoculum density. Inoculation of *B. glumae* with as low as 10^2 to 10^4 cfu/ml by a spraying method led to BPB (Hikichi, et al. 1994). During infection, *B. glumae* is thought to first get across stomata, enter lemma and paleae, reproduce in parenchyma (Tabei et al, 1989), and spread to the surrounding cells (Hikichi, 1993). Bacteria can make long-distant movement through host vascular system (Yuan, 2004).

Distribution of BPB. BPB disease was first reported as bacterial grain rot and seedling blight in Japan (Goto et al., 1956). It was later reported to be found in main rice growing countries including Korea (Jeong et al., 2003), Taiwan (Chien and Chang,

1987), Vietnam (Trung et al., 1993) and Latin America (Nandakumar et al., 2007). Recently, BPB was also found in Ecuador and South Africa in extensive rice growing regions (Riera-Ruiz et al., 2014, Zhou, 2014). In the United States, the infection area mainly covered the Gulf of Mexico (Rush, et al., 1998). In Louisiana, severe outbreaks of the disease happened in 1995, 1998 and 2000, causing almost 40% yield loss in the heavily infected fields (Nandakumar et al., 2008). Since *B. glumae* is a seed-borne pathogen, the wide spread of the pathogen is very likely due to the long-distance transport of contaminated rice seeds.

Detection of *B. glumae*. Due to the economic importance of BPB, accurate and efficient detection of *B. glumae* becomes very important. As described before, the symptoms of BPB are similar to several diseases caused by *Pseudomonas*. So, identification of *B. glumae* through symptom observation can be inaccurate (Schaad et al., 2001). A semi-selective medium, SPG medium, was developed for selective culturing of *B. glumae* (Tsuschima et al., 1986). However, it cannot differentiate at species level within the genus of *Burkholderia*. Several biochemical and serological identification methods, for example, Biolog^{TM} system and fatty acid methyl ester (FAME)-fingerprinting are still under use (Cottyn et al., 1996). However, all the biochemical and physiological methods available require pure culture of the bacteria (Schaad et al., 2001). An rRNA gene-based polymerase chain reaction (PCR) method was later developed for the identification and detection of *B.* glumae, which was very simple, rapid and sensitive (Takeuchi et al., 1997). Recently, quantitative polymerase chain reaction (qPCR) has been developed to detect and quantify *B. glumae* with great

sensitivity (Nandakumar et al., 2009). Rep-PCR fingerprint analysis has shown to be able to resolve differences among various *B. glumae* isolates (Sayler et al., 2006).

Disease control of BPB. To date, several bactericides, for example copper-containing chemicals and antibiotics, have been indicated to be able to suppress the plant pathogenic *Burkholderia* spp. Oxolinic acid, which targets the DNA gyrase, is known to be able to effectively suppress the disease development of BPB (Maeda et al., 2004). Yet oxolinic acid is not allowed for use in the United States. Furthermore, oxolinic acid resistant strains were found in Japan, which implied the limitation of this chemical as a reliable control agent (Maeda et al., 2004). Thus, it is imperative to develop alternative and long-lasting disease control methods for this disease.

Virulence factors and their regulation in *B. glumae*

Until now, several virulence factors have been identified to be important for the pathogenesis of *B. glumae*, e.g. toxoflavin, type III secretion system, lipase, flagella and catalase.

Toxoflavin. Toxoflavin is one of the most important and well-studied virulence factors in *B. glumae*. Toxoflavin-deficient *B. glumae* mutants cause negligible disease symptoms (Devescovi et al., 2007, Kim et al., 2004). Toxoflavin can act as an effective electron carrier to bypass the cytochrome electron transport system and ,as a result, causes hydrogen peroxide production from oxygen (Latuasan & Berends, 1961). So, the toxicity of toxoflavin on rice may due to the toxicity of hydrogen peroxide which can affect photosynthesis, photorespiration and other metabolic processes (Hung & Kao, 2004, Wan & Liu, 2008). Toxoflavin has been identified and characterized to be synthesized through a polycistronic operon including five genes, *toxABCDE* (Kim et al.,

2004; Suzuki et al., 2004). Insertional mutations in any of these genes can abolish toxoflavin production (Kim et al., 2004). A second polycistronic operon, *toxFGHI*, was characterized to be responsible for toxoflavin transport (Kim et al., 2004). Several regulatory elements have been known to play important regulatory roles in toxoflavin biosynthesis. The LysR transcriptional activator, ToxR, can bind to the promoter region of *toxA* and is predicted to be involved in the activation of the *tox* operon transcription (Shingu and Yoneyama, 2004). And toxoflavin produced through the action of ToxR in turn functions as the co-inducer of ToxR (Kim et al., 2004). Besides, *toxJ*, encoding a transcriptional activator, is another important regulatory element for toxoflavin biosynthesis (Kim et al., 2004).

The biosynthesis and transport of toxoflavin in *B. glumae* are also tightly regulated by quorum sensing (QS) (Kim et al., 2004), which is a cell-to-cell communication mechanism for bacteria to behave in a collective manner (Miller & Bassler, 2001). Until now, the only known QS system in *B. glumae* is mediated by TofI and TofR, which are homologous to LuxI and LuxR, respectively, in *Vibrio fischeri* (Fuqua et al., 1994, Kim et al., 2004). TofI is a quorum sensing auto-inducer synthase which synthesizes two different AHL molecules, N-octanoyl homoserine lactone (C8-HSL) and N-hexanoyl homoserine lactone (C6-HSL). TofR acts as the cognate receptor that AHL molecules bind to (Dunlap, 1999, Kim et al., 2004). In *B. glumae*, the binding of C8-HSL to TofR activates the transcription of *toxJ* which is required for the expression of *toxR*. ToxR, along with ToxJ, in turn activates the expression of *toxABCDE* and *toxFGHI* for toxoflavin biosynthesis and transport (Figure 1.2) (Kim et al., 2004). Although toxoflavin is an important virulence factor of *B. glumae*, some research

demonstrated that toxoflavin-deficient strains still can cause minor disease symptom in disease-susceptible rice cultivars (Ham et al., 2011). So, the pathogenesis of *B. glumae* is most likely to be affected by multiple virulence factors.

Flagella. Although flagella do not directly cause the disease, they are important for the bacterial motility in hosts, and thus important for bacterial pathogenesis (Hase, 2001, McNally et al., 2007). Until now, two forms of motility, swimming and swarming, have been observed. Swarming motility is defined as the fast (2-10 μm/s) and coordinated bacterial movement on the surface through rotating flagella (Henrichsen, 1972, Harshey, 2003), whereas swimming motility is defined as the bacterial movement individually in liquid phase (Moens & Vanderleyden, 1996). In *B. glumae*, swimming and swarming activities can be observed on Luria–Bertani (LB) agar plates with 0.4% and 0.7% agar, respectively (Kim et al., 2007). The nonmotile mutants, which are defective in flagella formation, cause minimum disease symptom on rice (Kim et al., 2007). As for the regulation of toxoflavin, flagella formation is also under the regulation of TofI/TofR

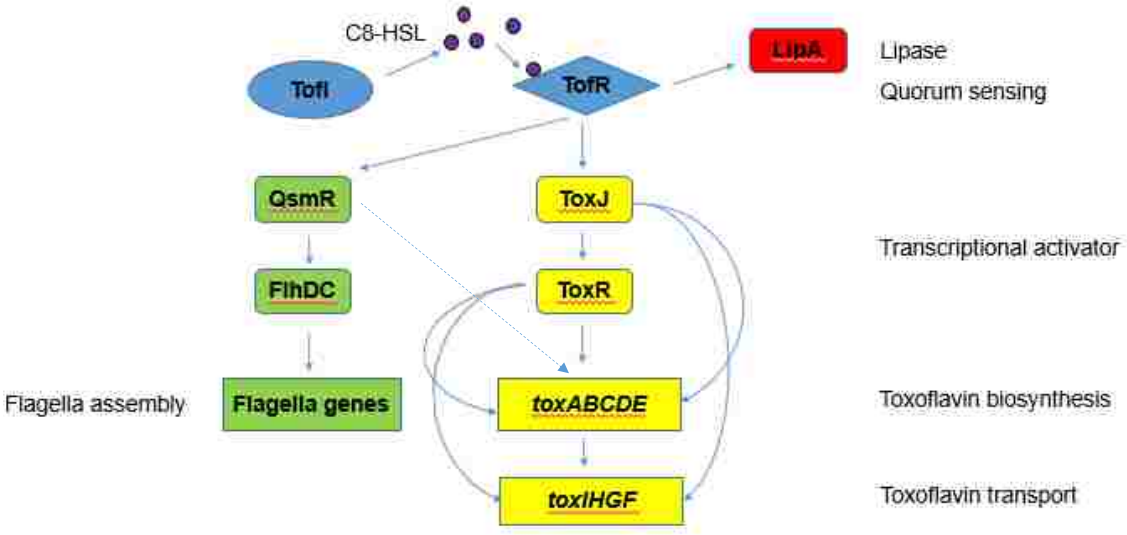


Figure 1.2. Schematic view of the regulatory cascade of toxoflavin, lipase and flagella by QS (Devescovi et al., 2007, Kim et al., 2004a, Kim et al., 2007).

QS system at 37 ˚C but not 28 ˚C (Kim et al., 2007). Additionally, flagella are also governed by quorum sensing master regulator R (QsmR), FlhC, and FlhD (Figure 1.2). Mutation of any of these regulatory genes causes the flagellum-deficient and nonmotile phenotype (Kim et al., 2007).

Lipase. Lipase is also considered as an important virulence factor. Mutation in *lipA*, encoding LipA lipase, leads to the attenuated disease symptoms on rice (Devescovi et al., 2007). Lipase has been known to be secreted through the type II secretion system (Kang et al., 2008). Lipase activity is also reported to be mediated by TofI/TofR QS system (Figure 1.2) (Devescovi et al., 2007, Kim et al., 2007).

Type III effectors. The type III secretion system has been widely used by various Gram-negative bacterial pathogens to directly inject proteinaceous virulence factors, type III effectors, into eukaryotic host cells (Abe et al., 2005)**.** Mutants of *B. glumae* which are defective in T3SS are significantly less virulent on rice panicles in comparison to the wild type strain (Kang et al., 2008), indicating that type III effectors also contribute to the full virulence of this pathogen. However, no type III effectors have been reported from the studies of *B. glumae*.

Other potential virulence factors. Catalase can also be an important virulence factor because insertional mutation in *katG*, encoding catalase, causes less severe disease symptoms on rice (Chun et al., 2009). This is presumably because bacteria become more susceptible to visible light without the protection of catalase (Chun et al., 2009). Interestingly, the expression of *katG* can be directly activated by QsmR (Chun et al., 2009). *B. glumae* has also been reported to produce exopolysaccharides (EPSs) on casamino acid peptone glucose medium. The importance of EPSs for the virulence of *B.*

glumae is still unknown. Two endo-polygalacutronases, PehA and PehB, have been reported to be important for virulence in *Burkholderia* spp. Interestingly, knockout of *pehA* or *pehB* individually did not show a significant effect on the virulence of *B. glumae*, but knockout of the both genes was not achieved (Degrassi et al., 2008). So, the importance of polygalacturonases for the virulence of *B. glumae* is still largely unknown.

Functions of LuxO regulator in *Vibrio* spp. TepR, encoded by bglu_1g09700, is homologous to LuxO in *Vibrio* spp. (Osti, 2014), and they may share similar regulatory functions. In *Vibrio harveyi*, light production is achieved through the expression of luciferase encoded by *luxCDABEGH* (Meighen, 1991). The expression of *luxCDABEGH* is regulated by LuxI/LuxR-mediated QS system in *V. fischeri*, (Nealson et al., 1970, Schaefer et al., 1996). *V. harveyi* possesses two independent QS systems (system 1 and system 2). Both systems are composed of an autoinducer N-(3-hydroxybutanoyl)-L-homoserine lactone (AI-1) or AI-2, and its cognate sensor LuxN or LuxQ, respectively (Bassler et al., 1993). These sensors belong to two-component signaling proteins, with both sensor kinase domain and response regulator domain (Bassler and Silverman, 1995). Bonnie *et al.* identified a novel regulator in *V. harveyi*, LuxO, which negatively controlled the expression of luminescence. Amino acid sequence analysis indicated LuxO was homologous to a two-component response regulator, NtrC (Bassler et al., 1994b, Bassler et al., 1994a). It was proposed that LuxO was able to interact with the two QS systems in *V. harveyi* by repression (Bassler et al., 1994b). It was found that, in low cell density, LuxO was phosphorylated and behaved as a repressor of the luminescence expression; in high cell density, LuxN and LuxQ send the signal to LuxO and inactivated its repression activity through

dephosphorylation (Bassler et al., 1994a). Sequence analysis showed LuxO in *V. harveyi* has a highly conserved 47th aspartate residue (Asp-47), which was of vital importance for the signal transduction (Bassler et al., 1994a). It was revealed that substitution of the conserved aspartate to glutamate (D47E) mimicked the phosphorylated LuxO (constitutive active form of LuxO), and substitution of the conserved aspartate to alanine or asparagine (D47A or D47N) mimicked the unphosphorylated LuxO (inactive form of LuxO) (Freeman & Bassler, 1999). Later, Carol *et al.* also characterized the negative regulatory function of LuxO in the luminescence in *V. fischeri* (Miyamoto et al., 2000).

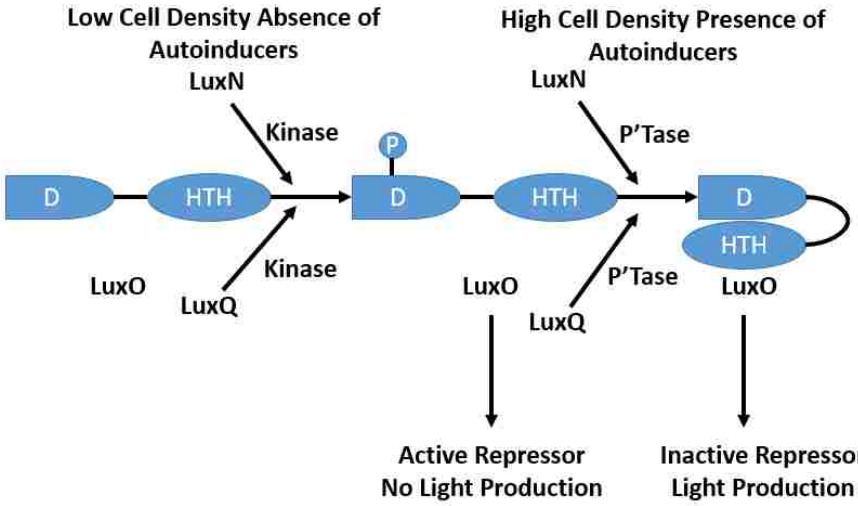


Figure 1.3. Cascade of the regulation of quorum sensing in *V. harveyi*. D stands for Asp-47; HTH denotes helix-turn-helix; P-LuxO indicates LuxO in phosphorylated form; P'Tase indicates phosphatase (Freeman & Bassler, 1999).

It is suggested that the negative regulatory role of the phosphorylated LuxO (active LuxO) was fulfilled by the activation of transcription of the small RNAs (sRNAs), qrr1-4, which functioned together with the sRNA-binding protein Hfq (Lenz et al., 2004). The sRNA-Hfq complex could affect the stability of the mRNA of quorum-sensing

master regulator LuxR (*V. harveyi*) and HapR (*V. cholerae*) (Lenz et al., 2004, Vance et al., 2003). And HapR (LuxR) is a positive regulator of a hemagglutinin/protease gene, *hapA* (Raychaudhuri et al., 2006). Besides the negative regulatory role in the luminescence and protease production, LuxO was also characterized to positively regulate the biofilm formation in *V. cholera* (Vance et al., 2003). LuxO was found to be negatively regulated by a TetR family protein, LuxT, by binding to the promoter of *luxO* gene (Lin et al., 2000). The regulatory role of LuxO is most likely to be well conserved in *Vibrio* spp. (Miyamoto et al., 2003). However, its functions in bacteria of other genus have never been reported.

Previous study of *tepR* in *B.* glumae. Previously, our lab identified *tepR* (locus_tag=bglu_1g09700), encoding a sigma 54-dependent response regulator (Osti, 2014). It was found to negatively mediate toxoflavin and protease production, and virulence in rice (Osti, 2014). Specifically, a markerless deletion mutation of *tepR* was constructed using the suicide vector pKKSacB from the wild type *B. glumae* strain 336gr-1 (Figure 1.4), named LSUPB401. For complementation, the coding region of *tepR*, together with 438bp in upstream and 233bp in downstream, was cloned to the plasmid pBBR1MCS-5 and introduced into LSUPB401 through triparental mating. Several phenotypic experiments were conducted to confirm the role of *tepR* in regulating toxoflavin, protease activities, swimming and swarming activities, and T3SS elicitation (Osti, 2014). It was indicated that *tepR* had negative roles in controlling toxoflavin production (Figure 1.5A), protease activities (Figure 1.5B), and positive roles in swimming and swarming activities (Osti, 2014). However, other regulatory functions of *tepR* was unknown. My research aims for a more comprehensive understanding of

the regulatory functions of the *tepR* and its regulation mechanisms through genetic and transcriptomic techniques.

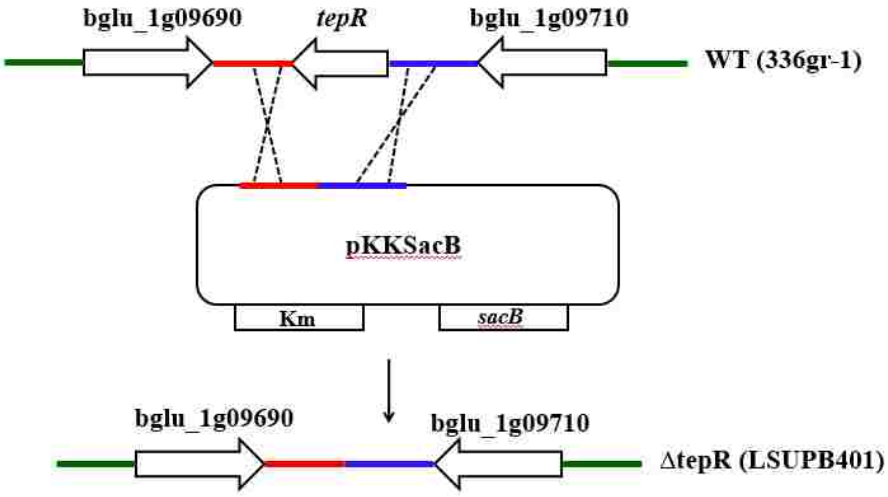


Figure 1.4. Schematic view of *ΔtepR* derivative of wild type *B. glumae* strain 336gr-1. Specifically, 479 bp upstream sequence (blue) and 358 bp downstream sequence (green) of *tepR* were cloned into pKKSacB and introduced into *B. glumae*. The deletion mutant was generated by double-crossover homologous recombination.

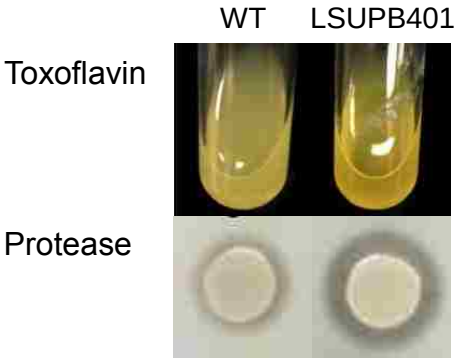


Figure 1.5. Increased production of toxoflavin and protease in LSUPB401 (a *ΔtepR* derivative of the wild type 336gr-1) (Osti, 2014).

Significance and Rationale

Rice (Oryza sativa L.) is one of the most important crops which serves as the basic dietary food for around half of the world's population (Zeigler & Barclay, 2008). The United States is one of the top five major exporters of rice (USA Rice Federation;

www.irri.org), and produces more than 19 billion pounds of rice annually. The major rice production in the United States is in the states of Arkansas, California, Louisiana, Mississippi, Missouri, and Texas (USA Rice Federation; www.irri.org). Rice production, milling and downstream marketing in the US constitute a billions of dollar industry which is of vital importance for the health of rural economic development (USA Rice Federation; www.irri.org).

However, rice production has been challenged by various diseases, among which BPB is one of the most devastating disease, caused by *B. glumae* (Ham et al., 2011). Efforts have been taken to develop chemicals that could control the disease, but, so far, no chemicals could control the disease efficiently and reliably (Ham et al., 2011). Our lab has been focused on using molecular genetic techniques to decipher the regulatory mechanisms of the bacterial virulence in *B. glumae* and finding out the candidate targets for disease control. My research study emphasizes the comprehensive understanding of the global regulatory functions of *tepR* in modulating bacterial behaviors and its regulation mechanisms. Specifically, the following objectives were pursued:

1. Investigation of the genes, pathways and systems that are significantly regulated by *tepR* in *B. glumae*.
2. Characterization of the regulatory mechanisms of *tepR* in *B. glumae*.

This study is very important in several ways. This is the first report of a comprehensive study of LuxO-homolog in organisms other than *Vibrio* spp. It will help to develop chemical agents that target the key regulatory factor, *tepR*, for disease suppression. TepR-homologs have been found in various plant pathogenic and animal pathogenic bacteria, indicating the global regulatory nature of TepR in the kingdom of bacteria. This study also uncovered a putative type VI secretion system, which is most

likely responsible for bacterial interspecies competition. This may direct to novel biological and chemical disease control strategies.

References

Abe A, Matsuzawa T, Kuwae A, 2005. Type-III effectors: sophisticated bacterial virulence factors. *C R Biol* **328**, 413-28.

Bassler BL, Wright M, Showalter RE, Silverman MR, 1993. Intercellular signalling in Vibrio harveyi: sequence and function of genes regulating expression of luminescence. *Mol Microbiol* **9**, 773-86.

Bassler BL, Wright M, Silverman MR, 1994a. Multiple signalling systems controlling expression of luminescence in Vibrio harveyi: sequence and function of genes encoding a second sensory pathway. *Mol Microbiol* **13**, 273-86.

Bassler BL, Wright M, Silverman MR, 1994b. Sequence and function of LuxO, a negative regulator of luminescence in Vibrio harveyi. *Mol Microbiol* **12**, 403-12.

Chun H, Choi O, Goo E*, et al.*, 2009. The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *J Bacteriol* **191**, 4152-7.

Cottyn B, Outryve M, Cleene M, *et al.*, 1996. Bacterial diseases of rice. II. Characterization of pathogenic bacteria associated with sheath rot complex and grain discoloration of rice in the Philippines. *Plant Disease* 4, 438-45.

Degrassi G, Devescovi G, Kim J, Hwang I, Venturi V, 2008. Identification, characterization and regulation of two secreted polygalacturonases of the emerging rice pathogen Burkholderia glumae. *FEMS Microbiol Ecol* **65**, 251-62.

Devescovi G, Bigirimana J, Degrassi G*, et al.*, 2007. Involvement of a quorum-sensing-regulated lipase secreted by a clinical isolate of Burkholderia glumae in severe disease symptoms in rice. *Appl Environ Microbiol* **73**, 4950-8.

Dunlap PV, 1999. Quorum regulation of luminescence in Vibrio fischeri. *J Mol Microbiol Biotechnol* **1**, 5-12.

Freeman JA, Bassler BL, 1999. A genetic analysis of the function of LuxO, a two-component response regulator involved in quorum sensing in Vibrio harveyi. *Mol Microbiol* **31**, 665-77.

Fuqua WC, Winans SC, Greenberg EP, 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* **176**, 269-75.

Ham JH, Melanson RA, Rush MC, 2011. Burkholderia glumae: next major pathogen of rice? *Mol Plant Pathol* **12**, 329-39.

Harshey RM, 2003. Bacterial motility on a surface: many ways to a common goal. *Annu Rev Microbiol* **57**, 249-73.

Hase CC, 2001. Analysis of the role of flagellar activity in virulence gene expression in Vibrio cholerae. *Microbiology* **147**, 831-7.

Henrichsen J, 1972. Bacterial surface translocation: a survey and a classification. *Bacteriol Rev* **36**, 478-503.

Hung KT, Kao CH, 2004. Hydrogen peroxide is necessary for abscisic acid-induced senescence of rice leaves. *J Plant Physiol* **161**, 1347-57.

Jeong Y, Kim J, Kim S, Kang Y, Nagamatsu T, Hwang I, 2003. Toxoflavin produced by Burkholderia glumae causing rice grain rot is responsible for inducing bacterial wilt in many field crops. *Plant Disease* **87**, 890-5.

Kim J, Kang Y, Choi O*, et al.*, 2007. Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in Burkholderia glumae. *Mol Microbiol* **64**, 165-79.

Kim J, Kim JG, Kang Y*, et al.*, 2004. Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in Burkholderia glumae. *Mol Microbiol* **54**, 921-34.

Latuasan HE, Berends W, 1961. On the origin of the toxicity of toxoflavin. *Biochim Biophys Acta* **52**, 502-8.

Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL, 2004. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. *Cell* **118**, 69-82.

Lim J, Lee TH, Nahm BH, Choi YD, Kim M, Hwang I, 2009. Complete genome sequence of Burkholderia glumae BGR1. J Bacteriol 191, 3758-9.

Lin YH, Miyamoto C, Meighen EA, 2000. Purification and characterization of a luxO promoter binding protein LuxT from Vibrio harveyi. *Protein Expr Purif* **20**, 87-94.

Luo J, Xie G, Li B, Lihui X, 2007. First report of Burkholderia glumae isolated from symptomless rice seeds in china. *Plant Disease* **91**, 1363-.

Maeda Y, Kiba A, Ohnishi K, Hikichi Y, 2004. Implications of amino acid substitutions in GyrA at position 83 in terms of oxolinic acid resistance in field isolates of Burkholderia glumae, a causal agent of bacterial seedling rot and grain rot of rice. *Applied and Environmental Microbiology* **70**, 5613-20.

Mcnally A, La Ragione RM, Best A, Manning G, Newell DG, 2007. An aflagellate mutant Yersinia enterocolitica biotype 1A strain displays altered invasion of epithelial cells, persistence in macrophages, and cytokine secretion profiles in vitro. *Microbiology* **153**, 1339-49.

Meighen EA, 1991. Molecular biology of bacterial bioluminescence. *Microbiol Rev* **55**, 123-42.

Miller MB, Bassler BL, 2001. Quorum sensing in bacteria. *Annu Rev Microbiol* **55**, 165-99.

Miyamoto CM, Dunlap PV, Ruby EG, Meighen EA, 2003. LuxO controls luxR expression in Vibrio harveyi: evidence for a common regulatory mechanism in Vibrio. *Mol Microbiol* **48**, 537-48.

Miyamoto CM, Lin YH, Meighen EA, 2000. Control of bioluminescence in Vibrio fischeri by the LuxO signal response regulator. *Mol Microbiol* **36**, 594-607.

Moens S, Vanderleyden J, 1996. Functions of bacterial flagella. *Crit Rev Microbiol* **22**, 67-100.

Nandakumar R, Bollich PA, Shahjahan AKM, Groth DE, Rush MC, 2008. Evidence for the soilborne nature of the rice sheath rot and panicle blight pathogen, Burkholderia gladioli. *Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie* **30**, 148-54.

Nandakumar R, Rush M, Shahjahan A, O'reilly K, Groth D, 2005. Bacterial panicle blight of rice in the southern United States caused by Burkholderia glumae and B-gladioli. *Phytopathology* **95**, S73-S.

Nandakumar R, Rush MC, Correa F, 2007. Association of Burkholderia glumae and B-gladioli with panicle blight symptoms on rice in Panama. *Plant Disease* **91**, 767-.

Nandakumar R, Shahjahan AKM, Yuan XL*, et al.*, 2009. Burkholderia glumae and B. gladioli Cause Bacterial Panicle Blight in Rice in the Southern United States. *Plant Disease* **93**, 896-905.

Nealson KH, Platt T, Hastings JW, 1970. Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol* **104**, 313-22.

Osti S, 2014.Characterization of a σ^{54}-dependent response regulatior, *tepR*, in the rice-pathogenic bacterium *Burkholderia glumae* and development of biocontrol strategies for bacterial panicle blight of rice (Master's thesis). *LSU Electronic Thesis & Dissertation Collection*.

Raychaudhuri S, Jain V, Dongre M, 2006. Identification of a constitutively active variant of LuxO that affects production of HA/protease and biofilm development in a non-O1, non-O139 Vibrio cholerae O110. *Gene* **369**, 126-33.

Riera-Ruiz C, Vargas J, Cedeno C*, et al.*, 2014. First Report of Burkholderia glumae Causing Bacterial Panicle Blight on Rice in Ecuador. *Plant Disease* **98**, 988-9.

Sayler RJ, Cartwright RD, Yang YN, 2007. Genetic characterization and real-time PCR detection of Burkholderia glumae, a newly emerging bacterial pathogen of rice in the United States. (vol 90, pg 603, 2006). *Plant Disease* **91**, 1050-.

Schaefer AL, Val DL, Hanzelka BL, Cronan JE, Jr., Greenberg EP, 1996. Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified Vibrio fischeri LuxI protein. *Proc Natl Acad Sci U S A* **93**, 9505-9.

Shingu Y, Yoneyama K, 2004. Essential regulator gene toxR for toxoflavin biosynthesis of *Burkholderia glumae*. *J Gen Pathol* **70**, 108-14.

Takeuchi T, Sawada H, Suzuki F and Matsuda I, 1997. Specific detection of *Burkholderia plantarii* and *B. glumae* by PCR using primers selected from the 16S-23S rDNA spacer regions. *Ann Phytopath Soc Japn* **63**, 455-62.

Tsushima, S., S. Wakimoto and S. Mogi. 1986. Selective medium for detecting *Pseudomonas glumae* Jurita et Tabei, the causal bacterium of grain rot of rice. *Ann Phytopath Soc Japn* **52**, 253-59.

