PATHOGENIC AND GENOMIC CHARACTERIZATION OF *Burkholderia* SPECIES CAUSING RICE BACTERIAL PANICLE BLIGHT IN PENINSULAR MALAYSIA

By

ADAM ZAFDRI BIN MD ZALI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

August 2023

FP 2023 23

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

PATHOGENIC AND GENOMIC CHARACTERIZATION OF *Burkholderia* SPECIES CAUSING RICE BACTERIAL PANICLE BLIGHT IN PENINSULAR MALAYSIA

By

ADAM ZAFDRI BIN MD ZALI

August 2023

Chairman : Dzarifah binti Mohamed Zulperi, PhD
Faculty : Agriculture

Rice (*Oryza sativa*) is a vital crop globally, contributing significantly to Malaysia's GDP through local and international markets. However, bacterial panicle blight (BPB), caused by *Burkholderia glumae* and *Burkholderia gladioli*, poses a major threat to rice quality and production. BPB symptoms include straw-colored panicles with florets displaying darker bases and reddish-brown edges, leading to kernel abortion and yield losses of up to 75% in severely affected areas. Climate factors, such as rising temperatures and humidity, exacerbate the disease's spread, making it a critical concern for Malaysia's rice industry.

This study aimed to (i) isolate and identify *Burkholderia spp.* associated with BPB in Peninsular Malaysia, (ii) assess their pathogenicity on MR219 rice, and

(iii) analyze their genomes using bioinformatics. From June 2021 to January 2022, 27 isolates were collected from symptomatic rice panicles.

Phenotypic characterization confirmed they were Gram-negative rodproducing diffusible yellow pigments. Molecular analyses identified 22 isolates as *B. glumae* and 5 as *B. gladioli*. Pathogenicity tests revealed varying virulence levels, with isolates BGS.AS1, UPMBG7, and UPMBG8 being most virulent.

Whole-genome sequencing of four strains revealed high alignment rates to reference strains, with B. glumae coding 5,453 to5,641 genes and *B. gladioli* coding 6,831 to 6,967 genes. Gene annotations identified pathways for flagella assembly and quorum sensing. Genome deposition on NCBI can be retrieved using accession numbers of JAMYCS000000000 (K6 strains), JAMYCT000000000 (P3 strains), JANIEE000000000 (UPMBG7 strains) and JANIEF000000000 (UPMBG8 strains). These findings provide insights into BPB pathogens, supporting improved management strategies to mitigate disease impacts and enhance rice production in Malaysia.

Keywords: Bacterial panicle blight, *Burkholderia gladioli*, *Burkholderia glumae*, pathogenic characterization, whole genome sequence.

SDG: GOAL 13: Climate action, GOAL 9: Industry innovation and infrastructure, GOAL 12: Responsible consumption and production

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master sains

PENCIRIAN PATOGENIK DAN GENOMIK SPESIES *Burkholderia* MENYEBABKAN HAWA BULIR BAKTERIA PADI DI SEMENANJUNG MALAYSIA

Oleh

ADAM ZAFDRI BIN MD ZALI

Ogos 2023

Pengerusi : Dzarifah binti Mohamed Zulperi, PhD
Fakulti : Pertanian

Rice (*Oryza sativa*) adalah tanaman penting di dunia, termasuk di Malaysia, di mana ia menyumbang kepada ekonomi negara. Namun, *Burkholderia glumae (B. glumae*) dan *Burkholderia gladioli (B. gladioli)* merupakan penyebab hawar bulir bakteria pada padi (BPB), yang mengancam kualiti dan hasil padi. Penyakit ini ditandai oleh bulir jerami dengan lekukan gelap di pangkal bunga dan pinggiran coklat kemerahan, menyebabkan biji padi gugur sebelum matang. BPB menyebabkan penurunan hasil sehingga 75% di kawasan terjejas. Perubahan iklim, termasuk peningkatan suhu global, memperburuk penyebaran BPB.

Kajian ini bertujuan untuk (i) mengasingkan dan mengenal pasti *Burkholderia spp.* yang berkaitan dengan BPB di Semenanjung Malaysia, (ii) menilai tahap kepatogenannya pada padi varieti MR219, dan (iii) menganalisis genomnya

menggunakan bioinformatik. Dari Jun 2021 hingga Januari 2022, sebanyak 27 isolat telah dikumpulkan daripada tangkai padi menunjukkan simptom penyakit.

Pencirian fenotip mengesahkan bahawa mereka adalah batang Gram-negatif yang menghasilkan pigmen kuning. Analisis molekular mengenal pasti 22 isolat sebagai *B. glumae* dan 5 sebagai *B. gladioli*. Ujian patogenisiti mendedahkan tahap virulens yang berbeza, isolat BGS.AS1, UPMBG7, dan UPMBG8 menunjukkan virulens yang paling tinggi.2.

Penjujukan genom keseluruhan bagi empat strain menunjukkan kadar penjajaran tinggi dengan strain rujukan, di mana *B. glumae* mengandungi 5,453 hingga 5,641 gen, manakala *B. gladioli* mengandungi 6,831 hingga 6,967 gen. Penjelasan gen mengenal pasti laluan bagi pemasangan flagella dan quorum sensing. Pemendapan genom di NCBI boleh diperoleh menggunakan nombor akses berikut: JAMYCS000000000 (strain K6), JAMYCT000000000 (strain P3), JANIEE000000000 (strain UPMBG7) dan JANIEF000000000 (strain UPMBG8). Penemuan ini memberikan pandangan baru tentang patogen BPB, menyokong strategi pengurusan yang lebih baik untuk mengurangkan kesan penyakit dan meningkatkan pengeluaran padi di Malaysia.

Kata Kunci: Bakteria panicle blight; *Burkholderia gladioli*; Burkholderia glumae; Ciri-ciri patogenik; Penjujukan genom keseluruhan

SDG: MATLAMAT 13: Mitigasi iklim; MATLAMAT 9: Industri, inovasi dan infrastruktur; MATLAMAT 12: Penggunaan dan pengeluaran yang bertanggungjawab.

ACKNOWLEDGEMENTS

Alhamdulillah, all praises Allah S.W.T for the strength, patience, guidance, and blessing in completing this thesis. It would not have been possible to finish writing this MSc thesis without the help and support of the kind people around me, especially my parent, supervisor, friends and UPM staff.

First and foremost, I sincerely thank Assoc. Prof. Dr. Dzarifah Mohamed Zulperi, my research supervisor and mentor from the Faculty of Agriculture, for sharing expertise, advice, encouragement and patience throughout my thesis completion. I also would like to thank Dr. Muhammad Asyraf Md Hatta and Prof. Dr. Mohd Razi Ismail for their research assistance, guidance and support. I am grateful to the School of Graduate Studies of UPM for funding me with a Graduate Research Fellowship.

I also sincerely thank Dr. Erneeza Mohamad Hatta for her continuous guidance, motivation, enthusiasm, and immense knowledge. I am also really thankful to Plant Protection Department laboratory staff, Ms. Asmalina Abu Bakar, and Mr. Johari Mohd. Sarikat, Ms. Mastura and Mr. Mohamed Nazri Abdul Rahman for their guidance during my laboratory work.

Moreover, I would like to thank all of my department colleagues and laboratorymates fromBacteriologyLab, Molecular BiologyLab, Ms. Yasmin, Mr. Iqbal, Ms. Arisa and especially my friends for their tireless support in allowing me to complete my research.

Special appreciation towards my parents Mr. Md Zali Abu and Mrs. Rozita Aladdin, My wife, Nurfatihah Arinazzani and my family for all the prayers and support which help this research runs smoothly. Last, I would like to express my thank you to everyone who have help me and support me.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Dzarifah binti Mohamed Zulperi, PhD
Associate Professor
Faculty of Agriculture
Universiti Putra Malaysia
(Chairman)

Mohd Razi bin Ismail, PhD
Associate Professor
Faculty of Agriculture
Universiti Putra Malaysia
(Member)

Muhammad Asyraf bin Md Hatta, PhD
Senior Lecturer
Faculty of Agriculture
Universiti Putra Malaysia
(Member)

_____
ZALILAH MOHD SHARIFF, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 7 November 2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%	Percentage
µL	Microlitre
°C	degree Celsius
ANOVA	Analysis of Variance
Bp	base pair
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal DNA
FAO	Food and Agriculture Organization
OECD	Organization for Economic Cooperation and Development
MT	Million tonnes
H	Hour
Mb	Megabase pair
Kg	Kilogram
v/v	Volume/volume percentage
mL	Milliliter
min	Minutes
sec	Seconds
TBE	tris-Borate-EDTA
NB	nutrient broth
UV	ultra-violet
V	voltan/volt
PE	Paired-end
INS	Insertion
DEL	Deletion
ITX	Intra-chromosomal translocation
CTX	Inter-chromosomal translocation
SV	Structural Variation
CNV	Copy number variation
QS	Quorum sensing

CHAPTER 1

INTRODUCTION

1.1 Background of study

Rice is the world's second most important crop after wheat. Local rice production should be expanded to ensure the country's rice supply is sufficient and escalate the export of rice production as only 7% of rice production from the country of origin is traded. The National Agro-Food Policy (2011-2020) and Eleventh Malaysian Plan (2016-2020) proposed that local rice production should be increased to ensure and meet the country's demand in the future due to the increasing population in Malaysia (Rajamoorthy *et al.,* 2015). Tan (2019) states that Malaysia produced 2.60 million tonnes of rice in 2013, 1.83 million tonnes in 2014, 2.74 million in 2015 and 2016, 2.57 million tonnes in 2017 and 2.64 million tonnes in 2018. According to the Department of Statistics Malaysia, rice production in Malaysia accumulated to approximately 2.35 and 2.34 million tonnes in 2019 and 2020, a significant decrease in yield compared to 2017 and 2018.

This fluctuation rate of rice production in Malaysia is due to weather conditions, pests and disease outbreaks, especially the rice bacterial panicle blight (BPB) (Ramachandran *et al.,* 2021). This disease has been a severe problem for rice production for many years in rice-growing regions such as Malaysia, Japan, the United States and other countries (Peng, 2015). Initially, panicle blight was considered a physiological disorder of rice, but later, it was identified as caused

by *B. glumae* species (Nandakumar *et al*., 2009). BPB was first detected in Malaysia in Sungai Ache, Penang and Kg. Banir, Kelantan during2018 and 2019. This disease records up to 15 to 75% of rice yield losses in the severe field infected by B. glumae (Ramachandran *et al*., 2021). The symptom was described as a straw-coloured panicle having florets with a darker base and a reddish-brown margin around the floret, resulting in kernel abortion before it fills. In contrast, severely infected panicle often stays upright rather than bent over due to poor filling. Aside from that, sheath rot symptom is distinguished by long, vertical greyish lesions surrounded by a dark reddish-brown border (Zhou-qi *et al*., 2016). Several practices were applied to control the disease, including chemical control such as oxolinic acid, cocodrie, application of antibiotics such as ceftazidime, meropenem ciprofloxacin and also heat treatment of rice seed (Weny *et al*., 2019).

Burkholderia spp. is a rod-shaped, aerobic, motile and Gram-negative seed-borne pathogen that causes rice panicle blight disease (Gao et al., 2015). According to Peng (2015), the overall genome size of *B. glumae* is more than 8Mbp comprises two chromosomes and four plasmids with a total of 6305 genes present in the genome, and 68.2%of the genome is GC content. Meanwhile, the genome size for *B. gladioli* is around 9.05Mbp with an average of 7216 coding gene present (Seo *et al*., 2011). The pathogenesis of *Burkholderia* spp. is a complex mechanism involving numerous virulence factors primarily regulated by a global-regulatory quorum sensing (QS) system mediated by TofI andTofR, which are homologues to LuxI and LuxR, respectively (Peng, 2015). Toxoflavin (host- nonspecific phytotoxin), lipase

formation, is the pathogen's most significant virulence factor (Kim *et al*., 2013, Kim *et al*., 2014, Gao *et al*., 2015). The optimum growth range temperature for *Burkholderia* spp. is comparatively high at 30 °C to 35 °C. This disease is thought to be more common in tropical and semi-tropical countries and during growing seasons with higher-than-average temperatures. With appropriate environmental conditions, a severe epidemic of panicle blight disease in rice can spread and increase rapidly (Zhou-qi *et al*., 2016).

1.2 Problem Statement

Rice is known as one of the world's most valuable food crops. The spread of a rice disease will be detrimental to the global economy. The epidemic of the emerging disease panicle blight in rice caused by *Burkholderia* spp. has become even more critical in the rice industry, as the disease is one of the most severe rice diseases in Malaysia. The immediate agronomic effects of such an incident would be a significant decrease in crop yield, production stability and economic activities. In Malaysia, BPB disease can decrease rice yield by up to 50% in seriously infected areas due to grain weight loss, floret sterility, seed germination inhibition, stand reduction, and year-to-year dissemination due to the pathogen's seed-borne existence (Ramachandran *et al.,* 2021). BPB decreases yield and rice quality to suppliers, leading to rice supply disruptions and causing suppliers to increase rice prices due to high consumer demand. Previous reports from other countries have shown that *Burkholderia* spp. is the causal agent of BPB in rice, but the study of *Burkholderia* spp. disease severity on Malaysian rice cultivars had not been done at that time.

Thus, the study was useful for comparing and contrasting the similarities and differences among different strains or pathovars of *Burkholderia* spp., as well as understanding the evolution and diversity of these bacteria and their adaptation and co-evolution with their hosts. Moreover, BPB may significantly affect the international rice industry, especially in rice-producing countries like Malaysia (Nilsson, 2020). Reduction in rice production caused by BPB cannot be mitigated by any pesticide currently used in rice cultivation. Climate conditions such as extended elevated daily minimum temperature and heavy rainfall during the panicle emergence and flowering times of rice which increases relative humidity levels influenced the disease occurrences. As global temperatures rise, BPB may become more widespread and economically disastrous in rice-dependent countries. According to Goo *et al.* (2015), *Burkholderia* spp. can rapidly rearrange the genome in response to the interaction with the host, leading to genetically/pathogenetically diverse strains of this pathogen. However, Zhou-qi *et al*. (2016) pointed out that this host-pathogen interaction may relate to environmental factors such as temperature and humidity, as different rice-production regions have distinct local rice cultivars and *Burkholderia* strains. To date, although the whole genomes of *B. glumae* and *B. gladioli* from other countries have been sequenced and published, there is no documented record of a fully annotated whole genome sequence of *B. glumae* from Malaysia.

This study aims to fill this gap by providing a detailed documentation of the genome sequence, gene prediction and annotation of *Burkholderia* spp. that

cause BPB on rice from a Malaysian strain, using next-generation sequencing approaches.

1.3 Significance of Study

BPB is among the most devastating rice diseases that significantly impact rice production worldwide, owing to the lack of a control measure and its causal pathogen. This research aimed to explore *Burkholderia* spp.'s virulence factors based on annotation and gene prediction of the bacterium, structural variations within this pathogenic bacterium through mapping alignment approaches and determine *Burkholderia* spp. degree of virulency through greenhouse assay. The findings will establish the essential foundation for efficient BPB disease management strategies, improving Malaysia's rice production industry

1.4 Objectives of Study

The objectives of this study were:

1. To isolateand identify *Burkholderia* spp. associated with BPB disease of rice in Peninsular Malaysia using phenotypic and molecular characterization.

2. To determine the pathogenicity of *Burkholderia* spp. via pathogenicity assessment on MR219 rice cultivar.

3. To elucidate the structural variations, gene predictions and gene annotations of *Burkholderia* spp. via whole-genome analysis and bioinformatics pipeline

CHAPTER 2

LITERATURE REVIEW

2.1 Rice (*Oryza sativa*)

2.1.1 History of Rice Cultivation in Malaysia

Oryza sativa, commonly known as rice, occupies a well-defined place in the taxonomy of the plant kingdom. It belongs to the Kingdom Plantae, which encompasses all plants. Within the phylum Angiosperms, rice stands as a representative of flowering plants that produce seeds enclosed within fruits. Further classification places it in the class Monocots, distinguished by characteristics such as single cotyledons and parallel leaf venation. The order Poales encompasses rice, as it is a cereal crop. At the family level, *Oryza sativa* falls under Poaceae, also known as the grass family, highlighting its close botanical relationship with other important cereal crops. Within the genus *Oryza*, it holds the species designation *Oryza sativa*, representing the cultivated rice species that has been selectively bred and cultivated for millennia, serving as a fundamental staple food source for much of the world's population. Understanding *Oryza sativa*'s taxonomy is critical for both botanical classification and its practical applications in agriculture and food security (Table 2.1).

Table 2.1: Rice Taxonomy Classification from Kingdom to Species

Name	Rice
Kingdom	Plantae
Phylum	Spermatophyta
Subphylum	Angiospermae
Class	Monocotyledonae
Superorder	Lilianae (monocotyledons)
Order	Poales
Family	Poaceae
Genus	*Oryza*
Species	*Sativa*

Cultivation of rice in Malaysia began in the late eighteenth century when the first consistent records appeared in local and European sources until 1910 (Bray, 2014). In the early cultivation of rice in Malaysia, rice was only the staple food for the immigrants, such as Chinese and Indian, Sumatran and Siamese, who immigrated to Malaysia during the nineteenth century. Local rice growing methods were seen to be inefficient, and supplies of rice from adjacent British Empire countries, particularly Bengal and Burma, were plentiful and inexpensive (Fahmi *et al*., 2013). However, during the nineteenth century, the rice sector grew significantly and became one of Malaysia's most important

crops due to several cultural shifts in Malaysian rice farming. For instance, rice is grown with an irrigation and drainage system with intercropping approach (Rahmat *et al*., 2019). Direct planting is thought to have aided in sustained rice production, which is critical for food self- sufficiency (Bishwajit *et al*., 2013). Thus, this new tradition-breaking alternative has led to the development of the rice sector in Malaysia and succeeded in the early large-scale rice production in this country.

In 1961, the land area for rice cultivation in Malaysia increased dramatically, from 51,649 hectares to the most extensive rice fields in Malaysia, 766,180 hectares in 1972. The land area for rice cultivation has remained constant at an average of 677,000 hectares through 2016, representing a 0.2% increase in rice harvest area (Rahmat *et al*., 2019). The Malaysian government has officially acknowledged ten primary granary regions, as illustrated in Figure 2.1. Among these are Kemubu Agricultural Development Authority (KADA), Kemasin Semerak Integrated Agriculture Development Area, North Terengganu Integrated Agriculture Development Area (KETARA), Muda Agricultural Development Authority (MADA), Intergrated Agriculture Development Area (IADA) Penang, Pulau Pinang, IADA Kerian, IADA Seberang Perak, IADA Barat Laut Selangor, IADA Pekan and IADA Rompin. This granary area covers almost 397,790 hectares of land and contributes about 70% or rice production in Malaysia (Fahmi *et al*., 2013; Vaghefi *et al*., 2015).

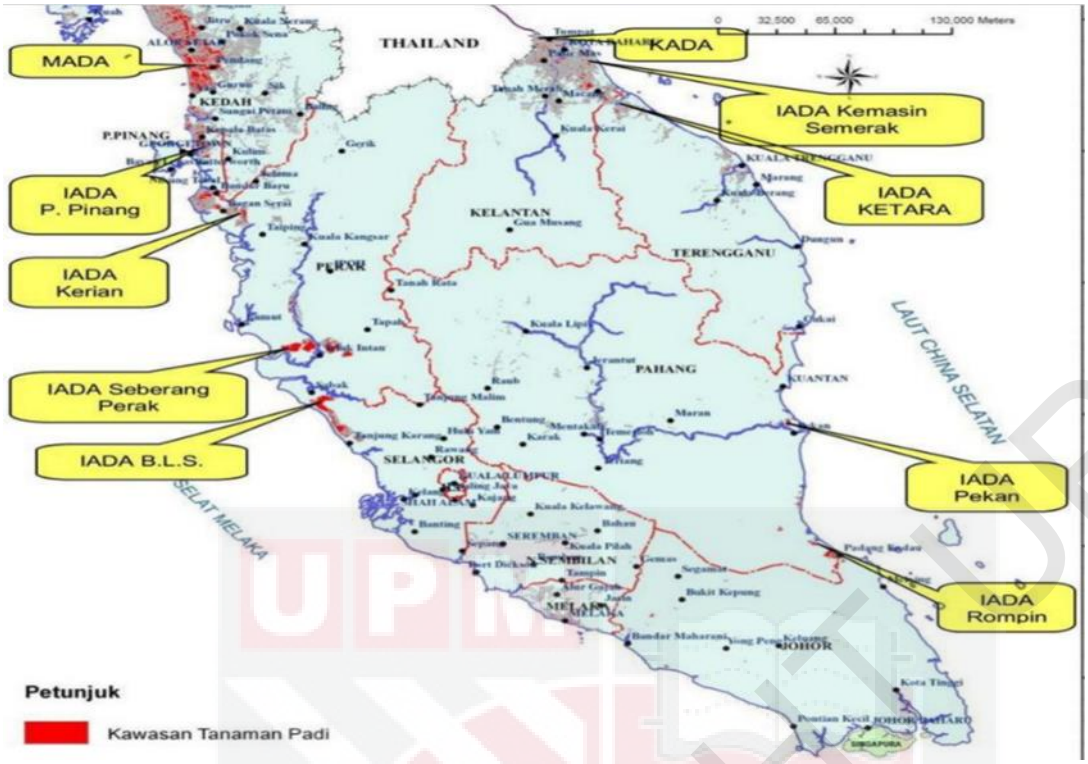

Figure 2.1: Ten main granary areas of rice fields in Pninsular Malaysia
(Source: Vaghefi *et al*., 2015)

In general, healthy and mature rice plants, with their upright culms, typically reach a height of up to 50 cm, as depicted in Figure 2.2. However, it's worth noting that certain floating swamp varieties have the potential to achieve greater heights up to 3 m tall (Jamil & Anwar, 2016). While the leaf sheath is smooth and glabrous on the surface, the leaf-blades surface is scaberulous and rough adaxially, about 12 to 65cm long and 4-18mm wide. The panicle, or inflorescence (flower cluster), comprises spikelets that bear flowers and produce the fruit or grain (Figure 2.2). The length, shape, and weight of the panicle, as well as the overall productivity of a given plant, vary significantly between varieties (Jonghan *et al*., 2013).