Tsushima S, 1996. Epidemiology of bacterial grain rot of rice caused by Pseudomonas glumae. *Jarq-Japan Agricultural Research Quarterly* **30**, 85-9.

Vance RE, Zhu J, Mekalanos JJ, 2003. A constitutively active variant of the quorum-sensing regulator LuxO affects protease production and biofilm formation in Vibrio cholerae. *Infect Immun* **71**, 2571-6.

Wan XY, Liu JY, 2008. Comparative proteomics analysis reveals an intimate protein network provoked by hydrogen peroxide stress in rice seedling leaves. *Mol Cell Proteomics* **7**, 1469-88.

Yuan X, 2004. Identification of Bacterial Pathogens Causing Panicle Blight of Rice in Louisiana (master's thesis). Retrieved from *http://etd.lsu.edu/docs/available/etd-01282004-111538/*

Zeigler RS, Barclay A, 2008. The Relevance of Rice. *Rice* **1**, 3-10.

Zhou XG, 2014. First Report of Bacterial Panicle Blight of Rice Caused by Burkholderia glumae in South Africa. *Plant Disease* **98**, 566-.

CHAPTER 2: GLOBAL REGULATORY ROLE OF *TEPR* IN *BURKHOLDERIA GLUMAE* AS REVEALED BY WHOLE TRANSCRIPTOMIC ANALYSES

Introduction

Whole transcriptome studies for characterization of the regulatory systems in eukaryotes have been available for more than a decade (Sorek & Cossart, 2010). Yet, transcriptome-wide studies in prokaryotes have been very hard, largely due to the lack of poly(A) tails in mRNAs, which are commonly used for enrichment of mRNAs (Sorek & Cossart, 2010). RNA sequencing (RNA-seq) has been a revolutionary technology for studying bacterial transcriptomes (Croucher & Thomson, 2010). It has advantages over microarray because RNA-seq can display the transcription features digitally with single-nucleotide resolution and does not require a sequenced genome as reference (Wang et al., 2009). Most RNA-seq studies include the following steps: RNA extraction, mRNA enrichment, library preparation, sequencing, gene expression profiling, annotation, and identification of novel transcripts. Following total RNA extraction, an mRNA enrichment step is necessary to get rid of all the ribosomal RNA and tRNA, which take up to 95% of the bacterial transcriptome (Neidhardt et al., 1987). Sequencing can be achieved by such cost-effective and high-throughput sequencing technologies as the Roche 454, Illumina Genome Analyzer, Applied Biosystems SOLiD and Ion torrent platforms (Wang et al., 2009, Quail et al., 2012). The output of RNA-seq usually contains raw sequence reads ranging from 25 bp to 200 bp. Millions of short reads of high quality can then be assembled into large contigs or scaffolds, and mapped to a reference genome (Schatz et al., 2010). The resultant abundance of transcripts can provide information for multiple purposes: identification of untranslated regions (promoters and terminators), identification of novel genes, identification of initiation codon (correction of gene

annotation), re-construction of operon structure, discovery of non-coding RNAs and antisense RNAs (Creecy & Conway, 2015, Quail et al., 2012, Sorek & Cossart, 2010). Several novel RNA-seq-based technologies have been developed. For example, differential RNA-seq (dRNA-seq) could also estimate whether the RNAs are primary or processed transcripts, based on a selective sequencing method (Sharma et al., 2010). Also, metatranscriptomics has been used to assess the transcript abundance of a microbial community, instead of studying single transcriptome from a pure culture (Frias-Lopez et al., 2008, Poretsky et al., 2005).

RNA-seq has been widely used to determine differential gene expression under different growth conditions and treatments (Kroger et al., 2013, Kim et al., 2014). To study the comprehensive functions of *tepR* in *B. glumae*, whole transcriptome analyses were conducted by RNA-seq, comparing the expression profile between the wild type (WT) *B. glumae* strain 336gr-1 and its *tepR* deletion derivative (*ΔtepR*), LSUPB401.

Although various computational tools are available for RNA-seq data processing, such as ABySS (Birol et al., 2009), RSAT (Mikkelsen et al., 2007) and Velvet (Zerbino & Birney, 2008), none of them is specifically designed for bacterial RNA-seq. However, bacterial genome structure is very different from that of eukaryote because of the frequent overlapped adjoining genes and operon structure which make the genome assembly very challenging (McClure et al., 2013). Rockhopper, used in this study, incorporates a novel algorithm which emphasizes the complexity of the bacterial genome. Specifically, a Burrows-Wheeler index (Burrows et al., 1994) ,using full-text minute space, is made for a reference genome (Langmead & Salzberg, 2012). With the index, an attempt is made to align each read exactly to the reference genome. When a

read can not be aligned well, the seed regions of the read are defined and aligned to the reference genome, and extended utilizing a dynamic program (McClure et al., 2013). The transcripts are then assembled and annotated. Transcript abundance levels can be quantified based on Reads Per Kilobase per Million mapped reads (RPKM) with upper quartile normalization (Bullard et al., 2010). Upper quartile normalization can increase the accuracy of the transcript abundance estimation, which eliminates the impact of those extremely low expression or non-expressed genes.

In this study, the resultant expression profile of RNA-seq was subjected to Gene Ontology (GO) enrichment. GO enrichment is a technique for explanation of a set of genes using the GO hierarchical classification system, where each gene is associated with functional characteristic terms organized in a graph structure (Young et al., 2010, Harris et al., 2004). GO terms are categorized into three terms, “biological process”, “molecular function” and “cellular components” (Ashburner et al., 2000). “Biological process” represents the biological objective that the gene or gene product correlates to; “molecular function” refers to the biochemical character of a gene product; “cellular component” is defined as the active site of a gene product (Ashburner et al., 2000). Each GO term is classified into different hierarchical levels. A higher level represents a broader definition, e.g. signal transduction, whereas a lower level indicates a more specific definition, e.g. cAMP biosynthesis (Ashburner et al., 2000). The output of the enrichment assay can be displayed in ranked GO terms with an associated significance value (Rhee et al., 2008).

To obtain an overview of the biological network, differentially expressed genes (DEGs) were also analyzed by the Kyoto Encyclopedia of Genes and Genomes (KEGG)

pathway enrichment (Kanehisa & Goto, 2000). KEGG ,by definition, is a collection of databases which links genetic and genomic information to graphical-displayed higher order functional categories (Kanehisa & Goto, 2000). KEGG pathway enrichment is a process to map the listed genes into the existing KEGG metabolic pathways and identify the pathways that are statistically significantly altered (Grosu et al., 2002). The output of KEGG pathway enrichment is usually displayed as a list of statistically enriched pathways with the corresponding significance value.

To validate the expression pattern of RNA-seq, the expression values of selected genes were quantified by quantitative PCR (qPCR). It is able to monitor the PCR product in real-time by detecting proportionally increased fluorescence. Several fluorescence detection methods have been developed, such as DNA-binding-based SYBR green and probed-based TaqMan (Sambrook and Green, 2012). SYBR green was used in this study which binds specifically to double-stranded DNA. The relative transcript abundance of qPCR can be analyzed by the 2(-Delta Delta C(T)) method (Livak & Schmittgen, 2001).

Materials and Methods

Bacterial strains and growth conditions. *B. glumae* strains used in RNA-seq (Table 2.1) were routinely recovered from the strain stock, kept in 30% glycerol at -70°C, and grown in an incubator at 37 ˚C on Luria-Bertani (LB) agar (25 g LB Miller broth (Fisher Scientific, Waltham, MA, USA) and 18 g agar in 1L distilled deionized water (ddH_2O)). Nitrofurantoin (Nit) was supplemented in the medium during the recovery of the *B. glumae* strains from storage to confirm the resistance and prevent any possible contamination of other bacteria. During sampling, strains were grown at 37 ˚C on an

incubator shaker at 200 rpm in LB broth (25 g LB Lennox broth (Fisher Scientific) in 1000 ml ddH2O) without any additional antibiotics.

Table 2.1. *B. glumae* strains used for RNA-seq study

Strain name	Genotype description[a]	References
336gr-1	Wild type (WT) strain isolated from diseased rice in Crowley, Louisiana, USA, Nit^R	(Nandakumar et al., 2009)
LSUPB401	A *ΔtepR* derivative of 336gr-1, Nit^R	(Osti, 2014)

[a]Nit^R indicates the resistance to Nitrofurantoin.

RNA extraction, quality assessment and sequencing. RNA extraction was conducted by the previous lab member Rebecca Melanson. One ml overnight culture of *B. glumae* strains, grown in 10 ml LB broth at 37°C with 200 rpm shaking, were centrifuged at 16,000g for 2 min and washed twice with 1 ml fresh LB broth. Ten µl of the washed culture was finally resuspended in 10 ml LB broth and incubated at 37 ˚C with 200 rpm shaking. Each culture was grown until the optical density at 600 nm (OD600) reached approximately 1.0. One ml of culture was centrifuged at 17,000 x g for 1 min and supernatant was removed. The bacterial pellets were frozen in liquid nitrogen for about 10 s and re-suspended in 1 ml of TRIzol® Reagent (Ambion® Life Technologies, Grand Island, NY, USA). Samples were stored at -70 ˚C before proceeding. For each culture, there were three biological replicates and each had two technical replicates.

Sample homogenate stored in TRIzol® Reagent at -70°C was thawed on ice. RNA was isolated using the Direct-zol™ RNA MiniPrep Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. In column crude RNA was again

purified by DNase provided in the Direct-zolTM RNA MiniPrep Kit according to the manufacturer's instructions. RNA was eluted with 35 µl DNase/RNase-free water. A second DNase treatment using the DNase Treatment and Removal kit (Ambion® Life Technologies, Grand Island, NY, USA) was conducted according to the manufacturer's instructions to get rid of any residual DNA contamination.

The concentration and purity of the extracted RNA was determined by the 260/280 and 260/230 ratios on a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). About 100 ng of total RNA from each sample was run on a 1% agarose gel to check for RNA degradation. The primers GL-13f (ACACGGAACACCTGGGTA) and GL-14r (TCGCTCTCCCGAAGAGAT) (Takeuchi et al., 1997), which specifically targets the spacer region between 16S and 23S ribosomal RNA (rRNA) of *B. glumae*, was used for PCR to check for DNA contamination.

Approximately 2-4 µg of total RNA from each replicate, stored in a Styrofoam ice chest with dry ice, was sent to the Johns Hopkins University Genetic Resources Core Facility (GRCF) in Baltimore, MD, USA, for quality assessment, rRNA depletion, and sequencing. The quality of the RNA samples was assessed with the Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). The elimination of rRNA was conducted using Ribo-ZeroTM (Bacteria) reagents (Epicentre, Madison, WI, USA). Samples was additionally purified using the Agencourt RNAClean XP kit (Catalog No. A63987, Beckman Coulter, Inc. Indianapolis, IN, USA). RNA-Seq libraries were prepared by the ScriptSeq^{TM} v2 RNA-Seq Library Preparation module of the ScriptSeq^{TM} Complete Kit (Bacteria) (Catalog No. BB1206, Epicentre). The quality of the libraries made were tested with the Agilent 2100 BioAnalyzer before conducting

sequencing. RNA sequencing was conducted on a single lane of Illumina HiSeq 2500, HCSv2.2.38, RTAv1.18.61 (Illumina, Inc., San Diego, CA, USA) to obtain 100 bp paired-end reads.

Assembly and annotation of transcriptomes. The quality of the RNA-seq raw data was evaluated by FastQC V0.11.2 (Babraham Bioinformatics, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The alignment, assembly and transcript abundance quantification of raw RNA-seq data was performed using the program Rockhopper version 2.0.3 (McClure et al., 2013). Since the genome of 336gr-1 has not been completely sequenced and annotated, the fully sequenced and annotated genome of *B. glumae* BGR1 obtained from NCBI's Genomes FTP site (ftp://ftp.ncbi.nih.gov/genomes/) was used as the reference for the alignment of RNA-seq raw reads (Table 2.2). Previous research had been using BGR1 genome for designing primers for targeted genes in 336gr-1 and no problems had been encountered (Chen et al., 2012), indicating their high similarity in genome sequence. So, alignment of the transcriptome of 336gr-1 to BGR1 would not be expected to cause noteworthy effects in the downstream analysis. Rockhopper, used in this study, created Burrows-Wheeler indices for 336gr-1 and LSUPB401 (Langmead & Salzberg, 2012). The seed regions and inexact alignment regions were set to be more than a third and less than 15% of the length of the reads, respectively. Transcript abundance was quantified based on Reads Per Kilobase per Million mapped reads (RPKM) with upper quartile normalization. The cut-off value for differentially expressed genes (DEGs) between 336gr-1 and LSUPB401 was greater than 1.5-fold change of transcript abundance with a p-value lower than 0.01.

Table 2.2. GenBank data files and protein feature files used for assembly, alignment and annotation in Rockhopper.

Genome	Name of the files	File description	Last modified date
Chromosome 1	NC_012724.gbk	sequence GenBank file	June 13, 2013
	NC_012724.ptt	Protein feature file	April 13, 2012
Chromosome 2	NC_012721.gbk	sequence GenBank file	June 13, 2013
	NC_012721.ptt	Protein feature file	April 13, 2012
Plasmid blgu_1p	NC_012723.gbk	sequence GenBank file	June 09, 2013
	NC_012723.ptt	Protein feature file	July 09, 2011
Plasmid blgu_2p	NC_012718.gbk	sequence GenBank file	June 09, 2013
	NC_012718.ptt	Protein feature file	July 09, 2011
Plasmid blgu_3p	NC_012720.gbk	sequence GenBank file	June 9, 2013
	NC_012720.ptt	Protein feature file	August 11, 2011
Plasmid blgu_4p	NC_012725.gbk	sequence GenBank file	June 09, 2013
	NC_012725.ptt	Protein feature file	July 09, 2011

Validation of RNA-seq expression pattern by quantitative reverse transcription PCR (qRT-PCR). RNA was extracted and purified by the same preparation procedures as in RNA-seq. ProtoScript® First Strand cDNA Synthesis Kit (New England Biolabs, Ipswich, MA, USA) was used to perform reverse transcriptase PCR on 5 or 6 µl RNA of 336gr-1 and LSUPB401. The resultant cDNAs were used as templates for qPCR. Primers used for qPCR were listed in Table 2.3, designed by the PrimerQuest tool on the Integrated DNA Technologies (IDT) website (http://www.idtdna.com/Primerqukest/Home/Index). For qPCR, 1 µl template cDNAs, 7

µl water, 1 µl 10 µM forward primer and 1 µl 10 µM reverse primer were mixed with 10 µl 2X Power SYBRR Green PCR Master Mix (Life Technologies, Grand Island, NY, USA) to perform a 20 µl qPCR reaction on IQ^{TM}5 Multicolor Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA), and the results were monitored by iQ™5 Optical System Software (Bio-Rad Laboratories, Hercules, CA, USA). Each reaction had three biological replicates and three technical replicates. The qPCR program for all primer sets was initial denaturation at 95°C for 30 s and followed by 40 cycles of 95°C for 15 s and 55°C for 30 s. To ensure the target gene was specifically amplified, the PCR cycling was followed by melting curve analysis from 50 °C to 95 °C. The expression fold changes of selected genes in 336gr-1 and LSUPB401 was estimated by $2^{-\Delta\Delta CT}$ formula (Livak & Schmittgen, 2001) using the constitutively expressed DNA gyrase subunit A gene *gyrA* as reference (Chen and Ham, unpublished).

Table 2.3. The genetic regions and primers used for qPCR verification of RNA-seq.

Internal region of the genes	Gene product	Product length	Primers (forward & reverse)
bglu_1g08720	DNA gyrase subunit A (gyrA) gene	140 bp	GyrA-q1: CACGACTACATCCTCTGTTTCTC GyrA-q2: ACGTTGATCTTCTCGCCTTC
bglu_1g08430	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	119 bp	RT1G08430F1: CGACATGGTGGAGTTCTTC RT1G08430R1: GGTAGACGTCGCCATAGA
bglu_2g09850	hypothetical protein	105 bp	RT2G09850F1: GGCTGCAAGTGTTTCTCT RT2G09850R1: GGAATGCGGATTCGATGT

(Table 2.3 continued)

Internal region of the genes	Gene product	Product length	Primers (forward & reverse)
bglu_2g19340	chaperonin Cpn10	96 bp	RT2G19340F1: GTCGGTGATCAGGTCATTT RT2G19340R1: GCCCATCACGTCTTCTTC
bglu_2g19330	chaperonin GroEL	96 bp	RT2G19330F1: ATCTGAGCCCGTACTTCAT RT2G19330R1: GAACCGCGGAAATCTTCTT
bglu_2g16650	hypothetical protein	127 bp	RT2G16650F1: CATGGGCATGACGTATCTG RT2G16650R1: CAGTATCTTCGCGAGCAC
bglu_1g22650	catalase KatE	109 bp	RT1G22650F1: GATGTCGAACGACCTGTATG RT1G22650R1: CGCGTCGTAATCGAACTT
bglu_2g06440	riboflavin biosynthesis protein RibD	106 bp	RT2G06440F1: CTCGATCCTGACTGGCAT RT2G06440R1: CCGAGCATCGATCAACAC
bglu_1g11120	heat shock protein Hsp20	166 bp	RTHsp20F1: GACCACTTATCGTTTCGG RTHsp20R1: GATCTCGATCGCTTCGTC
bglu_1g10250	quorum sensing master regulator QsmR	125 bp	RTqsmrF1: CGAGATCGAGAATCGTCCA RTqsmrR1: GGGATGCCTGTCAATTCTG
bglu_2g16690	RebB protein	91 bp	RTReb1F1: CCGTCAACGACCAGATCAC RTReb1R1: TGGAACAGGTTGCCCATC

(Table 2.3 continued)

Internal region of the genes	Gene product	Product length	Primers (forward & reverse)
bglu_2g06400	methyltransferase UbiE (ToxA)	116 bp	RTtoxAF1: AGTACGACAGCTCGATACG RTtoxAR1: CCGGAAACGTCACCAATC
bglu_1g03960	type VI secretion system protein TssA	82 bp	RTTssAF1: TCGACGACGAACTCGAAC RTTssAR1: ATGCGCACCGAGTCATAG
bglu_2g16670	RebA protein	94 bp	RT2G16670F1: CAACGATCAGATCACCGATTC RT2G16670R1: CGTGGCCTGGTAGAGATT

Annotation of DEGs with unknown functions. Among the 238 DEGs, 36 genes were annotated as “hypothetical protein” with 31 of them being up-regulated in LSUPB401. The functions of these hypothetical proteins were predicted and annotated using the HMMER 3.1b1 (Eddy, 1998, Mistry et al., 2013). Specifically, the amino acid sequence of each hypothetical protein gene was subjected to BLAST 2.2.28+ using the program “blastp” against the database “Non-redundant protein sequences (nr)” (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The sequences of the top thirty hits were retrieved and aligned with MUSCLE v3.6 (Edgar, 2004). The alignment file was inputted to make the profile HMM using the program “hmmbuild”. The resultant profile HMM was subjected to the “hmmsearch” against the “Reference Proteomes” database (http://hmmer.janelia.org/search/hmmsearch) with the options setting “--E 1 --domE 1 --incE 0.01 --incdomE 0.03 --seqdb refseq”. The function of the top hit with the lowest E-

value was used as the new annotation for each of the hypothetical proteins. The cutoff E-value was 1.0E-3.

Gene ontology (GO) enrichment. Blast2GO 3.0 was used for whole transcriptome GO annotation (Conesa et al., 2005). Build-in Blast search was first performed for the whole-transcriptome of *B. glumae* BGR1. The transcriptome file of *B. glumae* BGR1 in FASTA format was downloaded from MicrobesOnline (http://www.microbesonline.org)(Dehal et al.,2010).The Blast program used was “blastx”; blast database was “nr”; blast expectation value (E-value) was 1.0E-3; blast mode was “QBlast-NCBI”. “Gene Ontology Mapping” was performed to retrieve all the GO terms associated to the hits obtained by the BLAST search results using the default “non-redundant reference protein” database, i.e. PSD, UniProt, Swiss-Prot, TrEMBL, RefSeq, GenPept and PDB. “InterProScan” was then performed with all the available database (BlasProDom, FPrintScan, HMMPIR, HMMSmart, HMMTiger,etc).This process retrieved all the information of known domain or motif in a sequence-wise manner. The corresponding GO terms were then transferred and merged with the existing GO terms through the function “Merge InterProScan GOs to Annotation”. The output was exported as the whole transcriptome GO database of *B. glumae*. The GO terms for the DEGs in this study were also selected for further GO enrichment assay. The GO enrichment was conducted by using the Singular Enrichment Analysis (SEA) from the web-based program agriGO version 1.2 (Du et al., 2010). Since the GO annotation database was not available in the web server, the GO terms of DEGs and whole transcriptome together with their gene locus tags, formatted in two columns as required, were inputted in the “Customized annotation” and “Customized annotated reference” respectively. The

significance of the enriched term was analyzed by Fisher's exact test with Yekutieli (FDR under dependency) adjustment. The cut-off p-value was set at 0.05; the minimum number of mapping entries was set at 5. The enrichment result was categorized in terms of biological process, molecular function, and cellular component.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. To obtain an overview of the gene biological network, the DEGs were also subjected to KEGG pathway enrichment assay using KOBAS 2.0 (Wu et al., 2006). Significant pathways were selected at the p-value ≤ 0.10 using Hypergeometric test/ Fisher's exact test. The minimum number of the displayed terms was set at 5. The representative enriched pathway networks were reconstructed with the program Visualization and Analysis of Networks containing Experimental Data (VANTED) V 2.2.0 (Junker et al., 2006). KEGG pathway maps of *B. glumae* were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/).

Classification of *T6SS-A* in *B.* glumae. The sequences of T6SS gene cluster (bglu_1g03850 – bglu_1g03990) that was significantly down-regulated in *B. glumae* LSUPB401 and reference T6SS gene clusters sequences in *B. pseudomallei* in GenBank format were downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov/). All the sequences downloaded were aligned with Mauve 2.3.1 and visualized by the build-in alignment viewer (Darling et al., 2010).

Examination of *tssA* expression in different bacteria growth stage by qRT-PCR. RNA was extracted and purified by the same preparation procedures as in RNA-seq except an overnight *B. glumae* culture was diluted to OD_{600}=0.05 and harvested at 0 hours. The culture was harvested every two hours. Procedure for cDNA synthesis and

qPCR followed the previously described method. The primers used for detecting *tssA* were RTTssAF1 and RTTssAR1 (Table 2.3).

Generation of heat maps. All the heat maps in this study were generated by MultiExperiment Viewer (MeV) v4.9 (Saeed et al., 2003). The data was displayed as $\log_2$ (RPKM) or $\log_{10}$ (RPKM) and mapped to colors using the "green-black-red" scheme.

Results

RNA-Seq raw data quality assessment and assembly. The quality of RNA-seq raw data was analyzed by FastQC, which provided quality control statistics for high-throughput raw sequence data (Babraham Bioinformatics, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The quality of each raw sequence data was analyzed in terms of "Per base sequence quality score", "Per sequence quality score", "Per base sequence content" and "Overrepresented sequences" etc., among which "Per base sequence quality score" was most important parameter. The summary of "Per base sequence quality" of each raw sequence data is presented in Appendix A. For each sample, the median quality scores and inter-quartile range (25-75%) scores of all read positions were greater than 30, indicating the excellent quality calls of each sample. Thus, no further data trimming was necessary. The total reads of each sample was greater than 14.3 million reads (Appendix A). Based on the formula Coverage = (Average read length) × (Number of reads)/ (Length of genome), total reads were enough to cover the genome of *B. glumae* (≈6.8 Mb) more than 200 times, which were "deep" enough for further analyses (Ajay et al., 2011). Some raw sequence data were indicated to have overrepresented Illumina Single End adapter sequence (AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCC)

or TruSeq Adapter (AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATG). The adapter sequences did not affect the reference-based assembly in this study since they did not hit any reference genome sequence of *B. glumae* BGR1 (data not shown).

Table 2.4. Assembly report of RNA-seq raw data.

B. glumae sample[a]	Total reads	Aligned total reads	Percentage of aligned total reads
336gr-1 R1	14312452	10572181	73.9
336gr-1 R2	19170445	14028786	73.2
336gr-1 R3	18207519	12992523	71.4
LSUPB401 R1	17476629	13057231	74.7
LSUPB401 R2	16735198	12172974	72.7
LSUPB401 R3	15252900	10930324	71.7

[a]R1, R2, and R3 represent three biological replicate of the sample.

B. glumae sample[a]	Aligned reads[b]					
	bglu_1g	bglu_2g	bglu_1p	bglu_2p	bglu_3p	bglu_4p
336gr-1 R1	8964213	1488626	48721	23067	31168	16386
336gr-1 R2	11869408	2024288	51818	24359	37991	20922
336gr-1 R3	11052326	1828906	42752	20472	31544	16523
LSUPB401 R1	10785737	2166736	42558	26548	21068	14584
LSUPB401 R2	9929335	2129695	45864	28592	23099	16389
LSUPB401 R3	8998425	1807445	48726	39674	20476	15578

[b]bglu_1g is chromosome 1 and bglu_1p is plasmid 1 and so forth.

Each raw RNA-seq data was mapped to the *B. glumae* BGR1 reference genome by Rockhopper version 2.0.3. The aligned total reads all samples ranges from 11 million to 14 million (Table 2.4), which provided greater than 160x coverage of the *B. glumae* genome. The average of percentage of aligned total reads was 72.9%. Majority of reads

were aligned to Chromosome 1 (3,906,507 bp). The rest of reads were mostly aligned to Chromosome 2 (2,827,333 bp). Around 1% of reads were aligned to the four plasmids, bplu_1p (133,591 bp), bplu_2p (141,792 bp), bplu_3p (141,067 bp), and bplu_4p (134,346 bp). The details of assembly information is displayed in Table 2.4.

Identification of DEGs from transcriptomic-wide profile between WT and LSUPB401. DEGs were defined by greater or equal to 1.5-fold change with 99% confidence level. A total of 238 DEGs were identified, with 199 DEGs being up-regulated and 39 being down-regulated in LSUPB401, respectively (Figure 2.1, Appendix B). A hundred and thirty seven DEGs were located on Chromosome 1 (bglu_1g) and 94 were located on Chromosome 2 (bglu_2g). Two, two and three DEGs were located on Plasmid 1 (bglu_1p), Plasmid 3 (bglu_3p) and Plasmid 4 (bglu_4p), respectively (Appendix B). A total of 38 genes were differentially expressed with greater or equal to 4-fold change. A hundred and six genes were differentially expressed with greater or equal to 2-fold change but lower than 4-fold. Ninety four genes were differentially expressed with greater or equal to 1.5-fold change but lower than 2-fold.

Among the DEGs that were up-regulated in LSUPB401, bglu_1g08430, encoding a 5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase, had a 7.1-fold change. The chaperonin Cpn10 gene, bglu_2g19340, was 5.3-fold up-regulated in LSUPB401. A serine metalloprotease gene, bglu_1g16590, was also significantly up-regulated with a 4.5-fold change. Amino acid sequence analyses by SignalP-4.1 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al., 2011) indicated that there was a putative signal peptide of the protein, with the C-score (raw cleavage site score), S-score (signal peptide score) and Y-score (indicates combined cleavage site

score) of 0.801, 0.869 and 0.987, respectively (Figure 2.2). The signal peptide was predicted to be released on the 27th amino acid between AQA and QT. One TenA family transcriptional activator gene, bglu_2g06890, had a 4.3-fold increase in LSUPB401. Besides, *katE*, encoding catalase, was 4.6-fold up-regulated in LSUPB401.

Table 2.5. Summary of the regulation in LSUPB401 for the genes responsible for toxoflavin production and its known regulatory systems.

Gene	Product	Function	Significance[a]
bglu_2g06330	ToxJ	Transcriptional activator	None
bglu_2g06350	ToxI	Toxoflavin transport	Up
bglu_2g06360	ToxH		Up
bglu_2g06370	ToxG		Up
bglu_2g06380	ToxF		Up
bglu_2g06390	ToxR	Transcriptional activator	None
bglu_2g06400	ToxA	Toxoflavin biosynthesis	Up
bglu_2g06410	ToxB		Up
bglu_2g06420	ToxC		Up
bglu_2g06430	ToxD		Up
bglu_2g06440	ToxE		Up
bglu_2g14470	TofR	Quorum sensing	None
bglu_2g14480	TofM		None
bglu_2g14490	TofI		None
bglu_1g10250	QsmR		Up

[a]Significance indicates the significant regulation of the gene in LSUPB401.

The gene clusters *toxABCDE* (bglu_2g06400-6440), encoding toxoflavin biosynthesis system and *toxIHGF* (bglu_2g06350-6380), encoding toxoflavin transport system, were consistently up-regulated in LSUPB401 (Table 2.5). However, toxoflavin transcriptional regulators *toxJ* (bglu_2g06330) and *toxR* (bglu_2g06390) were not differentially expressed (Table 2.5). The production of toxoflavin in *B. glumae* has known to be tightly regulated by quorum sensing (QS) mediated by *tofI*, *tofR* and *tofM* (Chen et al., 2012b). In this study, the expression of *tofI* (bglu_2g14490), *tofR* (bglu_2g14470) and *tofM* (bglu_2g14480) were not significantly differentially expressed in LSUPB401 (Table 2.5).

The quorum sensing master regulator gene, *qsmR* (bglu_1g10250), was significantly up-regulated in LSUPB401 by 2.7 folds.

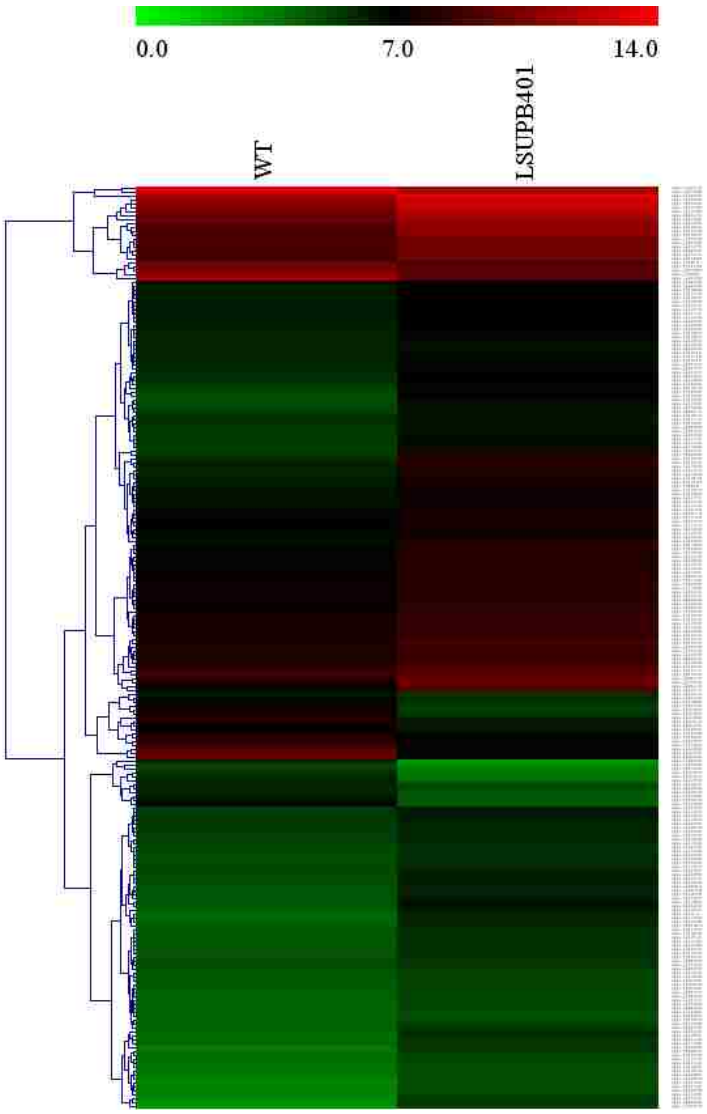


Figure 2.1. Heat map showing a total of 238 DEGs with greater than 1.5 fold change at 99% confidence level between WT (336gr-1) and LSUPB401 (*ΔtepR*). Colors in cells represent $\log_{10}$ RPKM of the genes in WT and LSUPB401.

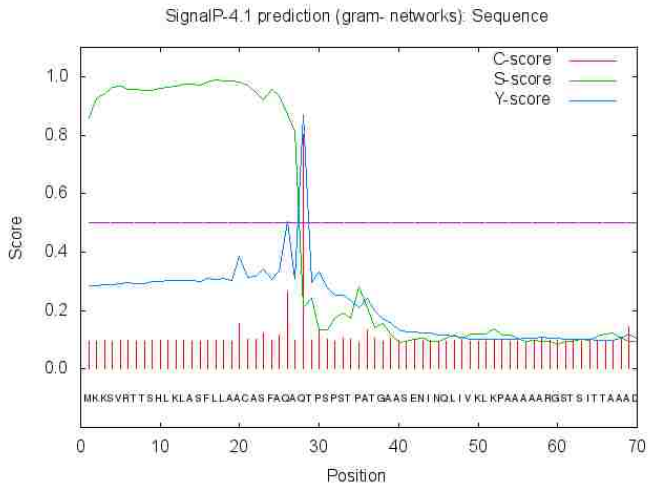


Figure 2.2. Signal peptide prediction of bglu_1g16590. C-score indicates raw cleavage site score; S-score indicates signal peptide score, Y-score indicates combined cleavage site score.

Among the DEGs that were down-regulated in LSUPB401, the expression of *tepR* (bglu_1g09700) was negligible in LSUPB401 (Appendix B). This was reasonable since LSUPB401 is the *tepR* knockout mutant. Gene, bglu_1g03930, encoding a type VI secretion system protein TssF, was the most significantly down-regulated gene with a 5.8-fold change (Appendix B). bglu_1g22700, encoding a LuxR family transcriptional regulator, was 4-fold down-regulated in LSUPB401 (Appendix B). In all 238 DEGs, 36 of them had unknown functions and were annotated as hypothetical proteins.

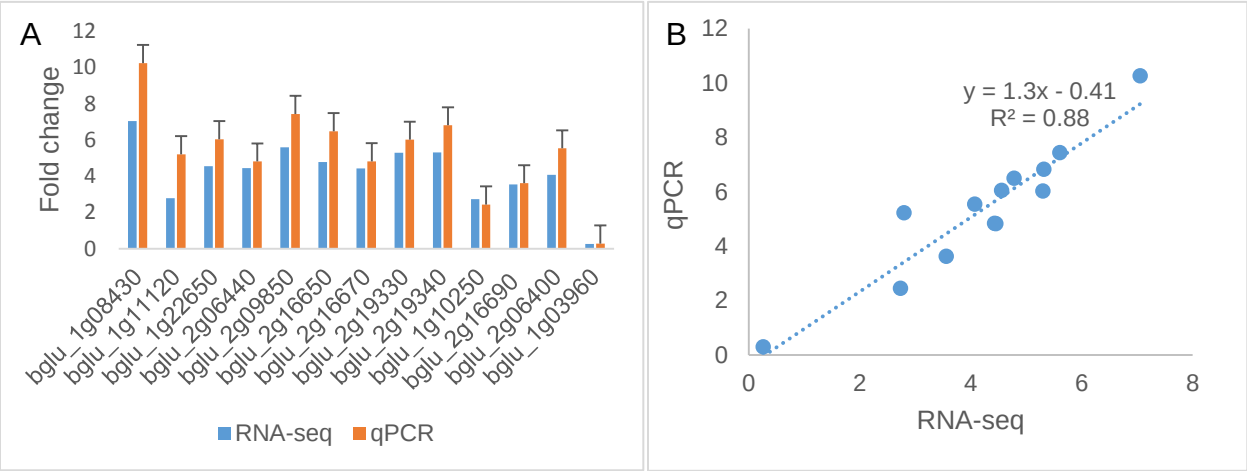


Figure 2.3. Validation of RNA-seq result by qRT-PCR. (A) Fold change of each gene in

qRT-PCR was calculated by the $2^{-\Delta\Delta Ct}$ method and compared with that in RNA-seq. Error bars represent standard error of the mean (SEM) of three independent experiments. (B) Linear correlation of fold change between RNA-seq (x-axis) and qRT-PCR (y-axis).

Validation of RNA-seq expression pattern by qRT-PCR. A total of 13 selected genes were used for the validation of RNA-seq expression (Table 2.3). RNA samples used for qPCR validation were independently prepared by the same method as in RNA-seq. The fold changes of the gene expression in qPCR were contrasted with RNA-seq data by scatter-plotting method (Figure 2.3). The result indicated that gene expression pattern in RNA-seq linearly correlated to that in qPCR with a correlation coefficient (R^2) of 0.88 (Figure 2.3).

Annotation of DEGs with unknown functions. Among the 36 DEGs of unknown functions, 25 genes were successfully annotated with profile hidden markov (HMM) searches (Table 2.6), with a cutoff E-value of 1E-3. The amino acid sequence of the eight genes remaining had less than 30 hits by subjecting to the “blastp” search. Thus, these genes were excluded from the downstream HMMER search. The left 3 genes of unknown functions could not be annotated by HMMER suite, since the hits were all still proteins of unknown functions. For the successfully annotated genes, 8 genes (bglu_1g04010, bglu_1g31460, bglu_1g02000, bglu_1g23130, bglu_1g26310, bglu_2g06380, bglu_2g08910, bglu_2g09660) were predicted to encode membrane proteins and 1 (bglu_1g19190) encodes inner-membrane translocator. Three genes (bglu_2g16630, bglu_2g16640, bglu_2g16650) were annotated as RebB protein.

Table 2.6. Functional annotation for DEGs encoding hypothetical proteins using the HMMER suite.

Regulation in LSUPB401	Gene	Gene Product	Putative Function[a]	E-value
Down-regulated	bglu_1g04010	hypothetical protein	membrane protein	3.1E-133
	bglu_1g31460	hypothetical protein	membrane protein	5.8E-43
	bglu_2g22240	hypothetical protein	small protein yjiX	5E-47
	bglu_3p0560	hypothetical protein	N/A	N/A
	bglu_4p0300	hypothetical protein	N/A	N/A
Up-regulated	bglu_1g01690	hypothetical protein	flagellar protein FliT	1.2E-54
	bglu_1g02000	hypothetical protein	membrane protein	1.5E-67
	bglu_1g08680	hypothetical protein	teneurin-3	5.9E-261
	bglu_1g10720	hypothetical protein	lytic transglycosylase	1.1e-09
	bglu_1g18240	hypothetical protein	N/A	N/A
	bglu_1g19040	hypothetical protein	cupin	5E-99
	bglu_1g19070	hypothetical protein	renal dipeptidase	2.7E-22
	bglu_1g19190	hypothetical protein	inner-membrane translocator	4.3E-126
	bglu_1g21320	hypothetical protein	unknown	N/A
	bglu_1g21920	hypothetical protein	N/A	N/A
	bglu_1g22630	hypothetical protein	PF14027 family protein	2e-38
	bglu_1g23120	hypothetical protein	unknown	N/A

(Table 2.6 continued)

Regulation in LSUPB401	Gene	Gene Product	Putative Function[a]	E-value
Up-regulated	bglu_1g23130	hypothetical protein	membrane protein	3.8E-33
	bglu_1g25590	hypothetical protein	transcriptional regulator	0.00065
	bglu_1g26310	hypothetical protein	outer membrane protein	3.5E-81
	bglu_1g27760	hypothetical protein	DNA repair ATPase	2.9E-105
	bglu_1g30010	hypothetical protein	ester cyclase	6.6E-70
	bglu_1g31060	hypothetical protein	endonuclease	1.1E-110
	bglu_2g05270	hypothetical protein	N/A	N/A
	bglu_2g06380	hypothetical protein	membrane protein	9.1E-76
	bglu_2g06900	hypothetical protein	N/A	N/A
	bglu_2g08910	hypothetical protein	membrane protein	0.00000001
	bglu_2g09660	hypothetical protein	membrane protein	0.0000014
	bglu_2g09850	hypothetical protein	unknown	N/A
	bglu_2g09930	hypothetical protein	N/A	N/A
	bglu_2g16620	hypothetical protein	N/A	N/A
	bglu_2g16630	hypothetical protein	rebB protein	0.000042
	bglu_2g16640	hypothetical protein	rebB protein	4.6E-72
	bglu_2g16650	hypothetical protein	rebB protein	1.9E-35
	bglu_2g16720	hypothetical protein	peptidoglycan-binding protein	3.3E-148

(Table 2.6 continued)

Regulation in LSUPB401	Gene	Gene Product	Putative Function[a]	E-value
	bglu_2g18820	hypothetical protein	deacylase	2.4E-128

[a]"N/A" indicated the genes precluded from the HMMER search, the amino acid sequence of which had less than 30 hits by subjecting to the "blastp" search. Thus, these genes were excluded from the downstream HMMER search. Genes with putative function of "unknown" indicated the genes still annotated as proteins of unknown functions by HMMER suite.

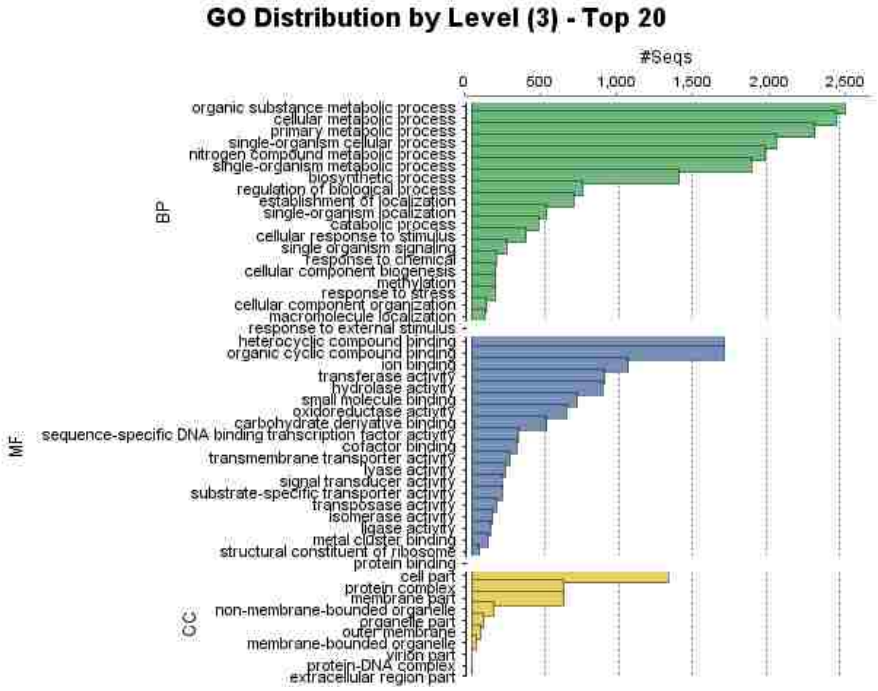


Figure 2.4. Distribution of the top 20 GO terms of whole transcriptome of *B. glumae* at GO hierarchy level 3. "BP" indicates biological process; "MF" molecular function; "CC" cellular component. Number of sequences (genes) is displayed in x-axis. GO terms at level 3 is displayed in y-axis.

Whole transcriptome GO annotation and enrichment. A total of 6305 genes in *B. glumae* were annotated by Blast2GO 3.0, of which 6301 genes could be annotated by InterProScan, 6139 genes had Blast hits, and finally 4574 genes could be mapped to GO terms (Appendix C). Around 900 genes were annotated with 2 or 3 GO terms, and

roughly 600 genes were annotated with 4 or 5 GO terms (Appendix C). The GO terms distribution of the mapped genes by level 3 showed that, for “biological process”, genes associated with the GO term “organic substance metabolic process” and “cellular metabolic process” were the most; for “molecular function”, genes associated with the GO term “heterocyclic compound binding” and “organic cyclic compound binding” were the most abundant; in terms of “cellular component”, genes associated with the GO term “cell part” were the most (Figure 2.4).

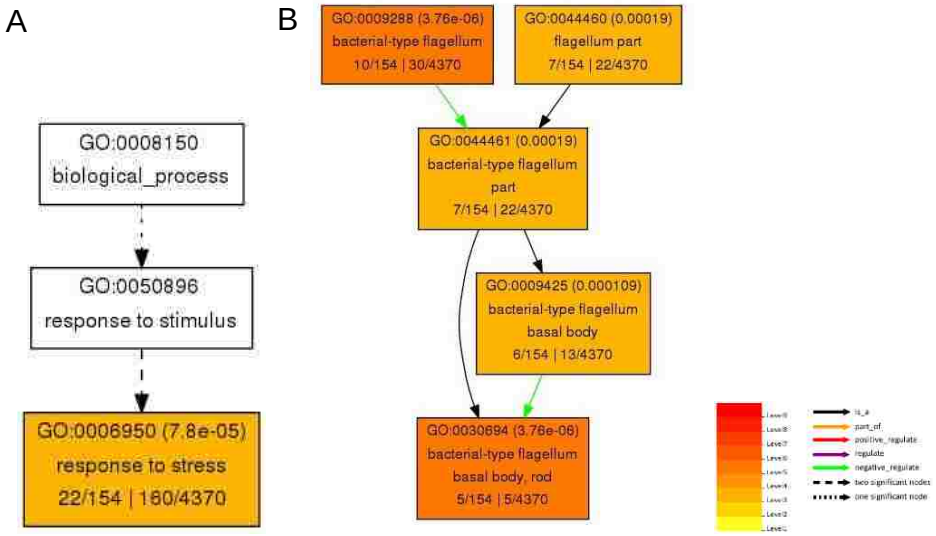


Figure 2.5. Hierarchical over-represented GO terms in category of biological process (A) and cellular component (B). Boxes in the graph contains the GO terms and associated identification number and statistical information. The significantly enriched GO terms (p-value ≤ 0.05) are applied by color, whereas other terms are shown in blank color. The diagram indicates the correlation between color and the enrichment level of the terms. Solid, dashed, and dotted lines represents two, one and zero enriched terms at both ends connected by the line, respectively. The direction of the hierarchical rank goes from top to bottom.

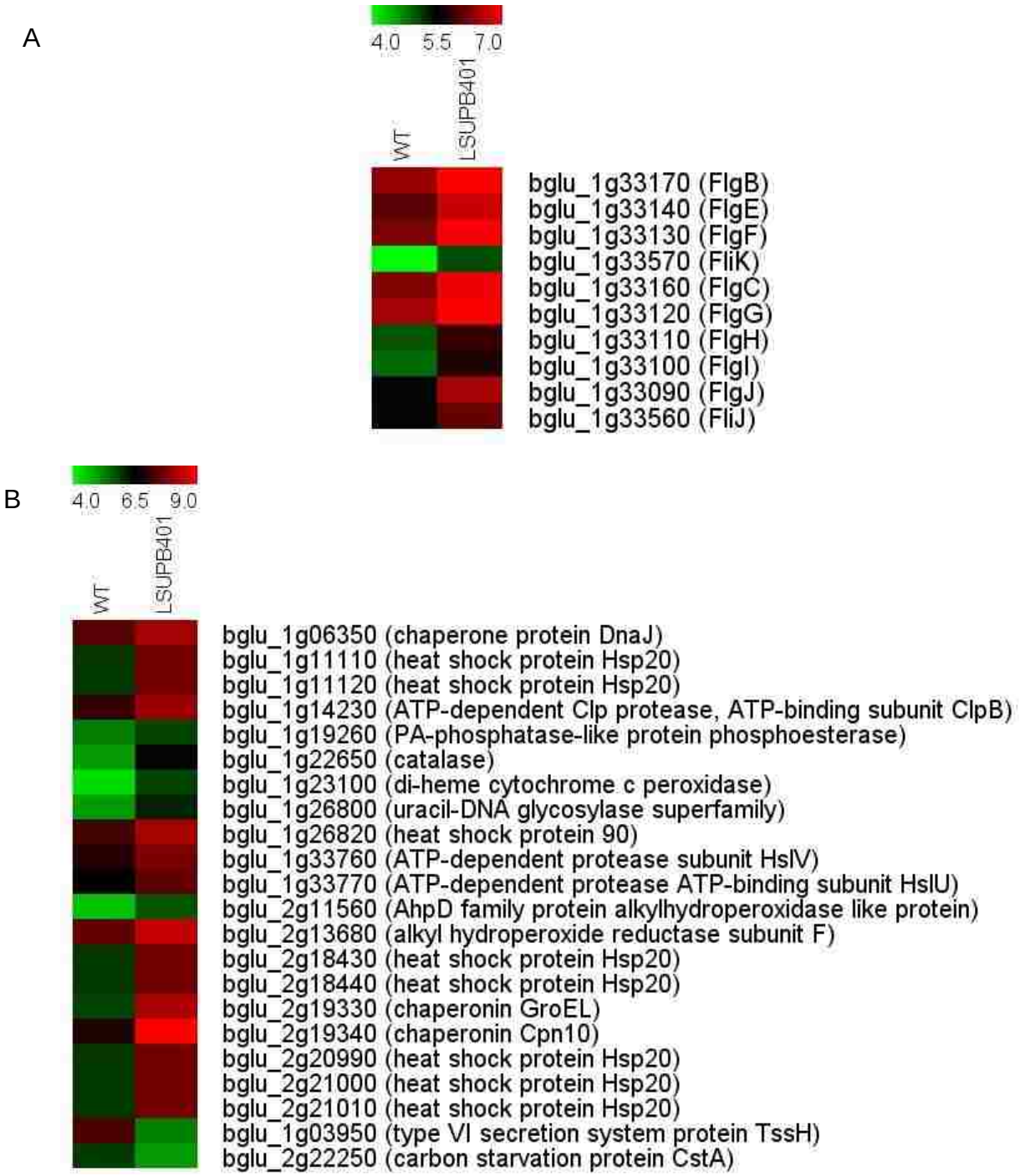


Figure 2.6. The heat map showing the DEGs enriched in the GO terms “bacterial-type flagellum” (A) and “stress to response” (B). Colors in cells represent $\log_2$ RPKM of the genes in WT and LSUPB401.

All the GO terms of the 238 DEGs were retrieved and subjected to GO enrichment analysis by AgriGO (http://bioinfo.cau.edu.cn/agriGO/) to find the over-

represented GO terms. In this study, 2 GO terms were significantly enriched. In the category of biological process, term “stress to response” was significantly enriched with a p-value of 7.8e-05 (Figure 2.5). Among the 22 genes associated with stress response, 20 of them were up-regulated in LSUPB401 (Figure 2.6). Interestingly, 14 of them, consistently highly expressed in LSUPB401, were functionally annotated as heat-shock protein or chaperone (Figure 2.6), indicating that *tepR* may negatively regulate heat-shock proteins globally in *B. glumae*. In the category of cellular component, all terms enriched were associated with bacterium-type flagellum with a p-value of 3.76e-06 (Figure 2.5). A total of 10 genes involved were consistently up-regulated in LSUPB401 (Figure 2.6), indicating that *tepR* may regulate the structure or functions of flagella.

KEGG pathway enrichment. All the DEGs were subjected to KEGG pathway analysis using KOBAS 2.0. A total of 5 terms were enriched with a cutoff p-value of 0.1. Other than flagella assembly, which was also over-represented in GO analysis, KEGG pathway enrichment demonstrated that thiamine metabolism, cysteine and methionine metabolism, ABC transporters system and bacterial secretion system were significantly enriched (Table 2.7). All the DEGs enriched in thiamine metabolism, cysteine and methionine metabolism were mapped to their corresponding KEGG pathway (Figure 2.7). A total of 8 DEGs were enriched in the ABC transporters system, among which 5 genes belonged to a specific branched-chain amino acid ABC transporter system. It is composed of one substrate-binding protein (bglu_1g16840), two permease proteins (bglu_1g26830 and bglu_1g26840) and two ATP-binding proteins (bglu_1g26810 and

Table 2.7. KEGG pathway enrichment assay.

Pathway ID	Term	Expression	Count	Background number	p-value
bgl02040	Flagella assembly	Up	13	42	2.80E-10
bgl00730	Thiamine metabolism	Up	4	14	0.00076
bgl00270	Cysteine and methionine metabolism	Up	4	35	0.014
bgl02010	ABC transporters system	Down	8	172	0.070
bgl03070	Bacterial secretion system	Down	6	65	0.0070

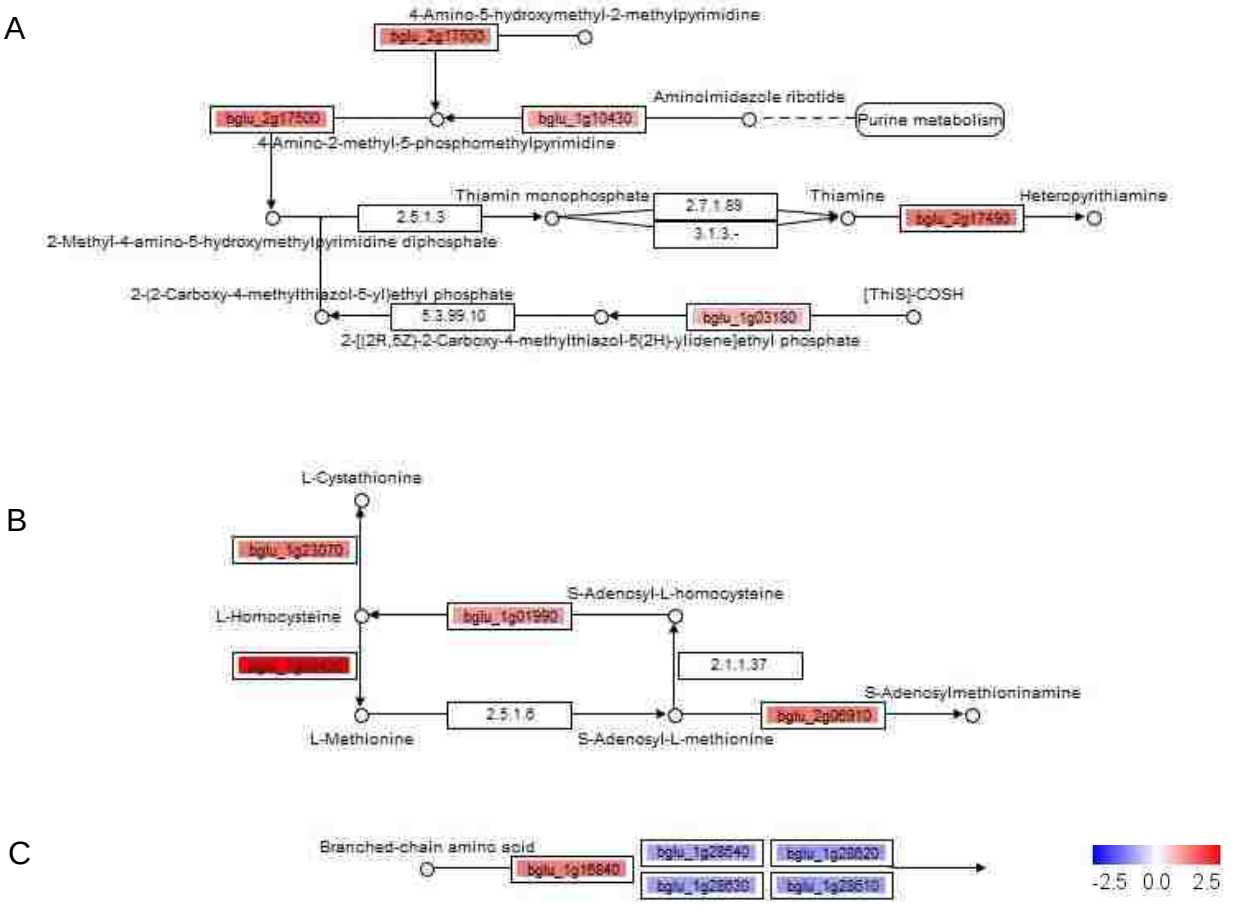


Figure 2.7. Representative enriched KEGG pathway or module in *B. glumae*, (A) Thiamine metabolism, (B) Cysteine and methionine metabolism, and (C) ABC transporter system. All the networks were reconstructed using the VANTED program. Pathway source data was downloaded from the KEGG database (http://www.kegg.jp/). Circle node represents chemical product; rectangle node represents gene locus tag or KEGG enzyme code; round rectangle node represents cognate pathway. Color in rectangles node indicates log2 (WT RPKM/LSUPB401 RPKM). Rectangle node in white color shows the enzyme without any DEGs associated.

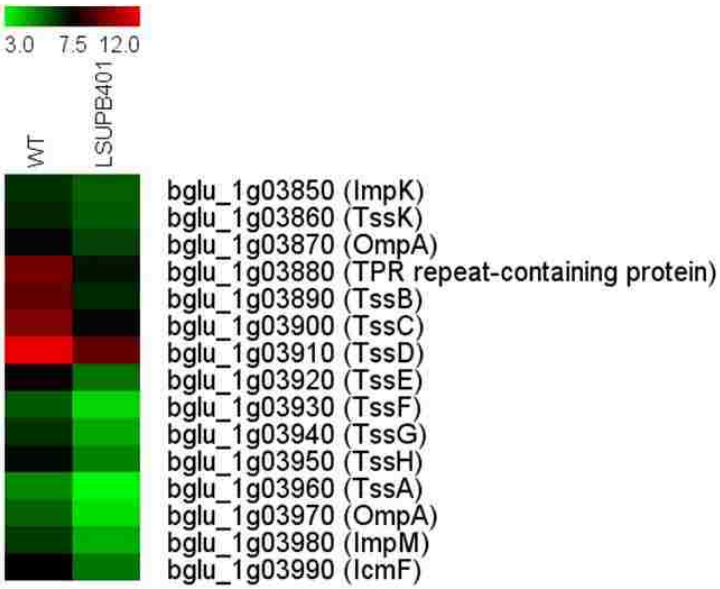


Figure 2.8. The heat map showing the DEGs enriched in the T6SS gene cluster. Colors in cells represent $\log_2$ RPKM of the genes in WT and LSUPB401.

bglu_1g26820). Substrate-binding protein gene was significantly up-regulated in LSUPB401, whereas permease proteins and ATP-binding proteins were significantly down-regulated. The 6 genes enriched bacterial secretion system were constantly down-regulated and belonged to a putative type VI secretion system (T6SS), named as T6SS-A. Further analyses confirmed that this T6SS gene cluster was shut down in LSUPB401. The heat map showing the DEGs within the cluster is displayed in Figure 2.8.

Genetic classification of *T6SS-A* in *B. glumae*. Since T6SS in *B. glumae* had never been studied, the nucleotide sequence of *T6SS-A* was compared with the T6SS in *B. pseudomallei*, which causes melioidosis in humans and animals (Wiersinga et al., 2012). In this study, Mauve alignment was used for the comparative study of the distinct T6SS gene clusters. It is able to display the gene structure, direction, and homology of the aligned sequence simultaneously. The white boxes in the upper part of each gene

cluster were genes in the forward direction and vice versa. The colored boxes were the aligned regions, connected by lines across different clusters. A similarity profile was shown in the colored box where the height corresponded to the average level of conservation of the sequence (Darling et al., 2010). Based on the nomenclature proposed by Schell et al., *T6SS-A* in *B. glumae* was grouped to T6SS-6 due to their high similarity in terms of gene composition, gene direction and sequence homology (Figure 2.9), which was also present in *B. pseudomallei*, *B. mallei* and *B. thailandensis* (Schell et al., 2007).