Figure 2.2: Mature and healthy *Oryza sativa*. Panicles bend over as the grains fills
(Source: Zhou-qi et al., 2016)

2.1.2 Economic Importance of Rice

Rice holds paramount economic importance in Malaysia, significantly contributing to the nation's Gross Domestic Product (GDP) and overall economic stability. This agricultural staple, deeply ingrained in Malaysia's culture and diet, serves as a vital source of livelihood for numerous farmers, playing a pivotal role in rural development (Abdul Rahim, 2018). As a staple food, it ensures food security, diminishing reliance on imports and stabilizing prices, which, in turn, helps maintain economic equilibrium (Ismail & Sarker, 2014). Additionally, rice production fosters employment generation across various sectors, from farming to processing, transportation, and marketing, thereby alleviating rural unemployment rates (Rahman, 2020). Furthermore, the sector exhibits export potential, primarily through high-quality aromatic rice

varieties, bolstering foreign exchange earnings (Ismail & Sarker, 2014). Overall, rice's multifaceted contributions make it an integral component of Malaysia's GDP and socioeconomic well-being. Malaysia's Gross Domestic Product (GDP) increase up to 21.22% growth in 2020 compared to past year (2015). An increase in GDP in Malaysia over the years indicates that the country's economy has grown and expanded (Table 2.2). GDP measures the total value of goods and services produced in a country over a specific period, usually a year (Md-Yassin, 2018). Therefore, an increase in GDP over the years means that the country's economy has produced more goods and services in each year which could be due to several factors such as increased investment, improvement in productivity, economic diversification and international trade (Abdurrahman & Wahid, 2019; Baharuddin *et al*., 2019).

According to Table 2.3 Domestic of Statistic Malaysia (2021) show that the agriculture sector contributed RM116.08 billion (8.2%), while the rice sector contributed RM2.521 billion (0.18%) in 2020. Compared to the past five year of rice GDP, from the statistics show a steady increment of GDP contribute from rice sector. The increment of GDP in the rice sector indicate an increase in the value of goods and services produced within that sector. Several factors could contribute to an increase in GDP in the rice sector, such as rice production increases, when more rice available for sale, leading to an increase in the value of goods produced in the rice sector. Other than that, the adoption of new technologies, such as better seeds, fertilizers, and irrigation systems, could increase rice yields, leading to higher production and an increase in GDP in the rice sector. Lastly, Malaysia country increases its rice exports, it could

generate more income for the country, which would contribute to an increase in GDP (Sadek et al., 2019; Ismail et al., 2019; Hamid et al., 2021).

Table 2.2: Malaysia's GDP from 2015 to 2020

Type of expenditure	2015	2016	2017	2018	2019	2020
	Value (RM Billion)					
Gross Domestic Product (GDP)	1,176.9	1,249.6	1,372.3	1,447.7	1,513.1	1,416.6

(Source: Department of Statistics Malaysia, 2021)

Table 2.3: Malaysia's Agriculture GDP from 2015 to 2020

Type of economic activity	2015	2016	2017	2018	2019	2020
	Value (RM Billion)					
Agriculture	97.539	105.756	117.995	108.757	109.549	116.084
Oil Palm	37.853	42.767	51.240	42.033	40.131	50.727
Rubber	3.348	3.548	4.872	2.894	3.324	2.590
Rice	2.205	2.246	2.225	2.407	2.434	2.521

(Source: Department of Statistics Malaysia, 2021)

According to the table 2.4 and figure 2.3 and 2.4 above shows that rice production and consumption in Malaysia has increase drastically compared to the past year. From the statistics we could see that rice have become important sector in our country. Some of the economic importance of rice in Malaysia is as food security. Rice is a vital source of calories and nutrition for the country's population, and it is estimated that Malaysians consume an average of 80kg of rice per person per year (Mohamad *et al*., 2017). Therefore, the production of rice plays a crucial role in ensuring food security for the country's population. Other than that, rice cultivation provides employment opportunities for more than 100,000 people and generating a gross output value of RM 7.1 billion in 2019 (Department of Statistics Malaysia, 2020). The rice industry is also an essential source of income for rural communities in Malaysia, with many farmers relying on rice production as their primary source of income (Omar *et al*., 2019). Next, rice is also an important source of income for many farmers

in Malaysia. According to the Ministry of Agriculture and Agro-Based Industry, the total value of rice production in Malaysia was RM5.6 billion (USD1.3 billion) in 2019 (Ministry of Agriculture and Agro-Based Industry, 2020). Also, Malaysia is a net importer of rice, but the country also exports rice to other countries. In 2020, Malaysia exported 98,806 metric tons of rice, worth RM117.4 million (USD28.3 million) (United States Department of Agriculture, 2021). Lastly, the rice industry in Malaysia is heavily subsidized by the government, which provides financial assistance to farmers and millers to promote production and ensure stable prices. The government also collects revenue from import duties and taxes on rice (Othman *et al*., 2019).

Table 2.4: Malaysia's Rice Production from 2014 to 2019

Year / Element	2014	2015	2016	2017	2018	2019
Area Harvested (ha)	614,579	681,559	688770	685548	699980	684416
Yield (hg/ha)	29855	40223	39775	37496	37704	42550

Table 2.4: Continued

Production						
Quantity (tonnes)	1834831	2741404	2739606	2750513	2639202	2912203

(Source: FAOSTAT, 2021)

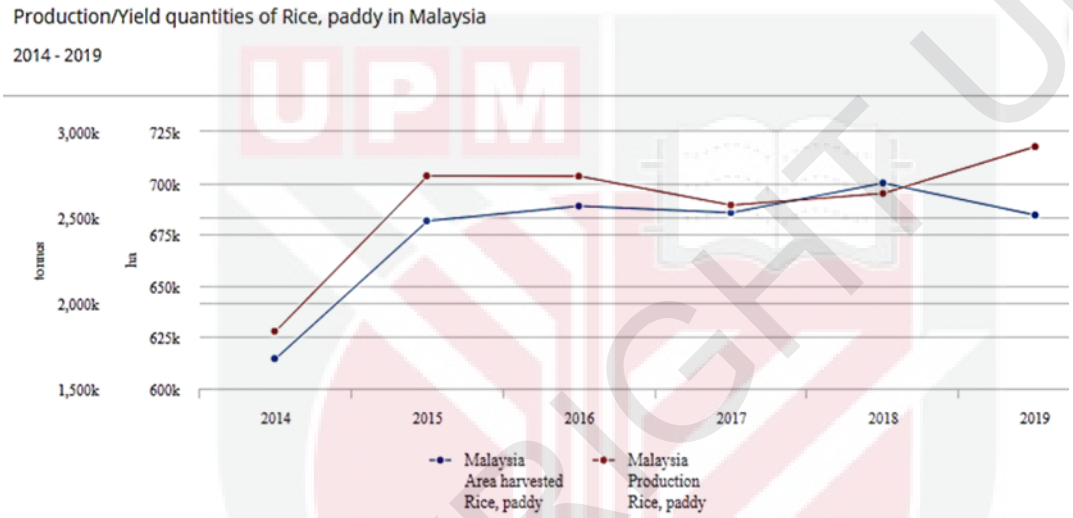

Figure 2.3: Production and yield quantities of rice in Malaysia from 2014 - 2019

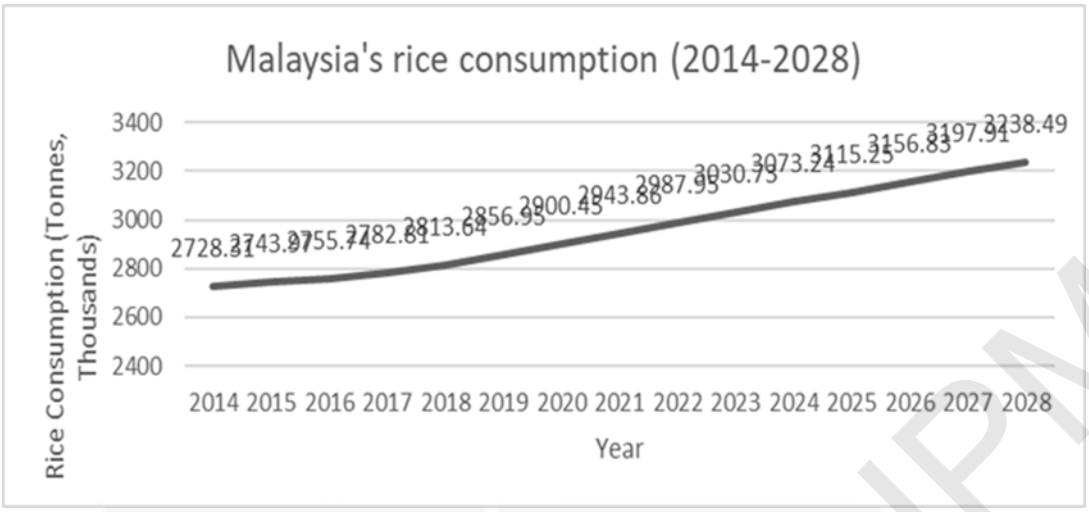

Figure 2.4: Malaysia's Rice Consumption from 2014 to 2028
(Source: OECD-FAO Agricultural Outlook, 2021)

Thus, the economic importance of rice in Malaysia is significant, with the rice industry providing employment opportunities, generating revenue, and ensuring food security for the country's population. Despite facing challenges, the rice industry in Malaysia continues to thrive and is an essential part of the country's culture and heritage.

2.1.3 Rice cultivars – MR219

The rice breeding program aims to increase yield productivity and rice resistance toward pests or fertilizer. Malaysia Agricultural Research and Development Institute (MARDI) has instructed the rice varietal program since 1969. To date, 51 rice varieties have been released since then, with each variety having a different speciality regarding resistance to disease, taste and yield quantity (Ab Razak *et al.,* 2020). Weny *et al.* (2019) state that 20-25% of yield production loses due to pests, diseases and weeds. The pest problem becomes severe when producing a high-yield agronomic package is practised with a high fertilizer rate and a broad range of pesticides to combat numerous pests. Sabran & Abas (2021) demonstrated that a pesticide purchase of RM 200/ha is necessary to contain pests and diseases over the crop, resulting in a higher preparation cost for farmers with large farm sizes.

MARDI had been developing the MR varieties since 1995 and MR219 in January 2001. This variety was created to direct the seeding planting system (Tan *et al.,* 2017). The objective of this seed was on the panicle component characteristics, especially on grain size and quantity of grains per panicle. A single grain of MR219 can weigh up to 28-30mg, and a panicle can contain up to 200 grains. As a result, this is essential for improving yield output depending on these two primary components. While in other studies state that other unique characteristics of MR219 are relatively tall with strong culms, moderate maturation period, moderate resistance to some rice diseases but tolerant toward panicle blasts and can be classified as long grain rice.

The average yield produced by MR219 is 8.1 tonnes/ha; if good water and fertilizer management is applied, this variety can grow up to 10 mt/ha (Sunian *et al.,* 2017). The MR269 variety, developed in 2002, has significant beneficial compare to MR219 in terms of resistance toward pests and disease but is slightly disadvantageous in yield produce. Misman and Zakaria (2019) stated that the MR269 variety has high resistance toward leaf blast disease caused by the fungus *Magnaporthe oryzae*. They also found that MR269 is moderately resistant toward panicle blast, bacterial leaf blight and sheath blight. Other notable characteristics of MR269 are slender and long grain, reasonably tall plant, and intermediate amylose content. According to the data presented in Table 2.5, MR219 boasts an impressive yield of 8.1 tonnes/ha, while MR269 yields between 7 to 7.9 tonnes/ha. This highlights MR219's remarkable productivity and underscores its significance in enhancing rice production efficiency (Mohd Kasim *et al.,* 2020).

Table 2.5: Rice Cultivar Summary

Rice Variety	Rice Specialty	Yield Potential (t/ha)
MR219	Highest rice yield production	8.1
MR269	High rice yield production	7-7.9

2.2 Bacterial Panicle Blight Disease

2.2.1 Symptom of Bacterial Panicle Blight of Rice

Bacterial panicle blight (BPB) has long been a source of concern for rice producers in many rice-growing countries, including Malaysia. This disease is not restricted to certain rice cultivars and has been identified as a global problem in rice production. BPB can harm a wide variety of rice cultivars, making it a difficult and pervasive problem for rice producers around the world. Weny *et al.* (2019) stated that BPB occurs in most rice-producing countries, such as Japan, Vietnam, Thailand, the United States, and India, but BPB does not present in Australia. On the other hand, until 1996-1997, BPB was thought to be caused by a physiological disorder and abiotic factors such as high temperature and water stress on the plant (Peng, 2015). Later, *B. glumae* and *B. gladioli* were isolated and discovered by Japan in 1950 and recognized as the primary cause of BPB (Nandakumar *et al*., 2009). The symptom of bacterial blight can be observed in seedling blight, the sheath rot of the flag leaves, and the panicle branches of the plants with significant yield losses (Zhou-qi *et al.,* 2016).

According to Nandakumar et al. (2009) and Sayler *et al.* (2006), BPB is distinguished by an upright, straw-colored panicle containing a floret with a dark grey and a reddish-brown line (border of the lesion). The rachis of the panicle remains green even after the glumes desiccate and turn tan. Furthermore, because the grain does not fill, the severely damaged panicle stays erect instead of bending over (Figure 2.5).

Figure 2.5: BPB disease on rice. The affected panicle remains upright as the grain does not fill
(Source: Zhou, 2019)

The leaves and spikelets contain small circular to oval tan lesions ranging from 1 to 5 mm with brown borders and grey or straw-coloured centres, resulting in kernel abortion before filling. Aside from that, an individual flag leaf sheath of an infected tiller may have a lesion several centimetres long with a grey and necrotic centre and a reddish-brown border resulting in sheath rot (Ortega and Rojas, 2021).

Another symptom that can be observed is bacterial panicle blight appearing on individual plants irregularly or in circular or oval patterns in the field. On the other hand, common panicle blanking, induced by abiotic stress such as excessive heat or water stress, grows more uniformly in the area and does not create an apparent circular pattern (Zhou, 2019). According to Shew *et al.*

(2019) found that BPB is produced by the simultaneous occurrence of high daily minimum temperatures (32 °C) and relative humidity (77%), both of which are expected to rise under the present global warming scenario.

Temperatures are predicted to rise in many parts of the world due to climate change. There will be places with higher warmth and humidity, resulting in increased BPB infection pressure in rice production worldwide. According to the findings of this study, even little temperature changes (+1°C) throughout the rice-growing season can have a substantial economic influence on the rice economy. As a result of global warming, BPB might become one of the most economically damaging rice illnesses in the following decades (Shew *et al.*, 2019). Zhou-qi (2016) also stated that the *B. glumae*-rice interaction might depend on environmental factors such as temperature and humidity, which may contribute to why this disease occurred.

2.2.2 Losses Caused by Bacterial Panicle Blight

High yield losses caused by *B. glumae* in rice are due to a lack of effective BPB prevention and control measures for susceptible rice culture (Mulaw *et al.*, 2018). Weny et al. (2021) revealed that grain abortion, grain rot, grain weight reduction, inhibition of seed germination, and floret sterility occur due to BPB infection, and the yield loss report in a severely infected field in the USA was up to 75% in several locations. Ramachandran et al. (2021) claimed that BPB caused by *B. glumae* had damaged rice granaries in several states in Peninsular Malaysia, accounting for up to 50% of disease incidence. On the other hand, Shew *et al.* (2019) discovered that yield loss caused by *B. glumae*

on different rice cultivars consisting of very susceptible rice (VS), susceptible rice (S), moderately susceptible rice (MR), and moderately resistant rice (MS).

From this study, they concluded that even moderately resistant rice varieties were still exposed to more than 10% yield losses. In 2017, a 6-20% incidence of bacterial panicle blight was reported where 43 of the panicle sample showed a symptom of reddish-brown with partially filled grains were collected from 15 towns in Guangxi and Guangdong provinces, China. These incidents have been reported to have caused significant yield losses in the country in 2017 (Hou *et al.,* 2020). In addition, a 2013 report from Zhou (2014) found that 25% of the rice plant cultivar Presidio and WAB56-104 showed positive BPB symptoms, and affected panicles remained upright with partially filled or abort at Ukulima farm in Limpopo Province, South Africa. According to the research, the BPB impacts the quality and quantity of grain, resulting in a significant decrease in grain weight and yield production (Zhou, 2014).

2.2.3 Control Strategies BPB Disease

Efforts have been made to create chemicals to manage the disease, but none are efficient or reliable (Ortega and Rojas, 2021). Successful disease control generally relies on adopting management techniques to decrease harm to a tolerable and acceptable level. To effectively manage rice BPB, rice farmers must first utilize pathogen-free seeds as an exclusion strategy, then plant with partially resistant cultivars, use available chemicals or biocontrol agents, and apply appropriate cultural practices (Peng, 2015). Integrating various

current management methods is critical to the practical and long-term control of BPB (Zhou, 2019).

Fory et al. (2014) stated that one successful strategy for excluding the disease from a disease-free geographic area is the use of certified seeds that are free of the BPB pathogens. Also, various molecular detection techniques, such as PCR, developed to analyze rice seed lots can help with this procedure. Since 2007, a similar plant quarantine regulation has been developed and implemented in China to prevent the possible importation of BPB diseases from other nations (Ortega and Rojas, 2021). On the other hand, seed treatment can be used as a last option to decrease and even eliminate seed-borne BPB pathogen populations and manage future crop disease to an acceptable level. BPB pathogens can be eradicated by treating rice seeds for six days at 65°C dry heat (Mizobuchi *et al.,* 2018).

Significant research efforts have been made worldwide to produce resistant cultivars as a viable, long-term approach for managing BPB in rice. However, no specific genes or quantitative trait loci (QTLs) for complete BPB resistance have been discovered. Mizobuchi *et al*. (2016), on the other hand, recently found two tropical *japonica* cultivars, Kale and Jaguary, with a high degree of resistance and numerous *indica* cultivars with a moderate level of resistance toward BPB disease. Another mechanism of control for counter BPB disease is through chemical and biological control. As for chemical control, oxolinic acid can be used in seed treatment or foliar application, and this is the only chemical that can control BPB disease. However, some countries do not commercially

allow the use of this acid, such as the USA, due to populations of *B. glumae* resistant to oxolinic acid that have been discovered in rice in Japan since 1998 (Goo *et al.,* 2015). The development of oxolinic acid resistance in *B. glumae* populations is due to an amino acid substitution at position 83 in *GyrA* (GyrA83) (Goo *et al.,* 2015). On the other hand, copper and copper-containing bactericides have also been reported to control BPB disease in rice effectively (Mizobuchi *et al.,* 2018). Earlier findings also found that bactericides such as kasugamycin, probenazole, and pyroquilon were applied to control rice seedling and grain rot (Riera-Ruiz *et al.,* 2014).

Few research have been performed to comprehend and establish cultural practices that may minimize the occurrence and severity of BPB in rice. According to Wamishe *et al.* (2014), rice plants are more susceptible to BPB disease when nitrogen fertility levels are high. Thus, the report recommended avoiding excessive use of nitrogen rate in pesticides or herbicides to reduce the damage caused by BPB. In addition, another successful strategy for mitigating disease damage is early planting or the use of early maturing rice varieties to escape the hottest periods of the growing season. Furthermore, avoiding high seeding rates can assist in reducing the occurrence and severity of the illness (Riera-Ruiz *et al.,* 2014).

2.3 *Burkholderia* species

2.3.1 History and Biology of *Burkholderia* spp.

Burkholderia species belong to the domain Bacteria, class Betaproteobacteria, order Burkholderiales, and family Burkholderiaceae. Within this family, Burkholderia forms a diverse and versatile genus comprising numerous species. These species are further classified based on genetic and phenotypic characteristics, such as DNA sequences and metabolic traits, into distinct groups. The taxonomy of *Burkholderia* has undergone significant revisions in recent years, leading to the recognition of novel species and the reclassification of some previously identified ones. These revisions have contributed to a deeper understanding of *Burkholderia*'s ecological and pathogenic diversity, enabling scientists to better study and address their roles in various environments, including soil, plants, and in some cases, as opportunistic human pathogens.

Table 2.6: *Burkholderia* Taxonomy Classification from Kingdom to Species

Name	Bacteria
Kingdom	Bacteria
Phylum	Proteobacteria
Class	Beta proteobacteria
Order	Burkholderiales
Family	Burkholderiaceae
Genus	*Burkholderia*
Species	*glumae*/*gladioli*

In 1967, a bacterial pathogen causing rice grain rot was known as the *Pseudomonas* species. However, in 1992, the non-fluorescent bacteria in *Pseudomonas* were classified as genus *Burkholderia* and others based on 16S rRNA sequences, DNA-DNA homology, cellular lipid and fatty acid compositions, and behavioral features. As a result, in 1992, *Pseudomonas glumae* was renamed *Burkholderia glumae* (Ortega and Rojas, 2021).

In recent years, *B. glumae* and *B. gladioli* (*Burkholderia* spp.) have caused a developing rice disease in the world's major rice-producing countries. However, it is not well known in terms of their infection process and pathway. Furthermore, in many field crops, the bacterial wilt produced by *Burkholderia* spp. is symptomatically indistinguishable from the bacterial wilt caused by *Ralstonia solanacearum* (Zhou-qi *et al.,* 2016). Therefore Magbanua *et al.* (2014) investigated the variations in gene expression between resistant and susceptible rice cultivars when they interacted with pathogenic *Burkholderia* spp., which may give helpful information for avoiding rice disease caused by this pathogen.

Burkholderia spp. is a bacterium that causes yield loss on susceptible rice cultivars in Malaysia and other rice-growing countries worldwide via bacterial panicle blight disease. Besides causing bacterial panicle blight on rice, Burkholderia spp. has also been attributed to bacterial wilt in other crops such as potato, tomato, and eggplant, resulting in significant agricultural and economic losses (Ramachandran *et al.,* 2021). For instance, Nandakumar *et al.* (2009) isolated 420 bacterial strains from BPB-infected rice sheath and

panicle from the United States from 2004 to 2009 rice season and identified 319 and 21 of the bacterial strains are *B. glumae* and *B. gladioli*, respectively. Furthermore, isolating human pathogenic *B. glumae* strains from surgical specimens of lung lesions defines *B. glumae* as having human pathogenic potential. It broadens the spectrum of *Burkholderia* species involved in the infection of patients with chronic granulomatous disease (GCD), primarily an immunodeficiency disease, ultimately leading to an increased risk of invasive human infections (Weinberg *et al.,* 2007). On the other hand, genetically varied strains of *Burkholderia* spp. have been identified from asymptomatic rice plants and those with BPB symptoms, indicating that *Burkholderia* spp. is capable of rapid evolution (Zhou, 2014).

Burkholderia species is a Gram-negative Betaproteobacteria bacterium rod-shaped. This bacterium is also a non-fluorescent, aerobic bacteria with 1-3 polar flagella (Peng, 2015). The optimal temperature for *Burkholderia* species growth is approximately 30-35 °C. However, it may thrive at temperatures as high as 41 °C (Wamishe *et al.,* 2014).

On different mediums, it can generate a yellow-green, water-soluble pigment. Thus, the colony appears grayish-white or yellow due to the pigment. *Burkholderia* spp. arginine dehydrolase, oxidase, indole production, and urease reactions are all negative (Weny *et al.,* 2019). This pathogen shows a positive reaction in the catalase test, arabinose, mannitol, and gelatin liquefaction (Ramachandran *et al.,* 2021). This pathogen infects seeds, invades plumules via stomata and wounds, and replicates in parenchymal

intercellular spaces during seed germination. *Burkholderia* spp. growth in plumules results in poisonous chemicals such as toxoflavin, which causes bacterial panicle blight disease (Zhou-qi *et al.,* 2016). Quorum sensing regulates genes involved in toxoflavin production and transport. According to Magbanua *et al.* (2014), the total genome size of *Burkholderia* spp. ranging from 6 to 9Mbp and according to Wang *et al.* (2020) demonstrated that *B. glumae* strain BGR1 is made up of two chromosomes and four plasmids, with chromosome 1 having 3,906,529 base pairs, 3,290 predicted coding sequences (CDS), 144 pseudogenes, three rRNA operons, and 56 tRNAs and chromosome 2 having 2,827,355 base pairs, 2,079 CDS, 192 pseudogenes, two rRNA operons, and eight tRNAs. While BSR3 (*B. gladioli*) strains comprised two chromosomes of 4,413,616 bp and 3,700,833 bp and four plasmids of 276,215, 129,399, 128,650, and 403,586 bp.

The data showed that the genome is larger than BGR1, with 7717 genes and an average GC content of 67.41% (Seo *et al.*, 2011; Lim *et al.,* 2009).

2.3.2 Virulence Factor of *Burkholderia* species

B. glumae and *gladioli* pathogenesis is a complicated process involving many virulence factors primarily controlled by a worldwide regulatory quorum-sensing (QS) system regulated by the LuxI and LuxR homologs, TofI and TofR, respectively (Francis *et al.,* 2013). Phytotoxins (toxoflavin and fervenulin) and lipases are these pathogen's most prominent pathogenic factors. Other virulent components contributing to the overall pathogenicity include *PehA* and *PehB* polygalacturonases, *KatG* catalase, and the *Hrp* type III secretion system (Hrp-

T3SS) (Kim *et al*., 2014). Endopolygalacturonase and exopolysaccharides (EPSs) are other hypotheses that may have a role in the pathogenesis of these pathogens (Lee et al., 2016). Furthermore, bacterial motility (flagella) has been shown to operate as virulence factors controlled by the quorum sensing (QS) system (Peng, 2015). While for the phytopathogen process, *Burkholderia* spp. must colonize and grow in host plant tissues, which are thought to have low nutritional, oxygen, and iron levels. In addition to these extreme environmental circumstances, host plants' immune systems make it challenging for pathogens to grow. Thus, pathogens commonly modify their gene expression to generate physiological changes and interactions with host plants to resist these severe circumstances and cause disease in host plants (Kim *et al.,* 2014).

According to Zhou (2014) and Hou et al. (2020), virulence *Burkholderia* can produce distinctive yellow pigments of toxoflavin and fervenulin, which are crucial for the pathogenicity of grain rots, which can result in chlorosis symptoms on rice panicles. On the other hand, some thioflavin-deficient *Burkholderia* mutants could not cause disease symptoms in rice. Toxoflavin production is temperature sensitive and achieves a maximum production at 37 °C, with no observable toxoflavin generated between 25 °C to 28 °C. In addition, toxoflavin function as an effective electron transporter and causes the formation of hydrogen peroxide from oxygen. Hydrogen peroxide's toxicity production can interfere with the photosynthesis, photorespiration, and other metabolic activities of rice (Francis *et al.,* 2013). Toxoflavin was found and characterized through a polycistronic operon composed of five genes (*tox*A,

*tox*B, *tox*C, *tox*D, and *tox*E) and four genes (*tox*F, *tox*H, and *tox*I). ToxR controls the expression of both the t*oxABCDE* operon and the t*oxFGHI* operon and is involved in activating *tox* operon transcription (Lee *et al.,* 2016). According to a current study, any mutational insertion on *toxABCDE* can block the synthesis of toxoflavin, and *toxFGHI* was shown to be responsible for toxoflavin transport (Mulaw *et al.,* 2018). Furthermore, toxoflavin biosynthesis and transportation need the transcriptional activator ToxJ, which QS controls this expression. However, according to some research, toxoflavin toxin causes pathogenesis in rice and may also be utilized to prevent various fungal infections in rice, such as *Fusarium graminearum* (Jung *et al.,* 2013).

Lipases have also been implicated in the pathogenicity *of Burkholderia* spp. (An *et al.,* 2014). *lip*A, an active extracellular lipase, is the most significant virulent-relative lipase. *Lip*B is another key of lipase that is involved in the biosynthesis of *lip*A and is required to generate active *lip*A (An *et al.,* 2014). *lip*B also significantly impacts protein stability for proteolytic destruction (Devescovi *et al.,* 2007). Other research, however, discovered that a mutation in *lip*A of lipase led to a reduction in disease symptoms in rice (Devescovi et al., 2007).

Gram-negative bacterial pathogens have commonly utilized the type III secretion system to inject proteinaceous virulence components (Ham *et al.,* 2010). According to Kang *et al.* (2008), a proteomic analysis of *B. glumae* indicated that the T3SS of *B. glumae* comprises 34 proteins that accumulate in a HrpB-dependent manner. In contrast, mutants of *B. glumae* with T3SS

defects are considerably less pathogenic on rice panicles than in the wild-type strain. This result indicates that Type III effectors play a critical role in producing virulence in rice. On the other hand, flagella is also an essential factor for pathogenic bacteria (Ham *et al.,* 2010). Since the flagella allow the bacterium to reach infection sites in a susceptible host and offers a considerable selection advantage during the early infection establishment phase (Lee *et al.,* 2016).

2.3.3 Epidemiology and Detection of *Burkholderia* species

Burkholderia spp. can live in various environments, including soil, water, and plants. This bacterium also prefers hot nights and high humidity, particularly during the rice-growing season (Li *et al*., 2016b). According to Pedraza *et al*. (2018), this disease is a seed and soil-borne pathogen that is thought to replicate on the phylloplane of rice plants during the growth stage and lives inside and outside of rice seeds, rice tissue, or weeds in the field.

Due to *Burkholderia* spp. is a seed-borne disease, the pathogen's widespread dissemination is very plausible owing to the long-distance transfer of infected rice seeds. In addition, Peng (2015) found that the disease prevalence cause by *B. glumae* can be influenced by various factors, including temperature, humidity, disease severity of the previous year, and host vulnerability. Shew *et al.* (2019) stated that even minor temperature changes (+1°C) throughout the rice-growing season could have a significant economic impact on the rice industry as *Burkholderia* spp. favor high temperatures, especially during the flowering stage and can lead to an outbreak of BPB. The density of the bacteria inoculum also influences the severity of the disease. *Burkholderia* was inoculated with as little as 10^2 to 10^4 CFU/ml through a foliar technique can also result in BPB in the rice area (Li *et al.,* 2016b). During infection, *Burkholderia* spp. is assumed first to bypass stomata, penetrate lemma and paleae, reproduce in the parenchyma, and disseminate to adjacent cells (Li *et al.,* 2016a).

Due to the major loss caused by these pathogens in rice production, it is vital to identify *Burkholderia* spp. accurately and efficiently. King's B agar medium, a semi-selective medium, was designed for the selective culture of *Burkholderia*. However, this media cannot distinguish at the species level within the genus *Burkholderia*. Thus, biochemical and molecular characterization tests are still needed to identify the species level of *Burkolderia* spp. Nandakumar *et al*. (2009) used PCR and specific primers developed for the β-subunit polypeptide of DNA gyrase gene (gyrB) and 16S ribosomal RNA gene to differentiate and identify *B. glumae* and *B. gladioli* strain from other *Burkholderia* spp. in Malaysia rice granaries. This rRNA gene-based PCR method was straightforward, rapid, and sensitive for detecting and identifying *Burkholderia* spp. Sayler *et al*. (2006) developed a real-time PCR (RT-PCR) method for isolating, detecting, and characterizing *Burkholderia* spp. isolated from Arkansas, United States, using specific primers designed for 16S-23S rDNA internal transcribed spacer sequence (ITS). The specific primers were designed only to amplify all *B. glumae* strains but not *Burkholderia* spp.

While in another study conducted by Moon *et al.* (2017) performed PCR-restriction fragment length polymorphism amplification using a 16S rRNA species-specific primer, which only amplified 470-bp fragment of 16S rDNA from *B. gladioli* strains. In addition, Kim *et al*., (2012) designed an RT-PCR technique for identifying *B. glumae* by employing species-specific primers created for the *rhs* family gene (YD repeat protein) of *B. glumae*. The same study found that 138-bp DNA fragmentation was amplified from the infected

grain sample by RT-PCR, and no single DNA fragment was amplified from the negative control. These results indicate that using specific species primer from *rhs* family gene using Bio-PCR, we could detect and characterize *B. glumae* from other species and *Burkholderia* spp. (Kim *et al.,* 2012).

Quantitative polymerase chain reaction (qPCR) has recently been utilized for detecting and quantifying *Burkholderia* spp. with high sensitivity (Sayler *et al*., 2006). In addition, Rep-PCR fingerprint analysis demonstrated the ability to distinguish between different *Burkholderia* strains (Sayler *et al.,* 2006). Recently, Cui *et al.* (2016) developed a multiplex PCR (mPCR) technique to identify six common rice bacterial infections, including *B. glumae, B. gladioli, Xanthomonas oryzae* pv. *oryzicola*, *Xanthomonas oryzae* pv. *oryzicola*, *Pseudomonas fuscovaginae*, and *Acidovorax avenae* subsp. *Avenae.* To validate the effectiveness and reliability of mPCR, 150 target and non-target bacterial strains from various rice-producing locations worldwide were tested. Utilizing mPCR allows for the amplification of several gene segments simultaneously rather than separate test runs for each.

These findings suggest that the mPCR technique accurately identifies pathogens in early detection and can help concurrently identify multiple pathogens from the same rice disease sample, which could play an essential role in disease prevention and treatment (Cui *et al.,* 2016).

2.4 Next Generation Sequencing

2.4.1 First Generation Sequencing

Determining the sequence of DNA has enabled researchers to investigate the structure and function of proteins and ribonucleic acid (RNA) and gain insight into the underlying causes of disease. Following Rosalind Franklin's fundamental DNA crystallography and X-ray diffraction research, Watson and Crick discovered the DNA structure in 1953. From this finding, in 1965, tRNA was the first molecule to be sequenced by Robert Holley, followed by RNA sequence from bacteriophage MS2 (McGinnis *et al*., 2016). Since then, several research organizations began using these methods to sequence DNA, and Fredrick Sanger and his colleagues achieved a breakthrough in 1977 by developing chain-termination processes (Thompson and Milos, 2011). For more than 40 years, Sanger sequencing has used the dideoxy chain termination method as the primary methodology for generating DNA sequences and the gold standard for addressing genomic variations. The first-generation sequencing was named after Frederick Sanger and his colleagues, who invented it in 1977. In Sanger sequencing, the unique procedure was by employing a DNA sequence of interest as a template for a PCR that adds modified nucleotides to the DNA strand during the extension stage, known as dideoxynucleotides (ddNTPs) (Gupta & Gupta, 2014). As high-throughput next-generation sequencing technologies now allow genome-wide sequencing applications, genomics research has turned toward DNA sequencing as a primary molecular tool (Mardis, 2011).

Sanger sequencing can sequence single-strand DNA at one time up to 1000bp with 99.999% per base accuracies, but the sequencing process needs a long time and a high cost per base. In addition, the advantage of Sanger sequencing is that it can target primer for sequencing, but the disadvantage is the sequencing process's need for cloning and the high amount of data per run (Ambardar *et al.,* 2016).

The principle of Sanger sequencing is similar to the polymerase chain reaction process. The difference in Sanger sequencing with conventional PCR is the addition of deoxyribonucleotides (ddNTPs) (Heather & Chain, 2016). The DNA polymerase adds the ddNTPs to the expanding DNA strand during the extension step of standard PCR by catalyzing the creation of a phosphodiester bond between the free 3'-OH group of the last nucleotide and the 5'-phosphate of the next nucleotide. In addition, Sanger sequencing has four ddNTPs: ddATP, ddTTP, ddGTP, and ddCTP. All four of the ddNTPs have a unique fluorescent label. After that, within the sequencing machine, all the chain-terminated oligonucleotides are sorted by size using single capillary gel electrophoresis. Lastly, A computer scans each band of the capillary gel in turn, utilizing ddNTPs' unique fluorescence to determine the identity of each terminal ddNTP. In summary, a laser stimulates the fluorescent tags in each band, and a computer detects the light produced consequently.

Since each of the four ddNTPs is labelled with a unique fluorescent label, the light emitted may be directly linked to identifying the terminal ddNTPs. The

result is called a chromatogram, which displays each nucleotide's fluorescence peak along the template DNA's length (Mardis, 2011).

2.4.2 Second Generation Sequencing

Compared to the highly automated and multistep pipelines required for clone-based sequencing, second-generation sequencing saves substantial time due to highly simplified sample preparation before sequencing, with little requirement for related equipment. The second-generation platform attempts to amplify a single strand of a DNA library from the DNA source of interest and sequence the amplified strands (Heather & Chain, 2016). A DNA library is a collection of DNA fragments cloned onto vectors for researchers to identify and extract the DNA fragments of interest for further investigation (Son and Taylor, 2011). The DNA libraries are created by annealing platform-specific linkers to blunt-ended fragments produced directly from DNA samples with a sequencing adapter (Mostafa *et al.,* 2015). The primary goals of second-generation sequencing are to create millions to billions of short reads in parallel, to do low-cost sequencing, and to detect sequencing output without using a capillary electrophoresis instrument (Kchouk *et al.,* 2017).

Second-generation sequencing platforms can be grouped into two comprehensive approaches, which are sequencing by ligation (SBL) and sequencing by synthesis (SBS) (Goodwin *et al.,* 2016). In the SBL perspective, the DNA ligase used to identify the nucleotide present at a particular location in a DNA sequence instead creates a second strand of the DNA. A pool of probe oligonucleotides labelled with fluorescent dyes is introduced and

hybridizes to the target DNA sequence (Del Vecchio *et al.,* 2017). When the probe oligonucleotide's bases match those of the target DNA sequence, it will ligate to an adjacent oligonucleotide for imaging. Based on the fluorescence emission spectrum generated by the probe, the nucleotide on the unknown target sequence may thus be identified (Goodwin *et al.,* 2016; Slatko *et al.,* 2018). From an SBS perspective, this approach is an extension of Sanger sequencing without the dideoxy terminators, in conjunction with multiple synthesis cycles, imaging, and techniques for incorporating more nucleotides into the expanding chain. The presence of thousands of identical copies of a DNA fragment in a specific region ensures that the signal can be distinguished from background noise (Ambardar *et al.,* 2016). Different platforms have different library preparation that refers to DNA template for sequencing. Second-generation sequencing consists of three main steps: DNA is fragmented to a specified length, an adapter is attached to the fragment to permit adhesion to solid surfaces, and templates are amplified to generate enough copies for the sequencers to identify them. The libraries can be either in single-end reads or paired-end reads (Del Vecchio *et al.,* 2017). The second generation is classified into five major sequencing platforms that have emerged: Roche/454 launched in 2005, Illumina/Solexa genome analyzer in 2006, ABI/SOLiD in 2007, Ion Torrent in 2010, and 2015 Beijing Genomics Institute (BGI). Platforms differ mainly in their sequencing chemistries and principle, which result in variations in throughput, read length, error rate, genome coverage, cost, run time, and applications.

2.4.3 Third Generation Sequencing

Second-generation sequencing methods often need a PCR amplification step, which leads to a time-consuming and costly operation. Thus, third-generation long-read sequencing provides new insight into sequencing by allowing for sequencing without PCR amplification of the DNA, allowing for shorter preparation times while reducing the possibility of errors and eliminating the need for costly, computationally time-consuming assembly steps (Kanzi *et al.,* 2020). The third-generation sequencing is capable of producing lengthy reads surpassing several kilobases (>1 kb to 2 Mb) for the solution of assembly problems, copy number changes, structural variants, mapping certainty, improve *de novo* assembly and repetitive sections of complex genomes (Xiao & Zhou, 2020). It enables billions of distinct fragments to be sequenced independently simultaneously. Aside from that, the signal is collected in real-time, which implies that the signal is monitored live throughout the enzymatic reaction of adding nucleotide in the complementary strand, whether it is fluorescent (Pacific Biosciences) or electric current (Nanopore) (Kanzi *et al.,* 2020).

Third-generation sequencing can be divided into two approaches: single-molecule real-time sequencing (SMRT) developed by Quake laboratory (Kchouk *et al.,* 2017), which Pacific Biosciences have used (PacBio) and Oxford Nanopore sequencing technology (ONT), and synthetic approach which relies on existing short read technologies to construct long reads *in silico* used by Illumina (TruSeq) and 10x Genomics. Single-molecule real-time sequencing (SMRT) is based on a synthesis-by-synthesis technique,

and real-time detects fluorescently tagged nucleotides when incorporations occur. ONT is a portable sequencer that employs a grid of membrane-embedded biological nanopores and a paradigm-shifting technology to enable rapid, real-time long-read sequencing of nucleic acids, resulting in higher resolution structural genomic variations and repeat content (Bowden et al., 2019). Instead of generating genuine long reads, the synthetic methods approach library preparation that uses barcodes to facilitate computer assembly of a more extensive segment (McCoy *et al.,* 2014). Such lengthy reads can traverse complicated or repetitive areas with a single continuous read, removing uncertainty in genomic element locations or size. Long reads are particularly valuable for transcriptome study since they may cover whole mRNA transcripts, allowing researchers to determine the precise connection of exons and distinguish gene isoforms (Zhou *et al.,* 2016).

2.4.4 Whole-Genome Sequencing

The whole-genome sequencing method analyses a cell's entire genomic DNA sequence at once, giving them the most comprehensive characterization of the genome and its biological implications. For instance, Vandamme *et al.* (2017) investigated the polyphasic taxonomic approach of four *Burkholderia singularis*, a rare and novel species of *Burkholderia,* through comparative whole genome analysis. They demonstrated that this species' G+C content is much lower than other

Burkholderia species, which is only 64.34%, probably due to altered DNA base excision repair. Through the WGS analysis, they found that non-coding regions within the DNA genome that were previously thought to be non-functional are regions that produce RNA molecules required for cell growth and genetics. In addition, the first complete genome sequence of plant pathogenic *B. glumae* was generated in 2009. The species consist of two chromosomes and four plasmids with a genome size of 7,284,636 bp, including 6,202 protein-encoding genes. The study found that pathogenicity-related genes such as toxoflavin biosynthesis and transport are located and found in the chromosome 2 sequence (Lim *et al.,* 2009).

Thomas *et al.* (2017) developed 103 sets of *Salmonella enterica* whole genome sequences and used the genome sequence data to identify the organism's antimicrobial resistance and virulence genes. Baltrus and Clark published a complete genome sequence of *Pseudomonas coronafaciens*

isolated from rice in 2020 using long reads sequenced of Nanopore MinION in Flongle flow cells.

They discovered this pathogen contains a single circular chromosome and a single circular plasmid with 5,220 genes indicating 4976 coding proteins, 63 tRNAs, four noncoding proteins, and three complete sets of rRNAs. Remenant *et al.* (2012) published a genome sequence of *Rastolnia solanacearum* that caused a tomato bacterial wilt outbreak in the Southeastern United States and annotated 360 unique genes. Among the unique gene annotated, 82% of the gene encodes for uncharacterized proteins, and 9% encodes proteins of extrachromosomal origin that demonstrate adaptability to Caribbean environmental circumstances (Remenant *et al.,* 2012).

CHAPTER 3

ISOLATION, PHENOTYPIC AND MOLECULAR CHARACTERIZATION OF *Burkholderia* SPECIES ASSOCIATED WITH BACTERIAL PANICLE BLIGHT DISEASE OF RICE IN PENINSULAR MALAYSIA

3.1 Materials and Methods

3.1.1 Sample Collection

Rice panicles with BPB symptoms were randomly sampled from rice growing areas in Peninsular Malaysia, including Perak, Kedah, and Selangor during June 2021 to January 2022. In this study, random sampling was used to ensure that every individual element of the affected population had an equal and unbiased possibility to be included in the sample. This study used random sampling to make statistically accurate inferences about the population under examination, enhancing the external validity of the findings.The selection of field were chosen according to the reported from DOA of Kedah, Perak and Selangor. In each field, the sample were taken follow random sampling table and two samples were taken from each plot of the field. The amount of sample collected is deemed adequate for the study as it meets the statistically determined sample size, ensuring that the results are statistically significant and capable of providing meaningful insights into the research question. In Kedah, sampling was conducted in Yan and Kodiang; in Selangor, samples were collected from Sekinchan, while in Perak, samples were collected from Seberang Perak. Five rice varieties were collected (MR220CL2, MR306, MR269, MR297 and MR307) (Table 3.1). Twenty-seven rice samples were collected in a plastic bag and labelled based on location and variety.