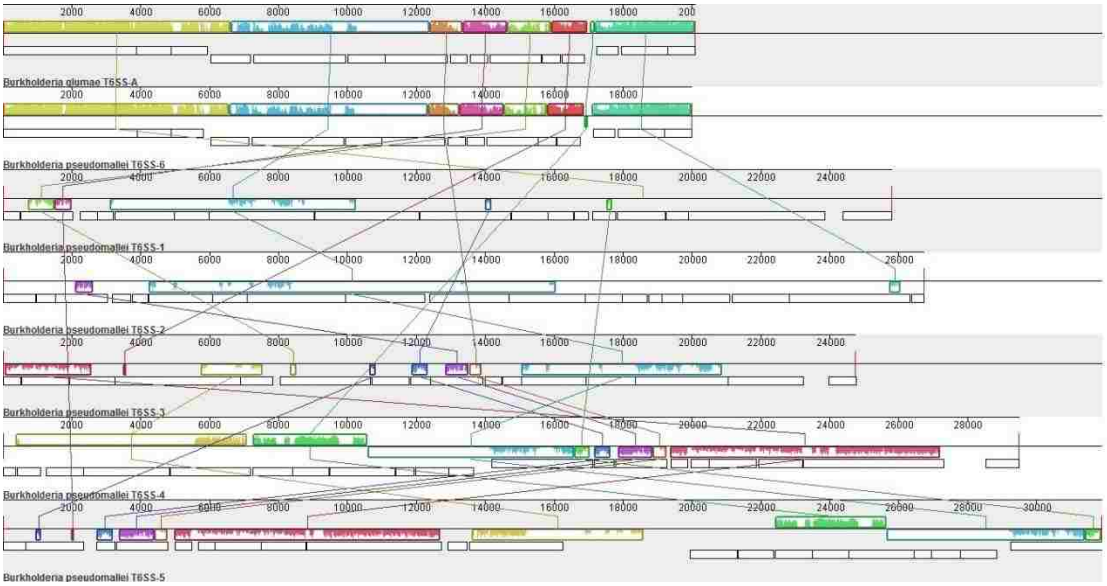
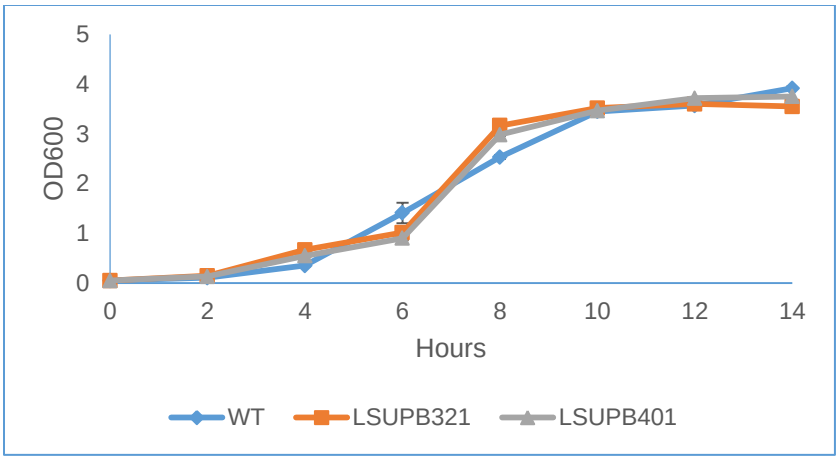


Figure 2.9. Multiple alignment of *T6SS-A* in *B. glumae* and T6SSs in *B. pseudomallei*. The nucleotide sequences of *T6SS-A* (bglu_1g03850 - 03990) in *B. glumae*, and T6SS-1 (BPSS1496-1512), T6SS-2 (BPSS0515-0533), T6SS-3 (BPSS2107-2091), T6SS-4 (BPSS0166-0185), T6SS-5 (BPSS0095-0116), T6SS-6 (BPSL3097-3111) in *B. pseudomallei* were compared using the MAUVE aligner version 2.3.1 (Darling et al. 2010). The figure was created by MAUVE build-in viewer. Locally Collinear Blocks (LCBs) are displayed in different colors. The degree of similarity is indicated by height of colored area, where higher height indicates higher similarity. Areas that are completely white within an LCB are not aligned and may contain sequence elements that are present in a particular genome.

A



B

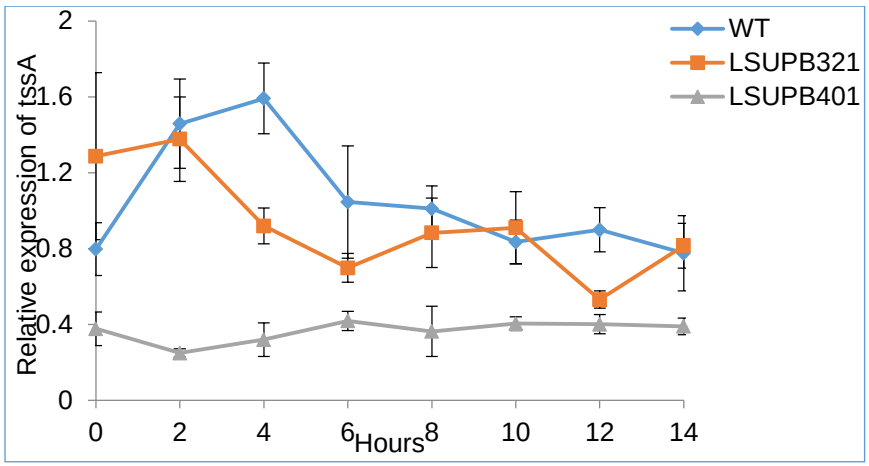


Figure 2.11. Relative expression of *tssA* in different cell density in *B. glumae*. (A) Growth curve of WT (diamond shape), LSUPB321 (square shape) and LSUPB401 (triangle shape). (B) Relative expression of *tssA* in WT (rhombus shape), LSUPB321 (square shape) and LSUPB401 (triangle shape). Error bars indicated ±SD of three replicates.

Knockout of *tepR* abolished the bacterial growth stage-dependent expression of *tssA*. From this study, it was found that *T6SS-A* gene cluster was shut down in LSUPB401 under OD_{600}≈1.0 (Figure 2.8), indicating the positive regulation of *T6SS-A* by *tepR*. However, whether the positive regulation of *T6SS-A* by *tepR* is dependent or independent of bacterial growth stage was unknown. Because *T6SS-A* is in a polycistronic operon structure and all the genes were less expressed in LSUPB401 (Figure 2.8), expression level of *tssA* was used as indicator of the expression level of

T6SS-A. The expression values of *tssA* in WT, LSUPB321 (*Δtofl*), and LSUPB401 were quantified from early exponential stage (OD_{600}=0.05) to late stationary phase. LSUPB321 (*Δtofl*) was included to see whether the expression of *tssA* was QS-dependent or not. The result showed that WT, LSUPB321, and LSUPB401 had similar growth patterns (Figure 2.11. A). The expression of *tssA* in WT increased dramatically during the Lag growth phase and early Log growth phase, and dropped to the starting expression level in the late stationary growth phase (Figure 2.11. B). The expression curve of *tssA* in LSUPB321 was similar to WT except that it was shifted towards the left (Figure 2.11. B). The expression of *tssA* in LSUPB401 was constantly shut down in all time points measured.

Discussion

In this study, the whole transcriptome profile was compared between WT and LSUPB401. A total of 238 DEGs were identified with 199 DEGs being up-regulated in LSUPB401 (Figure 2.1). Therefore, *tepR* was mainly a negative regulator of gene expression. The most significant negatively regulated gene was bglu_1g08430, encoding a 5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase. It catalyzes the chemical reaction of transferring the substrate L-homocysteine to the product L-methionine. It plays a key role in cysteine metabolism and methionine metabolism (Guest et al., 1964). Chaperonin Cpn10 (also called GroES) gene, bglu_2g19340, was also significantly negatively regulated by *tepR*. Chaperonins are a group of proteins that could prevent the misfolding of other proteins (Ellis & Hemmingsen, 1989). Many chaperonins belong to the family of heat shock proteins which response to stressful conditions (De Maio, 1999). A serine metalloprotease gene,

bglu_1g16590, was also negatively regulated by *tepR*. This is noteworthy as our previous study indicated LSUPB401 had significantly more extracellular protease activity than WT (Figure 1.5). Whether bglu_1g16590 is responsible for extracellular protease production is under investigation. Extracellular protease is known to be an important virulence factor in numerous bacteria (Lantz, 1997, Butler et al., 2006, Lebrun et al., 2009), but whether extracellular protease is important for pathogenesis in *B. glumae* has never been studied. Extracellular protease has been known to be mainly secreted by a type I secretion system (T1SS) or a type II secretion system (T2SS) (Sandkvist, 2001, Delepelaire, 2004). Based on the current knowledge, only proteins secreted through T2SS require signal peptides. To predict whether the metalloprotease, encoded by bglu_1g16590, is secreted through T2SS or T1SS, the peptide sequence was searched using the online server of SignalP-4.1 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al., 2011). The result indicated there was a putative signal peptide predicted to be released on the 27th amino acid between the amino acid AQA and QT. So, it is very likely that the metalloprotease is secreted through T2SS. One TenA family transcriptional activator gene, bglu_2g06890, was negatively regulated by *tepR*. TenA was at first identified as a transcriptional enhancer to stimulate the production of extracellular protease in *Bacillus subtilis* (Pang et al., 1991). Later, TenA homolog was found in a wide range of bacteria and archaea (Itou et al., 2004), but its function in *Burkholderia* has not been studied. Based on this study, it is possible that in *B. glumae*, *tepR* could negatively regulate the production of extracellular protease indirectly through the transcriptional activation of the metalloprotease gene by TenA. *katE* was significantly less expressed in *tepR* knockout

mutant. In *Escherichia coli,* two catalase enzymes, hydroperoxidase I (HPI) and HPII, are encoded by *katG* and *katE* respectively (Loewen et al., 1985). In *B. glumae*, the catalase activity of KatG has been studied which is regulated by quorum sensing (Chun et al., 2009). However, the function and regulation mechanism of *katE* has never been studied. It is possible the catalase activity of KatE is negatively regulated by *tepR*. As discussed in Chapter 1 (Figure 1.5), LSUPB401 produced significantly more toxoflavin than WT in both liquid and solid medium. This was validated in this transcriptomic study. The gene clusters *toxABCDE* (bglu_2g06400-6440), encoding toxoflavin biosynthesis system and *toxIHGF* (bglu_2g06350-6380), encoding toxoflavin transport system, were constantly up-regulated in LSUPB401 (Table 2.5). Nevertheless, *toxJ* (bglu_2g06330) and *toxR* (bglu_2g06390), the toxoflavin transcriptional activators, were not significantly differentially expressed (Table 2.5). It is known that the production of toxoflavin in *B. glumae* is tightly regulated by quorum sensing mediated by *tofI*, *tofR* and *tofM* (Chen et al., 2012, Kim et al., 2004). However, in this study, the expression of *tofI*, *tofR* and *tofM* is not differentially regulated in LSUPB401 (Table 2.5). Based on the previous scenario, TofI/TofR-mediated quorum sensing positively regulates transcriptional regulators, *toxJ* and *toxR*, which further activates the expression of *tox* gene clusters for toxoflavin biosynthesis and transport. Apparently, this was not the case in this study, where the known regulatory components, *toxJ* and *toxR*, were not differentially expressed (Table 2.5). Thus, the repression of toxoflavin through *tepR* is seemingly independent from the known regulatory system. There might be an unknown regulatory system which modulates the expression of toxoflavin in *B. glumae*. The quorum sensing master regulator gene, *qsmR*, was also negatively regulated by *tepR* in this study. Based on

the result of Kim *et al.*, *qsmR* is activated by TofR which, in turn, activates the flagellum biosynthesis genes and universal stress protein genes (Kim et al., 2007, Kim et al., 2012). Also, a *qsmR*-defective mutant showed significantly lower expression of toxoflavin and the quorum sensing signal, C8-HSL (N-octanoyl-L-Homoserine lactone) (Kim et al., 2007). So, *tepR* may act as a repressor of *qsmR* and thus modulate the downstream phenotypes.

Thirty-six DEGs were previously annotated as hypothetical proteins. By using HMMER suite, 25 hypothetical protein encoding DEGs were newly annotated. Eight of them were annotated as membrane protein. Three of them were annotated as RebB protein. To date, the functions of RebB protein are poorly understood, but they appeared to be necessary for the production of R-bodies. R-bodies are essential for the "killing trait" for endosymbiont bacteria to attack their host Paramecium (Pond et al., 1989; Raymann et al., 2013). However, the mechanism for the "killing trait" is still unknown. In *B. glumae*, the function of RebB proteins and R-bodies have never been characterized.

By GO enrichment, 2 GO terms, that were significantly associated with *tepR*, were uncovered. Term "stress to response" and "bacterium-type flagellum" was significantly enriched with a p-value of 7.8e-05 and 3.76e-06 respectively (Figure 2.5). Among the 22 genes enriched in "stress to response", 14 genes, higher expressed in LSUPB401, were functionally annotated as chaperone or heat-shock protein, a subgroup of chaperones (Figure 2.6). *tepR* may negatively regulate heat-shock proteins or chaperone proteins globally in *B. glumae*, but the mechanism of the regulation is unknown. Noteworthy, in the kingdom of bacteria, several sigma factors, RpoB, RpoE,

RpoH, and RpoS, have been found to be important for bacteria to respond to heat stress (Kazmierczak et al., 2005, Vanaporn et al., 2008, Raman et al., 2001). As previously mentioned, *tepR* encodes a Sigma-54 dependent response regulator. It is possible that RpoN (Sigma-54) is an important regulator of *tepR* which further modulates the bacterial stress response in *B. glumae*. In *E.coli*, the phosphorylated NtrC, a homolog of LuxO, can interact with RpoN, which further activates the transcription in response to the nitrogen limitation (Austin & Dixon, 1992). But whether RpoN is involved in modulating bacterial heat stress response has never been reported. RpoN is also known to be involved in the flagella-dependent motility of various bacteria (Kazmierczak et al., 2005). In this study, as revealed by GO analysis, a total of 10 genes involved in flagella assembly were consistently negatively regulated by *tepR* (Figure 2.6). So, it is possible *tepR* regulated the flagella genes through RpoN. The expression of RpoN, encoded by bglu_1g31210, was not differentially expressed between WT and *tepR* deletion mutant (q-value=1, data not shown). So, *tepR* might not modulate the expression of RpoN in the transcriptional level.

From KEGG pathway analyses, a total of 5 functional terms were enriched: flagella assembly, thiamine metabolism, cysteine and methionine metabolism, ABC transporters system, and bacterial secretion system. Consistent with the GO enrichment, flagella assembly was enriched in KEGG pathway analyses. KEGG pathway analyses also revealed the negative regulatory role of *tepR* upon thiamine, cysteine and methionine metabolism. They all are involved in sulfur compound. It is possible *tepR* is also involved in modulating the availability of sulfur compounds. It was also indicated that a branched-chain amino acid ABC transporter system was under the regulation of

tepR, which is composed of one substrate-binding protein (bglu_1g16840), two permease proteins (bglu_1g26830 and bglu_1g26840) and two ATP-binding proteins (bglu_1g26810 and bglu_1g26820). The substrate-binding proteins, present in periplasm for gram-negative bacteria, can bind to the targeting substrate specifically (Higgins, 2001). Commonly, permease proteins are structured as six putative α-helical transmembrane parts and perform as a channel for substrates to be transported through (Higgins, 2001). In this study, the substrate-binding protein gene was significantly up-regulated in LSUPB401 whereas permease proteins and ATP-binding proteins were significantly down-regulated. This suggests that greater amount of branched-chain amino acid, i.e. valine, leucine, isoleucine, threonine and histidine, can bind to the transporter system, but they are less capable of passing through the channel. Overall, this transporter system may not be able to function properly in LSUPB401. ABC transporters are also involved several cellular processes, e.g. resistance to antibiotics, acquisition of nutrients, sensing environmental signals, growth under stressful conditions and bacterial virulence (Basavanna et al., 2009). KEGG pathway analyses also uncovered an uncharacterized putative type VI secretion system (T6SS), named as *T6SS-A*, that was significantly positively regulated by *tepR* (Figure 2.8). However, the function of T6SS in *B. glumae* has never been studied. The T6SSs in *B. pseudomallei* were chosen for reference to predict the function of *T6SS-A* in *B. glumae*. This was not only because *B. pseudomallei* has a close phylogenetic relationship to *B. glumae*, but also because it has six distinct T6SS gene clusters, the most among *Burkholderia* spp., named as T6SS-1 to T6SS-6. In *B. mallei* and *B. thailandensis*, they lack T6SS-5 and T6SS-4, respectively (Schell et al., 2007). Due to the large genetic variance of different

T6SS clusters in terms of gene composition, direction, and sequence, phylogenetic nomenclature based on amino acid sequence of one gene or a few concatenated genes may not reflect their relationship very well. Based on the nomenclature proposed by Schell *et al.*, *T6SS-A* in *B. glumae* is classified to the family of T6SS-6 because of their high similarity in terms of gene composition, gene direction and gene homology (Figure 2.9), which was also present in *B. pseudomallei*, *B. mallei* and *B. thailandensis* (Schell et al., 2007). The guide tree achieved from Mauve alignment was consistent with the conclusion (Figure 2.10). T6SS-6 in *B. pseudomallei* was renamed as *tss-1* in another publication (Shalom et al., 2007) and its homolog in *B. thailandensis* was named as T6SS-1 (Schwarz et al., 2010).

In this study, it was found that *T6SS-A* gene cluster was positively regulated by *tepR* under OD_{600}≈1.0. However, whether *tepR* modulated the expression of *T6SS-A* in a population density-dependent or –independent manner was unknown. Because the genes encoding *T6SS-A* are in an operon structure, they are likely to be expressed and regulated synchronously as validated in this study (Figure 2.8). Here *tssA* was used as a representative gene to indicate the expression pattern of *T6SS-A*. Expression of *tssA* in WT increased dramatically during the Lag phase and early Log phase of bacteria growth, and decreased to the initial expression level in the late stationary phase (Figure 2.11.B). Thus, the expression of *tssA* was likely density-dependent and showing a pattern of quorum sensing. The density-dependent expression pattern was not completely abolished in the quorum sensing-defective mutant, LSUPB321 (*ΔtofI*), except the curve was shifted towards the left (Figure 2.11. B). Thus, there seemed to be quorum sensing system(s) in addition to the TofI/TofR-mediated quorum sensing. In

LSUPB401, the expression of *tssA* was constantly shut down independent of the bacterial population density. Apparently, *tepR* is a dominating positive regulator of *tssA*, where the density-dependent expression pattern was completely abolished in LSUPB401. Based on these results, it is very likely that *tepR* is involved in a previously uncharacterized quorum sensing system which functions together with TofI/TofR-mediated quorum sensing system.

References

Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH, 2011. Accurate and comprehensive sequencing of personal genomes. *Genome Res* **21**, 1498-505.

Ashburner M, Ball CA, Blake JA*, et al.*, 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9.

Austin S, Dixon R, 1992. The prokaryotic enhancer binding protein NTRC has an ATPase activity which is phosphorylation and DNA dependent. *Embo j* **11**, 2219-28.

Basavanna S, Khandavilli S, Yuste J*, et al.*, 2009. Screening of Streptococcus pneumoniae ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infect Immun* **77**, 3412-23.

Birol I, Jackman SD, Nielsen CB*, et al.*, 2009. De novo transcriptome assembly with ABySS. *Bioinformatics* **25**, 2872-7.

Bullard JH, Purdom E, Hansen KD, Dudoit S, 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *Bmc Bioinformatics* **11**, 13.

Burrows,M. and Wheeler,D.J. (1994) A Block Sorting Lossless Data Compression Algorithm, Technical Report 124. Digital Equipment Corporation, Palo Alto, CA.

Butler SM, Festa RA, Pearce MJ, Darwin KH, 2006. Self-compartmentalized bacterial proteases and pathogenesis. *Mol Microbiol* **60**, 553-62.

Chen RX, Barphagha IK, Karki HS, Ham JH, 2012. Dissection of Quorum-Sensing Genes in Burkholderia glumae Reveals Non-Canonical Regulation and the New Regulatory Gene tofM for Toxoflavin Production. *Plos One* **7**, 15.

Chun H, Choi O, Goo E*, et al.*, 2009. The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *J Bacteriol* **191**, 4152-7.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M, 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-6.

Creecy JP, Conway T, 2015. Quantitative bacterial transcriptomics with RNA-seq. *Curr Opin Microbiol* **23**, 133-40.

Croucher NJ, Thomson NR, 2010. Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* **13**, 619-24.

Darling AE, Mau B, Perna NT, 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *Plos One* **5**, e11147.

De Maio A, 1999. Heat shock proteins: facts, thoughts, and dreams. *Shock* **11**, 1-12.

Dehal PS, Joachimiak MP, Price MN*, et al.*, 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* **38**, D396-400.

Delepelaire P, 2004. Type I secretion in gram-negative bacteria. *Biochim Biophys Acta* **1694**, 149-61.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z, 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* **38**, W64-70.

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7.

Ellis RJ, Hemmingsen SM, 1989. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem Sci* **14**, 339-42.

Frias-Lopez J, Shi Y, Tyson GW*, et al.*, 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**, 3805-10.

Grosu P, Townsend JP, Hartl DL, Cavalieri D, 2002. Pathway Processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res* **12**, 1121-6.

Guest JR, Friedman S, Foster MA, Tejerina G, Woods DD, 1964. Transfer of the methyl group from N5-methyltetrahydrofolates to homocysteine in Escherichia coli. *Biochem J* **92**, 497-504.

Harris MA, Clark J, Ireland A*, et al.*, 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-61.

Higgins CF, 2001. ABC transporters: physiology, structure and mechanism--an overview. *Res Microbiol* **152**, 205-10.

Itou H, Yao M, Watanabe N, Tanaka I, 2004. Structure analysis of PH1161 protein, a transcriptional activator TenA homologue from the hyperthermophilic archaeon Pyrococcus horikoshii. *Acta Crystallogr D Biol Crystallogr* **60**, 1094-100.

Junker BH, Klukas C, Schreiber F, 2006. VANTED: a system for advanced data analysis and visualization in the context of biological networks. *Bmc Bioinformatics* **7**, 109.

Kanehisa M, Goto S, 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30.

Kazmierczak MJ, Wiedmann M, Boor KJ, 2005. Alternative sigma factors and their roles in bacterial virulence. *Microbiol Mol Biol Rev* **69**, 527-43.

Kim H, Goo E, Kang Y, Kim J, Hwang I, 2012. Regulation of universal stress protein genes by quorum sensing and RpoS in Burkholderia glumae. *J Bacteriol* **194**, 982-92.

Kim J, Kang Y, Choi O*, et al.*, 2007. Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in Burkholderia glumae. *Mol Microbiol* **64**, 165-79.

Kim J, Kim JG, Kang Y*, et al.*, 2004. Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in Burkholderia glumae. *Mol Microbiol* **54**, 921-34.

Kim S, Park J, Lee J*, et al.*, 2014. Understanding pathogenic Burkholderia glumae metabolic and signaling pathways within rice tissues through in vivo transcriptome analyses. *Gene* **547**, 77-85.

Kroger C, Colgan A, Srikumar S*, et al.*, 2013. An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* **14**, 683-95.

Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-U54.

Lantz MS, 1997. Are bacterial proteases important virulence factors? *J Periodontal Res* **32**, 126-32.

Lebrun I, Marques-Porto R, Pereira AS, Pereira A, Perpetuo EA, 2009. Bacterial toxins: an overview on bacterial proteases and their action as virulence factors. *Mini Rev Med Chem* **9**, 820-8.

Livak KJ, Schmittgen TD, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-8.

Loewen PC, Switala J, Triggs-Raine BL, 1985. Catalases HPI and HPII in Escherichia coli are induced independently. *Arch Biochem Biophys* **243**, 144-9.

Mcclure R, Balasubramanian D, Sun Y*, et al.*, 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res* **41**, e140.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M, 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**, e121.

Mikkelsen TS, Ku M, Jaffe DB*, et al.*, 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60.

Nandakumar R, Shahjahan AKM, Yuan XL*, et al.*, 2009. Burkholderia glumae and B. gladioli Cause Bacterial Panicle Blight in Rice in the Southern United States. *Plant Disease* **93**, 896-905.

Pang AS, Nathoo S, Wong SL, 1991. Cloning and characterization of a pair of novel genes that regulate production of extracellular enzymes in Bacillus subtilis. *J Bacteriol* **173**, 46-54.

Pepke S, Wold B, Mortazavi A, 2009. Computation for ChIP-seq and RNA-seq studies. *Nature Methods* **6**, S22-S32.

Petersen TN, Brunak S, Von Heijne G, Nielsen H, 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-6.

Pond FR, Gibson I, Lalucat J, Quackenbush RL, 1989. R-body-producing bacteria. *Microbiol Rev* **53**, 25-67.

Poretsky RS, Bano N, Buchan A*, et al.*, 2005. Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* **71**, 4121-6.

Quail MA, Smith M, Coupland P*, et al.*, 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *Bmc Genomics* **13**, 341.

Raman S, Song T, Puyang X, Bardarov S, Jacobs WR, Jr., Husson RN, 2001. The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in Mycobacterium tuberculosis. *J Bacteriol* **183**, 6119-25.

Raymann K, Bobay LM, Doak TG, Lynch M, Gribaldo S, 2013. A genomic survey of Reb homologs suggests widespread occurrence of R-bodies in proteobacteria. *G3 (Bethesda)* **3**, 505-16.

Rhee SY, Wood V, Dolinski K, Draghici S, 2008. Use and misuse of the gene ontology annotations. *Nat Rev Genet* **9**, 509-15.

Saeed AI, Sharov V, White J*, et al.*, 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-8.

Sandkvist M, 2001. Type II secretion and pathogenesis. *Infect Immun* **69**, 3523-35.

Schatz MC, Delcher AL, Salzberg SL, 2010. Assembly of large genomes using second-generation sequencing. *Genome Res* **20**, 1165-73.

Schell MA, Ulrich RL, Ribot WJ*, et al.*, 2007. Type VI secretion is a major virulence determinant in Burkholderia mallei. *Mol Microbiol* **64**, 1466-85.

Schwarz S, West TE, Boyer F*, et al.*, 2010. Burkholderia type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog* **6**, e1001068.

Shalom G, Shaw JG, Thomas MS, 2007. In vivo expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. *Microbiology* **153**, 2689-99.

Sharma CM, Hoffmann S, Darfeuille F*, et al.*, 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250-5.

Sorek R, Cossart P, 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**, 9-16.

Takeuchi T, Sawada H, Suzuki F and Matsuda I, 1997. Specific detection of *Burkholderia plantarii* and *B. glumae* by PCR using primers selected from the 16S-23S rDNA spacer regions. *Ann Phytopath Soc Japn* **63**, 455-62.

Vanaporn M, Vattanaviboon P, Thongboonkerd V, Korbsrisate S, 2008. The rpoE operon regulates heat stress response in Burkholderia pseudomallei. *FEMS Microbiol Lett* **284**, 191-6.

Wang Z, Gerstein M, Snyder M, 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

Wiersinga WJ, Currie BJ, Peacock SJ, 2012. Melioidosis. *N Engl J Med* **367**, 1035-44.

Wu J, Mao X, Cai T, Luo J, Wei L, 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res* **34**, W720-4.

Young MD, Wakefield MJ, Smyth GK, Oshlack A, 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14.

Zerbino DR, Birney E, 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-9.

CHAPTER 3: VALIDATION OF THE PHENOTYPES REGULATED BY *TEPR* AS REVEALED BY RNA-SEQ

Introduction

Bacterial panicle blight, caused by *Burkholderia glumae*, is one of the major diseases of rice in the southern United States. The pathogenesis of *B. glumae* has been known to involve several virulence factors including toxoflavin, type III secretion system, lipase, flagella and catalase (Ham et al., 2011). The transcriptomic study discussed in Chapter 2 indicated that *tepR* regulated genes for a multitude of bacterial cellular processes. Specifically, *tepR* negatively regulated *tox* gene clusters (Kim et al., 2004b), a serine metalloprotease gene (bglu_1g16590), flagella genes, stress response genes, and thiamine, cysteine and methionine metabolism genes, but positively regulates branched-chain amino acid transporter system genes and type VI secretion system (T6SS) genes. It was not surprising that *tepR* could regulate toxoflavin and metalloprotease genes in a negative regulatory manner, since LSUPB401 (*ΔtepR*) indeed produced significantly more toxoflavin and extracellular protease (Osti, 2014). However, the regulatory roles of *tepR* in other bacterial cellular processes have not been previously studied.

As revealed in Gene Ontology enrichment analyses, genes responsible for "stress to response" were significantly higher expressed in LSUPB401. It is therefore hypothesized that *tepR* may negatively regulate the survival of *B. glumae* under stressed conditions. Besides, KEGG pathway enrichment indicated that *tepR* positively regulated *T6SS-A*, which is homologous to T6SS-6 in *B. pseudomallei* (Schell et al., 2007) and T6SS-1 in *B. thailandensis* (Schwarz et al., 2010). Since, T6SS-1 in *B. thailandensis* is responsible for cell contact-dependent competition (Schwarz et al.,

2010), it is possible that *T6SS-A* in *B. glumae* shares similar positive roles in bacterial interspecies competition. In this study, it was hypothesized that LSUPB401, defective in *T6SS-A*, lost its contact-dependent competition capacity against other bacteria. A general review of the bacterial stress response and T6SS is displayed as below.

Stress response. The survival of bacteria *in vivo* is challenged by various environmental stresses, and bacteria have evolved to have multiple stress response strategies. One of the most common environmental stresses bacteria may encounter is nutrient limitation. The classic example of nutrient limitation is amino acid depletion, which activates the stringent response (Chatterji & Ojha, 2001). The typical character of stringent response is the reduced expression of genes responsible for growth but increased expression of genes for survival through transcriptional switching (Sharma & Chatterji, 2010). Through transcriptional switching, the activation of housekeeping sigma factor 70 is diminished whereas the activation of alternative sigma factors are favored (Sharma & Chatterji, 2010), which further activates the expression of numerous genes required to maintain cell viability in a stressed environment (Fang et al., 1992, Lange & Hengge-Aronis, 1991). Induced by stringent response, the expression of guanosine pentaphosphate (ppGpp) is also increased which plays myriad effects on bacterial cell physiology and antimicrobial susceptibility (Wu et al., 2010, Magnusson et al., 2005). Besides nutrient limitation stress, bacteria may also encounter oxidative stress from reactive oxygen species (ROSs) that are generated during aerobic growth (Storz & Imlay, 1999). ROSs can cause severe damage to DNA, RNA, proteins and lipids (Storz & Imlay, 1999, Cabiscol et al., 2000). To prevent damage from oxidative stresses, bacteria have built up response mechanisms to eliminate ROSs through

enzymes like catalase and superoxide dismutase, and small proteins such as thioredoxin and glutaredoxin (Cabiscol et al., 2000). In *B. glumae*, *katG* is known to encode a catalase. It is quorum sensing-dependent and activated by the quorum sensing master regulator, QsmR (Chun et al., 2009). A *katG*-defective mutant causes less severe disease on rice, presumably because it is more susceptible to visible light (Chun et al., 2009). Bacteria may also be under stress when exposed to unfavorably high temperature which causes the inactivation or denaturation of bacterial proteins. Bacteria have evolved to develop a global regulatory system for heat tolerance, the heat-shock response system. This response is known to be present in both Gram-negative and Gram–positive bacteria (Derre et al., 1999, Craig, 1985). The heat shock response is characterized by the involvement of a large group of proteins called heat shock proteins (HSPs). Many of them are chaperon proteins (e.g. GroEL and DnaK) which play important roles in protein folding, repairing and degradation (Zeilstra-Ryalls et al., 1991, Calloni et al., 2012). The expression of HSPs is controlled by different mechanisms. One common regulatory mechanism of HSPs is achieved through transcriptional activation by alternative sigma factors (Rosen & Ron, 2002). For example, in the Gram-negative bacteria, *Pseudomonas* spp. and *Burkholderia* spp., the synthesis of HSPs is under the regulation of sigma factor 32 (RpoH) and sigma factor 24 (RpoE) (Vanaporn et al., 2008, Suh et al., 1999), whereas in Gram-positive bacteria, such as *Bacillus* spp., it is regulated by sigma factor B (Zuber & Schumann, 1994). HSPs can also be regulated through transcriptional repressors, e.g. HrcA, found in *Bacillus subtilis* (Zuber & Schumann, 1994). HrcA can bind to the conserved inverted repeat (CIRCE) region which is present upstream of the heat-shock protein genes

(Baird et al., 1989). For years, the HrcA-CIRCE regulatory system was thought to be only present in Gram-positive bacteria until such hairpin-loop structure was first found in *Agrobacterium tumefaciens* (Segal & Ron, 1993). In *E. coli*, the inverted repeat sequence (TTAGCACTC-N9-GAGTGCTAA) is well conserved in all the HSP genes tested (Segal & Ron, 1996). In *B. glumae*, 11 universal stress proteins (Usps genes *usp1* to *usp11*) have been identified and are delicate to heat shock stress (Kim et al., 2012). Usps1 and Usps2 are regulated by Sigma factor S and by QsmR-involved quorum sensing (Kim et al., 2012).

Type VI Secretion System (T6SS). In Gram-negative bacteria, protein is secreted by transport from the synthesis site in the cytoplasm, passage through the cell envelope (including the inner membrane, periplasm and outer membrane) to the cell exterior (Holland, 2010). This process is accomplished by at least six distinct proteinaceous machines, named Type I-VI secretion systems (Holland, 2010). One bacterium could utilize some or all of these secretion systems, largely depending on their life style and niche (Coulthurst, 2013). These protein secretion systems play important roles in the bacterial virulence and the host-pathogen interaction (Gerlach & Hensel, 2007). One of the latest discovered protein secretion systems in Gram-negative bacteria is the type VI secretion system (T6SS). T6SS was first identified in *Vibrio cholerae* (Pukatzki et al., 2006) and *Pseudomonas aeruginosa* (Mougous et al., 2006). In *Vibrio cholera*, one highly conserved locus, named *vas* locus, was identified to be required for the secretion of Hcp (haemolysin co-regulated protein) and VgrG (valine glycine repeat protein) (Pukatzki et al., 2006). A T6SS is now recognized as a cluster of 13 essential conserved genes (Figure 3.1, Figure 3.2). The encoded proteins are

named from TssA to TssM, which are predicted to be components of the protein secretion system apparatus (Blondel et al., 2009). At the same time, other components can also be functional in T6SSs, for example the σ54-dependent transcriptional activator VasH (Boyer et al., 2009).

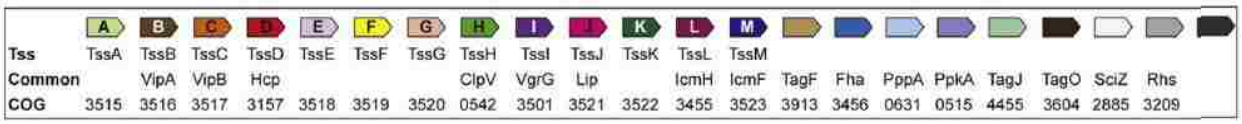


Figure 3.1. The composition of a T6SS. Each protein is designated to Tss name, common name, and Clusters of Orthologous Groups (COG) number (Coulthurst, 2013).

One significant finding is that the structure of T6SS partially resembles an inverted bacteriophage, as shown in Figure 3.3 (Filloux, 2013). The Hcp (TssD) and VgrG (TssI) are the extracellular part of the system that can structure as a needle injection device like the bacteriophage tail. Specifically, Hcp forms a hexametric ring that stacks into a tube-like composition, which resembles a λ gpV tail tube protein (Ballister et al., 2008, Pell et al., 2009). The VgrG protein is directly connected to Hcp and extends as the needle tip of the tube. This resembles the fused gp5-gp27 phage T4 protein (Leiman et al., 2009). Furthermore, TssE shows high sequence homology to the gp25 phage protein which is in a plastic foundation structure (Lossi et al., 2011). Other proteins, though not homologous to bacteriophage, also have important roles in the system. VipA and VipB (TssB and TssC, respectively) can assemble into a large tube with a structure of a sheath-like tail. It is predicted to contract and facilitate the secretion of the protein out of the cell (Basler & Mekalanos, 2012). The installed T6SS is considered to extend through the bacterial cytoplasm, inner membrane, periplasm and outer membrane, and finally secretes effector protein(s) into the cell of the target (Coulthurst, 2013).

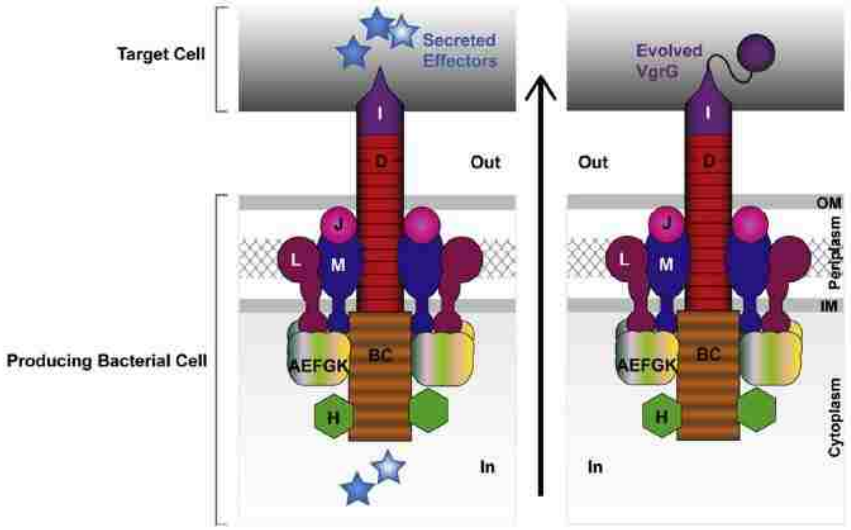


Figure 3.2. The structural machinery of T6SS. “A-M” indicates the 13 essential genes, TssA-TssM, for a functional T6SS. On the left is the delivery of an effector through T6SS; on the right is the interaction of T6SS with target cell through an evolved VgrG (Coulthurst, 2013).

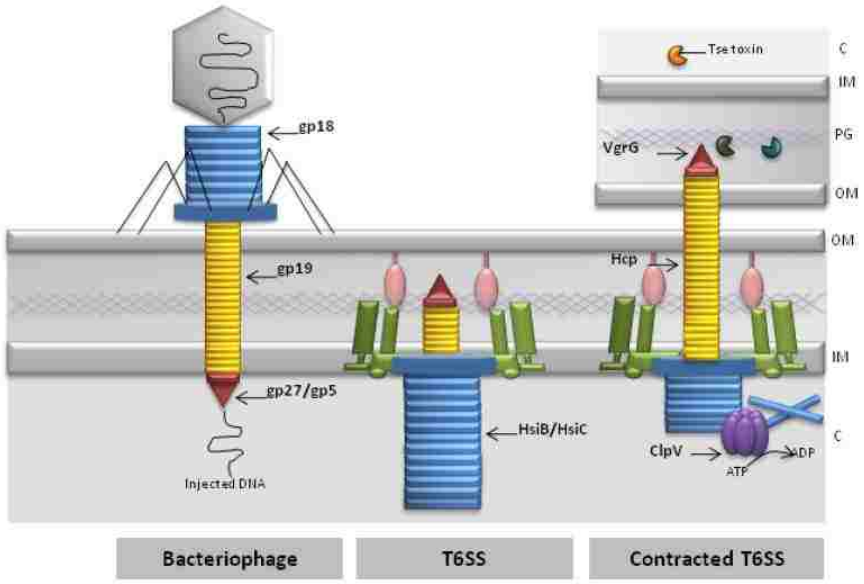


Figure 3.3. Structure comparison between bacteriophage and H1-type VI secretion system (H1-T6SS) in *Pseudomonas aeruginosa*. “C” indicates cytoplasm; “IM” indicates inner membrane; “OM” indicates outer membrane; “PG” indicates peptidoglycan layer. The left illustrates the bacteriophage which is injecting through the bacterial surface; the middle is the T6SS that structurally resembles an inverted bacteriophage; the right shows the contracted sheath pushes the T6SS machinery through the cell envelope and delivers the effector (Tse toxin) to the target bacterial cell (Filloux, 2013).

T6SSs have been found to be involved in host interaction and antibacterial activity (Hood et al., 2010). In *P. aeruginosa*, one T6SS could deliver different toxins, Tse1-3, into the periplasm of other bacteria (Hood et al., 2010, Russell et al., 2011). Until now, several T6SS-related toxin delivery systems have been reported (Benz et al., 2012, Murdoch et al., 2011, Schwarz et al., 2010). T6SSs have been hypothesized to provide an important access to the pathogen to infect the host, and competition advantage over other bacterial organisms within the same environmental niche (Filloux, 2013).

Schwarz *et al.* first described the six distinct T6SSs (named T6SS-1 to T6SS-6) in *Burkholderia thailandensis*, and examined their evolutionary relationship with the well-known T6SS in *V. cholera* and *Aeromonas hydrophila*, based on one of the core conserved gene VipA (TssB) (Schwarz et al., 2010). The phylogeny result shows that only T6SS-5 is clustered as the same subtree with *V. cholera* and *A. hydrophila* systems, whereas all other T6SS systems are clustered in other subtrees (Schwarz et al., 2010). Intriguingly, T6SS-5 is also the only T6SS known to be involved for virulence, whereas the remaining T6SSs are dispensable for virulence, since the disruption of the T6SSs-5 in *B. thailandensis* does not abolish the infection in mice (Schwarz et al., 2010).

As mentioned in Chapter 1, TepR is homologous to the quorum-sensing response regulator LuxO in *Vibrio* spp. LuxO plays key roles in *V. cholera* in regulating the T6SS in a density-dependent manner (Shao & Bassler, 2014). Specifically, at low cell density, LuxO activates the transcription of several sRNAs, *qrr1* to *qrr4*. These sRNAs can pair with the mRNA of the T6SS cluster and directly interfere with their translation (Figure 3.4). Also, sRNAs inhibit the translation of quorum-sensing regulator

HapR, which acts as an activator of T6SS and protease production (Shao & Bassler, 2014, Shao & Bassler, 2012, Rutherford et al., 2011). At high cell density *qrr1* to *4* are not transcribed, and HapR is expressed at a relatively high level (Fig. 6). T6SS, as a result, is no longer repressed (Shao & Bassler, 2014). The regulator TsrA, acting together with LuxO, can repress T6SS (Zheng et al., 2010). Mutation in LuxO or TsrA leads to the expression of haemolysin coregulated protein, Hcp (Zheng et al., 2010), which has commonly been considered as an indicator of the expression of T6SS (Shao & Bassler, 2014). These findings for LuxO provide important reference value for studying the regulatory mechanisms for *T6SS-A* through the LuxO homolog, TepR.

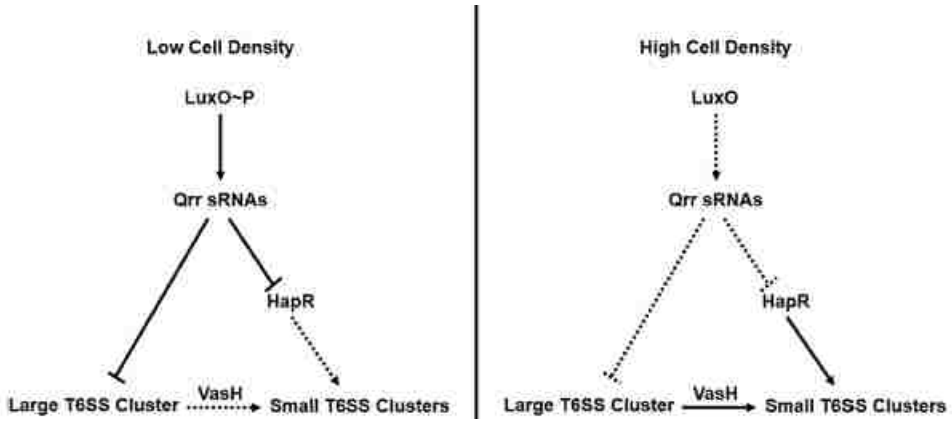


Figure 3.4. Regulation T6SS by LuxO under different cell density in *V. cholera* (Shao and Bassler, 2014)

Materials and Methods

Heat stress tolerance assay. 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB1068 (*ΔtepR*(ptepR)), LSUPB145 (*ΔtofI*) and LSUPB139 (*ΔtofI-tofR*) of *B. glumae* were tested for their tolerance to heat stress following Suh's method (Suh et al., 1999) with some modifications. Bacteria grown in LB broth at 37˚ C with shaking at 200 rpm overnight were harvested and washed twice with phosphate-buffered saline (PBS) buffer. The concentration of the overnight culture (CFU/ml) was determined by counting the number

of colonies in the spots of 10 µl culture from a 10-fold dilution series. Five hundred µl of the overnight culture was transferred to a water bath at 50˚ C. The number of the viable cells was counted using the same spotting method over 10 minute intervals. The survival rate was determined by dividing the CFU/ml after the heat challenge with the CFU/ml before the challenge.

Stationary-phase stress tolerance assay. 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB1068 (*ΔtepR*(pBBtepR)), LSUPB145 (*ΔtofI*) and LSUPB139 (*ΔtofI-tofR*) were tested for their tolerance to stationary-phase stress. Bacteria strains were grown in LB broth at 37 ˚C with shaking at 200 rpm for 8 hours. The cultures were washed twice with fresh LB broth and adjusted to 10^8 CFU/ml (OD_{600}=0.1). The population density of the culture was determined every 6 hours by spotting 10 µl cultures with series of dilution (Goo et al., 2012).

Bacterial interspecies competition assay. Spontaneous nalidixic acid resistant *Pantoea stewartii* DC283 (Coplin et al., 1986), spontaneous rifampicin resistant *pseudomonas syringae* pv. *syringae* B728a (Loper et al., 1987), spontaneous nalidixic acid resistant *Pantoea ananatis* RSPAM1 isolated from rice seed (Jingyu, unpublished) and naturally nitrofurantoin resistant *B. glumae* 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB561 (*ΔtepR*/*toxA*$^-$), and LSUPB139 (*ΔtofI-tofR*) stains were used for bacterial competition assays. The antibiotics used in this assay were kanamycin (Km) at 50 µg/ml, nalidixic acid (Nal) at 20 µg/ml, rifampicin (Rif) at 80 µg/ml, and nitrofurantoin (Nit) at 100 µg/ml. Luria-Bertani (LB) agar was used for growing *B. glumae* strains. Nutrient agar was used for growing DC283, B728a, and RSPAM1. The procedure followed Basler et al.'s method (Basler et al., 2013) with some modifications. Overnight cultures

of bacteria on LB agar or nutrient agar were washed and resuspended in PBS buffer to OD_{600}=0.1 and mixed in 1:1 ratio. 10 µl of the suspension mixture was spotted on nutrient agar and incubated for 24 hours at 30˚ C. The mixture was then resuspended in 1 ml PBS buffer and serially diluted. 10 µl of the mixture was spotted on media with selective antibiotics. After 24 hours incubation at 30˚ C, CFU were counted.

Phylogenetic homologs of *tepR* in other pathogenic bacteria. The amino acid sequence of TepR in *B. glumae* and 34 TepR homologs (defined as greater than 30% amino acid identity to TepR) in bacteria of the same or different genus were chosen for phylogenetic study. The amino acid sequences were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/). The sequences was aligned with MUSCLE v3.6 (Edgar, 2004) and culled with Gblocks version 0.91b (Castresana, 2000) with the setting -b1=19 –b2=29 –b3=8 –b4=10 –b5=a. Phylogenetic tree was constructed by FastTree 2.1.3 using an approximately maximum-likelihood algorism (Price et al., 2010).

Results

The heat stress tolerance of *B. glumae* was negatively and positively regulated by *tepR* and quorum sensing, respectively. The stationary phase cultures of 336gr-1, LSUPB401 and the quorum sensing-defective mutants, LSUPB145 and LSUPB139, were shifted from 37˚C to 50˚C to measure their tolerance to heat stress. The survival rates of all the strains tested were not significantly different (p-value<0.05) after 10 minutes of heat stress treatment (Figure 3.5). However, after 20 minutes, the survival rate of LSUPB401 was significantly higher than WT and QS-defective mutants (Figure 3.5). After 30 minutes, the survival rate of LSUPB401 was still significantly higher than WT and the other strains tested. LSUPB139 survived significantly less than

WT whereas survival of LSUPB145 was similar to WT (Figure 3.5). The survival rate of the *tepR* deletion mutant, LSUPB401, was restored to about the same as WT by the complementation of a plasmid-born *tepR* in LSUPB1068. The result indicated *tepR* and quorum sensing played key roles in heat stress tolerance in *B. glumae* in a negative and positive manner, respectively.

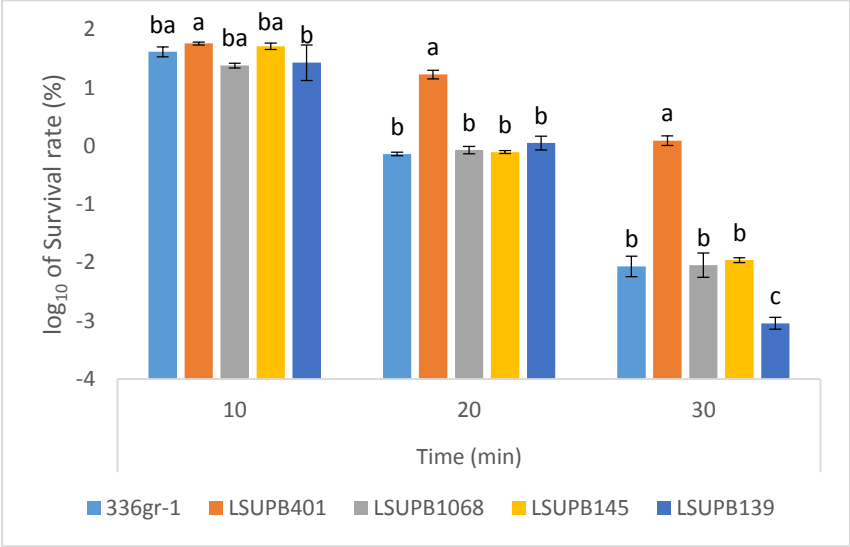


Figure 3.5. Heat stress tolerance of LSUPB401 and quorum sensing mutants. Survival rates were calculated as the CFU after the heat stress treatment divided by the CFU before the treatment. The $\log_{10}$ survival rate of 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB1069 (*ΔtepR*(ptepR)), LSUPB145 (*Δtofl*) and LSUPB139 (*Δtofl-tofR*) was display. The bacteria were treated at 50˚C for 0, 10, 20, 30 min, and viable cells were counted. Letters above the bars indicate the statistically significant differences (P< 0.05) among treatments obtained by Tukey's Studentized Range (HSD) Test. Error bars indicate ± standard deviation (SD) of three replicates.

Quorum sensing, but not *tepR*, was necessary for the survival of *B. glumae* in stationary-phase. The population of WT, LSUPB401, LSUPB1068, LSUPB145 and LSUPB139 grown in LB broth were normalized to 10^8 CFU/ml at 0 hour. Their population density were measured at 6 hours interval. The population were not significantly different (P< 0.05) between WT and all mutants tested during the exponential growth phase and early stationary growth phage (Figure. 3.6). However,

significant differences appeared beginning at 18 hours. WT, LSUPB401 and LSUPB1068 maintained the high population in stationary phase, whereas the population of both quorum sensing mutants, LSUPB145 and LSUPB139, started to decline rapidly. By the last quantification time point at 30 hours, WT, LSUPB401 and LSUPB1068 had the greatest populations, indicating that *tepR* did not have a significant impact on the growth curve of *B. glumae*. In contrast, the rapid drop in the population of quorum sensing mutants indicated that quorum sensing is essential for maintaining the population density in late stationary phase in *B. glumae* 336gr-1.

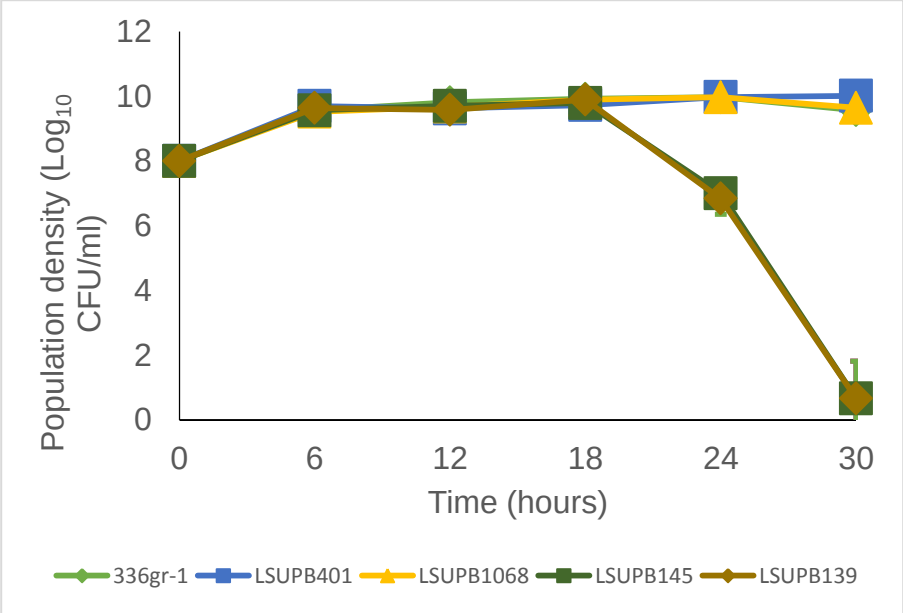
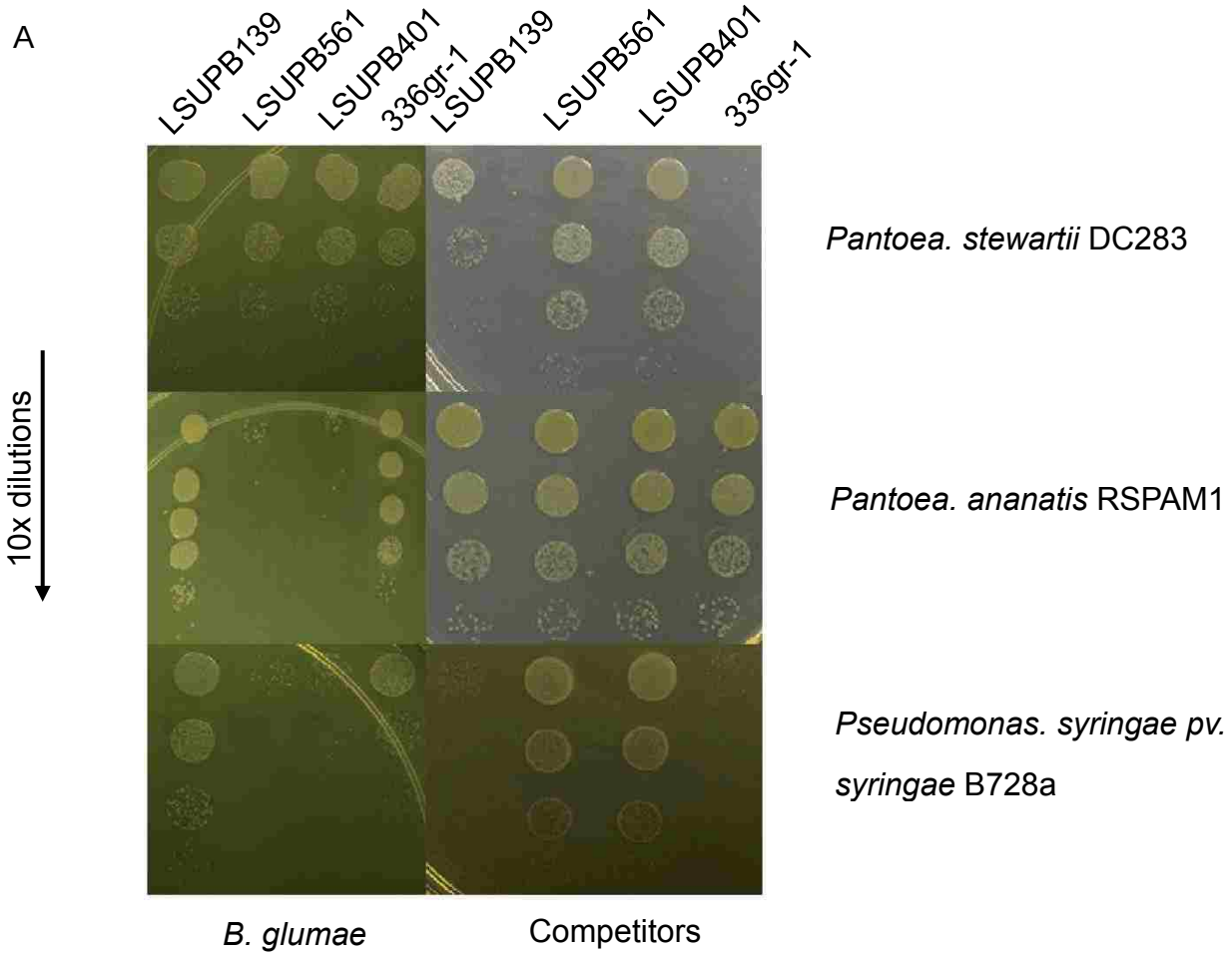


Figure 3.6. The Viability of 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB1069 (*ΔtepR*(ptepR)), LSUPB145 (*ΔtofI*) and LSUPB139 (*ΔtofI-tofR*) during stationary-phase stress. The viable cells were counted every 6 hours. Error bars indicate the ± standard deviation (SD) of three replicates.

The survival of *B. glumae* during interspecies competition was positively regulated by *tepR*. *B. glumae* strain 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB561 (*ΔtepR*/*toxA*$^-$), and LSUPB139 (*ΔtofI-tofR*) were tested for competition against three bacteria, *P. stewartii* DC283, *P. syringae* pv. *syringae* PssB728a and *P. ananatis*

RSPAM1. LSUPB561 (*ΔtepR*/*toxA*$^{-}$), a toxoflavin null mutant, was included in this study to make sure the difference in competition capacity between WT and LSUPB401 was not due to the over-expressed toxoflavin in LSUPB401. LSUPB139 (*ΔtofI-tofR*) was included to study whether or not quorum sensing regulated bacterial interspecies competition. The competition capacity is analyzed by Log CFU ratio (*B.glumae*/competitor). In case of competition against DC283 and PssB728a, both WT and LSUPB139 predominated, though LSUPB139 was less competitive than WT (Figure 3.7). In contrast to WT and the QS-defective mutant, both LSUPB401 and



B

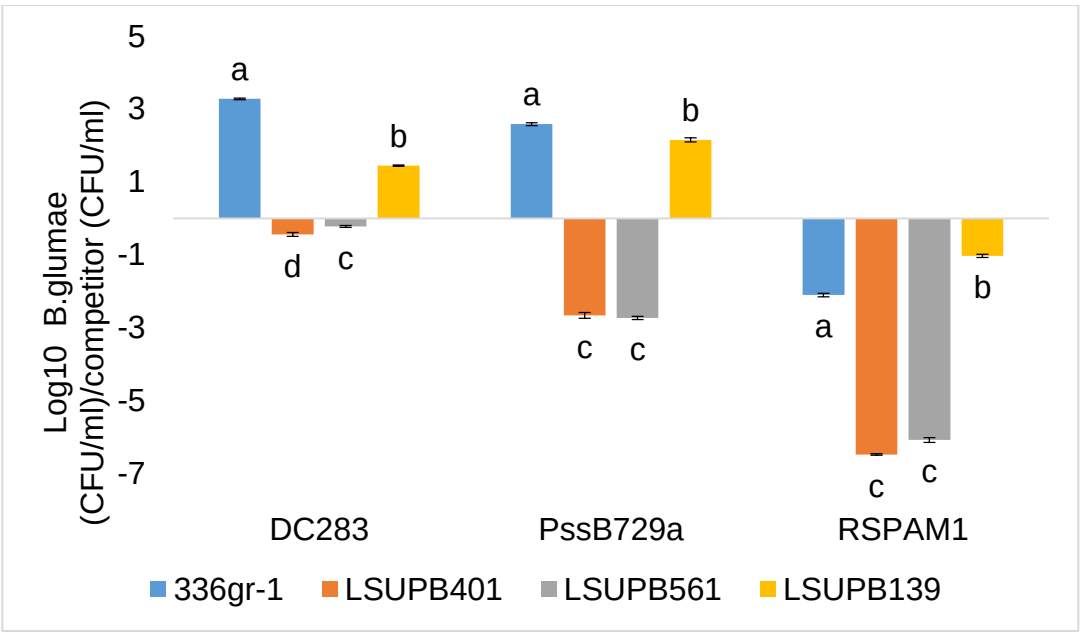


Figure 3.7. Positive regulation of bacterial interspecies competition by *tepR*. (A) Enumeration of the surviaval rate of *B. glumae* strains and the competitors in mixed cultures by 10-fold dilution series on selective medium. (B) Summary of the competition assay. The *B. glumae* strains tested were 336gr-1 (WT), LSUPB401 (*ΔtepR*), LSUPB561 (*ΔtepR*/toxA$^-$), and LSUPB139 (*ΔtofI-tofR*). The competitors were *Pantoea. stewartii* DC283, *Pantoea. ananatis* RSPAM1, *Pseudomonas. syringae pv. syringae* B728a. The bars indicate the mean of Log CFU of *B.glumae* /competitor. Error bars represent ± SD of three replicates.