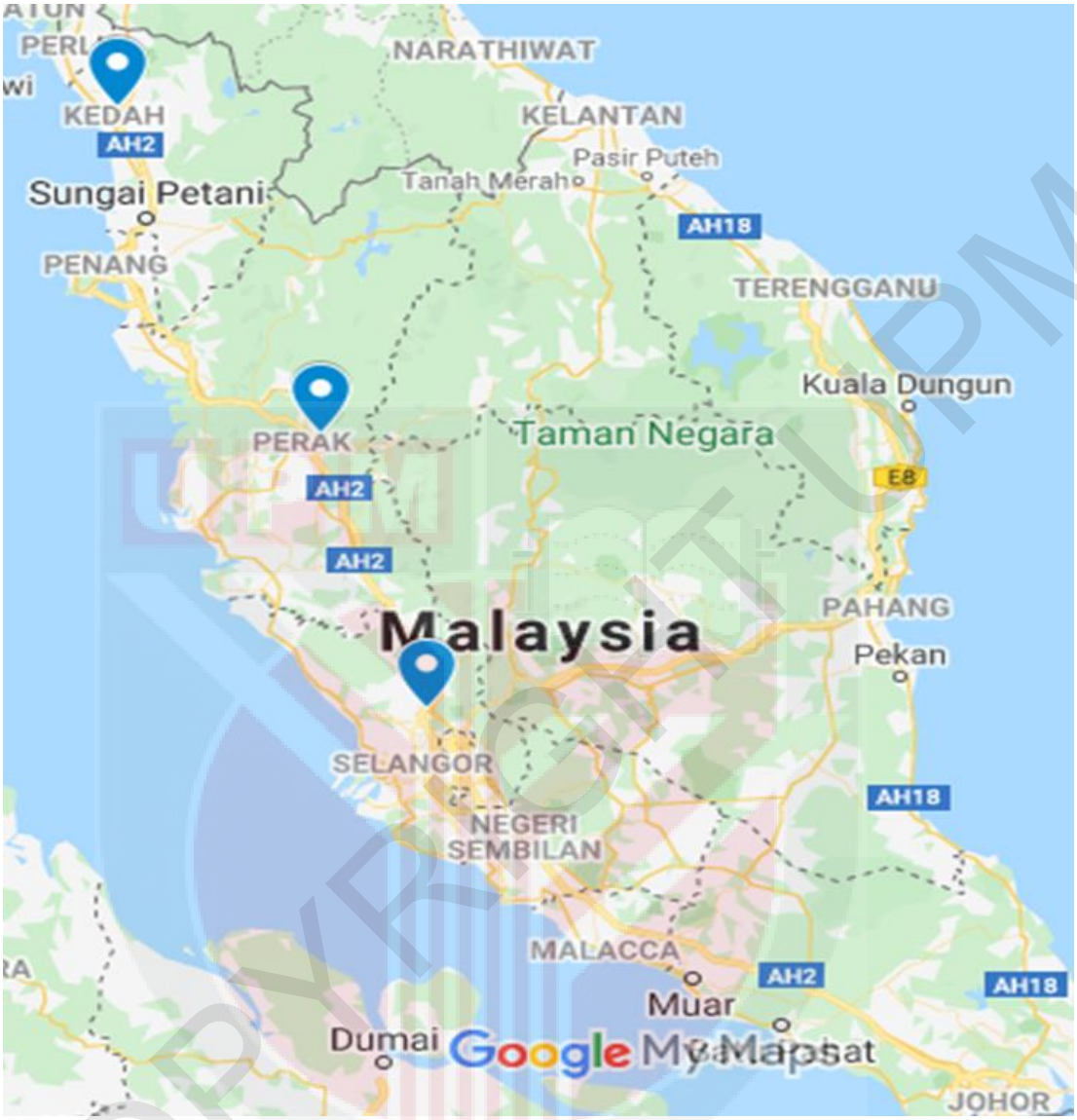

Figure 3.1: The pin location indicates the state of sample collection in Peninsular Malaysia

The samples transport back to the Bacteriology Lab, Department of Plant Protection, Faculty of Agriculture, and stored at 4 °C. The samples were used for initial diagnosis, isolation, and identification of pathogenic bacteria causing bacterial panicle blight in the location stated. Figure 3.1 and Table 3.1 show the distribution location of the sample collection.

Table 3.1: Rice collected from different locations area in Peninsular Malaysia BPB symptoms from June 2021 to January 2022

States	Sampling Areas	Varieties	Total Samples	Date
Kedah	Yan (5.7633° N, 100.3702° E)	MR220CL2 MR306	2 2	June 2021
	Kodiang (6.3933° N, 100.3057° E)	MR269	4	June 2021
Selangor	Sekinchan (3.5041° N, 101.1033° E)	MR297 MR307 MR269	4 5 3	November 2021
Perak	Seberang Perak (4.1013° N, 100.9521° E)	MR297	7	December 2021

3.1.2 Isolation of Bacterial Pathogen

The isolation of bacterial pathogens was carried out per the methods described by Weny *et al.* (2019) with minor modifications. The rice panicles and sheaths were excised and surface sterilized using distilled water containing 1% sodium hypochlorite (NaOCl). After a few seconds, wash the sample gradually with sterile distilled water and culture the sample on nutrient agar and King's B for 24 to 48 hours at 41 °C. Isolated colonies were subcultured onto King's B agar until pure isolates of suspected bacteria were obtained. Pathogenic *Burkholderia* spp. formed cream colonies with yellow pigment on King's B agar due to toxoflavin, a significant virulence factor for *Burkholderia* species. The pure culture colonies were grown on nutrient broth and stored in 20% (v/v) glycerol at -70°C as frozen stock for further use.

3.1.2.1 Morphological Characterization of Bacterial Isolates

The morphological characteristics of the pure culture colonies on the King's B agar were identified through microscopic and macroscopic identification. As for macroscopic identification, the colonies' form or shape, the colonies' colour, the colonies' elevation, the colonies' margin, texture and pigment of the colonies were observed thoroughly. The microscopic observation was identified based on Mizobuchi *et al.* (2016), and the morphology of the pure culture was studied under a light compound microscope

3.1.3 Phenotypic Characterization of Bacterial Isolates

Phenotypic bacterial identification is fundamentally based on comparing the phenotypic characteristics of unknown bacteria to those of pure colonies culture.

The reliability of the identification is in direct proportion to the number of similar features. There are two essential stages in phenotypic characterization: determining the genus of the pure culture obtained through Gram-staining. The second stage is determining the species level through biochemical tests (Hindmarsh & Geertsma, 2017).

3.1.3.1 Gram-Staining

The gram-staining technique was used to identify the shape and whether the organism was Gram-positive or Gram-negative bacteria. Four reagents were required in the method: crystal violet, Lugol's iodine, acetone alcohol solution, and safranin. Initially, using an inoculating loop, one loop of bacterial colonies was inoculated, spread onto a glass slide, and smeared the bacteria over a Bunsen burner flame. After that, crystal violet was dropped onto the slide for 60 seconds and washed with distilled water. Lugol's iodine solution is dropped onto the slide for 30 seconds before being washed with distilled water and decolourized with an acetone alcohol solution, which is then washed with distilled water after a few seconds. Finally, safranin is dropped onto the slide for one minute before being rinsed with distilled water. Gram staining is based on the ability of the bacteria cell wall to retain the crystal violet dye during solvent treatment. Gram-positive bacteria showed blue or purple-coloured colonies when observed under the microscope as the organism retains the crystal violet dye on its cell wall. Gram-negative bacteria showed red or pink-coloured bacteria under a microscope as they could not retain crystal violet dry and counterstain with red safranin to smear the decolorized bacteria (Takenaka *et al.,* 2012).

3.1.3.2 Oxidase Test

An oxidase test assessed the bacterial capacity to manufacture the cytochrome oxidase enzyme. A few drops of Kovac's reagent (1% tetramethyl-para-phenylene diamine dihydrochloride) are applied on a piece of filter paper.

A pure culture is selected and spread on the filtered paper using an inoculating loop. A colour change in the inoculated paper area will be observed within 10-30 seconds. Positive results reveal that colourless Kovac's reagent was oxidized to purple (bacteria capable of producing cytochrome C oxidase enzyme). In contrast, negative results showed that Kovac's reagent did not change colour (Md. Zali & Iberahim, 2021).

3.1.3.3 Catalase Test

The catalase test determines an organism's capacity to produce catalase enzymes. A few hydrogen peroxides (H_2O_2) solutions are dropped onto a clean glass slide. The pure bacterial colony is inoculated from King's B agar and spread on the reagent drop. The reaction was quick, and the result can be observed within 3-5 seconds. Bubble formation indicated a positive result, while negative results showed no bubble formation (Md. Zali & Iberahim, 2021).

3.1.3.4 Potassium Hydroxide (KOH) Test

The potassium hydroxide test determined the difference between Gram-negative and Gram-positive organisms. A few drops of 3% KOH solution are dropped onto a clean glass slide. Fresh bacterial culture was inoculated into the solution for 10 seconds. Bacteria became thick and stringy for Gram-negative bacteria, and no changes for Gram-positive bacteria (Aryal, 2019).

3.1.3.5 Sulfur, Indole and Motility (SIM) Test

SIM test was done to test the capacity of an organism to reduce sulfur, produce indole, and the degree to swim through the SIM agar. A pure colony from King's B agar was streaked using inoculating loop and stabbed to the bottom of the SIM medium tube. The SIM medium was incubated for 24 to 48 hours at 35 °C. After that, observe for motility and sulfur (H_2S) production. As for H_2S positive result is shown by a blackening of the medium along the inoculation line, and the absence of blackening indicates a negative H_2S test. As for the positive motility test, a spreading zone of growth flaring from the inoculation line can be observed and growth restriction to the stab line indicates a negative motility test. After recording sulfur and motility test data, three drops of Kovac's reagent will be dropped on top of the medium for the indole test. A positive indole test is when the development of pink to red is formed on top of the medium surface, while a negative result shows no change of colour on top of the medium surface (Hindmarsh & Geertsma, 2017).

3.1.3.6 Carbohydrate Fermentation Test

The broth was prepared in the test tube with ten mL of carbohydrate sources (glucose, sucrose, mannitol and inositol). After that, a pure colony from King's B agar was inoculated into the broth using an inoculating loop. The positive result of the carbohydrate fermentation test is when there is a colour change in the broth, while the negative result is when the broth remains the same colour after incubating for 24 hours at 37°C (Mizobuchi *et al.,* 2018)

3.1.4 Molecular Characterization

3.1.4.1 DNA Extraction of Bacterial Isolates

DNA extraction of the bacterial isolates was done following the methods described by Kim *et al.* (2014). Each bacterial isolate from the frozen stock was isolated, cultured on nutrient broth, and shaken for 24 to 48 hours at 37°C. Then, genomic DNA was isolated from each nutrient broth isolates using a commercial genomic DNA purification kit according to the manufacturer's protocols (Geneaid Biotech Ltd., Taiwan).

One mL of the above mention overnight culture was centrifuged at 10,000 x g for 3 minutes, and the supernatant was discarded while the pellet was resuspended in 180L GT buffer. 20L of Proteinase K was added and incubated in a water bath at 60 °C for 10 minutes. After that, following up with the Lysis step, 200µL GB buffer will be added to the sample and vortex for about 10-15 seconds. Later, samples were incubated in the water bath for 10 minutes at 70°C.

For the DNA binding step, 200µL of absolute ethanol was added to the sample lysate and shaken vigorously using a vortex for 10-20 seconds. At the same time, 2mL collection tubes were placed in the GD column, and all the mixtures mentioned before were transferred into the GD column. Centrifuge the GD column at 10000 x g for 3 minutes.

The GD column was placed into the new collection tube, 400µL W1 buffer was added, and samples were centrifuged at 10000 x g for 3 minutes. Flow-through was discarded, and 600µL wash buffer was added before the tubes were centrifuged for 3 minutes at 10000 x g to dry the column matrix.Finally, the clean GD was transferred to a 1.5 mL microcentrifuge tube for elution. The previously prepared pre-heated elution buffer (100 µL) was added to the centre of the GD column and let to rest for 3 minutes before centrifuging again at 10000 x g for 1 minute. Finally, the purified DNA is stored at -20 °C for future use.

For the purity and quantification of extracted DNA, extracted DNA will be measured using a NanoDrop spectrophotometer (Thermo Scientific, USA). The spectrophotometer's measurements will be 260nm and 280nm wavelengths. A reading ratio of 1.7-2.0 is accepted as pure DNA if less than 1.7 means proteins, phenol, or other contaminants are present in DNA.

3.1.4.2 PCR Amplification of Bacterial Isolates Using 16S Ribosomal RNA Gene

All extracted bacterial genomic DNA was amplified using 16r rRNA gene universal primer pair: 27F (5'- AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTACGACTT-3') (Sayler *et al*., 2006). The PCR reaction was prepared using 20µl reaction volume consisting 1ng of 1µl of genomic DNA template, 1µl of 0.2µM forward primer (27F), 1µl of 0.2µM reverse primer (1492R), 10µl of REDiant 2X PCR Master Mix (First Base Laboratories, Malaysia) and 7µl of sterile distilled water. The PCR amplification was initiated at 95 °C for 2 min, followed by 30 cycles at 94 °C for 30 sec, 55 °C for 1 min, 72 °C for 2 min and 72 °C for 10 min as the final extension. The amplicon's expected size was ~1500 bp (Sayler *et al.,* 2006).

3.1.4.3 PCR Amplification of *Burkholderia gladioli* Using β-Subunit of DNA Gyrase (*gyrB*) Gene

The presence of *Burkholderia gladioli* was detected using PCR with a pair of primers to detect specific DNA fragments corresponding to the gyrB nucleotide sequence, gla-FW (5'-CTGCGCCTGGTGGTGAAG-3') and reverse: gla-RV (5'-CCGTCCCGCTGCGGAATA-3') (Mulaw et al., 2018). The PCR mixture was prepared using 20µl reaction volume consisting 1ng of 1µl of genomic DNA template, 1µl of 0.2µM forward primer (glA-FW), 1µl of 0.2µM reverse primer (glA-RV), 10µl of REDiant 2X PCR Master Mix (First Base Laboratories, Malaysia) and 7µl of sterile distilled water. The PCR amplification was initiated at 94°C for 2 min, followed by 35 cycles at 94°C for 1 min, 63°C for 1 min, and

72°C for 1 min and 72°C for 10 min as the final extension. The expected size of the amplicon was ~479 bp (Mulaw *et al.,* 2018).

3.1.4.4 PCR Amplification of *Burkholderia glumae* Using β-Subunit of DNA Gyrase (*gyrB*) Gene

Conventional PCR analyses for the *gyrB* gene regions of *Burkholderia glumae* were carried out using specific primers; glu-FW (5'-GAAGTGTCGCCGATGGAG-3') and reverse: glu-RV (5'-CCTTCACCGA CAGCACGCAT-3') (Mulaw et al., 2018). The PCR mixture was prepared using 20µl reaction volume consisting 1ng of 1µl of genomic DNA template, 1µl of 0.2µM forward primer (glu-FW), 1µl of 0.2µM reverse primer (glu-RV), 10µl of REDiant 2X PCR Master Mix (First Base Laboratories, Malaysia) and 7µl of sterile distilled water.

The PCR amplification was initiated at 94°C for 2 min, followed by 35 cycles at 94°C for 1 min, 63°C for 1 min, and 72°C for 1 min and 72°C for 10 min as the final extension. The amplicon's expected size was ~530 bp (Mulaw *et al.,* 2018).

3.1.4.5 PCR Amplification of *Burkholderia glumae* Using Specific Burkholderia *glumae* Species (*toxB*) Gene

To further verify, all 22 strains of *Burkholderia glumae* were determined as described by (Bangratz *et al.,* 2020). Closely related species PCR was carried out using specific primer pair of toxB_F (5'- GCATTTGAAACCGAGATGGT-3')

and toxB_RD(5'-TCGCATGCAGATAACCRAAG-3'). The PCR mixture was prepared using 20µl reaction volume consisting 1ng of 1µl of genomic DNA template, 1µl of 0.2µM forward primer (toxB_F), 1µl of 0.2µM reverse primer (toxB_RD), 10µl of REDiant 2X PCR Master Mix (First Base Laboratories, Malaysia) and 7µl of sterile distilled water. The PCR amplification was initiated at 95°C for 12 min, followed by 30 cycles at 94°C for 30 sec, 58°C for 30 sec, 72°C for 45 sec and 72°C for 7 min as the final extension. The primer was designed based on the gene within the toxoflavin biosynthesis operon. The amplicon's expected size was approximately 508 bp (Bangratz *et al.,* 2020). Summarize of primer used in our research presented in Table 3.2.

Table 3.2: List of primers used in this study.

Primers		Sequence (5'-3')	Size (bp)	Specify	References
16r rRNA	**27F**	AGAGTTTGATCCTGGCTCAG	1500	Universal primer	Sayler *et al*., 2006
	1492R	GGTTACCTTACGACTT			
Gyrase B subunit (*B. glumae*)	**glu-FW**	GAAGTGTCGCCGATGGAG	530	Specific *gyrb* gene of *B. glumae*	Mulaw *et al.,* 2018
	glu-RV	CCTTCACCGACAGCACGCAT			
Gyrase B subunit (*B. gladioli*)	**gla-FW**	CTGCGCCTGGTGGTGAAG	479	Specific *gyrb* gene of *B. gladioli*	Mulaw *et al.,* 2018
	Gla-RV	CCGTCCCGCTGCGGAATA			
ToxB	**toxB_F**	GCATTTGAAACCGAGATGGT	508	Specific Toxoflavin biosynthesis operon (*toxB*)	Bangratz *et al.,* 2020
	toxB_RD	TCGCATGCAGATAACCRAAG			

3.1.4.6 Detection of PCR Products

The detection and separation of the amplified DNA were checked by using 1% agarose gel electrophoresis (Agarose Biotechnology Grade). A 1% agarose gel was prepared by mixing and boiling 1g agarose powder in 100 ml of 1X Trise-Borate-EDTA (TBE) buffer. Aliquot 5µl of each PCR product was loaded onto 1% TBE agarose gel and electrophoresed at 80V (Power Supply model 1000/500, Major Science, USA) for 30 min. Gels were stained with floro+Red Nucleic Acid Stain (First Base Laboratories, Malaysia) to detect DNA fragments corresponding to PCR amplicons.

3.1.4.7 Sanger DNA Sequencing and Sequence Alignment

All unpurified PCR products were sent for purification and sequencing services (First Base Laboratories, Malaysia). DNA sequences obtained were assembled using Bioedit 7.2 software and were analysed against sequences in the NCBI database using BLASTn (Nucleotide Basic Local Alignment Search Tool). The DNA sequence were check for similarity, query coverage, and E-values to ensure a confident identification. Pay attention to the percent identity between query sequence and the top BLAST hits, as well as the extent of query coverage. When find a high degree of similarity, substantial query coverage, and low E-values associated with a reference sequence from a known species, can confidently declare the species identified. All sequences were then deposited in the GenBank database and obtained accession numbers.

3.1.4.8 Phylogenetic analyses

All sequences obtained were aligned and used to build a phylogenetic tree using the Maximum Likelihood method generated by MEGA Software version 11.0 (https://www.megasoftware.net/). MEGA software's "Find Best DNA Models" tools were used to find the best-fit substitutions model and calculate genetic distance and sequence similarity (Kumar *et al*., 2018). Branching confidence analysis was performed with 1000 bootstrap replicates to assess the robustness of the tree topologies. The edited 16r rRNA, *gyrB*, and *toxB* sequences were compared to reference and outgroup strains acquired from the Genbank database.

3.1.5 Hypersensitivity Test

As a preliminary assessment, all twenty-seven isolates identified as *B. glumae* and *B. gladioli* were tested for pathogenicity level using tobacco plants (Mulaw et al., 2018). Tobacco plants were chosen for hypersensitivity testing in our study because they are well-suited for this purpose. They are simple to cultivate, have a short growth cycle, and exhibit a significant hypersensitive reaction when exposed to pathogens or certain stimuli. This makes them a great model for researching how plants respond to threats. Tobacco plants were grown in a glasshouse at Institut Kajian Perladangan, UPM, for four weeks at an average regulated temperature (day: 35 °C, night: 25 °C) and relative humidity of 70-100%. 0.5 mL of 10^8 CFU/mL bacterial suspension from each bacterial isolate was injected into three sites of tobacco leaves using 1 mL sterile syringes. In contrast, control leaves were inoculated with sterile

distilled water. One week after inoculation, the diameters of the necrosis lesions were measured using the four-category disease scales described in Table 3.3. Twelve highly virulent strains were chosen for rice plant pathogenicity test following disease scoring.

Table 3.3: Scoring system for tobascco leaves based on the level of hypersensitivity reaction to each bacterial isolate in glasshouse inoculation tests

Scale	Symptom	No. of strains
0	No signs of disease were produced after inoculation	
1	Slightly browning around the injected site with less than 0.5 cm diameter	
2	Distinct necrosis on injected site with 0.5-1 cm diameter	
3	Lesion spreading from the injection with more than 1 cm diameter, and some have the whole leaf get wilted completely	

(Source: Mulaw *et al.,* 2018)

3.1.6 Pathogenicity Test

Ten bacterial strains of *B. glumae* (BGK.AZ1, BGK.AZ4, BGK.AZ7, BGP.A2, BGP.A4, BGP.A6, BGS.AS1, BGS.AS3, BGS.A4 and BGS.A5) and two bacterial strains of *B. gladioli* (UPMBG7 and UPMBG8) were cultured onto King's B agar and incubated for 24h at 35°C. The colonies were then harvested using a sterile inoculating loop and suspended in a sterile distilled water test tube. The bacterial suspension concentration was then adjusted to 10^8 CFU/mL for inoculation (Nandakumar *et al.,* 2009). Rice seedlings of MR219 were planted in the glasshouse at UPM's Institut Kajian Perladangan under average regulated temperatures (day: 35 °C, night: 25 °C) and humidity (70–100%) conditions.

Method for the pathogenicity test was by injecting 10mL of 10^8 CFU/ml bacterial suspension from *B. glumae* and *B. gladioli* isolates into the panicles and crowns of 75-day-old rice seedlings of the MR219 using a sterile syringe. Control rice seedlings inoculated with sterilized water. The experiment used Randomized Complete Block Design with three replicates of each isolate (Figure 3.2).

A	B	C	D	E	F	BLOCK 1
G	H	I	J	K	L	

C	F	A	G	B	L	BLOCK 2
J	D	I	E	H	K	

B	G	E	K	A	I	BLOCK 3
F	J	C	H	D	L	

Figure 3.2: A Sample Layout of Randomized Complete Block Design with Ten Treatments (A, B, C, D, E, F, G, H, I, J). Each block represented a replicate

3.1.6.1 Disease Severity Assessment

The disease reaction/severity on panicles of the infected plant was documented and rated up to 4 weeks after inoculation using the four-category disease scale based on Nandakumar *et al.* (2009) provided in Table 3.4 and Table 3.5 with some modifications.

Table 3.4: Scoring system for panicle blight disease based on symptoms caused by *Burkholderia* species in greenhouse inoculation tests and numbers of isolated strains falling into each category

Scale	Symptom
0	No signs of disease on the panicle or florets
1	0-20% of florets turn brown with most grain filling normally
2	21-40% of the florets turn brown with most grain filling normally
3	41-60% on most florets turns brown and some grain not filing
4	61-80% of panicles poorly emerging, florets discoloured or with dark brown base with many florets sterile
5	81-100% of panicles poorly emerging, florets discoloured or with dark brown base with many florets sterile and panicle tending to remain upright due to lack of grain filling.

(Source: Nandakumar et al., 2009)

The percentage of the disease severity was calculated using a formula based on Lalithiya *et al.* (2017) with some minor modifications:

$$\text{Severity}(\%) = \frac{\text{Number of units x disease scales}}{\text{Total units observed in a set x maximum grade}} \times 100$$

Table 3.5: Score to determine the virulent level of *Burkhorlderia* spp.

Disease Severity (%)	Virulent Level
0	Avirulent
1-20	Low virulent
21-40	Low virulent
41-60	Intermediate virulent
61-80	High virulent
81-100	High virulent

3.1.6.2 Statistical Analysis

All statistical analysis data were analyzed using Statistical Analysis Software (version 9.4). The value of the disease severity level index was calculated by using a one-way Analysis of Variance (ANOVA) and multiple comparison test (HSD/Tukey's test) to test for the significance differences between groups at a 95% confidence level ($P<0.05$). The percentage of disease severity was used to construct the area under the disease progress curve (AUDPC) proposed by Shanner & Finney (1977) for each rice variety using the following formula:

$$AUDPC=\Sigma\lfloor yi+1+yi/2\rfloor ni=1\times(xi+1+xi)$$

Where y_i= disease severity at the ith observation, xi= time at the ith observation, n = total number of observations.