LSUPB561 strains largely lost their competition capacity. For competition against RSPAM1, all 4 strains tested were less competitive than the competitor RSPAM1. Consistent with the competition against the other competitors, both LSUPB401 and LSUPB561 mutants were significantly less competitive than WT and LSUPB139. In all cases, both LSUPB401 and LSUPB561 had the least competition capacity indicating that *tepR* positively regulated the bacterial interspecies competition in *B. glumae*, which was not due to toxoflavin production. LSUPB139 also conferred relatively less competition capacity than WT in competition against DC283 and PssB728a. Thus, qorum sensing may also play a minor role for *B. glumae* to compete with other bacteria.

TepR gene is well conserved among *Burkholderia* spp. Phylogenetic analysis was conducted to compare TepR and its homologs in other bacteria of same or different genus. As indicated in Figure 3.8, TepR is well conserved in *Burkholderia* spp. and shows high similarity to WP_016945153.1 in *Xanthomonas campestris* ATCC 33913. 3 TepR homologs within *B. glumae* are grouped together with that in *Ralstonia solanacearum* etc. in another clade. Both clades are closed to LuxO in *Vibrio* spp. Other TepR homologs in *Yersinia kristensenii*, *Klebsiella pneumonia*, *Pantoea* spp., *Pectobacterium* spp., *Pseudomonas* spp., *Salmonella enterica*, *Escherichia coli*, together with two TepR homologs in *B. glumae*, are most distantly related to TepR.

Discussion

As discussed in Chapter 2, *tepR* regulates "response to stress" as revealed by the GO enrichment analysis. Among the 22 genes associated with the term, 14 genes, consistently up-regulated in LSUPB401, encode heat shock protein or chaperone protein and they have been known to be important for bacteria to live under stressed conditions (De Maio, 1999). To confirm whether LSUPB401 is indeed more tolerant to stressed conditions, heat stress and stationary phase stress challenges were performed on WT, LSUPB401, LSUPB1068 and quorum sensing-defective mutants. Quorum sensing-defective mutants were included in the assay because previous RNA-seq study indicated that several genes encoding heat shock proteins were shut down in LSUPB139 (*ΔtofI-tofR*) background (Chen and Ham, unpublished).

In this study, LSUPB401 was significantly more tolerant to 20 minutes or longer treatment at 50˚ C than WT (Figure 3.5). The survival rate was restored to the WT level in the complementation strain, LSUPB1068. The survival rate of LSUPB145 (*ΔtofI*) was

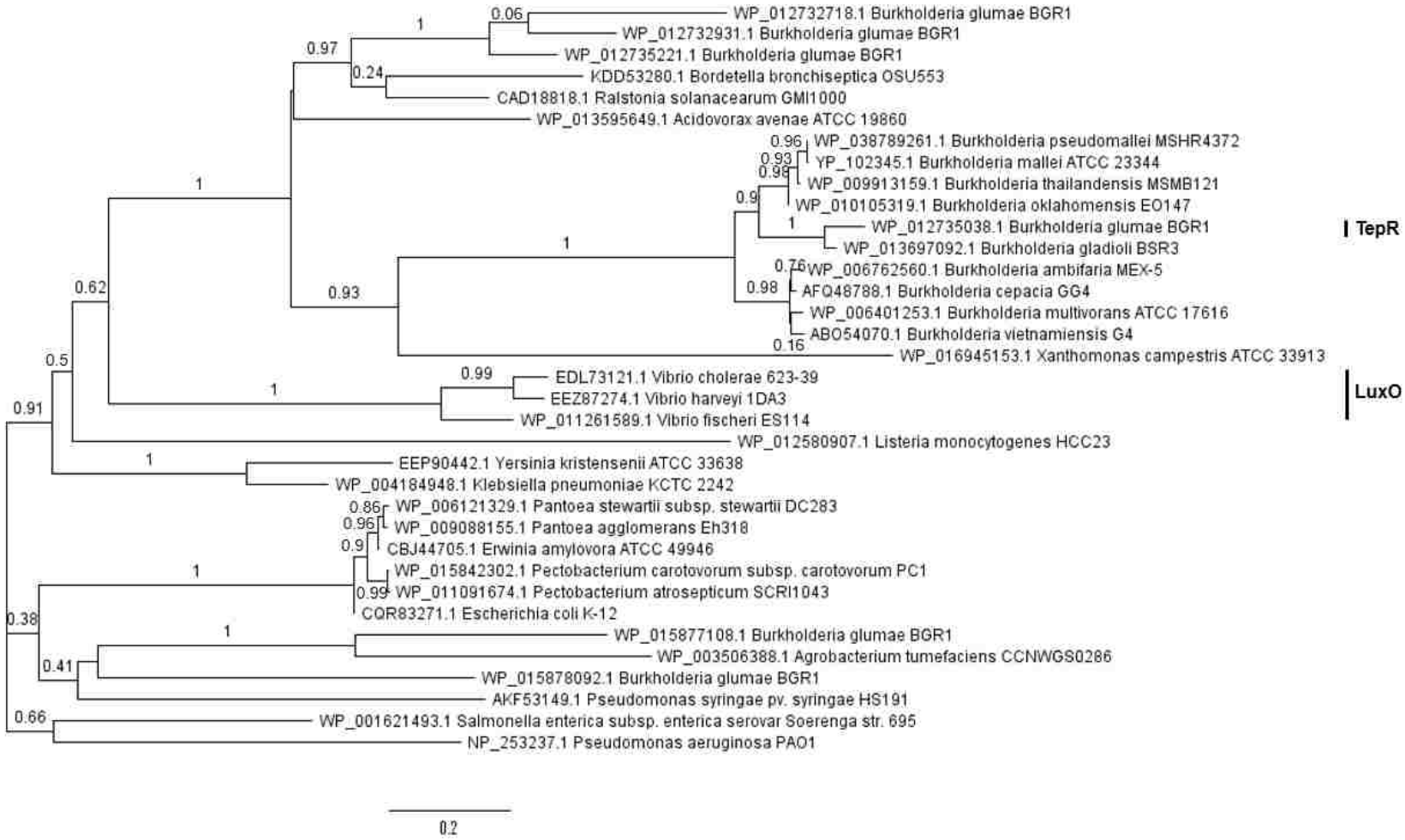


Figure 3.8. A phylogenetic tree of TepR and its homologs. The phylogenetic tree was constructed by FastTree 2.1.3 using a ML-NNIs algorism based on the amino acid sequence of TepR and 34 homologs. Node values indicate the posterior probability of each node. Scale bar represents 0.2 changes per position.

not significantly different from WT at any time point tested. The survival rate of LSUPB139 (*ΔtofI-tofR*) was not significantly different from WT at less than 30 minutes treatment but had reduced survival compared to WT thereafter. Thus, *tepR* negatively regulates bacterial survival in heat stress whereas TofI/TofR-mediated quorum sensing positively regulates bacterial survival in heat stress, even though TofI alone does not affect the heat tolerance of the *B. glumae*.

In the stationary phase stress assay, the population of WT and mutants were not significantly different (p-value< 0.05) until 18 hrs of incubation (Figure. 3.6). After 18 hours incubation, WT, LSUPB401 and LSUPB1068 maintained high bacterial populations, whereas the population of both quorum sensing mutants, LSUPB145 and LSUPB139, started to decrease rapidly. Hence, *tepR* does not play a major role in regulating bacterial survival in late stationary phase. In contrast, TofI/TofR mediated-quorum sensing is essential for maintaining the population in late stationary phage. Previous research by Goo *et al.* also found that in *B. glumae*, *B. pseudomallei* and *B. thailandensis*, the population of quorum sensing mutants dropped rapidly in late stationary phase which was due to the ammonia-mediated alkalization in nutrient-rich media (Goo et al., 2012). Quorum sensing also positively regulated oxalate production which counteracted alkalization toxicity (Goo et al., 2012). The rapid decline of the population of LSUPB145 and LSUPB139 in the late stationary phase might also due to the increased pH which was not favorable to bacterial survival. In the nature, bacteria may encounter various environmental stresses, e.g. cold stress, oxidative stress. Whether *tepR* regulates other stress responses or not is under investigation.

As revealed by KEGG pathway enrichment analysis, *tepR* positively regulated *T6SS-A* (Figure 2.8), which is homologous to T6SS-1 in *B. thailandensis* and it was found to be involved in cell contact-dependent competition (Schwarz et al., 2010). It is possible that *T6SS-A* in *B. glumae* may also be responsible for bacterial interspecies competition. LSUPB401, defective in *T6SS-A*, was therefore hypothesized to be less competitive than WT against bacterial competitors. To test this hypothesis, a competition assay was conducted between *B. glumae* strains and competitors, which were *Pantoea stewartii* DC283, *pseudomonas syringae* pv. *syringae* B728a, *Pantoea ananatis* RSPAM1. *tepR* positively regulated the bacterial interspecies competition as revealed by the significantly lower competition capacity in LSUPB401 (*ΔtepR*) than the WT. Note that the survival rate of LSUPB561 (*ΔtepR*::*toxA*) was not significantly different from LSUPB401 (*ΔtepR*), indicating that toxoflavin did not play a major role in the bacterial competition. This is not surprising, because toxoflavin has been known to be an effective electron carrier and facilitates the production of hydrogen peroxide which causes oxidative stress to plants (Latuasan & Berends, 1961). Based on the result of this study, quorum sensing might also play a minor role in competition with some bacteria and not with others. This study reveals a new gene which plays a positive regulatory role in bacterial interspecies competition. This study also strongly suggests that *T6SS-A* is responsible for the competitive ability in *B. glumae*. Besides bacterial competition, certain T6SSs were also found to be important for bacterial virulence (Schell et al., 2007, Hood et al., 2010, Shalom et al., 2007, Pilatz et al., 2006). However, *T6SS-A* in *B. glumae* may not play a significant role in bacterial virulence since

LSUPB401 causes more severe disease symptoms on rice than WT (Osti, 2014). Further experiments in progress to test this hypothesis.

Finally, to examine presence of TepR homologs in other bacteria and their phylogenetic relationship, a phylogenetic tree was made comparing 35 TepR homologs including TepR by FastTree 2.1.3 (Price et al., 2010). The amino acid sequence alignments were culled by Gblocks (Castresana, 2000) to get rid of the very poorly aligned regions. This is because all the TepR homologs are well conserved in the three functional domains, REC signal receiver domain, Sigma54 interaction domain and HTH_8 Fis transcriptional regulator domain but highly variable in the rest of the regions. Following Gblocks trimming, 334 out of 576 original aligned positions were kept in a total of 3 blocks. In this study, TepR homologs could be found in all the bacteria examined. TepR homologs are especially well conserved among *Burkholderia*. Whether TepR homologs have shared functions in other bacteria is unknown.

References

Baird PN, Hall LM, Coates AR, 1989. Cloning and sequence analysis of the 10 kDa antigen gene of Mycobacterium tuberculosis. *J Gen Microbiol* **135**, 931-9.

Ballister ER, Lai AH, Zuckermann RN, Cheng Y, Mougous JD, 2008. In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci U S A* **105**, 3733-8.

Basler M, Ho BT, Mekalanos JJ, 2013. Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions. *Cell* **152**, 884-94.

Basler M, Mekalanos JJ, 2012. Type 6 secretion dynamics within and between bacterial cells. *Science* **337**, 815.

Benz J, Sendlmeier C, Barends TR, Meinhart A, 2012. Structural insights into the effector-

Blondel CJ, Jimenez JC, Contreras I, Santiviago CA, 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in Salmonella serotypes. *Bmc Genomics* **10**, 354.

Boyer F, Fichant G, Berthod J, Vandenbrouck Y, Attree I, 2009. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *Bmc Genomics* **10**, 104.

Cabiscol E, Tamarit J, Ros J, 2000. Oxidative stress in bacteria and protein damage by reactive oxygen species. *Int Microbiol* **3**, 3-8.

Calloni G, Chen T, Schermann SM*, et al.*, 2012. DnaK functions as a central hub in the E. coli chaperone network. *Cell Rep* **1**, 251-64.

Castresana J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-52.

Chatterji D, Ojha AK, 2001. Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol* **4**, 160-5.

Chun H, Choi O, Goo E*, et al.*, 2009. The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *J Bacteriol* **191**, 4152-7.

Coplin DL, Frederick RD, Majerczak DR, Haas ES, 1986. MOLECULAR-CLONING OF VIRULENCE GENES FROM ERWINIA-STEWARTII. *Journal of Bacteriology* **168**, 619-23.

Coulthurst SJ, 2013. The Type VI secretion system - a widespread and versatile cell targeting system. *Res Microbiol* **164**, 640-54.

Craig EA, 1985. The heat shock response. *CRC Crit Rev Biochem* **18**, 239-80.

De Maio A, 1999. Heat shock proteins: facts, thoughts, and dreams. *Shock* **11**, 1-12.

Derre I, Rapoport G, Msadek T, 1999. CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. *Mol Microbiol* **31**, 117-31.

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7.

Fang FC, Libby SJ, Buchmeier NA*, et al.*, 1992. The alternative sigma factor katF (rpoS) regulates Salmonella virulence. *Proc Natl Acad Sci U S A* **89**, 11978-82.

Gerlach RG, Hensel M, 2007. Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* **297**, 401-15.

Goo E, Majerczyk CD, An JH*, et al.*, 2012. Bacterial quorum sensing, cooperativity, and anticipation of stationary-phase stress. *Proc Natl Acad Sci U S A* **109**, 19775-80.

Ham JH, Melanson RA, Rush MC, 2011. Burkholderia glumae: next major pathogen of rice? *Molecular Plant Pathology* **12**, 329-39.

Holland IB, 2010. The extraordinary diversity of bacterial protein secretion mechanisms. *Methods Mol Biol* **619**, 1-20.

Hood RD, Singh P, Hsu F*, et al.*, 2010. A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell Host Microbe* **7**, 25-37.

Kim H, Goo E, Kang Y, Kim J, Hwang I, 2012. Regulation of universal stress protein genes by quorum sensing and RpoS in Burkholderia glumae. *J Bacteriol* **194**, 982-92.

Kim J, Kang Y, Choi O*, et al.*, 2007. Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in Burkholderia glumae. *Mol Microbiol* **64**, 165-79.

Kim J, Kim JG, Kang Y*, et al.*, 2004. Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in Burkholderia glumae. *Mol Microbiol* **54**, 921-34.

Kim TJ, Young BM, Young GM, 2008. Effect of flagellar mutations on Yersinia enterocolitica biofilm formation. *Appl Environ Microbiol* **74**, 5466-74.

Lange R, Hengge-Aronis R, 1991. Identification of a central regulator of stationary-phase gene expression in Escherichia coli. *Mol Microbiol* **5**, 49-59.

Latuasan HE, Berends W, 1961. On the origin of the toxicity of toxoflavin. *Biochim Biophys Acta* **52**, 502-8.

Leiman PG, Basler M, Ramagopal UA*, et al.*, 2009. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**, 4154-9.

Loewen PC, Switala J, 1986. Purification and characterization of catalase HPII from Escherichia coli K12. *Biochem Cell Biol* **64**, 638-46.

Loewen PC, Switala J, Triggs-Raine BL, 1985. Catalases HPI and HPII in Escherichia coli are induced independently. *Arch Biochem Biophys* **243**, 144-9.

Loper JE & Lindow SE, 1987. Lack of Evidence for In Situ Fluorescent Pigment Production by *Pseudomonas syringae* pv. *syringae* on Bean Leaf Surfaces. *Phytopathology* **77,** 1449-54.

Lossi NS, Dajani R, Freemont P, Filloux A, 2011. Structure-function analysis of HsiF, a gp25-like component of the type VI secretion system, in Pseudomonas aeruginosa. *Microbiology* **157**, 3292-305.

Magnusson LU, Farewell A, Nystrom T, 2005. ppGpp: a global regulator in Escherichia coli. *Trends Microbiol* **13**, 236-42.

Mougous JD, Cuff ME, Raunser S*, et al.*, 2006. A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science* **312**, 1526-30.

Murdoch SL, Trunk K, English G, Fritsch MJ, Pourkarimi E, Coulthurst SJ, 2011. The opportunistic pathogen Serratia marcescens utilizes type VI secretion to target bacterial competitors. *J Bacteriol* **193**, 6057-69.

Osti S, Barphagha IK, Ham J, 2013. The novel regulator, tepR, influences toxoflavin production and virulence in the rice pathogenic bacterium Burkholderia glumae. *Phytopathology* **103**, 8-9.

Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR, 2009. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci U S A* **106**, 4160-5.

Pilatz S, Breitbach K, Hein N*, et al.*, 2006. Identification of Burkholderia pseudomallei genes required for the intracellular life cycle and in vivo virulence. *Infect Immun* **74**, 3576-86.

Price MN, Dehal PS, Arkin AP, 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *Plos One* **5**, e9490.

Pukatzki S, Ma AT, Sturtevant D*, et al.*, 2006. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc Natl Acad Sci U S A* **103**, 1528-33.

Rosen R, Ron EZ, 2002. Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom Rev* **21**, 244-65.

Russell AB, Hood RD, Bui NK, Leroux M, Vollmer W, Mougous JD, 2011. Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**, 343-7.

Rutherford ST, Van Kessel JC, Shao Y, Bassler BL, 2011. AphA and LuxR/HapR reciprocally control quorum sensing in vibrios. *Genes Dev* **25**, 397-408.

Schell MA, Ulrich RL, Ribot WJ*, et al.*, 2007. Type VI secretion is a major virulence determinant in Burkholderia mallei. *Mol Microbiol* **64**, 1466-85.

Schwarz S, West TE, Boyer F*, et al.*, 2010. Burkholderia type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog* **6**, e1001068.

Segal G, Ron EZ, 1993. Heat shock transcription of the groESL operon of Agrobacterium tumefaciens may involve a hairpin-loop structure. *J Bacteriol* **175**, 3083-8.

Segal R, Ron EZ, 1996. Regulation and organization of the groE and dnaK operons in Eubacteria. *FEMS Microbiol Lett* **138**, 1-10.

Shalom G, Shaw JG, Thomas MS, 2007. In vivo expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. *Microbiology* **153**, 2689-99.

Shao Y, Bassler BL, 2012. Quorum-sensing non-coding small RNAs use unique pairing regions to differentially control mRNA targets. *Mol Microbiol* **83**, 599-611.

Shao Y, Bassler BL, 2014. Quorum regulatory small RNAs repress type VI secretion in Vibrio cholerae. *Mol Microbiol* **92**, 921-30.

Sharma UK, Chatterji D, 2010. Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity. *FEMS Microbiol Rev* **34**, 646-57.

Storz G, Imlay JA, 1999. Oxidative stress. *Curr Opin Microbiol* **2**, 188-94.

Suh SJ, Silo-Suh L, Woods DE, Hassett DJ, West SE, Ohman DE, 1999. Effect of rpoS mutation on the stress response and expression of virulence factors in Pseudomonas aeruginosa. *J Bacteriol* **181**, 3890-7.

Triggs-Raine BL, Loewen PC, 1987. Physical characterization of katG, encoding catalase HPI of Escherichia coli. *Gene* **52**, 121-8.

Vanaporn M, Vattanaviboon P, Thongboonkerd V, Korbsrisate S, 2008. The rpoE operon regulates heat stress response in Burkholderia pseudomallei. *FEMS Microbiol Lett* **284**, 191-6.

Wu J, Long Q, Xie J, 2010. (p)ppGpp and drug resistance. *J Cell Physiol* **224**, 300-4.

Zeilstra-Ryalls J, Fayet O, Georgopoulos C, 1991. The universally conserved GroE (Hsp60) chaperonins. *Annu Rev Microbiol* **45**, 301-25.

Zheng J, Shin OS, Cameron DE, Mekalanos JJ, 2010. Quorum sensing and a global regulator TsrA control expression of type VI secretion and virulence in Vibrio cholerae. *Proc Natl Acad Sci U S A* **107**, 21128-33.

Zuber U, Schumann W, 1994. CIRCE, a novel heat shock element involved in regulation of heat shock operon dnaK of Bacillus subtilis. *J Bacteriol* **176**, 1359-63.

CHAPTER 4: REGULATION OF TOXOFLAVIN AND PROTEASE BY *TEPR* IS QSMR-DEPENDENT

Introduction

The gene *qsmR* was first identified and characterized by Kim et al. (Kim et al., 2007). Sequence analysis indicated QsmR shows 93% identity in amino acid to IcIR protein in *Escherichia coli* (Kim et al., 2007). Qsm*R* is termed after quorum sensing master regulator because its expression is activated by TofI/TofR quorum sensing (Kim et al., 2007). A *qsmR*-defective mutant did not produce a comparable amount of toxoflavin as the wild type *B. glumae* strain in LB liquid medium but no significant difference was observed on LB agar (Kim et al., 2007). QsmR also regulates the expression of *flhDC* which further activates the flagellum biosynthesis genes (Kim et al., 2007, Kutsukake, 1997). Mutant defective in *qsmR* causes failure to produce flagella in *B. glumae* at 37˚ C (Kim et al., 2007, Jang et al., 2014). At 28˚ C, flagella biosynthesis is independent from QsmR and QS indicating that flagella formation in *B. glumae* can be co-regulated by QsmR and temperature (Jang et al., 2014). QsmR was also characterized to positively regulate a total of 11 universal stress protein (usp) genes (Kim et al., 2012, Nystrom & Neidhardt, 1992). Among them, the expression of *usp1* and *usp2* is also under the regulation of sigma factor S which directly binds to their promoter region (Kim et al., 2012). An independent mutation experiment showed all the QsmR-dependent *usp* genes contributes to heat shock stress (Kim et al., 2012). QsmR controls the expression of type II secretion system (T2SS), which is encoded by 12 general secretion pathway (gsp) genes (Goo et al., 2010). Several important virulence factors are known to be secreted by T2SS, e.g. lipase and protease (Gauthier et al., 2000, Sandkvist, 2001). Furthermore, the expression of a catalase gene, *katG*, is also

activated by QsmR and *katG* is required for the full virulence of *B. glumae* on rice (Chun et al., 2009). In addition to being a positive regulator, QsmR is also reported to be a repressor of glucose uptake, primary metabolism and nucleotide synthesis (An et al., 2014).

As discussed in Chapter 2, the expression of *qsmR* was significantly higher in LSUPB401 (*ΔtepR*) than in WT, indicating that *tepR* negatively regulates the expression of *qsmR*. Note that the homolog of TepR, LuxO, was characterized as a quorum sensing-dependent repressor of bioluminescence in *V. harveyi*. One possible explanation for the *tepR*-dependent repression of toxoflavin and extracellular protease production in *B. glumae* is that *tepR* acts as a repressor in the expression of QsmR, which further modulates the quorum sensing-dependent phenotypes, for example toxoflavin production and extracellular protease activity. To test this hypothesis, *ΔqsmR* and *ΔqsmR*/*ΔtepR* derivatives of *B. glumae* 336gr-1 were constructed in this study. The toxoflavin production in LB broth was quantified and the protease activity on Nutrient Agar supplemented with skimmed milk was observed.

Materials and Methods

Bacterial strains, plasmids and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 4.1. *B. glumae* strains and *E.coli* strains were routinely grown in LB broth with shaken at 200 rpm or on LB agar plates at 37˚ C overnight. LB agar supplemented with 30% sucrose was used for screening the secondary crossover recombinant mutants, which lost the sucrose-sensitive gene, *sacB* (Chen et al., 2012). Antibiotics used in this study were 50 μg/ml kanamycin (Km), 100 μg/ml nitrofurantoin (Nit), and 20 μg/ml gentamycin (Gm).

plasmids and strains	Description	References
Escherichia coli		
DH10B	F- *araD139 D(ara, leu)*7697 *ΔlacX74 galU galK rpsL deoR ø80dlacZΔ*M15 *endA1 nupG recA1* *mcrA Δ(mrr hsdRMS mcrBC)*	(Grant et al., 1990)
S17-1λpir	*recA thi pro hsdR* [res-mod+][RP4::2-Tc::Mu-Km::Tn7] *λ pir* phage lysogen, Sm^R/Tp^R	(Simon et al., 1983)
Burkholderia glumae		
336gr-1	Wild type strain isolated from diseased rice in Crowley, Louisiana, USA	This study
LSUPB574	A *ΔqsmR* derivative of 336gr-1	This study
plasmid		
pBBR1MCS-5	A broad host range cloning vector, RK2 *ori*, *lacZα*, Gm^R	(Kovach et al., 1995)
pBBqsmRG	A subclone of pSCqsmRG for the 678-bp *qsmR* and downstream region inserted into pBBR1MCS-5 at the PstI and EcoRI sites, Gm^R	This study
pBBqsmRP	A subclone of pSCqsmRP for the 937-bp *qsmR* and upstream region inserted into pBBR1MCS-5 at the PstI sites, Gm^R	This study
pBBqsmR	A *qsmR* clone in pBBR1MCS-5, Gm^R	This study
pKNOCK-Km	A suicide vector; R6K γ-*ori*, RP4 *oriT*, Km^R	(Alexeyev, 1999)
pKNOCKqsmR	A *qsmR* clone in pKNOCK-Km	This study

(Table 4.1 continued)

plasmids and strains	Description	References
pKKSacB	A suicide vector; R6K γ-*ori*, RP4 *oriT*, *sacB*, Km^R	(Chen et al., 2012)
pKKqsmRU	A subclone of pSCqsmRU for the upstream flanking region of *qsmR* in pKKSacB, Km^R	This study
pKKSacBΔqsmR	plasmid with the upstream and downstream flanking regions of *qsmR* in pKKSacB, Km^R	This study
pRK2013::Tn7	A helper plasmid; ColE1 *ori*	(Ditta et al., 1980)
pSC-A-amp/kan	A blunt-end PCR cloning vector; f1 *ori*, pUC *ori*, *lacZ'*, Km^R, Amp^R	Stratagene
pSCqsmRU	A PCR clone of the 458-bp upstream flanking region of *qsmR* in pSC-A-amp/kan, Km^R, Amp^R	This study
pSCqsmRD	A PCR clone of the 469-bp downtream flanking region of *qsmR* in pSC-A-amp/kan, Km^R, Amp^R	This study
pSCqsmRP	A clone of the 606-bp region of *qsmR* with 376-bp upstream region, Amp^R, Km^R	This study
pSCqsmRG	A clone of the 408-bp region of *qsmR* with 366-bp downstream region, Amp^R, Km^R	This study

Table 4.1. Bacterial strains and plasmids used in this study.

Construction of a marker-less *qsmR* deletion mutant. The marker-less deletion of *qsmR* (locus tag= bglu_1g10250) in *B. glumae* 336gr-1 was achieved through a double-crossover allelic exchange method using the suicide vector, pKKSacB, harboring a *sacB* gene for sucrose sensitivity (Chen et al., 2012a). DNA constructs used

for making deletion mutations in this study are listed in Table 4.1. The *qsmR* deletion mutant was achieved through double-crossover homologous recombination in the flanking regions of the *qsmR* gene. Specifically, a 458-bp upstream region and a 469-bp downstream region were amplified with the primer sets qsmrBamHIUL/qsmrSpeIUR and qsmrSpeIDL/qsmrXbalIDR1 respectively (Table 4.2). Resultant PCR products of the upstream and downstream regions were then cloned to pSC-A-amp/kan to generate pSCqsmRU and pSCqsmRD using the Strataclone PCR Cloning Kit (Agilent Technologies, Santa Clara, CA). The upstream region was further cloned to pKKSacB with the restriction site BamHI and SpeI to generate pKKqsmRU. The downstream region was cloned to pKKqsmRU with the restriction site SpeI and XbaII to achieve pKKSacBΔqsmR. The allelic exchange in *B. glumae* to introduce the *qsmR* deletion was based on the previous described method (Chen et al., 2012). Briefly, pKKSacBΔqsmR was transformed to *E. coli* S17-1λ*pir* through electroporation. The construct was introduced to WT and LSUPB401 by triparental mating using the helper strain pRK2013::Tn7. Colonies were picked from LB agar plates supplemented with Km and Nit. Overnight liquid cultures of the selected colonies were spread on LB plate supplemented with 30% sucrose and incubated for 72 hours. The colonies were picked on both LB plates and Km-supplemented LB plates. Colonies that could only grow on LB but not Km-supplemented LB were selected. Deletion of the *qsmR* gene was confirmed by lack of the amplification with the primers qsmrDLcheckF and qsmrDLcheckR (Table 4.2). The map showing the schematic view of the DNA constructs for *qsmR* deletion mutation was displayed in Figure 4.1.

Table 4.2. Primers and PCR conditions used in this study.

Primer names	Primer sequences	Annealing and extension conditions
qsmrBamHIUL	GGGATCCGTCGATTTCATCGCCAATTT	Annealing: 51°C/30 s Extension: 72°C/45 s
qsmrSpeIUR	GGGACTAGTCGGTCGCTGCTTTATTCAGT	
qsmrSpeIDL	GGGACTAGTGAATGGCTGCTCGAGACT	Annealing: 51°C/30 s Extension: 72°C/45 s
qsmrXbalIDR1	GGGTCTAGATCATGTTCGATCTGGCTGAC	
QSMPF1	GTTGCGCAGCGTATCCTC	Annealing: 50°C/30 s Extension: 72°C/1 min
QSMPR2	ATTGGGAAATTTGGCTTTTTC	
QSMRGNF2	CGGTTCCGGGTTATTCATGTTC	Annealing: 50°C/30 s Extension: 72°C/1 min
QSMRGNR2	GTCACCCGGCTCGAGAT	
qsmrDLcheckF	CGTGCTAGAACCTGAGAGAC	Annealing: 50°C/30 s Extension: 72°C/1.5 min
qsmrDLcheckR	ATCGTCCAGAGCACTTTCT	

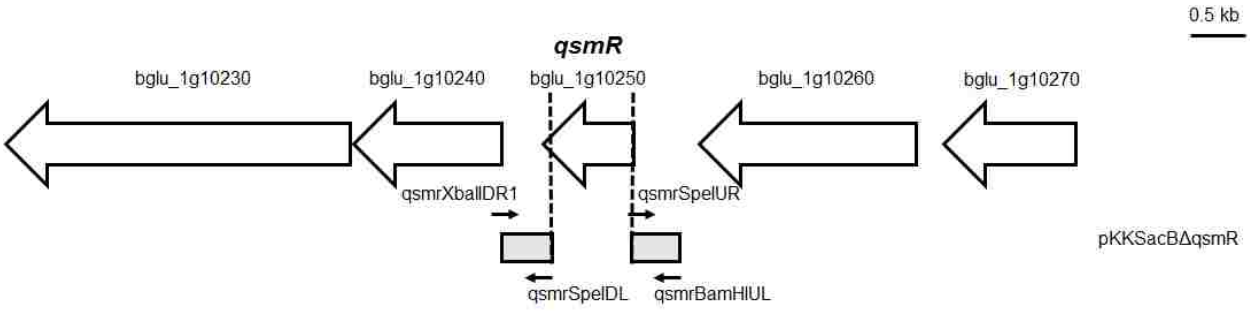


Figure 4.1. A schematic view of the pKKSacBΔqsmR construction for generating markerless deletion of *qsmR*. The grey boxes represent the flanking regions of *qsmR* cloned in pKKSacB. Between hatched lines is the region knocked out.

Complementation of *qsmR* deletion mutant. Briefly, 376 bp of the putative promoter region of *qsmR* together with 606 bp of the gene from the start codon was amplified with the primers QSMPF1 and QSMPR2 (Table 2) and cloned to pSC-A-amp/kan to generate pSCqsmRP. A fragment with 408 bp of the gene and 366 bp downstream of the stop codon was amplified with the primers QSMRGNF2 and QSMRGNR2 (Table 2) and cloned to pSC-A-amp/kan to generate pSCqsmRG. Plasmid pSCqsmRG was cut at the PstI and EcoRI sites and then subcloned to pBBRMCS-5 to get pBBqsmRP. Plasmid pSCqsmRP was cut at the PstI site and subcloned to pBBqsmRG to achieve pBBqsmR. pBBqsmR was cut at KpnI and SpeI sites, and then subcloned to pKNOCK-Km (Alexeyev, 1999) to achieve pKNOCKqsmR. The plasmid pKNOCKqsmR was conjugated to *B. glumae* 336gr-1 (WT) and LSUPB401 (*ΔtepR*) by triparental mating. The success of the complementation was confirmed by PCR using the primer set QSMPF1 and QSMPR2 (Table 4.2).

Toxoflavin extraction and quantification. The toxoflavin production of WT and mutants of *B. glumae* in liquid media were extracted and quantified following a previously described method (Chen et al., 2012) with slight modifications. *B. glumae* strains tested were recovered from glycerol stock stored at -70˚C and streaked on LB agar overnight. Strains were again cultured in LB broth for 18 hours at 37 ˚C with 200 rpm shaking. The concentrations of cultures were measured at OD_{600} in a spectrophotometer (Biomate 3, thermoelectromate corporation, USA). One ml culture was centrifuge at 17,000 x g for 1 min. Nine hundred µl supernatant was transferred to another 2 ml microcentrifuge tube and mixed with 900 µl chloroform by vortexing for 30 sec. The mixture was centrifuged at 12,000 x g for 10 min. The chloroform phase was

transferred to a new microcentrifuge tube and evaporated in a fume hood for 24 hours. Following evaporation, the solid residue was resuspended in 900 µl 80% methanol and absorbance was measured at OD_{260}. Statistical analysis of OD_{260}/OD_{600} ratio was performed for all the strains tested obtained by Tukey's Studentized Range (HSD) Test with a cut-off P-value of 0.05.

Protease activity. The protease activity was tested following Huber's method (Huber et al., 2001) with some modifications. Overnight cultures of *B. glumae* strains tested in LB broth were washed twice with fresh LB broth and finally diluted to OD_{600}=1.0. 5 µl diluted culture was spotted on Nutrient Agar (NA) plate containing 1.5% skimmed milk powder. The plate was incubated at 37 ˚C for 30 hours. Protease activity was indicated by the clear zone formed surrounding the bacterial culture spot.

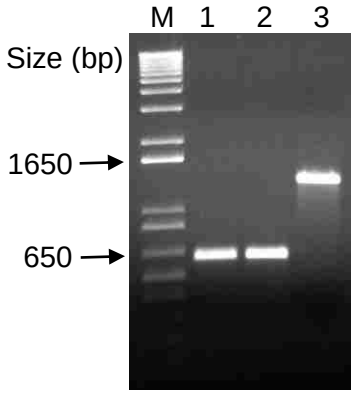


Figure 4.2. Confirmation of the deletion mutations in *Burkholderia glumae* strains. The success of *qsmR* deletion was confirmed by a diagnostic PCR with the primer set, qsmrDLcheckF and qsmrDLcheckR. The DNA for each lane is as follows: M, 1 Kb Plus DNA Ladder (Invitrogen, Santa Clara, CA, USA); 1, the genomic DNA of LSUPB1047 (*ΔqsmR*); 2, the genomic DNA of LSUPB1049 (*ΔqsmR/ΔtepR*); and 3, the genomic DNA of WT (336gr-1). The size of the PCR product for WT was 1321 bp, whereas they were 701 bp for *ΔqsmR* and *ΔqsmR/ΔtepR* mutants.

Results

Marker-less deletion of *qsmR* was constructed. The *qsmR* knockout was generated in the background of *B. glumae* WT and LSUPB401 using a pKKSacB system (Chen et al., 2012). The success of deletion was confirmed by PCR with primers qsmrDLcheckF and qsmrDLcheckR (Figure 4.2 and Table 4.2). The size of the PCR product for WT was 1321 bp, whereas they were 701 bp for *ΔqsmR* and *ΔqsmR/ΔtepR* mutants. The resultant *ΔqsmR* and *ΔqsmR/ΔtepR* mutants generated were named LSUPB1047 and LSUPB1049 respectively.

Knockout of *qsmR* abolished the toxoflavin production in WT and LSUPB401 background. As shown in Figure 4.3, the production of toxoflavin in LSUPB401 was significantly higher than in WT. This was consistent with the previous study (Osti, 2014). Both LSUPB1047 (*ΔqsmR*) and LSUPB1049 (*ΔqsmR/ΔtepR*) produced very little toxoflavin in LB broth, indicating that QsmR is an essential component for toxoflavin production independent from the regulation by *tepR*.

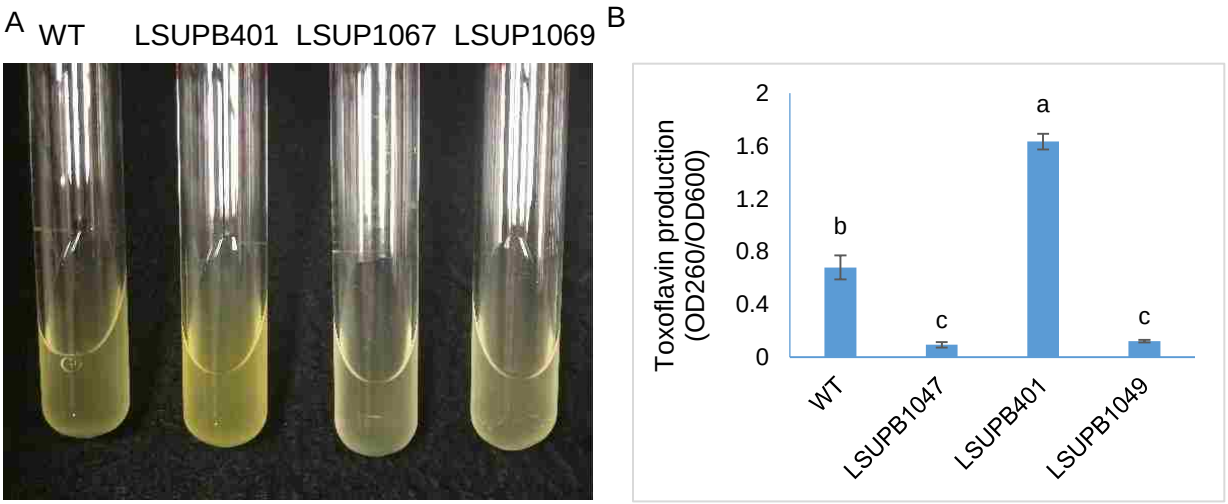


Figure 4.3. The necessity of QsmR for toxoflavin production. (A) Toxoflavin production in LB broth, characterized as a yellow pigment. (B) Quantification of toxoflavin production. Error bars indicate ± SD of three replicates.

The negative regulation of protease activity by *tepR* is dependent upon the positive regulation by quorum sensing. As shown in Figure 4.4, knockout of *qsmR* abolished the protease activity of *B. glumae*. As discussed in Chapter 1, LSUPB401 (*ΔtepR*) has greater protease activity. However, in this study LSUPB1049 (*ΔtepR* /*ΔqsmR*) lost protease activity, indicating the negative regulation of protease activity by *tepR* was dependent upon the positive regulator QsmR. To further figure out whether protease activity is under the regulation of TofI/TofR-mediated quorum sensing, the protease activity of quorum sensing-defective mutants, LSUPB145 (*ΔtofI*), LSUPB169 (*ΔtofR*), LSUPB139 (*ΔtofI-tofR*), LSUPB201 (*ΔtofI/tofR*) and LSUPB139 (pBBtofIMR), were tested on skimmed milk supplemented-nutrient agar without or with 1μM C8-HSL. Result indicated that protease activity was indeed dependent upon quorum sensing. No quorum sensing mutant showed any protease activity after 24 hours on skimmed milk supplemented-nutrient agar. The abolished protease activity in LSUPB139 was restored by pBBtofIMR. When supplemented with 1 μM C8-HSL, the protease activity of LSUPB145 was as about the same as WT.

Figure 4.4. Regulation of protease activity by quorum sensing. (A) QsmR-dependent protease activity. (B) TofI/TofR-dependent protease activity.

Discussion

Toxoflavin, a phytotoxin, is one of the most important virulence factors in *B. glumae* (Yoneyama et al., 1998). While it was not known that production of toxoflavin by *B. glumae* is positively regulated by QsmR-involved quorum sensing (Kim et al., 2004, Kim et al., 2007) and negatively regulated by *tepR* (Osti, 2014). This study shows that the negative regulation of toxoflavin production by *tepR* is still dependent upon positive regulation by QsmR. Overall, QsmR played the dominating regulatory role in toxoflavin production between the two counteractive regulatory systems. This was demonstrated by the fact when *qsmR* was knocked out in both WT and LSUPB401 (*ΔtepR*) background, toxoflavin production of the resulting mutants, LSUPB1047 (*ΔqsmR*) and LSUPB1049 (*ΔqsmR/ΔtepR*) largely abolished the ability to produce toxoflavin indicating the positive regulatory role of *qsmR*.

Various bacteria can produce extracellular or cell surface proteases that could be important virulence factors (Lantz, 1997, Lebrun et al., 2009). Our previous study indicated *B. glumae* could also produce extracellular protease, which was negatively regulated by of *tepR* (Osti, 2014), but other regulatory systems of extracellular protease activity were unknown. In this study, protease activity of *B. glumae* was positively regulated by QsmR. LSUPB1049 also lacked protease activity, indicating that the negative regulation of protease activity by *tepR* is still dependent upon the positive regulation of QsmR. Because QsmR is known to be regulated by TofI/TofR-medicated quorum sensing (Kim et al., 2007), it is very likely that protease activity is also under the regulation of TofI/TofR-medicated quorum sensing. To test this, the protease activity of quorum sensing-defective mutants was examined. The result showed TofI/TofR-

mediated quorum sensing indeed played key regulatory role in protease activity in *B. glumae*.

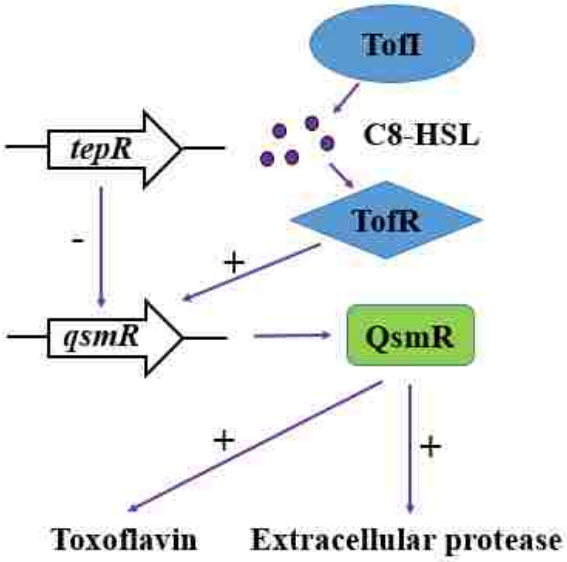


Figure 4.5. Schematic of toxoflavin and extracellular protease regulation in *B. glumae*. Mark "+" denotes positive regulation and "-" denotes negative regulation.

Based on the result of this study and previous study, both toxoflavin and protease activity were positively regulated by QsmR and TofI/TofR-mediated quorum sensing but negatively regulated by *tepR* (Figure 4.5). Between these two counteractive systems, QsmR and TofI/TofR-mediated QS played the dominating regulatory role in toxoflavin production and extracellular protease activity (Figure 4.5).

References

Alexeyev MF, 1999. The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria. *Biotechniques* **26**, 824-6, 8.

An JH, Goo E, Kim H, Seo YS, Hwang I, 2014. Bacterial quorum sensing and metabolic slowing in a cooperative population. *Proc Natl Acad Sci U S A* **111**, 14912-7.

Chen R, Barphagha IK, Karki HS, Ham JH, 2012. Dissection of quorum-sensing genes in Burkholderia glumae reveals non-canonical regulation and the new regulatory gene tofM for toxoflavin production. *Plos One* **7**, e52150.

Chun H, Choi O, Goo E*, et al.*, 2009. The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *J Bacteriol* **191**, 4152-7.

Ditta G, Stanfield S, Corbin D, Helinski DR, 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of Rhizobium meliloti. *Proc Natl Acad Sci U S A* **77**, 7347-51.

Gauthier YP, Thibault FM, Paucod JC, Vidal DR, 2000. Protease production by Burkholderia pseudomallei and virulence in mice. *Acta Trop* **74**, 215-20.

Goo E, Kang Y, Kim H, Hwang I, 2010. Proteomic analysis of quorum sensing-dependent proteins in Burkholderia glumae. *J Proteome Res* **9**, 3184-99.

Grant SG, Jessee J, Bloom FR, Hanahan D, 1990. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc Natl Acad Sci U S A* **87**, 4645-9.

Huber B, Riedel K, Hentzer M*, et al.*, 2001. The cep quorum-sensing system of Burkholderia cepacia H111 controls biofilm formation and swarming motility. *Microbiology* **147**, 2517-28.

Jang MS, Goo E, An JH, Kim J, Hwang I, 2014. Quorum sensing controls flagellar morphogenesis in Burkholderia glumae. *Plos One* **9**, e84831.

Kim H, Goo E, Kang Y, Kim J, Hwang I, 2012. Regulation of universal stress protein genes by quorum sensing and RpoS in Burkholderia glumae. *J Bacteriol* **194**, 982-92.

Kim J, Kang Y, Choi O*, et al.*, 2007. Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in Burkholderia glumae. *Mol Microbiol* **64**, 165-79.

Kim J, Kim JG, Kang Y*, et al.*, 2004. Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in Burkholderia glumae. *Mol Microbiol* **54**, 921-34.

Kovach ME, Elzer PH, Hill DS*, et al.*, 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175-6.

Kutsukake K, 1997. Autogenous and global control of the flagellar master operon, flhD, in Salmonella typhimurium. *Mol Gen Genet* **254**, 440-8.

Lantz MS, 1997. Are bacterial proteases important virulence factors? *J Periodontal Res* **32**, 126-32.

Lebrun I, Marques-Porto R, Pereira AS, Pereira A, Perpetuo EA, 2009. Bacterial toxins: an overview on bacterial proteases and their action as virulence factors. *Mini Rev Med Chem* **9**, 820-8.

Nystrom T, Neidhardt FC, 1992. Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in Escherichia coli. *Mol Microbiol* **6**, 3187-98.

Sandkvist M, 2001. Type II secretion and pathogenesis. *Infect Immun* **69**, 3523-35.

Yoneyama K, Kono Y, Yamaguchi I, Horikoshi M and Hirooka T, 1998. Toxoflavin is an essential factor for virulence of *Burkholderia glumae* causing rice seedling rot disease. *Ann Phytopath Soc Japn* **64**, 91-96.

CHAPTER 5: CONCLUSIONS

Burkholderia glumae causes bacterial panicle blight in rice leading to severe yield losses (Nandakumar et al., 2005; Nandakumar et al., 2009). Several virulence factors contribute to successful infection by *B.* glumae, e.g. toxoflavin, lipase, flagella, and type III secretion system (Ham et al., 2011). TofI/TofR-mediated quorum sensing has been well studied in modulating the production of virulence factors in *B. glumae* (Kim et al., 2004, Jang et al., 2014, An et al., 2014, Devescovi et al., 2007). Recently, our lab identified and characterized the *tepR* gene, encoding a sigma 54-dependent response regulator (Osti, 2014). It was found to negatively regulate toxoflavin and extracellular protease production (Osti, 2014). To gain insights into the comprehensive biological and cellular functions of *tepR* in *B. glumae*, a comparative transcriptomic study was conducted between the wild type *B. glumae* strain 336gr-1 and its *tepR* deletion derivative, LSUPB401. A total of 238 differentially expressed genes (DEGs), defined by a greater than or equal to 1.5-fold change with 99% confidence level, were identified in this study,199 of which were up-regulated and 39 were down-regulated in LSUPB401. As expected, the genes responsible for toxoflavin biosynthesis and transport were consistently up-regulated in LSUPB401. The *tofI*/*tofR*-mediated quorum sensing system, the known positive regulatory system of toxoflavin production, was not differentially expressed in this study. However, the quorum sensing master regulator gene, *qsmR*, was significantly up-regulated in LSUPB401. The fold change of 13 selected genes in RNA-seq were validated by quantitative PCR (qPCR) with a R^2 value of 0.88. The 36 hypothetical protein encoding DEGs were analyzed by the HMMER suite (Eddy, 1998, Mistry et al., 2013) and 25 of them were successfully annotated with putative functions.

Gene Ontology (GO) enrichment analysis of the DEGs indicated that genes responsible for flagella assembly and stress response were consistently up-regulated in LSUPB401. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment further revealed that *tepR* positively regulated the gene cluster encoding a putative type VI secretion system (T6SS), named as T6SS-A. Mauve alignment (Darling et al., 2010) indicated that T6SS-A was homologous to T6SS-6 in *B. pseudomallei*. Phenotypic assays validated that LSUPB401 was significantly more tolerant to heat stress but not stationary phase stress. Also, interspecies competition assays indicated that LSUPB401 largely lost its competition capacity against other bacterial competitors tested, i.e. *Pantoea stewartii* DC283, *Pseudomonas. syringae* pv. *syringae* PssB728a and *P. ananatis* RSPAM1, respectively. Whether the deficiency in competition capacity in LSUPB401 was due to the shutdown of the T6SS-A requires further investigation. Phylogenetic study indicated that TepR homologs are conserved in *Burkholderia* spp. and other plant and animal pathogenic bacteria. It is possible that the modulation of bacterial behavior through the regulatory action of *tepR* is well conserved in other pathogenic bacteria. Finally, it was found that knockout mutation of *qsmR* abolished the production of toxoflavin and extracellular protease in both wild type *B. glumae* strain 336gr-1 and LSUPB401 backgrounds. Thus, the negative regulatory role of *tepR* in toxoflavin and extracellular protease production was dependent upon the positive regulator *qsmR* in *B. glumae*. In summary, this study deciphered the important global regulatory role of *tepR* in *B. glumae*, which may direct to novel biological and chemical disease control strategies to minimize the yield loss due to the bacterial panicle blight in rice.

An JH, Goo E, Kim H, Seo YS, Hwang I, 2014. Bacterial quorum sensing and metabolic slowing in a cooperative population. *Proc Natl Acad Sci U S A* **111**, 14912-7.

Darling AE, Mau B, Perna NT, 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *Plos One* **5**, e11147.

Devescovi G, Bigirimana J, Degrassi G*, et al.*, 2007. Involvement of a quorum-sensing-regulated lipase secreted by a clinical isolate of Burkholderia glumae in severe disease symptoms in rice. *Appl Environ Microbiol* **73**, 4950-8.

Eddy SR, 1998. Profile hidden Markov models. *Bioinformatics* **14**, 755-63.

Jang MS, Goo E, An JH, Kim J, Hwang I, 2014. Quorum sensing controls flagellar morphogenesis in Burkholderia glumae. *Plos One* **9**, e84831.

Kim J, Kim JG, Kang Y*, et al.*, 2004. Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in Burkholderia glumae. *Mol Microbiol* **54**, 921-34.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M, 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**, e121.

Nandakumar R, Rush M, Shahjahan A, O'reilly K, Groth D, 2005. Bacterial panicle blight of rice in the southern United States caused by Burkholderia glumae and B-gladioli. *Phytopathology* **95**, S73-S.

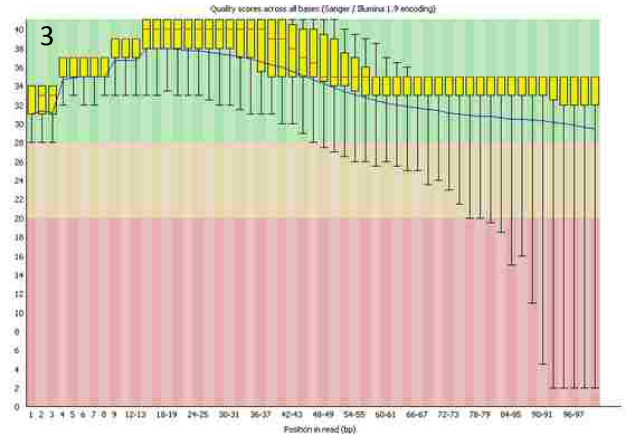
Nandakumar R, Shahjahan AKM, Yuan XL*, et al.*, 2009. Burkholderia glumae and B. gladioli Cause Bacterial Panicle Blight in Rice in the Southern United States. *Plant Disease* **93**, 896-905.

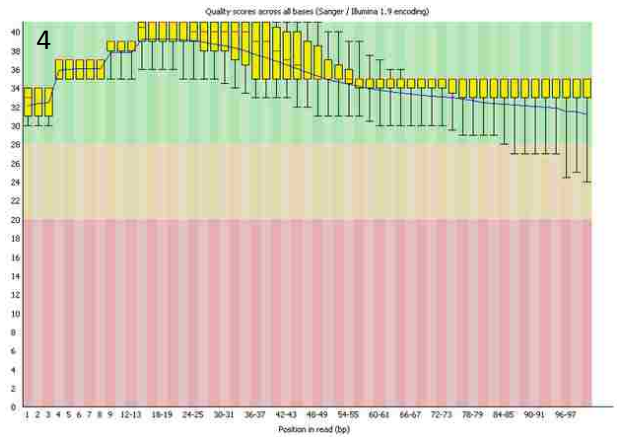
Osti S, Barphagha IK, Ham J, 2013. The novel regulator, tepR, influences toxoflavin production and virulence in the rice pathogenic bacterium Burkholderia glumae. *Phytopathology* **103**, 8-9.

APPENDIX A: PER BASE QUALITY OF RNA-SEQ RAW DATA

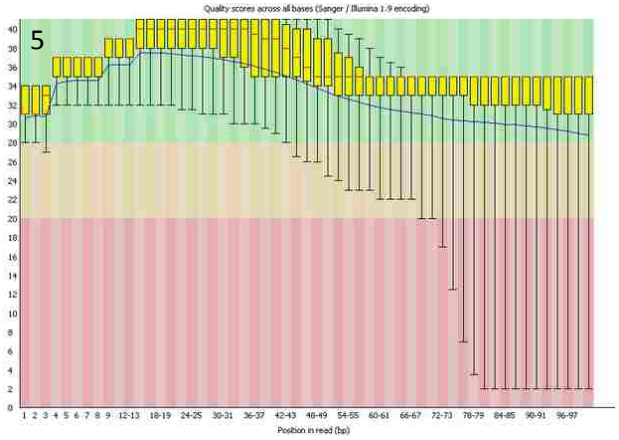
Replicate 1: *B. glumae* 336gr-1 (paired-end reads)

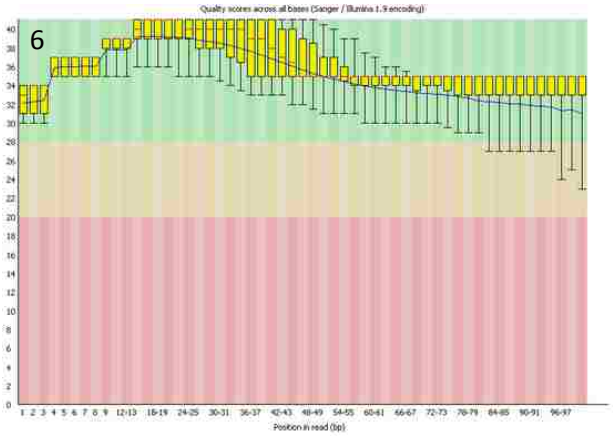
Replicate 2: *B. glumae* 336gr-1 (paired-end reads)



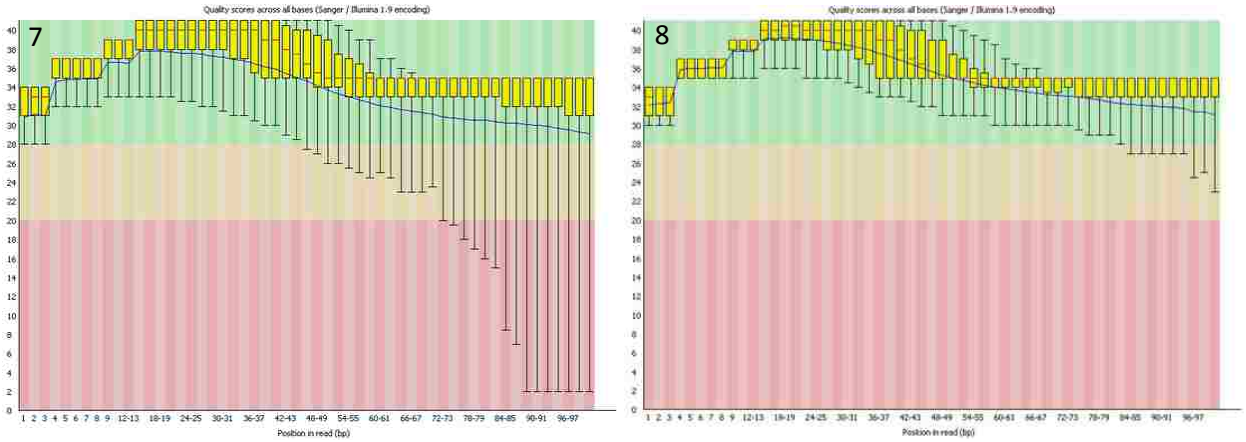


Replicate 3: *B. glumae* 336gr-1 (paired-end reads)

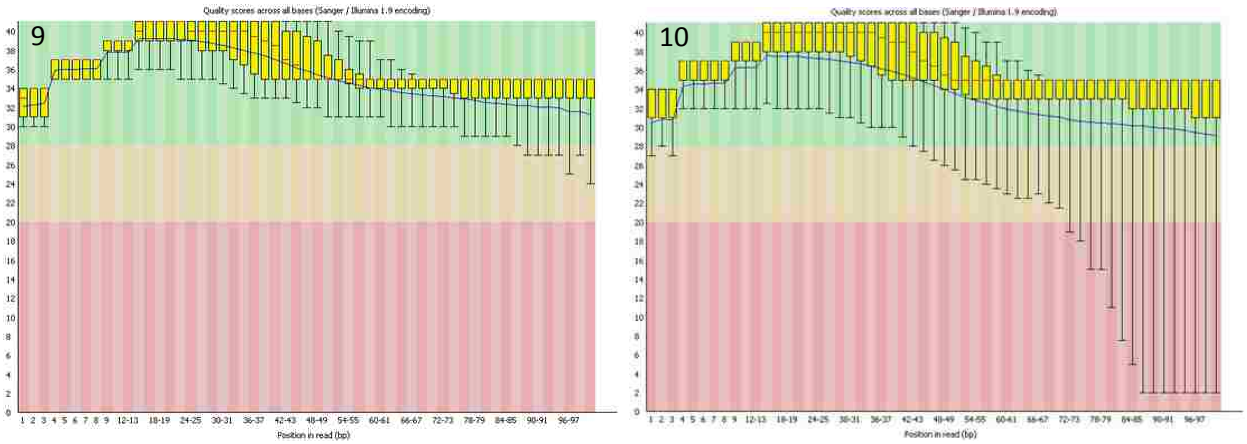




Replicate 1: *B. glumae* LSUPB401 (paired-end reads)



Replicate 2: *B. glumae* LSUPB401 (paired-end reads)



Replicate 3: *B. glumae* LSUPB401 (paired-end reads)

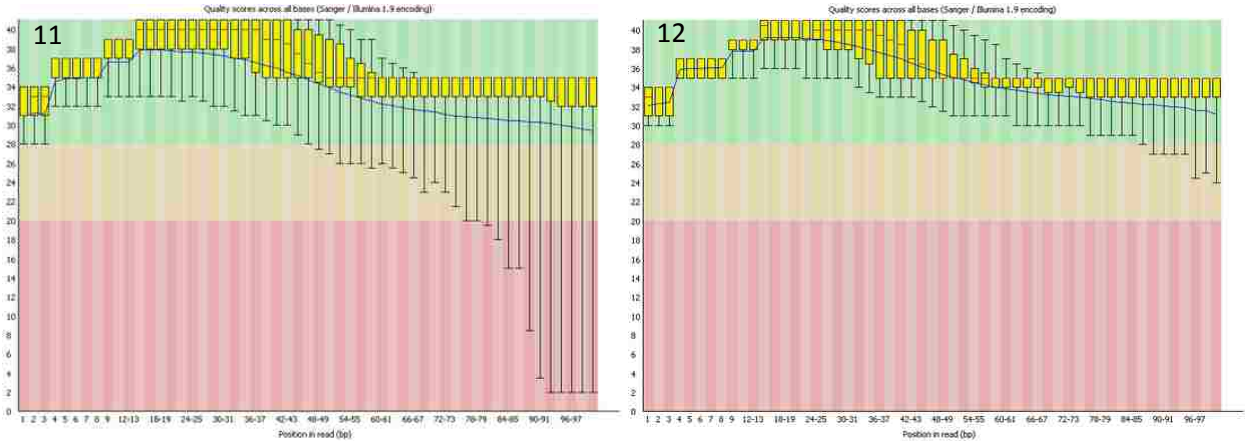


Figure A. The y-axis indicates the quality scores. Higher score means a better base call. The green color in the background indicates calls of very good quality; orange indicates calls of reasonable quality; red indicates calls of poor quality.