3.1.6.3 Re-isolation (Koch's Postulate)

To complete Koch's postulates, bacterial isolates were re-isolated from diseased rice plants after disease scoring. The bacterial isolates from the disease rice plant will be cultured onto King's B agar at 35 °C for 24h.

3.2 Results

3.2.1 Field Symptoms of BPB Disease

Bacteria panicle blight (BPB) was observed as grain rotting and seedling blight on panicle with florets exhibiting dark grey or reddish-brown base and panicle frequently remaining upright due to loss of grain weight. Other symptoms can be observed on the sheath of an infected tiller with a lesion several centimetres long with a gray and necrotic center and a reddish-brown border. Also, leaves and spikelets contain small circular to oval tan lesions ranging from 1 to 5 mm with brown borders (Figure 3.3). Twenty-three isolates of panicle blight of rice caused by *Burkholderia* spp. used in this study were collected from diseased rice samples from Selangor, Kedah and Perak. From the 23 collected rice samples, all infected samples (4 samples from Yan, 4 samples from Kodiang, 12 samples from Sekinchan and 7 samples from Seberang Perak) showed symptoms of BPB (Table 3.6).

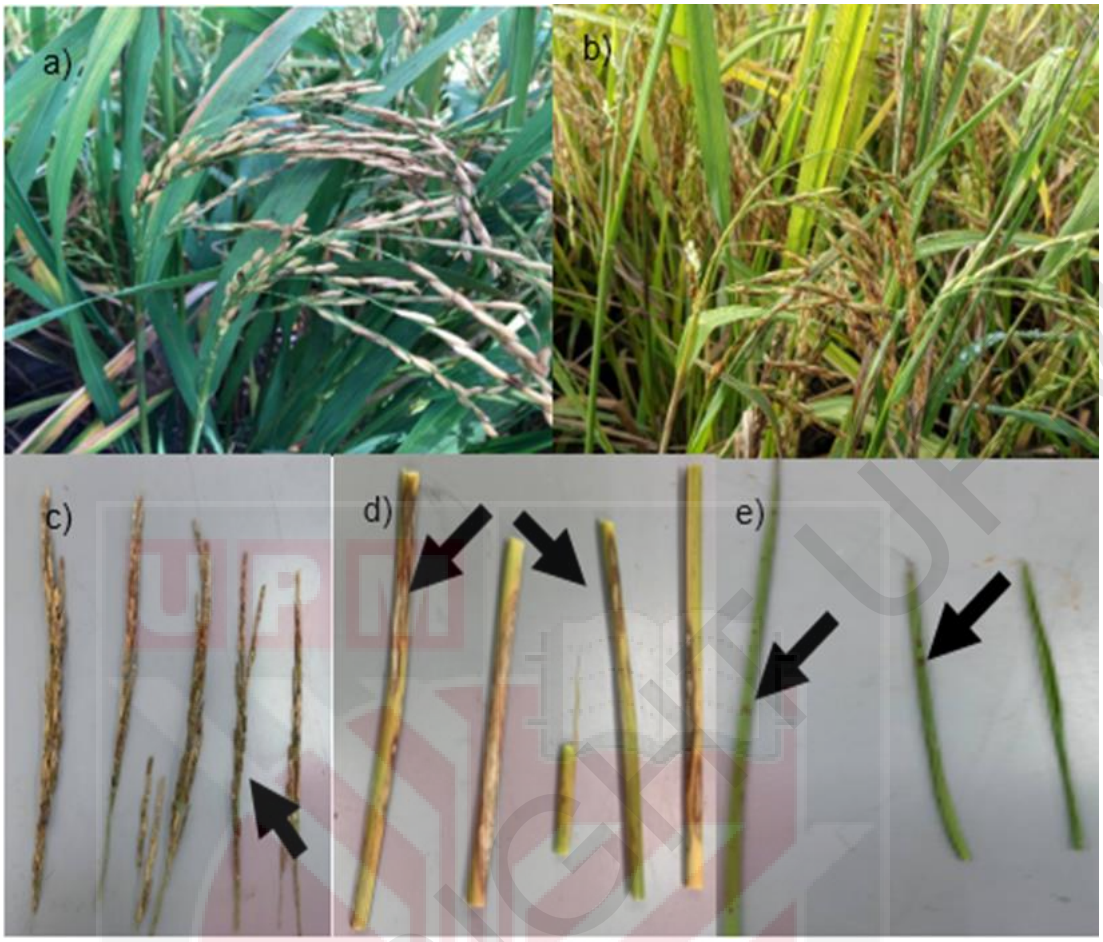

Figure 3.3: Symptoms of Leaf Blight on Rice Field. (a) and (b) Rice crop in Sekinchan and Yan showed BPB disease outbreak; (c) Panicles containing floret with dark grey and a reddish-brown lesion across the floret; (d) Leaf sheath has a lesion several centimetres long with grey necrosis; (e) Leaves contains small circular tan lesion with brown border

Table 3.6: Sampling state, rice varieties, and isolation source of 23 Isolates of *Burkholderia* spp. used in this study.

Number of isolates	Isolates name	Isolation source	Sampling state	Rice varieties
1	BGK.AZ1	Panicle	Yan, Kedah	MR220CL2
2	BGK.AZ2	Panicle	Yan, Kedah	MR220CL2
3	BGK.AZ3	Panicle	Yan, Kedah	MR306
4	BGK.AZ4	Panicle	Yan, Kedah	MR306
5	BGK.AZ5	Panicle	Kodiang, Kedah	MR269
6	BGK.AZ6	Leaves	Kodiang, Kedah	MR269
7	BGK.AZ7	Sheath	Kodiang, Kedah	MR269
8	BGK.AZ8	Sheath	Kodiang, Kedah	MR269
9	BGS.AS1	Panicle	Sekinchan, Selangor	MR297
10	BGS.AS2	Panicle	Sekinchan, Selangor	MR297
11	BGS.AS3	Panicle	Sekinchan, Selangor	MR297
12	BGS.AS4	Panicle	Sekinchan, Selangor	MR297
13	BGS.AS5	Panicle	Sekinchan, Selangor	MR269
14	BGS.AS6	Panicle	Sekinchan, Selangor	MR269
15	BGS.AS7	Sheath	Sekinchan, Selangor	MR269
16	BGP.A1	Panicle	Seberang Perak, Perak	MR297
17	BGP.A2	Panicle	Seberang Perak, Perak	MR297
18	BGP.A3	Panicle	Seberang Perak, Perak	MR297
19	BGP.A4	Panicle	Seberang Perak, Perak	MR297
20	BGP.A5	Panicle	Seberang Perak, Perak	MR297
21	BGP.A6	Panicle	Seberang Perak, Perak	MR297
22	BGP.A7	Leaves	Seberang Perak, Perak	MR297
23	UPMBG7	Panicle	Sekinchan, Selangor	MR307
24	UPMBG8	Panicle	Sekinchan, Selangor	MR307
25	UPMBG9	Panicle	Sekinchan, Selangor	MR307
26	UPMBG15	Panicle	Sekinchan, Selangor	MR307
27	UPMBG17	Panicle	Sekinchan, Selangor	MR307

3.2.2 Isolation and Morphological Characterization of Bacteria

Twenty-two isolates of *Burkholderia glumae* and five of *Burkholderia gladioli* were successfully isolated from symptomatic panicle blight of rice in Kedah, Perak and Selangor. All the 27 colonies (isolated from the 27 rice samples) were identified accordingly to Mizobuchi *et al.* (2016) to their morphological characteristics. After inoculating onto NA agar and subculturing onto King's B agar, the pure colonies were observed as circular and cream colour, flat with smooth margins. Also, they produced a diffusible yellow pigment on agar (toxoflavin) after incubating for 24 to 48 hr at 41°C.

However, *B. gladioli* have vigorous yellow pigment on agar compared to *B. glumae* (Figure 3.4).

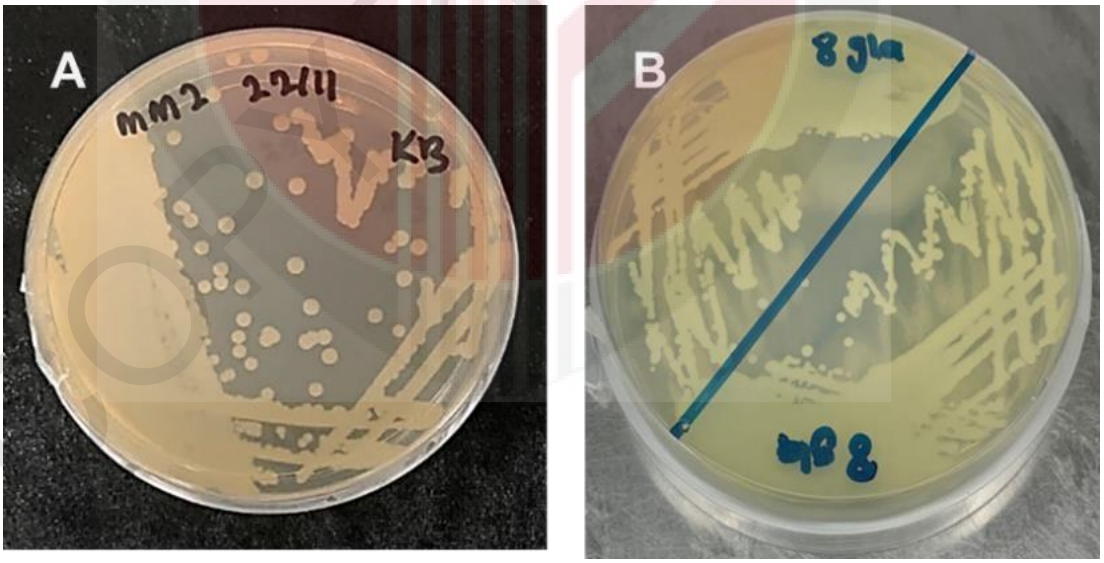

Figure 3.4: Colony morphology of a) *Burkholderia glumae* and b) *Burkholderia gladioli* isolated from infected rice on a King's B agar medium

3.2.3 Phenotypic Characterization of Bacterial Isolates

According to Nandakumar et al. (2009), the phenotypic characterization and isolation identification of isolates on King's B agar show a positive of *Burkholderia* spp. (27 isolates) as the results resembled *Burkholderia* spp. features (Table 3.7, Figure 3.5a and Figure 3.5b).

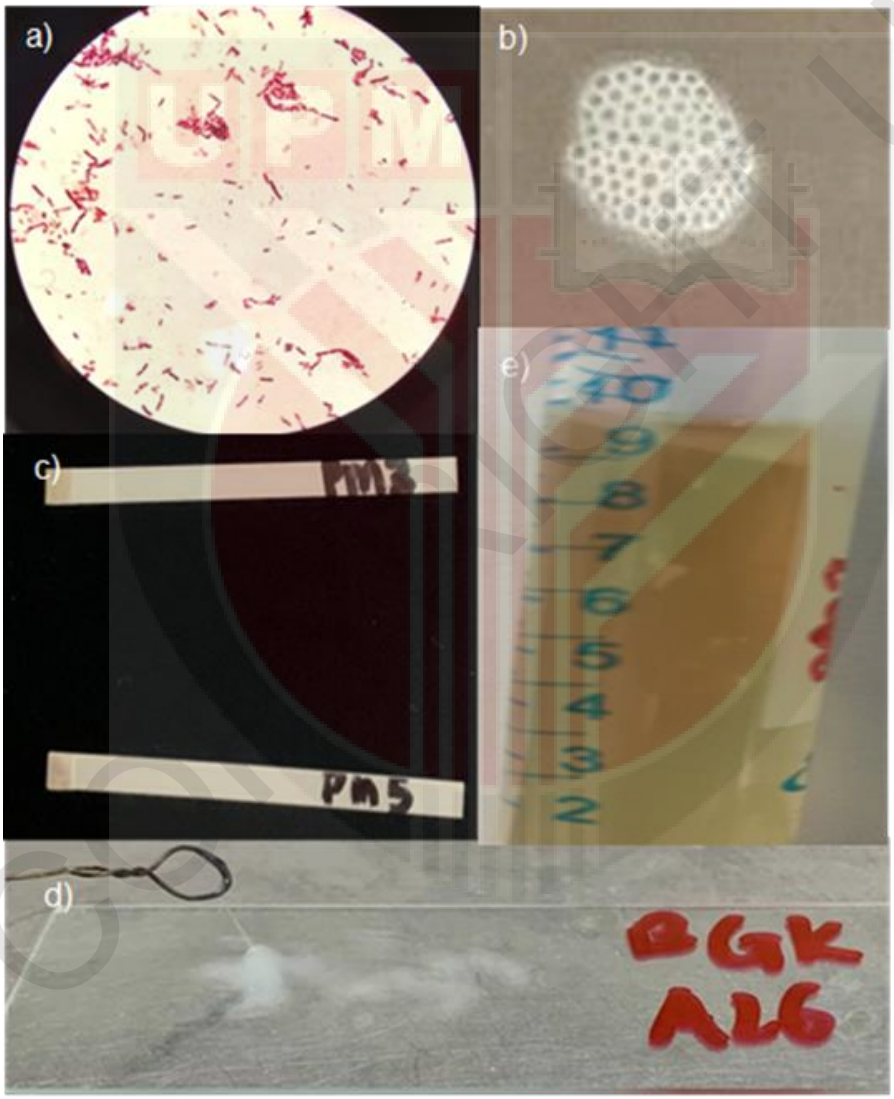

Figure 3.5: Phenotypic test performed on the 27 bacterial isolates (a) Gram-staining; (b) 3% catalase test; (c) Kovacs oxidase test; (d) 3% potassium hydroxide (KOH) test; (e) Sulfur Indole Motility test.

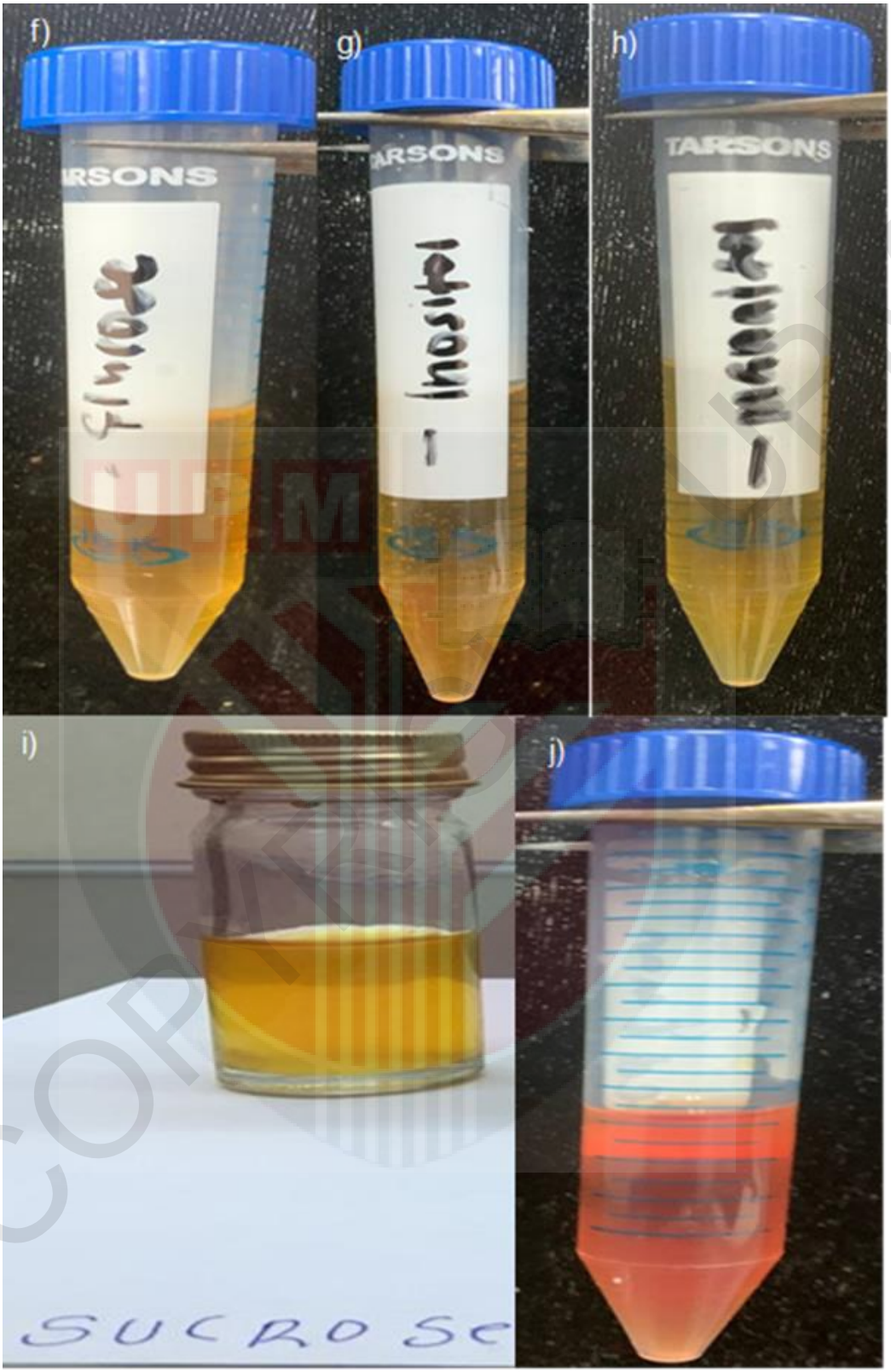

Figure 3.6: Carbohydrate Test; (f) Positive Glucose test; (g) Positive Inositol test; (h) Positive Mannitol test; (i) Positive Sucrose test; (j) Negative Control reaction

Table 3.7: Phenotypic characterization on 28 isolates of *Burkholderia* species isolated from infected rice samples

Number of isolates	Isolates-names	Gram-stain	Catalase	Oxidase	KOH	Indole Production	Motility	Sulfur production	Glucose	Inositol	Mannitol	Carbohydrate Sucrose
1	BGK.AZ1	-	+	-	+	-	+	-	+	+	+	+
2	BGK.AZ2	-	+	-	+	-	+	-	+	+	+	+
3	BGK.AZ3	-	+	-	+	-	+	-	±	±	+	+
4	BGK.AZ4	-	+	-	+	-	+	-	±	±	+	+
5	BGK.AZ5	-	+	-	+	-	+	-	+	+	+	+
6	BGK.AZ6	-	+	-	+	-	+	-	+	+	±	+
7	BGK.AZ7	-	+	-	+	-	+	-	+	+	±	+
8	BGK.AZ8	-	+	-	+	-	+	-	+	+	±	+
9	BGS.AS1	-	+	-	+	-	+	-	+	+	±	+
10	BGS.AS2	-	+	-	+	-	+	-	+	+	±	+
11	BGS.AS3	-	+	-	+	-	+	-	+	+	+	±
12	BGS.AS4	-	+	-	+	-	+	-	+	+	+	±
13	BGS.AS5	-	+	-	+	-	+	-	+	+	+	+
14	BGS.AS6	-	+	-	+	-	+	-	+	+	+	+
15	BGS.AS7	-	+	-	+	-	+	-	±	+	+	+
16	BGP.A1	-	+	-	+	-	+	-	+	+	+	+
17	BGP.A2	-	+	-	+	-	+	-	+	+	+	+
18	BGP.A3	-	+	-	+	-	+	-	+	+	+	+
19	BGP.A4	-	+	-	+	-	+	-	+	+	+	+
20	BGP.A5	-	+	-	+	-	+	-	+	+	+	+
21	BGP.A6	-	+	-	+	-	+	-	+	±	+	+
22	BGP.A7	-	+	-	+	-	+	-	+	±	+	+
23	UPMBG7	-	+	-	+	-	+	-	+	±	+	+
24	UPMBG8	-	+	-	+	-	+	-	+	±	+	±
25	UPMBG9	-	+	-	+	-	+	-	+	+	+	±
26	UPMBG15	-	+	-	+	-	+	-	+	+	+	±
27	UPMBG17	-	+	-	+	-	+	-	+	+	+	±

(Note: + = Positive reaction; - = Negative reaction, ± = weak reaction)

3.2.4 Molecular Characterization and Phylogenetic Analysis

3.2.4.1 PCR Amplification of Bacterial Isolates Using 16S Ribosomal DNA Gene

PCR amplification of all 27 isolates using 16S Ribosomal DNA Gene (rDNA) primer pair showed approximately 1500 bp amplicon (Figure 3.6). PCR amplification showed no contamination occurred as no amplicon was obtained in the negative control reaction. The PCR product was then sequenced using the Sanger sequencing method. The obtained sequence was aligned and compared with known sequences from the Genbank database using BLASTn. BLASTn query showed that the 16S rRNA gene sequence of all bacterial isolates share 95% to 99% identity with *Burkholderia* spp. (Table 3.8). The sequenced have also been deposited into Genbank for accession numbers (Table 3.9). Phylogenetic analysis based on the 16S rRNA gene (Figure 3.6) using Maximum Likelihood showed two clusters of *Burkholderia* spp. Cluster I comprised *B. glumae* strains isolated in this study and were 100% clustered with the *B. glumae* reference strains 411gr-6, GX and P1-22-1. *B. gladioli* (Cluster II) strains isolated in this study were 100% clustered with the *B. gladioli* reference strains CIP 105410, NBRC 13700, CFBP 2427, FDAARGOS 389, and B27. The outgroup sequence of 16S rRNA used in this study was *Pseudomonas aeruginosa* strain DSM 50071. The list of origin and characterization of *Burkholderia* spp. and reference strains used in this study based on the 16S rRNA gene sequences and their accession numbers in the Genbank database are listed in Table 3.9.