APPENDIX B: IDENTIFIED DIFFERENTIALLY EXPRESSED GENES FROM RNA-SEQ

[a]The genes listed are DEGs with at least 1.5 fold expression level change (up-regulated or down-regulated) and at the 99% confidence level or greater in LSUPB401 compared to WT.
[b]WT and LSUPB401 indicate the Reads Per Kilobase per Million mapped reads (RPKM) with upper quartile normalization of the genes in WT and LSUPB401 background respectively.
[c]'up' or 'down' of the fold change indicates that the gene was up-regulated or down-regulated respectively.
[d]q-values were adjusted p-values according to the False Discovery Rate (FDR) (Benjamini Hochberg) method.

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_1g08430	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	34	240	7.06 up	1.04E-172
bglu_2g09850	hypothetical protein	41	230	5.61 up	1.76E-136
bglu_2g19340	chaperonin Cpn10	109	580	5.32 up	2.14E-106
bglu_2g16650	hypothetical protein	190	909	4.78 up	2.83E-84
bglu_2g19330	chaperonin GroEL	56	297	5.3 up	4.95E-84
bglu_1g16590	Serine metalloprotease	28	126	4.5 up	2.87E-73
bglu_1g22650	Catalase	32	146	4.56 up	2.96E-70
bglu_2g06890	TenA family transcriptional activator	23	100	4.35 up	2.08E-67
bglu_2g16720	hypothetical protein	9	44	4.89 up	1.31E-65
bglu_2g16640	hypothetical protein	39	161	4.13 up	4.04E-64
bglu_2g06900	hypothetical protein	8	40	5 up	3.09E-62
bglu_2g16630	hypothetical protein	66	264	4 up	2.31E-61
bglu_2g06440	riboflavin biosynthesis protein RibD	154	686	4.45 up	1.31E-56
bglu_2g16620	hypothetical protein	15	60	4 up	1.43E-52
bglu_1g22640	Ankyrin repeat-containing protein	7	35	5 up	3.42E-52
bglu_2g06880	putative transferase	12	47	3.92 up	8.02E-51
bglu_2g16670	RebA protein	479	2124	4.43 up	2.71E-50
bglu_1g10720	hypothetical protein	19	71	3.74 up	3.20E-48
bglu_2g06430	serine/threonine kinase	220	1000	4.55 up	3.35E-47
bglu_1g10710	peptidoglycan domain-containing protein	70	260	3.71 up	5.74E-44
bglu_2g16700	RebB protein	1228	4350	3.54 up	2.05E-17

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_2g10060	poly (3-hydroxybutyrate) depolymerase	25	91	3.64 up	1.83E-43
bglu_2p0540	8-amino-7-oxononanoate synthase	17	59	3.47 up	1.41E-40
bglu_2g06410	GTP cyclohydrolase II	815	3629	4.45 up	2.82E-37
bglu_2g16660	RebB protein	339	1288	3.8 up	1.49E-36
bglu_2g06420	WD-repeat-containing protein	443	1847	4.17 up	1.87E-32
bglu_2g18450	alpha,alpha-trehalose-phosphate synthase	10	35	3.5 up	7.03E-32
bglu_1g11110	Heat shock protein Hsp20	63	200	3.17 up	1.04E-31
bglu_1g21320	hypothetical protein	55	162	2.95 up	7.51E-28
bglu_2g17550	cyclopropane-fatty-acyl-phospholipid synthase	18	56	3.11 up	3.55E-27
bglu_2g16680	RebB protein	537	1819	3.39 up	6.56E-27
bglu_2g18820	hypothetical protein	32	97	3.03 up	2.37E-26
bglu_2g10055	Expressed protein SR2101	15	47	3.13 up	4.01E-26
bglu_1g19170	acetylornithine aminotransferase	70	212	3.03 up	4.56E-26
bglu_1g22990	Non-ribosomal peptide synthase	32	107	3.34 up	9.53E-26
bglu_1g23010	nonribosomal peptide synthase	23	78	3.39 up	9.53E-26
bglu_1g23100	di-heme cytochrome c peroxidase	20	57	2.85 up	8.26E-25
bglu_1g21920	hypothetical protein	10	29	2.9 up	7.16E-24
bglu_1g11120	Heat shock protein Hsp20	40	112	2.8 up	1.98E-23
bglu_2g02710	Non-ribosomal peptide synthase-like protein	25	67	2.68 up	8.13E-22
bglu_1g23080	acetyl-transferase	70	197	2.81 up	2.33E-21
bglu_1g10250	IclR family regulatory protein	27	74	2.74 up	3.89E-21
bglu_2g09870	major facilitator superfamily protein	11	31	2.82 up	2.70E-20
bglu_1g21390	branched-chain amino acid aminotransferase	11	31	2.82 up	4.15E-20
bglu_2g16690	RebB protein	1774	6317	3.56 up	4.87E-20
bglu_2g09940	transcription factor jumonji	14	39	2.79 up	6.35E-19
bglu_2g17490	putative thiaminase I	18	48	2.67 up	7.32E-19
bglu_1g23090	Amidohydrolase family protein	51	137	2.69 up	1.06E-18
bglu_2g06870	flavin-nucleotide-binding protein	51	135	2.65 up	1.39E-18
bglu_1g23060	hydrolase/decarboxylase	35	92	2.63 up	7.71E-18

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_2g16710	RebB protein	1581	5326	3.37 up	4.33E-17
bglu_2g06910	S-adenosylmethionine decarboxylase	26	65	2.5 up	7.31E-17
bglu_1g16840	Extracellular ligand-binding receptor	14	36	2.57 up	8.18E-17
bglu_1g23020	polyketide synthase	44	141	3.2 up	8.57E-17
bglu_2g15435	Expressed protein GR10	393	1094	2.78 up	1.19E-16
bglu_1g21400	malate/L-lactate dehydrogenase	17	42	2.47 up	2.53E-16
bglu_2g21010	heat shock protein Hsp20	23	59	2.57 up	3.08E-16
bglu_2g09950	metallo-beta-lactamase superfamily protein	30	76	2.53 up	6.87E-16
bglu_2g17500	phosphomethylpyrimidine kinase	41	102	2.49 up	4.83E-15
bglu_1g23040	Long-chain-fatty-acid--CoA ligase	114	295	2.59 up	1.07E-14
bglu_2g09930	hypothetical protein	12	29	2.42 up	1.68E-14
bglu_2g06340	putative transposase	39	92	2.36 up	2.44E-14
bglu_2g17540	MutT/nudix family protein	200	537	2.69 up	3.19E-14
bglu_2g17510	UbiE/COQ5 family methyltransferase	46	111	2.41 up	3.89E-14
bglu_2p0620	putative exported avidin family protein	81	194	2.4 up	8.17E-14
bglu_2g17530	putative nucleoside 2-deoxyribosyltransferase	114	270	2.37 up	9.40E-14
bglu_1g05910	NAD/NADP transhydrogenase alpha subunit-like protein	29	70	2.41 up	1.16E-13
bglu_1g23070	Cysteine synthase	151	387	2.56 up	1.18E-13
bglu_2g05770	ATP-dependent carboxylate-amine ligase domain-containing protein	31	72	2.32 up	7.47E-13
bglu_1g23030	beta-ketoacyl synthase	95	253	2.66 up	1.04E-12
bglu_2g05760	phenazine biosynthesis PhzC/PhzF protein	45	103	2.29 up	1.91E-12
bglu_1g27760	hypothetical protein	22	51	2.32 up	6.27E-12
bglu_1g19160	oxidoreductase-like protein	80	203	2.54 up	9.81E-12
bglu_1g19140	class V aminotransferase	21	45	2.14 up	1.96E-11
bglu_2g06400	putative ubiquinone/menaquinone biosynthesis methyltransferase	3587	14627	4.08 up	2.26E-11
bglu_2g09960	MbtH-like protein	458	1142	2.49 up	2.44E-11

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_2g09860	beta-hydroxylase, aspartyl/asparaginyl family	13	30	2.31 up	2.96E-11
bglu_1g22200	Crystal protein ET79	2069	5524	2.67 up	3.42E-11
bglu_1g23050	pyrroloquinoline quinone (coenzyme pqq) biosynthesis protein c	706	1910	2.71 up	2.15E-10
bglu_1g10660	Cupin 2 barrel domain-containing protein	22	47	2.14 up	2.45E-10
bglu_2g19400	autoinducer-binding domain-containing protein	24	50	2.08 up	2.65E-10
bglu_2g17520	Thymidylate synthase	122	292	2.39 up	4.38E-10
bglu_1g08680	hypothetical protein	64	134	2.09 up	5.58E-10
bglu_2g05750	LmbE family protein	96	203	2.11 up	6.02E-10
bglu_2g19390	CAIB/BAIF family protein CoA transferase	18	37	2.06 up	9.51E-10
bglu_2g06370	putative RND efflux membrane-fusion protein	145	319	2.2 up	1.49E-09
bglu_2g21000	heat shock protein Hsp20	33	70	2.12 up	2.22E-09
bglu_1g19110	peptidase C26	81	180	2.22 up	3.72E-09
bglu_2g07570	PPE-repeat-containing protein	55	113	2.05 up	6.34E-09
bglu_1g30320	cytochrome c4	29	57	1.97 up	9.57E-09
bglu_2g18780	transferase	233	531	2.28 up	1.11E-08
bglu_1g16170	RND family efflux transporter MFP subunit	25	51	2.04 up	1.33E-08
bglu_1g33570	flagellar hook-length control protein FliK	16	33	2.06 up	1.37E-08
bglu_1g19130	Nucleoside diphosphate kinase	13	28	2.15 up	2.22E-08
bglu_1g21410	lysine2,3-aminomutase	43	85	1.98 up	3.14E-08
bglu_1g21360	beta-lactamase domain-containing protein	15	31	2.07 up	4.39E-08
bglu_1g19190	hypothetical protein	186	363	1.95 up	6.91E-08
bglu_1g26310	hypothetical protein	106	207	1.95 up	7.96E-08
bglu_1g33550	flagellum-specific ATP synthase FliI	25	49	1.96 up	8.84E-08
bglu_1g16930	Thiotemplate mechanism natural product synthetase	18	35	1.94 up	9.59E-08
bglu_1g33090	flagellar rod assembly protein/muramidase FlgJ	47	92	1.96 up	1.14E-07

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_2g02690	amino acid adenylation domain-containing protein	44	86	1.95 up	1.17E-07
bglu_2g02700	peptide synthetase	78	148	1.9 up	1.21E-07
bglu_2g17085	Expressed protein GR12	134	261	1.95 up	1.85E-07
bglu_1g16910	TubD protein	15	30	2 up	2.27E-07
bglu_2g18430	heat shock protein Hsp20	123	236	1.92 up	3.14E-07
bglu_2g20990	heat shock protein Hsp20	30	57	1.9 up	3.72E-07
bglu_2g18440	heat shock protein Hsp20	334	691	2.07 up	3.98E-07
bglu_1g33130	flagellar basal body rod protein FlgF	75	143	1.91 up	4.29E-07
bglu_1g21370	N-oxygenase	18	33	1.83 up	1.36E-06
bglu_2g05800	Threonine dehydrogenase	18	32	1.78 up	2.33E-06
bglu_2g11560	AhpD family protein alkylhydroperoxidase like protein	24	45	1.88 up	3.05E-06
bglu_2g05730	putative coenzyme PQQ synthesis protein c	131	266	2.03 up	3.67E-06
bglu_1g33110	flagellar basal body L-ring protein	32	57	1.78 up	4.12E-06
bglu_1g19040	hypothetical protein	68	122	1.79 up	5.79E-06
bglu_1g26810	Chorismate lyase family protein	27	49	1.81 up	5.96E-06
bglu_2g06350	putative outer membrane protein OprM	191	441	2.31 up	8.96E-06
bglu_1g33100	flagellar basal body P-ring protein	29	52	1.79 up	9.84E-06
bglu_1g01990	S-adenosyl-L-homocysteine hydrolase	529	1089	2.06 up	1.04E-05
bglu_2g01190	taurine catabolism dioxygenase tauD/tfdA	42	75	1.79 up	1.36E-05
bglu_1g33120	flagellar basal body rod protein FlgG	89	158	1.78 up	1.61E-05
bglu_1g33760	ATP-dependent protease subunit HslV	118	207	1.75 up	2.01E-05
bglu_2g05740	putative Branched-chain amino acid aminotransferase	111	216	1.95 up	2.23E-05
bglu_1g23190	LuxR family transcriptional regulator	227	415	1.83 up	2.24E-05
bglu_2g11820	glycoside hydrolase	63	110	1.75 up	2.65E-05
bglu_1g19050	N,N-dimethylformamidase large subunit	62	118	1.9 up	2.95E-05
bglu_1g26800	Uracil-DNA glycosylase family	31	55	1.77 up	3.38E-05

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_1g19060	acyl-CoA-ligase/synthetase	75	144	1.92 up	4.25E-05
bglu_2g06380	hypothetical protein	154	288	1.87 up	4.68E-05
bglu_1g23150	Auxin-binding protein	195	371	1.9 up	5.73E-05
bglu_1g33770	ATP-dependent protease ATP-binding subunit HslU	92	175	1.9 up	5.74E-05
bglu_2g05720	MhpE-like protein	94	172	1.83 up	6.26E-05
bglu_1g10430	thiamine biosynthesis protein ThiC	73	136	1.86 up	7.22E-05
bglu_2g18630	Insertion element IS402	409	710	1.74 up	7.85E-05
bglu_1g23170	ABC transporter ATP-binding protein	507	1078	2.13 up	9.18E-05
bglu_2g16920	CsbD-like protein	110	181	1.65 up	9.99E-05
bglu_1g33150	flagellar basal body rod modification protein	104	185	1.78 up	1.01E-04
bglu_1g26010	carbonate dehydratase	41	70	1.71 up	1.24E-04
bglu_2g08910	hypothetical protein	61	101	1.66 up	1.90E-04
bglu_2g06360	putative RND efflux transporter	211	414	1.96 up	1.93E-04
bglu_1g14230	ATP-dependent Clp protease, ATP-binding subunit ClpB	131	274	2.09 up	2.05E-04
bglu_1g33180	flagellar basal body P-ring formation protein FlgA	42	71	1.69 up	2.05E-04
bglu_1g23120	hypothetical protein	156	269	1.72 up	2.22E-04
bglu_1g26820	heat shock protein 90	143	290	2.03 up	2.48E-04
bglu_1g26710	cytochrome c family protein	66	105	1.59 up	2.63E-04
bglu_1g19080	Nucleoside diphosphate kinase	168	297	1.77 up	3.07E-04
bglu_1g23230	Phytanoyl-CoA dioxygenase	79	129	1.63 up	3.47E-04
bglu_2g09660	hypothetical protein	139	227	1.63 up	3.82E-04
bglu_1g06350	chaperone protein DnaJ	166	277	1.67 up	4.09E-04
bglu_2g13340	Cyanate transporter	50	82	1.64 up	4.14E-04
bglu_1g31060	hypothetical protein	18	29	1.61 up	4.75E-04
bglu_1g19210	Non-ribosomal peptide synthetase	61	102	1.67 up	4.84E-04
bglu_1g23160	efflux ABC transporter permease	279	560	2.01 up	5.04E-04
bglu_1g19230	metallophosphoesterase	216	358	1.66 up	5.04E-04
bglu_1g32670	Sodium/bile acid symporter family protein	78	131	1.68 up	6.37E-04
bglu_2g18640	non-ribosomal peptide synthase	149	260	1.74 up	6.87E-04
bglu_1g19220	multidrug efflux protein	69	122	1.77 up	6.87E-04

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_1g19250	nucleoside-diphosphate-sugar epimerase	147	253	1.72 up	8.99E-04
bglu_1g18240	hypothetical protein	190	318	1.67 up	9.48E-04
bglu_2g21670	L-carnitine dehydratase/bile acid-inducible protein F	21	32	1.52 up	0.001037733
bglu_1g33560	flagellar export protein FliJ	44	68	1.55 up	0.001043746
bglu_1g19240	CDP-alcohol phosphatidyltransferase	54	87	1.61 up	0.001052373
bglu_1g30010	hypothetical protein	62	95	1.53 up	0.001068033
bglu_2g13680	alkyl hydroperoxide reductase subunit F	180	351	1.95 up	0.001130729
bglu_1g11560	Serine carboxypeptidase family protein	28	45	1.61 up	0.001169927
bglu_1g19200	Nucleoside diphosphate kinase	62	98	1.58 up	0.001186519
bglu_2g05270	hypothetical protein	27	42	1.56 up	0.0015844
bglu_2g09380	putative lipoprotein transmembrane	40	63	1.58 up	0.001751673
bglu_1g33160	flagellar basal body rod protein FlgC	77	121	1.57 up	0.002169789
bglu_1g22630	hypothetical protein	22	32	1.45 up	0.002243924
bglu_1g33170	flagellar basal body rod protein FlgB	83	128	1.54 up	0.002528843
bglu_2g05710	chitinase	20	29	1.45 up	0.003212234
bglu_1g16950	thioesterase II	24	36	1.5 up	0.004290945
bglu_1g24640	FeS assembly protein IscX	127	187	1.47 up	0.004417579
bglu_2g02050	TetR family transcriptional regulator	77	117	1.52 up	0.004922734
bglu_1g33140	flagellar hook protein FlgE	66	104	1.58 up	0.004944054
bglu_1g19070	hypothetical protein	247	454	1.84 up	0.005076145
bglu_1g27700	lipoprotein	23	35	1.52 up	0.005555788
bglu_1g25590	hypothetical protein	1311	2026	1.55 up	0.005735745
bglu_2g16410	xylose isomerase domain-containing protein	130	200	1.54 up	0.006338905

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_1g23110	translation initiation inhibitor	239	386	1.62 up	0.006338905
bglu_1g19260	PA-phosphatase-like protein phosphoesterase	38	56	1.47 up	0.007103448
bglu_2g14960	TonB-dependent receptor	35	54	1.54 up	0.007209991
bglu_2g22070	OprD family outer membrane porin	44	65	1.48 up	0.008524613
bglu_1g19150	aminoglycoside phosphotransferase	26	37	1.42 up	0.008556779
bglu_1g23130	hypothetical protein	573	1022	1.78 up	0.009676573
bglu_1g21330	glutathione S-transferase	83	121	1.46 up	0.009975398
bglu_1g05920	NAD(P) transhydrogenase subunit beta	20	32	1.6 up	0.001461743
bglu_1g03180	thiazole synthase	38	64	1.68 up	1.26E-04
bglu_1g02010	5,10-methylenetetrahydrofolate reductase	92	176	1.91 up	2.81E-06
bglu_1g02000	hypothetical protein	177	286	1.62 up	4.43E-04
bglu_1g01980	flagellar biosynthesis sigma factor	21	34	1.62 up	4.52E-04
bglu_1g01970	flagellar biosynthesis protein FlhG	29	48	1.66 up	8.10E-05
bglu_1g01960	flagellar biosynthesis regulator FlhF	21	40	1.9 up	5.70E-07
bglu_1g01950	flagellar biosynthesis protein FlhA	32	49	1.53 up	0.004591879
bglu_1g01690	hypothetical protein	87	143	1.64 up	3.32E-04
bglu_1g01620	acetyltransferase	19	31	1.63 up	1.75E-04
bglu_1g00180	flagellar biosynthesis protein FliQ	41	62	1.51 up	0.006795091
bglu_2g22280	porin	190	321	1.69 up	0.002045894
bglu_4p0300	hypothetical protein	58	0	N/A	0
bglu_4p0370	bacteriophage protein gp37	49	0	N/A	0
bglu_3p0320	bacteriophage protein gp37	45	0	N/A	0
bglu_1g09700	Sigma-54 dependent DNA-binding response regulator	61	7	8.71 down	1.03E-282
bglu_1g03890	type VI secretion system protein TssB	644	112	5.75 down	5.74E-85
bglu_1g03880	TPR repeat-containing protein	801	137	5.85 down	4.55E-74

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_1g04010	hypothetical protein	114	21	5.43 down	6.72E-73
bglu_1g03970	OmpA/MotB	54	12	4.5 down	6.73E-52
bglu_1g03930	type VI secretion system protein TssF	61	13	4.69 down	5.25E-48
bglu_1g03920	type VI secretion system protein TssE	200	45	4.44 down	1.05E-46
bglu_1g22700	LuxR family transcriptional regulator	68	17	4 down	2.98E-45
bglu_1g03900	type VI secretion system protein TssC	856	166	5.16 down	2.23E-43
bglu_1g03940	type VI secretion system protein TssG	99	23	4.3 down	2.85E-43
bglu_1g03960	type VI secretion system protein TssA	34	9	3.78 down	4.98E-38
bglu_1g04000	peptidase M15B and M15C, D,D-carboxypeptidase VanY/endolysins	290	70	4.14 down	3.90E-37
bglu_1g03990	type VI secretion protein IcmF	188	41	4.59 down	4.37E-35
bglu_1g03950	type VI secretion system protein TssH	158	37	4.27 down	6.63E-35
bglu_1g03980	type VI secretion-associated protein	89	22	4.05 down	4.58E-33
bglu_1g03910	type VI secretion system protein TssD	3405	642	5.3 down	1.80E-31
bglu_4p0610	hypothetical protein	1293	474	2.73 down	2.59E-09
bglu_2g10840	putative LysM domain-containing protein	61	25	2.44 down	3.83E-09
bglu_1g04140	Cold shock-like protein CspD	11195	3653	3.06 down	3.34E-08
bglu_2g10830	Rhs element Vgr protein	34	15	2.27 down	1.46E-07
bglu_3p0560	hypothetical protein	1874	725	2.58 down	8.38E-07
bglu_1g03870	type VI secretion system protein TssJ	171	81	2.11 down	1.34E-05
bglu_1g10060	histidine ABC transporter permease HisQ	109	53	2.06 down	2.25E-04
bglu_1g28620	Branched-chain amino acid ABC transporter ATP-binding protein	144	73	1.97 down	2.26E-04

(Appendix B continued)

Gene[a]	Product	WT[a]	LSUPB401[b]	Fold change[c]	q-value[d]
bglu_1g26230	Small multidrug resistance protein	61	33	1.85 down	4.08E-04
bglu_2g15080	translation initiation factor IF-1	1396	671	2.08 down	6.58E-04
bglu_2g04360	monosaccharide-transporting ATPase	99	53	1.87 down	0.001426
bglu_2g04350	Periplasmic binding protein/LacI transcriptional regulator	300	150	2 down	0.0014617
bglu_1g28610	ABC transporter-like protein	338	173	1.95 down	0.0025267
bglu_1g31460	hypothetical protein	1246	635	1.96 down	0.0025288
bglu_2g22250	carbon starvation protein CstA	59	31	1.9 down	0.0030143
bglu_1g28630	Inner-membrane translocator	248	130	1.91 down	0.0047146
bglu_1g31950	C4-dicarboxylate transporter DctA	220	116	1.9 down	0.0053633
bglu_2g22240	hypothetical protein	77	46	1.67 down	0.0055526
bglu_2g15090	Cold shock protein	6407	2972	2.16 down	0.0073716
bglu_1g28640	branched chain amino acid family ABC transporter inner membrane subunit	166	88	1.89 down	0.0101435

APPENDIX C: STATISTICS ANALYSES OF WHOLE TRANSCRIPTOME GENE ONTOLOGY ANNOTATION

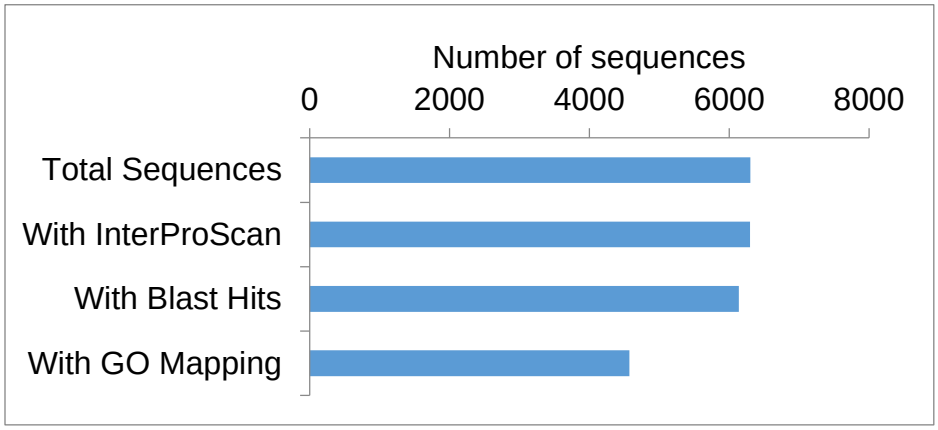


Figure C.1. Schematic of whole transcriptome GO annotation of *B. glumae* by Blast2GO.

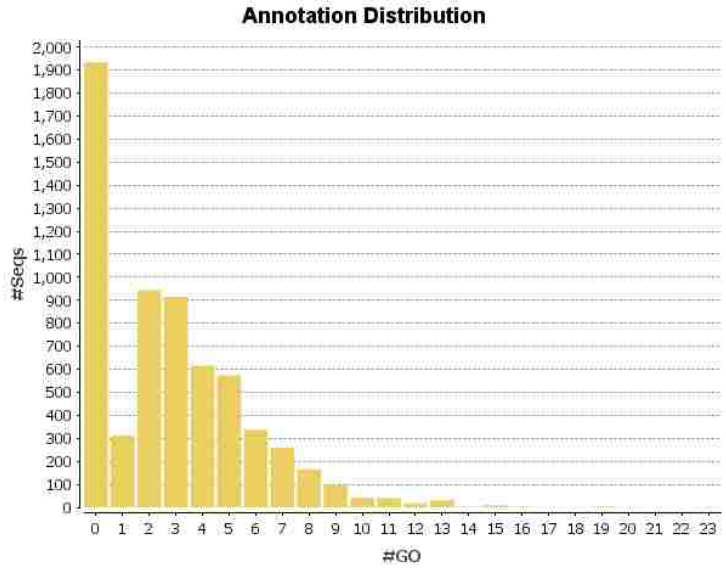


Figure C.2. GO annotation distribution of whole transcriptome of *B. glumae*. Number of GO annotation of sequences (genes) is displayed in x-axis. Number of sequences (genes) is displayed in y-axis.

VITA

Jingyu Peng was born in Yantai, China on September, 1990. He received his Bachelor's degree in Shandong Agricultural University majoring in Plant Protection. He was interested in applying molecular genetics to develop novel plant disease control strategies. To achieve his goal, he started to pursue his Master's degree in Dr. Jong Hyun Ham's lab in the Department of Plant Pathology and Crop Physiology at Louisiana State University in Baton Rouge, Louisiana in August, 2013. His research focused on the genetic and transcriptomic study of the rice pathogenic bacterium *Burkholderia glumae*. His thesis title is "Genetic and transcriptomic analyses of the rice pathogenic bacterium, Burkholderia glumae, reveal the important roles of the regulatory gene, tepR, for bacterial survival in environmental stresses". He received the travel award from Department of Plant Pathology and Crop Physiology Graduate Student Association to attend the 2015 American Phytopathological Society Annual Meeting. He hopes to receive his Master's degree in Plant Health (concentration: Plant Pathology) in December 2015.