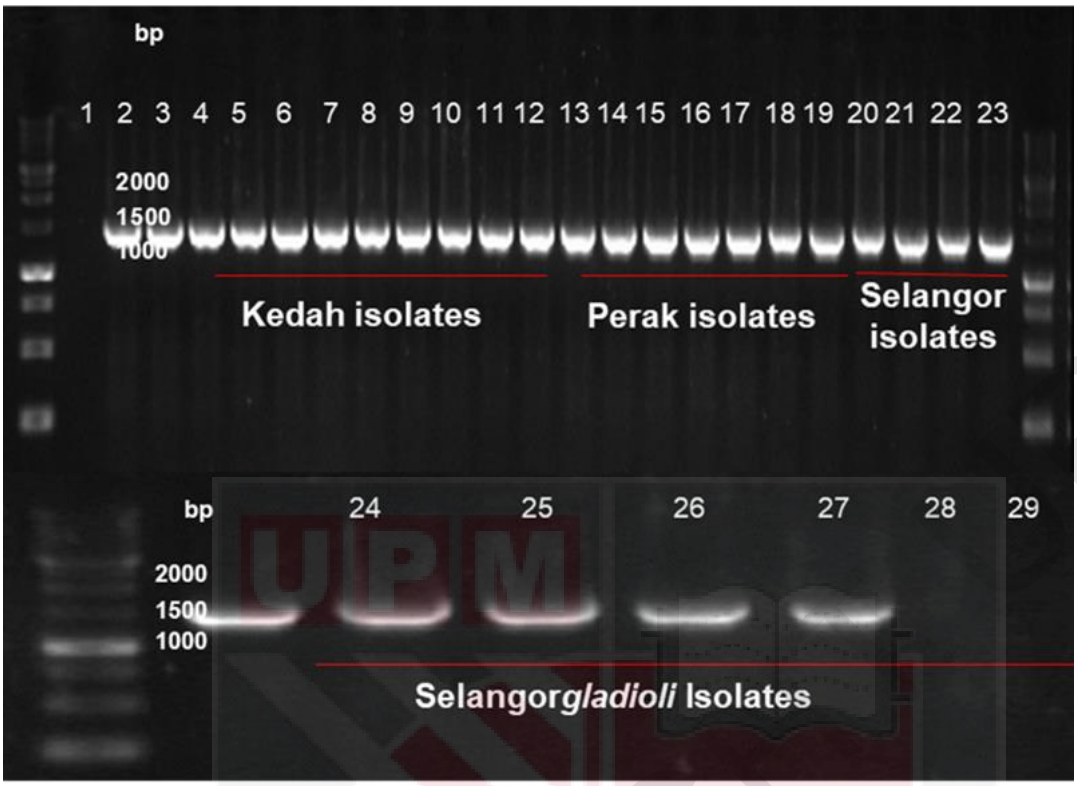

Figure 3.7: Gel electrophoresis showing representative amplicon bands of PCR product from 27F/1492R 16r rRNA gene primer region. The amplification fragments were ~1500 bp. Lane 2 to 9: *B. glumae* Strains from Kedah; Lane 10 to 16: *B. glumae* Strains from Perak; Lane 17 to 23: *B. glumae* strains from Selangor; Lane 24 to 28: *B. gladioli* strains from Selangor and Lane 1 and 29 represent the Negative control

Table 3.8: BLASTn result of 27 bacterial isolates using 16S rRNA gene primer.

State	Strain names	Identity (%)	NCBI accession No	Reference bacteria
Kedah	BGK.AZ1	99	OK632514	*Burkholderia glumae*
	BGK.AZ2	99	OK632515	*Burkholderia glumae*
	BGK.AZ3	95	OK632516	*Burkholderia glumae*
	BGK.AZ4	99	OK631741	*Burkholderia glumae*
	BGK.AZ5	99	OL347392	*Burkholderia glumae*
	BGK.AZ6	99	OL347393	*Burkholderia glumae*
	BGK.AZ7	96	OL347394	*Burkholderia glumae*
	BGK.AZ8	99	OL347395	*Burkholderia glumae*
Selangor	BGS.AS1	99	ON062124	*Burkholderia glumae*
	BGS.AS2	99	ON062125	*Burkholderia glumae*
	BGS.AS3	99	ON062126	*Burkholderia glumae*
	BGS.AS4	99	ON062127	*Burkholderia glumae*
	BGS.AS5	99	ON062128	*Burkholderia glumae*
	BGS.AS6	99	ON062129	*Burkholderia glumae*
	BGS.AS7	99	ON062130	*Burkholderia glumae*
Perak	BGP.A1	99	ON062131	*Burkholderia glumae*
	BGP.A2	99	ON062132	*Burkholderia glumae*
	BGP.A3	99	ON062133	*Burkholderia glumae*
	BGP.A4	99	ON062134	*Burkholderia glumae*
	BGP.A5	99	ON062135	*Burkholderia glumae*
	BGP.A6	99	ON062136	*Burkholderia glumae*
	BGP.A7	99	ON062137	*Burkholderia glumae*

Table 3.8: Continued

Selangor	UPMBG7	99	OM869953	*Burkholderia gladioli*
	UPMBG8	99	OM869954	*Burkholderia gladioli*
	UPMBG9	99	OM869955	*Burkholderia gladioli*
	UPMBG15	99	OM869956	*Burkholderia gladioli*
	UPMBG17	99	OM869957	*Burkholderia gladioli*

Table 3.9: Origin and characterization of *Burkholderia* spp. strains from Malaysia and reference strains using 16S rRNA primer in this study

Strains Name	State	Country	Host	Species	GenBank accession	Reference
BGK.AZ1	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OK632514	This study
BGK.AZ2	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OK632515	This study
BGK.AZ3	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OK632516	This study
BGK.AZ4	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OK631741	This study
BGK.AZ5	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OL347392	This study
BGK.AZ6	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OL347393	This study
BGK.AZ7	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OL347394	This study

Table 3.9: Continued

BGK.AZ8	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	OL347395	This study
BGS.AS1	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062124	This study
BGS.AS2	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062125	This study
BGS.AS3	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062126	This study
BGS.AS4	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062127	This study
BGS.AS5	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062128	This study
BGS.AS6	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062129	This study
BGS.AS7	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062130	This study
BGP.A1	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062131	This study

Table 3.9: Continued

BGP.A2	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062132	This study
BGP.A3	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062133	This study
BGP.A4	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062134	This study
BGP.A5	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062135	This study
BGP.A6	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062136	This study
BGP.A7	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON062137	This study
UPMBG7	Selangor	Malaysia	*Oryza sativa*	*Burkholderia gladioli*	OM869953	This study
UPMBG8	Selangor	Malaysia	*Oryza sativa*	*Burkholderia gladioli*	OM869954	This study
UPMBG9	Selangor	Malaysia	*Oryza sativa*	*Burkholderia gladioli*	OM869955	This study

Table 3.9: Continued

UPMBG15	Selangor	Malaysia	*Oryza sativa*	*Burkholderia gladioli*	OM869956	This study
UPMBG17	Selangor	Malaysia	*Oryza sativa*	*Burkholderia gladioli*	OM869957	This study
411gr-6	Busan	South Korea	*Oryza sativa*	*Burkholderia glumae*	CP021158	Lee et al., (2021)

Figure 3.8: Phylogenetic tree constructed using Maximum-likelihood tree based on 16S rRNA gene sequence; bootstrap values after 1000 replicates are shown as a percentage. The scale bar below indicates 0.02 substitutions per nucleotide position. The accessions numbers of strain isolates, reference, and outgroup strains were Indicated in the figure. *Pseudomonas aeruginosa* is used as an outgroup

3.2.4.2 PCR Amplification of *Burkholderia gladioli* Using β-Subunit of DNA Gyrase (gyrB) Gene

From the PCR amplification using Universal primer of all *Burkholderia* spp. tested using 27F/1492R primer, approximately produce ~1500 bp amplicon. *B. gladioli* isolates were confirmed using a specific subspecies primer that only replicates the gyrB region of B. gladioli and produces an amplicon of ~470 bp (Figure 3.8). The PCR reaction showed no bands were produced in the negative control with a template of *B. glumae* DNA. BLASTn result using the *gyrB* gene sequence revealed that all strains shared 99% similarity to *B. gladioli* as shown in Table 3.10. All sequence was aligned and submitted into the Genbank database for the accession number (Table 3.10). The list of origin and characterization of *B. gladioli*, *B. glumae,* and reference strains used in this study based on the *gyrB* gene sequences and their accession numbers in the Genbank database are listed in Table 3.12.

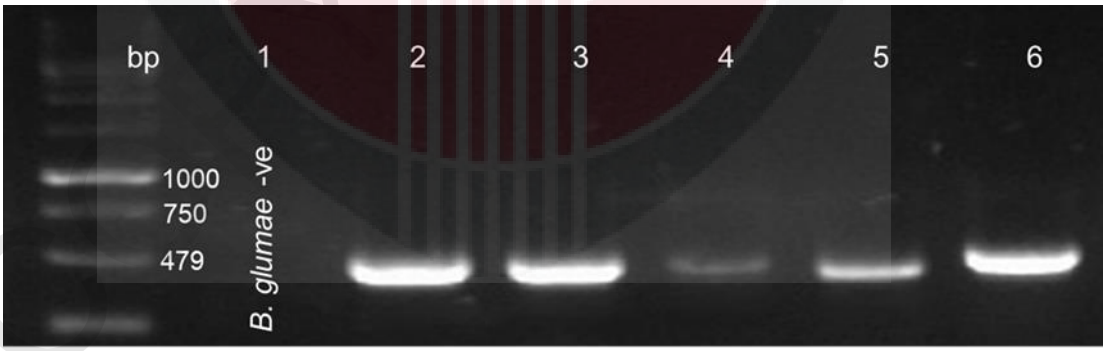

Figure 3.9: Gel electrophoresis showing representative amplicon bands of PCR product from Gla-FV/GLa-RV *gyrb* gene primer region. The amplification fragments were ~470 bp. Lane 1 represents the Negative control. Lane 2 to 6: *B. gladioli* strains from Selangor Lane.

Table 3.10: BLASTn result of bacterial isolates using *gyrB B. gladioli* gene primer

State	Strain names	Identity (%)	NCBI accession No	Reference bacteria
Selangor	UPMBG7	99	OM824438	*Burkholderia gladioli*
	UPMBG8	99	OM824439	*Burkholderia gladioli*
	UPMBG9	99	OM824440	*Burkholderia gladioli*
	UPMBG15	99	OM824441	*Burkholderia gladioli*
	UPMBG17	99	OM824442	*Burkholderia gladioli*

3.2.4.3 PCR Amplification of *Burkholderia glumae* Using β-Subunit of DNA Gyrase (gyrB) Gene

B. glumae isolates were also confirmed using a specific subspecies primer that only replicates the gyrB region of *B. glumae* and produces an amplicon 0f ~520 bp (Figure 3.9). The PCR reaction also showed no bands were produced in the negative control with a template of *B. gladioli* DNA. BLASTn result using the *gyrB* gene sequence revealed that all strains shared 99% similarity to *B. glumae* as shown in Table 3.11. All sequence was aligned and submitted into the Genbank database for the accession number (Table 3.11).

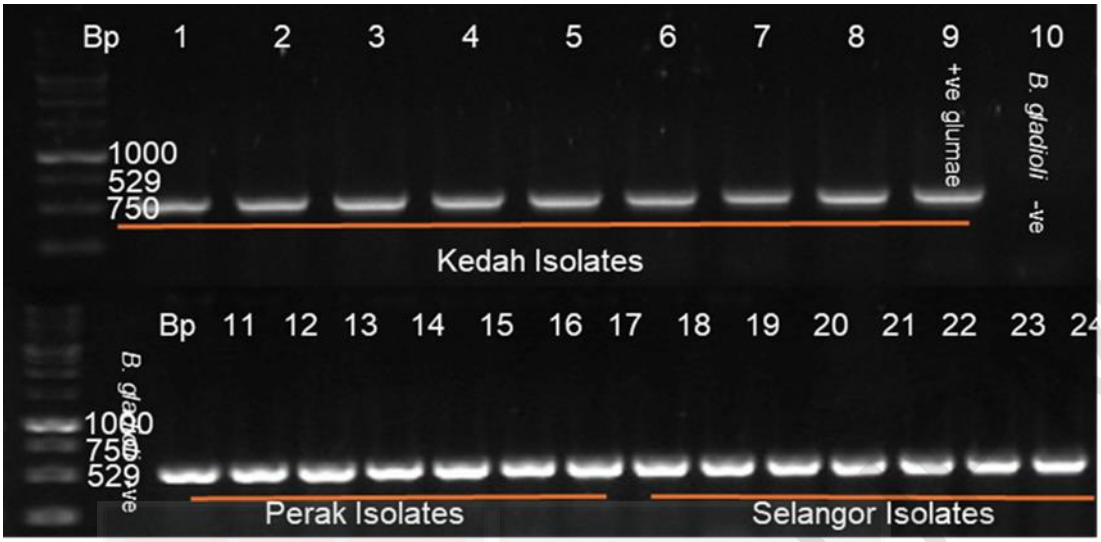

Figure 3.10: Gel electrophoresis showing representative amplicon bands of PCR product from Glu-FV/GLU-RV *gyrB* gene primer region. The amplification fragments were ~520 bp. Lanes 10 and 11 represent the Negative control; Lane 1 to 8: *B. glumae* strains from Kedah; Lane 12 to 18: *B. glumae* strains from Perak; Lane 19 to 25: *B. glumae* strains from Selangor. Lane 9 represents Positive control

Table 3.11: Origin and characterization of *Burkholderia* spp.

State	Strain names	Identity (%)	NCBI accession No	Reference bacteria
Kedah	BGK.AZ1	99	OL461777	*Burkholderia glumae*
	BGK.AZ2	99	OL461778	*Burkholderia glumae*
	BGK.AZ3	99	OL461779	*Burkholderia glumae*
	BGK.AZ4	99	OL461780	*Burkholderia glumae*
	BGK.AZ5	99	OL461781	*Burkholderia glumae*
	BGK.AZ6	99	OL461782	*Burkholderia glumae*
	BGK.AZ7	99	OL461783	*Burkholderia glumae*
	BGK.AZ8	99	OL461784	*Burkholderia glumae*

Table 3.11: Continued

Selangor	BGS.AS1	99	ON086704	*Burkholderia glumae*
	BGS.AS2	99	ON086705	*Burkholderia glumae*
	BGS.AS3	99	ON086706	*Burkholderia glumae*
	BGS.AS4	99	ON086707	*Burkholderia glumae*
	BGS.AS5	99	ON086708	*Burkholderia glumae*
	BGS.AS6	99	ON086709	*Burkholderia glumae*
	BGS.AS7	99	ON086710	*Burkholderia glumae*
Perak	BGP.A1	99	ON086711	*Burkholderia glumae*
	BGP.A2	99	ON086712	*Burkholderia glumae*
	BGP.A3	99	ON086713	*Burkholderia glumae*
	BGP.A4	99	ON086714	*Burkholderia glumae*
	BGP.A5	99	ON086715	*Burkholderia glumae*
	BGP.A6	99	ON086716	*Burkholderia glumae*
	BGP.A7	99	ON086717	*Burkholderia glumae*

Table 3.12: Origin and characterization of *Burkholderia* spp. strains from Malaysia and reference strains using *gyrB* primer in this study.

Strains	State	Country	Host	Species	Accession	Reference
BGK.AZ1	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461777	This study
BGK.AZ2	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461778	This study
BGK.AZ3	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461779	This study
BGK.AZ4	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461780	This study
BGK.AZ5	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461781	This study
BGK.AZ6	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461782	This study
BGK.AZ7	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461783	This study
BGK.AZ8	Kedah	Malaysia	Oryza sativa	*Burkholderia glumae*	OL461784	This study
BGS.AS1	Selangor	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086704	This study
BGS.AS2	Selangor	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086705	This study
BGS.AS3	Selangor	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086706	This study

Table 3.12: Continued

BGS.AS4	Selangor	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086707	This study
BGP.A5	Perak	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086715	This study
BGP.A6	Perak	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086716	This study
BGP.A7	Perak	Malaysia	Oryza sativa	*Burkholderia glumae*	ON086717	This study
UPMBG7	Selangor	Malaysia	Oryza sativa	*Burkholderia gladioli*	OM824438	This study
UPMBG8	Selangor	Malaysia	Oryza sativa	*Burkholderia gladioli*	OM824439	This study
UPMBG9	Selangor	Malaysia	Oryza sativa	*Burkholderia gladioli*	OM824440	This study
UPMBG15	Selangor	Malaysia	Oryza sativa	*Burkholderia gladioli*	OM824441	This study
UPMBG17	Selangor	Malaysia	Oryza sativa	*Burkholderia gladioli*	OM824442	This study
BGR48S	Seoul	South Korea	*Oryza sativa*	*Burkholderia glumae*	CP100285	Kang *et al.*, (2022)

Table 3.12: Continued

DOA-BG14	Nakhon Pathom	Thailand	*Oryza Sativa*	*Burkholderia glumae*	KX213523	Unpublished
BP2-004	Nakhon Pathom	Thailand	*Oryza sativa*	*Burkholderia glumae*	LC474870	Unpublished
MAFF 302533	Kochi	Japan	*Oryza sativa*	*Burkholderia gladioli*	AB190628	Unpublished
LMG19584	Kochi	Japan	Oryza sativa	*Burkholderia gladioli*	AB220898	Unpublished
K 96243	Paris	France	Unknown	*Burkholderia pseudomallei*	EU024223	Tayeb *et al.,* (2008)

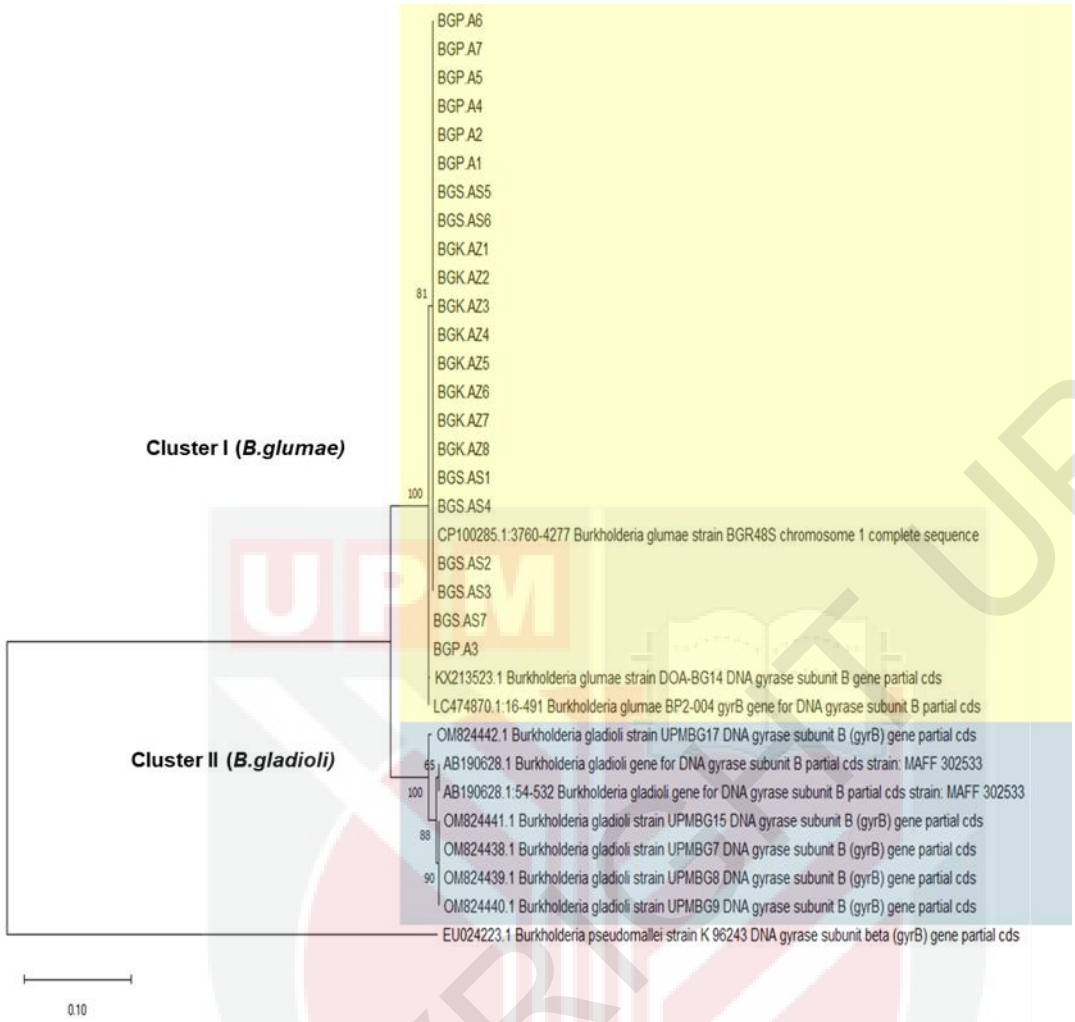

Figure 3.11: Phylogenetic tree constructed using Maximum-likelihood tree based on the gyrb gene sequence; bootstrap values after 1000 replicates are shown as a percentage. The scale bar below indicates 0.10 substitutions per nucleotide position. The accession numbers of strain isolates, reference, and outgroup strains were indicated in the figure. *Burkholderia pseudomallei* are used as an outgroup.

3.2.4.4 PCR Amplification of *Burkholderia glumae* Using Specific *Burkholderia glumae* Species (*toxB*) Gene

For further confirmation of *B. glumae* isolates obtained from this study, PCR amplification using specific species primer gene coding for toxB (Toxoflavin biosynthesis operon) produced an amplicon of ~500 bp (Figure 3.11). PCR products was sequenced, aligned, and compared to other known sequences from the Genbank database using BLASTn. BLASTn results indicate that 22

B. glumae isolates have 99% similarity with *B. glumae* (Table 3.13). Phylogenetic analysis based on the *toxB* gene (Figure 3.12) using Maximum-likelihood showed 100% of the *B. glumae* isolated used in this study under the same clade with *B. glumae* reference strains 411 gr-6, 257sh-1 and HN1. The outgroup strain used is *Burkholderia pseudomallei* strain Yap1. The list of origin and characterization of *B. glumae* and reference strains used in this study based on the *toxB* gene sequences, and their accession numbers in the Genbank database are listed in Table 3.14.

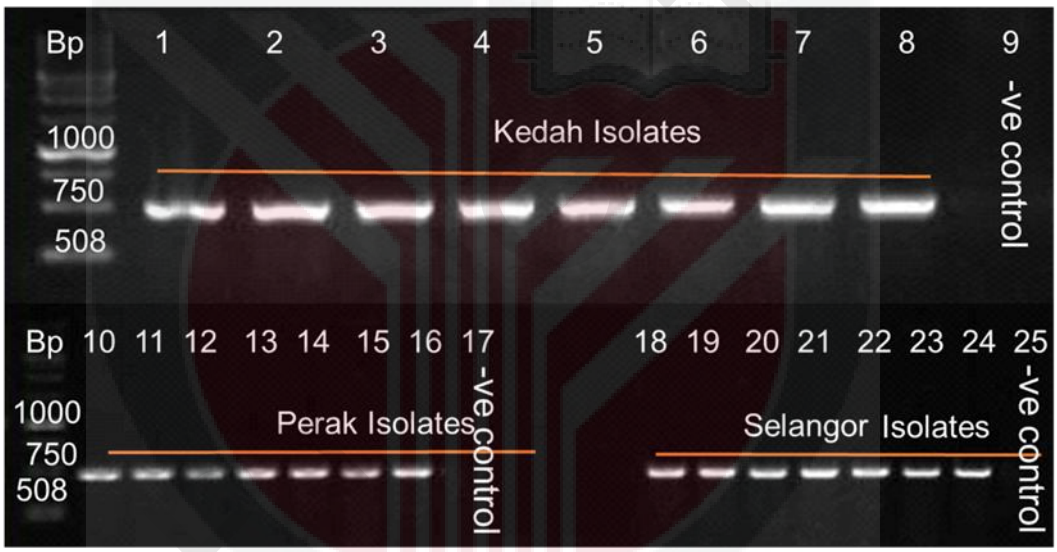

Figure 3.12: Gel electrophoresis showing representative amplicon bands of PCR product from toxB_F/toxB_RD *toxB* gene primer region. The amplification fragments were ~500 bp. Lanes 9,17,25 represent the Negative control; Lane 1 to 8: *B. glumae* strains from Kedah; Lane 10 to 16: *B. glumae* strains from Perak; Lane 18 to 24: *B. glumae* strains from Selangor

Table 3.13: BLASTn result of 22 bacterial isolates using *toxB* gene primer.

State	Strain names	Identity (%)	NCBI accession No	Reference bacteria
Kedah	BGK.AZ1	99	ON075583	*Burkholderia glumae*
	BGK.AZ2	99	ON075584	*Burkholderia glumae*
	BGK.AZ3	99	ON075585	*Burkholderia glumae*
	BGK.AZ4	99	ON075586	*Burkholderia glumae*
	BGK.AZ5	99	ON075587	*Burkholderia glumae*
	BGK.AZ6	99	ON075588	*Burkholderia glumae*
	BGK.AZ7	99	ON075589	*Burkholderia glumae*
	BGK.AZ8	99	ON075590	*Burkholderia glumae*
Selangor	BGS.AS1	99	ON086719	*Burkholderia glumae*
	BGS.AS2	99	ON086720	*Burkholderia glumae*
	BGS.AS3	99	ON086721	*Burkholderia glumae*
	BGS.AS4	99	ON086722	*Burkholderia glumae*
	BGS.AS5	99	ON086723	*Burkholderia glumae*
	BGS.AS6	99	ON086724	*Burkholderia glumae*
	BGS.AS7	99	ON086725	*Burkholderia glumae*
Perak	BGP.A1	99	ON086726	*Burkholderia glumae*
	BGP.A2	99	ON086727	*Burkholderia glumae*
	BGP.A3	99	ON086728	*Burkholderia glumae*
	BGP.A4	99	ON086729	*Burkholderia glumae*
	BGP.A5	99	ON086730	*Burkholderia glumae*
	BGP.A6	99	ON086731	*Burkholderia glumae*
	BGP.A7	99	ON086732	*Burkholderia glumae*

Table 3.14: Origin and characterization of *Burkholderia glumae* strains from Malaysia and reference strains using *toxB* gene

Name	State	Country	Host	Species	GenBank accession	Reference
BGK.AZ1	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075583	This study
BGK.AZ2	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075584	This study
BGK.AZ3	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075585	This study
BGK.AZ4	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075586	This study
BGK.AZ5	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075587	This study
BGK.AZ6	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075588	This study
BGK.AZ7	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075589	This study
BGK.AZ8	Kedah	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON075590	This study
BGS.AS1	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086719	This study
BGS.AS2	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086720	This study
BGS.AS3	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086721	This study

Table 3.14: Continue

BGS.AS4	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086722	This study
BGS.AS5	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086723	This study
BGS.AS6	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086724	This study
BGS.AS7	Selangor	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086725	This study
BGP.A1	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086726	This study
BGP.A2	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086727	This study
BGP.A3	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086728	This study
BGP.A4	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086729	This study
BGP.A5	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086730	This study
BGP.A6	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086731	This study
BGP.A7	Perak	Malaysia	*Oryza sativa*	*Burkholderia glumae*	ON086732	This study
BGR48S	Seoul	South Korea	*Oryza sativa*	*Burkholderia glumae*	CP100285	Kang *et al.*, (2022)

Table 3.14: Continue

411 gr-6	Busan	South Korea	*Oryza sativa*	*Burkholderia glumae*	CP021158	Lee *et al.,* (2021)
257 sh-1	Los Angeles	USA	*Oryza sativa*	*Burkholderia glumae*	CP035901	Lee et al., (2021)
HN1	Shanghai	China	*Oryza sativa*	*Burkholderia glumae*	CP052864	Unpublished
Yap1	Atlanta	USA	Blood	*Burkholderia pseudomallei*	CP038227	Unpublished

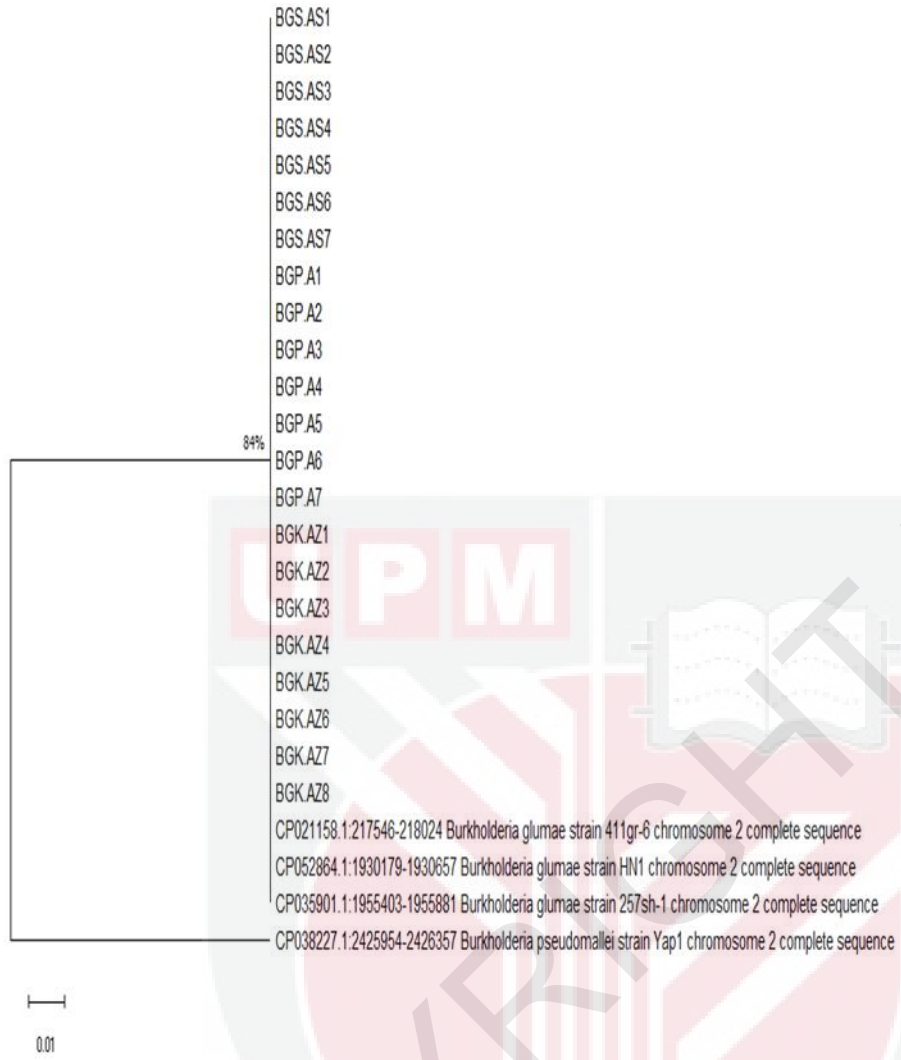

Figure 3.13: Phylogenetic tree constructed using Maximum-likelihood tree based on *toxb* gene sequence; Bootstrap values after 1000 replicates are shown as a percentage. The scale bar below indicates 0.01 substitutions per nucleotide position. The accession numbers of strain isolates, reference, and outgroup strains were indicated in the figure. *Burkholderia pseudomallei* are used as an outgroup

3.2.5 Hypersensitivity Test

Tobacco exposure to pathogenic bacteria demonstrated that each of the 27 isolates studied has a distinct level of pathogenicity (Table 3.15). Approximately 12 isolates (44%) of the 27 are highly pathogenic toward tobacco plants. Nine isolates (33%) were somewhat moderately virulent. The remaining 6 isolates are weakly virulent, and plant injections with control conditions stay healthy with no apparent hypersensitive reaction (Figure 3.13a-c). Pathogenicity tests were later performed on high virulent isolates.

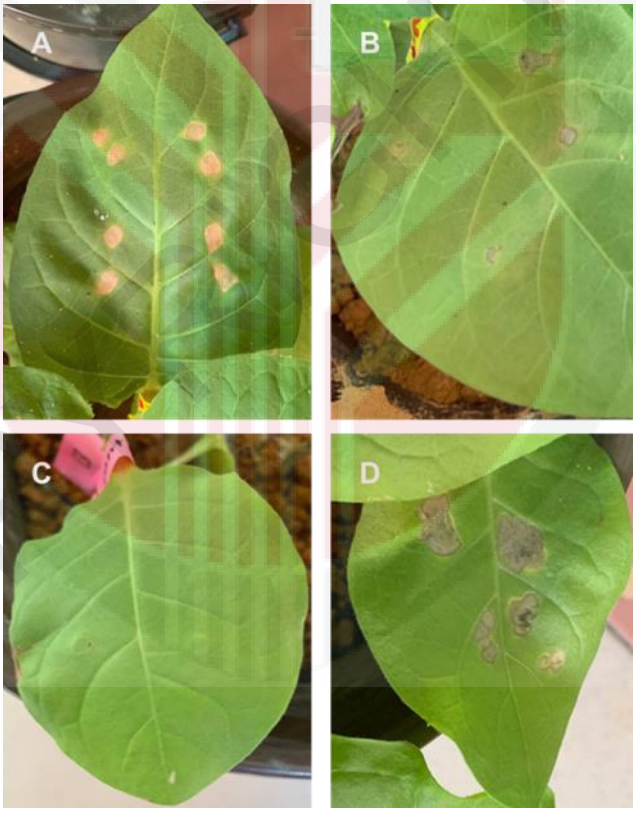

Figure 3.14: Hypersensitive necrosis on tobacco leaves caused by *Burkholderia* spp. (a); Highly virulent level of *Burkholderia* spp. on tobacco plant and (b) Weakly virulent level of *Burkholderia* spp. on tobacco plant (c) Control tobacco plant.

Table 3.15: Virulence level assessed by inoculation of isolates into tobacco leaves to determine their hypersensitive necrosis level.

State	Strain names	Virulent Level
Kedah	BGK.AZ1	3
	BGK.AZ2	1
	BGK.AZ3	2
	BGK.AZ4	3
	BGK.AZ5	2
	BGK.AZ6	1
	BGK.AZ7	3
	BGK.AZ8	2
Selangor	BGS.AS1	3
	BGS.AS2	1
	BGS.AS3	3
	BGS.AS4	3
	BGS.AS5	3
	BGS.AS6	2
	BGS.AS7	1
Perak	BGP.A1	2
	BGP.A2	3
	BGP.A3	1
	BGP.A4	3
	BGP.A5	2
	BGP.A6	3
	BGP.A7	2
Selangor	UPMBG7	3
	UPMBG8	3
	UPMBG9	2
	UPMBG15	1
	UPMBG17	2
Control	Distilled Water	0

0 indicates Healthy plant; 1 Weakly virulent level; 2 Moderate virulent level; 3 Highly virulent level after one-week observation, respectively

3.2.6 Pathogenicity Test

After 4-7 days of inoculation on MR219 cultivars, 36 pots began to develop bacterial panicle blight, whereas the control rice remained healthy. Up to 28DAI, the symptoms were evaluated and documented every 24 hours to notice the browning/lesion on panicle florets. The symptomatic plants first appeared when florets became brown across the floret base and gradually developed into classic panicle blight and florets with a completely dark brown base and sterile florets (Figure 3.14a-d). Besides BPB, no other symptom was seen among the plants with *Burkholderia* spp. infected on the rice. After 28DAI, all plants developed BPB disease, with floret blight expanding consistently from the bottom to the top of the florets. The disease incidence ranges from 27% to 77% due to severe panicle loss and severity progress.

Figure 3.15: BPB Symptom induced by *B. glumae* and *B. gladioli* at 28DAI. (a) Control panicle; (b) Severe infected BPB on MR219; (c) Sheath rot of BPB; (d) Weakly infected of *Burkholderia* spp. on rice.

The statistical analysis of 12 *Burkholderia* spp. pathogenicity tests on rice cultivar MR219 revealed a significant variation in pathogenicity across isolates with a P-value<0.001 (One-Way ANOVA P<0.05). MR219 cultivar proved to be very susceptible to *Burkholderia* spp. *B. gladioli* UPMBG7 reported the highest degree of severity (P<0.05) at 28DAI (77.43%) and an AUDPC score of 1123.57, while *B. glumae* BGS.AS1 has the second highest degree of severity (P<0.05) at 28DAI (68.15%) and an AUDPC score of 948.28. However, *B. glumae* BGK.AZ1 had the lowest severity (27.60%) and AUDPC

score at 28DAI. Table 3.16 displayed the percentage of disease severity and AUDPC value for the cultivars MR219.

According to Figure 3.15, disease severity increased steadily at 7DAI until 28DAI for all isolates. According to the AUDPC graph in Figure 3.16, MR219 was less vulnerable to all *B. glumae* strains from Kedah isolates and BGS.AS3 strain from Selangor isolates.

Table 3.16: Percentage of disease severity and Area Under Disease Progress Curve (AUDPC) on MR219 cultivars for 12 selected isolates *Burkholderia* species on panicles.

Variety	Strain	AUDPC value	Percent Disease Severity (%) after 28 days inoculation (28DAI±SD)	Disease score at 28DAI	Isolate Virulence level
MR219	BGK.AZ1	410.76	27.60±3.98^{D}	2	Low
	BGS.AS3	658.15	47.35±6.25^{C}	3	Intermediate
	BGK.AZ4	666.48	49.19±7.56^{C}	3	Intermediate
	BGP.A2	790.44	49.82±3.06^{BC}	3	Intermediate
	BGS.AS5	779.77	61.01±3.27^{ABC}	4	High
	BGP.A4	927.79	61.28±3.45^{ABC}	4	High
	BGS.AS4	804.10	63.81±5.70^{ABC}	4	High
	BGP.A6	1053.75	63.94±3.23^{ABC}	4	High
	BGK.AZ7	761.79	67.81±4.86^{AB}	4	High
	BGS.AS1	948.28	68.15±7.61^{A}	4	High
	UPMBG8	1081.73	72.36±3.42^{A}	4	High
	UPMBG7	1123.58	77.43±4.87^{A}	4	High

± Average standard deviation from triplicate samples. Means in the same coumns followed by the same letter are not significantly different (P= 0.05)

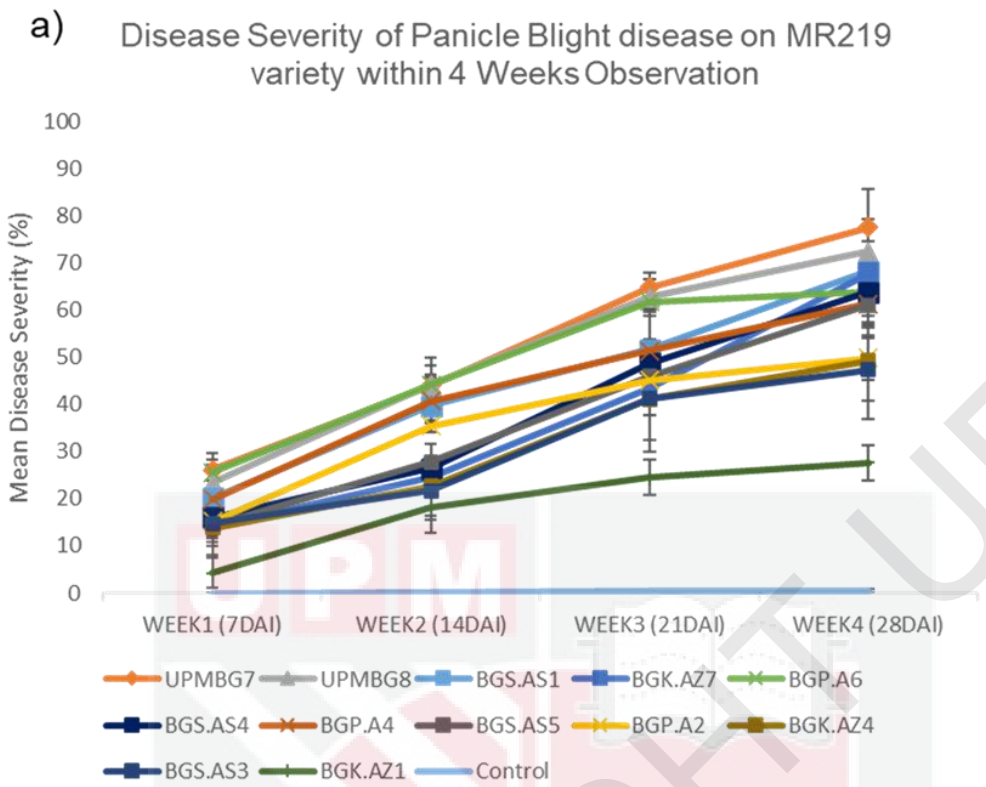

Figure 3.16: The percentage of disease severity on MR219 cultivar after 28dai. (a) Percentage of disease severity of MR219 variety

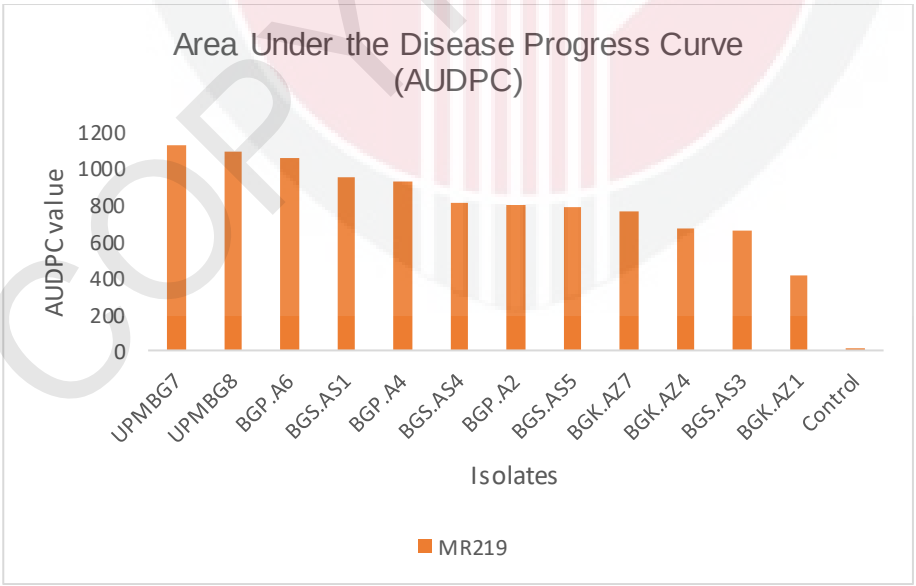

Figure 3.17: Area Under the Disease Progress Curve (AUDPC) of 12 *Burkholderia* isolates inoculated on MR219 cultivar.

Each strain was then re-isolated on King's B agar for pure culture, checked their morphological and biochemical test, and resembled *Burkholderia* spp. Thus, these results showed a positive result to Koch's postulates as the *Burkholderia* spp. caused BPB disease.

3.3 Discussion

This study provides comprehensive identification and characterization methods for detecting *Burkholderia* spp. related to BPB disease of rice. Identification and characterization of *Burkholderia* spp. using phenotypic and molecular approaches is critical for a preliminary investigation. Based on the field symptom observation on rice panicles suspected with the BPB disease were parallel to the previous report in that panicle containing floret with dark gray and a reddish-brown line (border of the lesion) across the floret between the darker and straw-colored areas, leaf sheath has gray, necrotic center lesion with reddish-brown border and up to several centimeters long and leaves showed small circular tan lesions ranging size from 1 to 5 mm with brown borders (Zhou-qi *et al.,* 2016; Nandakumar et al., 2009; Sayler *et al.,* 2006; Ortega and Rojas 2021). However, BPB symptoms caused by *Burkholderia* spp. could not be differentiated due to the symptoms' resemblance.

The purpose of collecting rice cultivars MR220CL2, MR306, MR269, MR297, and MR307, all of which were found to be infected with bacterial panicle blight (BPB) in the field during the sampling reported by the Department of Agriculture (DOA), the rice cultivar was collect, to investigate and isolate the

pathogen related to the BPB infection. Based on isolation and phenotypic characterization, 27 *Burkholderia* spp. from rice in Kedah, Perak, and Selangor were successfully isolated and demonstrated similar and parallel reactions to *Burkholderia* spp. characteristics (Nandakumar *et al.,* 2009, Ramachandran *et al.,* 2021). Two types of media were used to isolate pathogenic bacteria. NA media was used as a general growth medium for bacteria, while King's B medium was used to isolate the *Enterobacteriaceae* family (Mizobuchi *et al.,* 2016). *Burkholderia* spp. morphological features on NA and King's B medium is consistent with the previous finding on morphological of *Burkholderia* spp. related to rice BPB disease (Nandakumar *et al.,* 2009).

Further tests were conducted by performing multiple biochemical tests on the 27 *Burkholderia* isolates. The reason for conducting a battery of biochemical tests is to establish a robust method for differentiating *Burkholderia* strains from other types of bacteria family. These tests are important because they provide critical information about the metabolic, structural, and physiological characteristics of bacteria. Other than that, biochemcical test were done to identified the isolated strain up to genus level. These tests examine the physiological responses of bacteria to various substrates and conditions, providing a biochemical "fingerprint" that can be used for identification. Therefore, employing a comprehensive set of biochemical tests ensures that *Burkholderia* strains are correctly identified. In summary, these biochemical tests serve as a reliable means to distinguish Burkholderia strains from other bacterial types, contributing targeted approaches in identification. Gram-staining was performed in the early investigation to identify the bacteria

isolates' size, shape, and arrangement. An oxidase test was conducted to identify the ability of organisms to produce cytochrome c oxidase enzyme, where the enzyme can oxidize reagents such as Kovac's reagent to purple color. Bacteria that can oxidize positively are usually from the aerobic family, where they can use oxygen as their terminal electron acceptor. Certain aerobe bacteria cannot oxidize, such as *Burkholderia* spp. due to the absence of cytochrome c oxidase enzyme. In contrast, *Burkholderia* spp. can produce catalase enzyme as the catalase enzyme will breakdown the hydrogen peroxide into oxygen and water and produce air bubbles (Tankeshwar, 2012). In the KOH test, the KOH solution dissolves the thin peptidoglycan layer in Gram-negative bacteria's cell wall, causing cell wall lysis and the formation of a sticky and thin strand of slime thread. *Burkholderia* spp. does not produce hydrogen sulfide as the sulfur is not reduced, and no black precipitation was produced in the SIM medium. As for indole production, *Burkholderia* spp. do not have methyl indole present due to the medium remaining yellow after adding the Kovacs reagent. The motility test is critical because *Burkholderia* spp. are motile due to multiple flagella (Peng, 2015). The results revealed that *Burkholderia* spp. is motile, as seen by a hazy growth that spreads across the medium from the stab line. In carbohydrates, fermentation showed that *Burkholderia* spp. is able to ferment the sugar present in the broth as the broth change from red to yellow due to the change of acid pH of broth (Aryal, 2019). There is not a single specific biochemical test that can definitively identify all *Burkholderia* strains. Identification of bacterial strains typically relies on a combination of tests and techniques to achieve accurate results. However,

among the tests that had been done, the oxidase test is often considered specific for differentiating *Burkholderia* strains (Peng, 2015). However, for precise identification and differentiation of *Burkholderia* species, including *B. glumae* and *B. gladioli*, advanced molecular techniques such as PCR assays targeting species-specific genes or genomic sequencing are recommended. These methods can provide more accurate and specific results for distinguishing closely related bacterial species like those within the *Burkholderia* genus.

A molecular characterization investigation was performed for further confirmation. According to prior research, the most often utilized genotypic approach for bacterial identification is the comparative analysis of 16S rRNA sequences (Sayler et al., 2006). The 16S rRNA sequencing revealed 96-99% similarity with *Burkholderia* spp. However, because the sequence exhibits substantial similarities throughout the *Burkholderia* group, using 16rDNA alone to identify bacteria among *Burkholderia* species is insufficient. Because the 16r rRNA is less effective in recognizing *Burkholderia* species, primers to detect *B. glumae* and *B. gladiioli*, as discovered by Maeda *et al.* (2006), are extremely sensitive to direct detection of the target species.

Primer pairs Glu-FW/Glu-RV and Gla-FW/Gla-RV were tested by Maeda *et al.* (2006) on *B. glumae, B. gladioli,* and *B. plantarii* that targeted their specific *gyrB* gene region. Results showed that by using Glu-FW/Glu-RV primers, only *B. glumae gyrB* gene region is amplified and implied for Gla-FW/Gla-RV that only amplified *gyrB* gene region of *B. gladioli*. These primer pairs were also

validated when 22 isolates of *B. glumae* and 5 isolates of *B. gladioli* PCR yielded at the expected amplicon size and 100% identical to their respective species obtained from the Genbank database BLAST.

As for double validation of *B. glumae* isolate obtained, toxB-F/toxB-RD (*toxB* Toxoflavin biosynthesis operon gene region) primers were used (Bangratz *et al.,* 2020). This primer only amplified a specific *toxB* gene region found in chromosome 1 of *B. glumae.* The sequence *is* specific and ideal for detecting and identifying *B. glumae.* This was proved when 22 isolates of *B. glumae* from this study produced the expected amplicon size and 100% similar sequence to *B. glumae* sequence from the Genbank database. For all primer used in this study, the top BLAST hits did not show a mixture of different *Burkholderia* species. Instead, the top BLAST hits corresponded accurately to their respective species.

As for phylogenetic analysis achieved with MEGA11 on all primers used in this study (16r rRNA, *gyrB* and *toxB*), gene sequence successfully clustered *B. glumae* and *B. gladioli,* respectively, to their reference strains. None of the strains in this study clustered with the outgroup or different reference species sequence. Based on the result, identifying *Burkholderia* species using the *gyr*B gene is much more efficient than the 16S rRNA gene sequence as the phylogenetic analysis showed a higher genetic variation in *gyrB* with a nucleotide base substitution of 0.1 compared to 16S rRNA with nucleotides base substitution of 0.02. Other than that, 16S rRNA gene sequences have

high similarity sequences among *Burkholderia* group (Coenye and Vandamme, 2003).

A hypersensitivity test (*Nicotiana tabacum*) was performed on *B. glumae* and *B. gladioli* for preliminary pathogenicity assessment, and results showed all 27 isolates tested are pathogenic at a different level. The tobacco leaves showed local necrosis occurred when inoculated with pathogenic bacteria. According to Mulaw *et al.* (2018), the pathogenic *Burkholderia* spp. caused necrosis on tobacco leaves due to the presence and production of yellow pigment toxoflavin from these pathogenic bacteria. In addition, Lee *et al.* (2016) stated that differential regulation/expression of toxoflavin production effect the virulence of *Burkholderia* spp. This finding was parallel to our results as some strains are highly virulent while some are weak virulence toward tobacco leaves. Thus, we only choose highly virulent strains of tobacco leaves to prevent errors during our pathogenicity test on rice plants for further assessment (Mulaw *et al.,* 2018).

As for the pathogenicity test on the MR219 rice cultivar, there were significant differences in all 12 strains of *Burkholderia* spp. regarding virulency variation. During the pathogenicity test, the decision to exclude the sampling of rice variety MR219 from our field research, while subsequently employing it in our controlled glasshouse experiments, emerges from a strategic alignment with our research objectives and the dynamic context of rice cultivation. At the time of field sampling, it's important to note that the affected areas harbored rice cultivars other than MR219. Our primary goal during field sampling was to

assess the impact of bacterial panicle blight (BPB) on the predominant rice varieties in the specific infected regions. MR219 gained prominence as a widely cultivated rice variety during the time of sampling and studies, particularly in the context of blast disease resistance studies.

Our decision to incorporate MR219 into our glasshouse experiments was prompted by the recognition that MR219 had become increasingly relevant and significant within the agricultural community due to its perceived resistance to blast disease. This development offered an excellent opportunity to investigate the actual resistance mechanisms and characteristics of MR219 under controlled conditions toward BPB. By employing MR219 in the glasshouse, we aimed to unravel the intricate dynamics of BPB resistance and susceptibility within this specific cultivar, contributing valuable insights to the broader body of knowledge concerning MR219's performance against BPB. It is possible to observe differences in pathogenicity when isolating a pathogen from one cultivar and testing it on a different cultivar. Pathogenicity can vary due to several factors, including the specific interactions between the pathogen and host cultivar, genetic variations in host plants, and the presence of resistance genes in some cultivars. As a result, the same pathogen may cause varying levels of disease severity or exhibit different symptoms when introduced to different cultivars. This variability in pathogenicity highlights the importance of studying pathogen-host interactions to understand disease dynamics fully and develop targeted disease management strategies tailored to specific cultivars and pathogens (Sundaramoorthy *et al.,* 2017).

Within the context of our research, the interpretation of the Area Under the Disease Progress Curve (AUDPC) values presents a nuanced understanding of disease progression and the responses of different cultivars to the pathogen. Our findings reveal a spectrum of disease severity, as reflected by the range of AUDPC values spanning from 410.762 to 1123.575. The strain with the lowest AUDPC value of 410.762 demonstrates the least disease progression, suggesting a reduced susceptibility to the host. Conversely, the strain with the highest AUDPC value of 1123 exhibits the most substantial disease progression, indicating greater susceptibility. The range of AUDPC values from the studies suggests varying degrees of resistance or susceptibility to the disease, with the cultivars having an AUDPC of 410.762 indicating partial resistance toward *B. glumae*, while those with higher values are generally less resistant. The intermediate AUDPC values observed among other strain represent varying levels of disease severity and, by extension, differing degrees of resistance or susceptibility. This variability underscores the multifaceted nature of disease dynamics, influenced by factors such as cultivar genetics, environmental conditions, and pathogen virulence. Moreover, the AUDPC values hold practical implications for disease management and breeding strategies. Cultivars with lower AUDPC values may be considered less virulent, while those with higher values may necessitate additional disease control measures. While the AUDPC values provide valuable insights, further statistical analyses and experiments may be required to validate these findings and delve deeper into the underlying factors contributing to distinct disease outcomes. In summary, the AUDPC values serve as a valuable tool for

characterizing disease progression, evaluating resistance, and guiding strategies for the effective management and breeding of crops.

The difference in virulency variation of these species is probably due to genetic variation, and physiology properties (Alam *et al.,* 2016; Sundaramoorthy *et al.,* 2017). From the disease severity index analysis, our studies find that *B. gladioli* (UPMBG7 & UPMBG8) have more severe virulence than *B. glumae* strain. *B. gladioli* strain UPMBG7 and UPMBG8 produce more toxoflavin production as shown in King's B agar, where yellow pigment toxoflavin production of *B. gladioli* on agar is more vigorous compared to *B. glumae* (Park *et al.,* 2006; El-Tarabily *et al.,* 2009; Saikia *et al.,* 2013). Other than that, natural and artificial environments may play a crucial role in causing severity toward rice plants. In the natural environment, many pathogenic bacteria may live/be infected on the rice plant. However, in an artificial environment, one plant infected one pathogenic bacterium, as only BPB disease was observed on the rice plant after 28DAI (Ishihara & Kiba, 2015; Sundaramoorthy *et al.,* 2017). Lastly, dominancy may affect the virulence severity of this pathogen. Based on our study, it appears that dominancy, in terms of disease severity, may indeed be relevant within the 27 samples. In the natural environment where multiple pathogens can be found on rice, our observation is that *B. gladioli* is more severe when isolated into rice using the glasshouse environment. In contrast, in the controlled glasshouse setting, *B. glumae* shows less disease severity compared to *B. gladioli*. This suggests that the interactions between these pathogens and their host plants, as well as the environmental conditions, may influence their dominancy and disease outcomes. This observation highlights

the complexity of pathogen-host interactions and underscores the importance of studying these dynamics for effective disease management. However, in our study, *B. glumae* is more prevalence than *B. gladioli* in the BPB disease outbreak. To validate this hypothesis, a proper test should be done to understand the virulency and toxoflavin production of *Burkholderia* species in causing plant disease.

In addition, our studies found that the most pathogenic isolate among *B. glumae* strains was BGS.AS1 Selangor isolates and the least pathogenic isolates were BGK.AZ1 Kedah isolates.

3.4 Conclusion

Pathogenic bacteria isolated through phenotypic and molecular methods had been confirmed to be *Burkholderia* spp.and the causal agent of bacterial panicle blight disease of rice (*Oryza sativa*). The traditional method for identifying species via molecular characterization showed that two different species were isolated, which were *B. glumae* and *B. gladioli.*

The pathogen is 100% clustered in the same clade toward their respective reference strains in phylogenetic analysis for gene 16r rRNA, gyrase B gene, and toxoflavin biosynthesis operon-specific gene (*toxB*). Lastly, pathogenicity assays showed that *Burkholderia* spp. has variable virulent levels and symptoms as observed in the field with outbreak disease.

CHAPTER 4

GENOMIC CHARACTERIZATION VIA RESEQUENCING AND GENOME ASSEMBLY

4.1 Introduction

Burkholderia glumae and *Burkholderia gladioli* are soil-borne bacterial plant pathogens that pose a significant threat to rice by causing panicle blight in rice. These bacterial species are known to produce a range of virulence factors that enable them to cause disease in their hosts. Whole genome sequencing (WGS) and bioinformatic pipelines have emerged as powerful tools for studying the genetic mechanisms that underlie bacterial pathogenicity. In particular, WGS has enabled the detection of structural variations (SVs) within bacterial genomes that can contribute to differences in genes between strains. Furthermore, bioinformatic pipelines have been developed to annotate and identify virulence genes in bacterial genomes, which can provide important insights into the pathogenesis of Burkholderia glumae and Burkholderia gladioli.

WGS has been applied to the study of several *Burkholderia* species, including *B. glumae* and *B. gladioli*. In *B. glumae*, WGS has revealed the presence of several genomic islands (GIs) that are associated with virulence, including a GI that encodes a type III secretion system (T3SS) and another GI that encodes a toxoflavin biosynthesis cluster (Lim et al., 2014). Similarly, WGS of

B. gladioli has identified several key virulence factors, including a T3SS and a flagella assembly gene cluster (Li et al., 2020).

In addition to identifying evolutionary and diversity, WGS has also enabled the detection of SVs within bacterial genomes. These SVs can arise from a variety of genetic events, including deletions, duplications, insertions, inversions, and translocations, and can contribute to diversity in gene among strains. For example, WGS has revealed several SVs in *B. glumae* genomes, including a deletion of a portion of the T3SS gene cluster and an inversion of a portion of the toxoflavin biosynthesis cluster (Lim et al., 2014).

Bioinformatic pipelines have been developed to annotate and identify virulence genes in bacterial genomes. One such pipeline is the Rapid Annotation Subsystem Technology (RAST), uses a subsystems approach to annotation, which allows for the rapid identification of metabolic pathways, virulence factors, and other functional modules in newly sequenced genomes (Overbeek et al., 2014). Another pipeline, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pipeline, has been used to provides pathway visualization and analysis, facilitating the interpretation of large-scale genomic data (Kanehisa et al., 2017).

In this thesis, we aim to understanding structural variation in *Burkholderia* species through reference alignment analysis and evaluate a bioinformatic pipeline for the annotation and identification of virulence genes in *Burkholderia* species plant pathogens using WGS data. We will focus on the use of comparative genomics, functional annotation, and machine

learning algorithms to identify and annotate virulence genes and their associated functions.

The research study on whole genome sequencing of *B. glumae*, particularly as the first fully annotated and gene-predicted strain in Malaysia, addresses a significant research gap in the understanding of this bacterial species. This gap primarily revolves around the limited availability of genomic data for *B. glumae* in the Malaysian context and the broader scientific community. By undertaking this sequencing project and subsequently resequencing and aligning our strain with the previously sequenced strain, we have contributed valuable insights into the evolution and mutation dynamics within this bacterial species.

This research not only fills a critical void in the genomic data for *B. glumae* in Malaysia but also offers a unique opportunity to explore how this pathogen may have adapted or evolved within the region. It allows for the identification of specific genetic variations, potentially related to virulence or antibiotic resistance, which can have implications for disease management and treatment strategies. Furthermore, comparing our strain's genomic data to the previously sequenced strain can shed light on the evolutionary relationships and potential genetic divergence within this pathogen.

In summary, our study not only provides a valuable resource for researchers and policymakers dealing with *B. glumae* in Malaysia but also contributes to the broader understanding of the evolutionary dynamics and genomic

adaptations of this pathogen. This research has the potential to inform future strategies for disease control and management.

The performance of the annotation will be evaluated using benchmark datasets, and its utility will be demonstrated through the analysis of real-world WGS data from *Burkholderia* species plant pathogens. Our ultimate goal is to provide a reliable and efficient tool for the identification and annotation of virulence genes in bacterial plant pathogens, facilitating the understanding of pathogenesis and the development of new strategies for disease management.

4.2 Materials and Methods

4.2.1 DNA Extraction of *Burkholderia glumae* (K6 & P3) and *Burkholderia gladioli* (UPMBG7 & UPMBG8)

The selection of strains for whole genome sequencing was based on their geographical origin within Malaysia. Specifically, a *B. gladioli* strain was chosen from Selangor, a *B. glumae* K6 strain from Kedah, and a *B. glumae* P3 strain from Perak. It is important to note that strain selection occurred prior to conducting pathogenicity tests. Consequently, the choice of strains was made without prior knowledge of their virulence levels. This selection process, while not based on virulence, serves as an unbiased representation of the bacterial population from the isolated strains. For DNA extraction, pure colonies of selected *Burkholderia* spp. was first inoculated in sterile one mL of nutrient broth and incubated overnight at 35 °C. Later, DNA extraction of selected *B. glumae* and *B. gladioli* was done following the methods described by Kim *et al.* (2014). genomic DNA of *Burkholderia* spp. will be extracted using a commercial genomic DNA purification kit (Presto™ Mini gDNA Bacteria Kit) according to the manufacturer's protocols (Geneaid Biotech Ltd., Taiwan). DNA concentration and purity will be measured using NanoDrop spectrophotometer (Thermo Scientific, USA) at wavelength 260 and 280nm for sample quality control. The purified DNA extracted was later sent for library preparation and whole genome shotgun sequencing analysis (Apical Scientific Sdn. Bhd., Selangor, Malaysia).

4.2.2 Library Preparation, Quality Control and Whole-Genome Sequencing

A DNA library is a collection of DNA fragments cloned onto vectors for researchers to identify and extract the DNA fragments of interest for further investigation (Son and Taylor, 2011). The method for performing DNA library preparation will be done according to Pasquali *et al.* (2019). Illumina TruSeq® DNA library preparation kit will generate paired-end genomic sequencing libraries. The genomic DNA will be randomly fragmented (fragmentation to 350bp) by sonication technique and generated double-stranded DNA fragments consisting of 3' or 5' overhangs. After fragmentation, the overhanging end will be converted into blunt ends (end-repaired) with the help of T4 DNA ligase and Klenow polymerases, where the Klenow polymerase exhibit 3' to 5' exonuclease activity to eliminate overhangs and polymerase activity filled in the 5' overhangs. After that, A-tailed steps where an 'A' base has been connected to the 3' end using the polymerase activity of the Klenow fragment (3' to 5' exo minus). This 'A' overhang enables adapters with a single thymine overhang to base pair with the DNA fragments. Then ligation by the DNA adaptor will occur on the DNA fragments. After that, the purification process will be done to remove unligated adapters and self-ligated adapters. Before detecting the expected size and yield of the adapter-modified DNA fragment, the adapter-modified DNA fragments were subjected to quality control to inspect the concentration and purity of the product (Figure 4.1). The library was checked with Qubit 3.0 Fluorometer, real-time PCR for quantification, and bioanalyzer for size distribution detection (Minoche *et al.,* 2011).

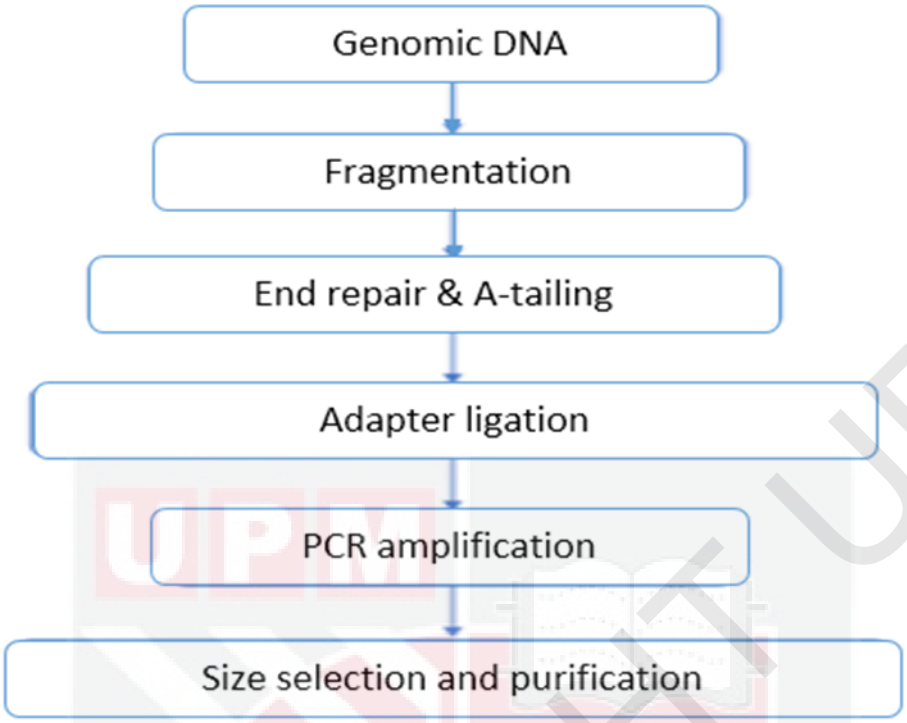

Figure 4.1: Workflow for DNA library preparation for Illumina sequencing of *Burkholderia* species

Quantified DNA Libraries with different indices were pooled and put onto high throughput sequencing Illumina HiSeq4000 platform according to the manufacturer's instructions (Illumina, San Diego, CA, USA) which uses sequencing-by-synthesis technology, and the genome was sequenced using a 2x150 paired-end sequencing kit. The sequence data obtained from the sequencer were checked for their sequencing data quality distribution and sequencing data filtration to remove adapter contamination and low-quality reading using FastQC software V0.20.0 (Babraham Bioinformatics, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

4.2.3 Genome Deposition of K6, P3, UPMBG7 and UPMBG8

The draft genome sequence for *B. glumae* (K6 & P3) and *B. gladioli* (UPMBG7 & UPMBG8) was deposited in GenBank (link: https: //www .ncbi .nlm .nih.gov/genbank/) under the accession number of JAMYCS000000000 (K6 strains), JAMYCT000000000 (P3 strains), JANIEE000000000 (UPMBG7 strains) and JANIEF000000000 (UPMBG8 strains).

4.2.4 Reference Alignment/Mapping Analysis

The alignment of clean data was performed using the program Burrows-Wheeler Aligner (BWA-MEM) version v0.7.17 with default parameter (mem -t 4 -k 32 -M) and duplicates were removed using Picard tools version v2.18.92 (Jung & Han, 2022). Since the clean data are not completely sequenced and annotated, the fully sequenced and annotated genome of *B. glumae* BGR1 (Accession number PRJNA33901) and *B. gladioli* BSR3 (Accession number PRJNA64503) was used as a reference for their respective alignment of clean data. The coverage was computed by using SAMtools software version v1.8, and the alignment information will be saved in BAM format files and subjected to a mate-pairing step, a read group addition step, and a marking step for duplicate reads generated by PCR. Lastly, the aligned data (BAM files) undergo the process of structural variation calling analyses (Jung & Han, 2022).

4.2.4.1 Structural Variation Analysis of *Burkholderia glumae* and *Burkholderia gladioli*

After clean data are aligned to the reference genome, bases different from the reference genome will be identified. SAMtools software detected individual single nucleotide polymorphism (SNPs) and insertion and deletion (InDel) from a small fragment smaller than 50bp. This analysis differentiates whether the different bases are SNPs / InDel or errors during sequencing. After that, the location of the SNP/InDel in the functional regions of the genome of the isolate will be annotated using ANNOVAR software. ANNOVAR (https://annovar.openbioinformatics.org/en/latest/) is a widely used software in variation annotation with multiple capabilities, including gene-based annotation, region-based annotation, filter-based annotation as well as other functionalities (Sathya *et al.,* 2014).

Structural variants (SV) is a catch-all term for medium-scale variations with mutation size over 50bp, including deletions, duplications, insertions, inversions and translocations identified between the genomes of individuals within a species (Lin *et al.,* 2015). The best situation for SV identification arises when both the sample and reference genomes are completely known at base resolution. First, BreakDancer software version v1.4.4 was downloaded and installed for identified genome structural variation from the sample sequence. Then, the BreakDancer program was conducted for insertion (INS), deletion (DEL), inversion (INV), intra-chromosomal translocation (ITX), and inter-chromosomal translocation (CTX) between reference strain and samples sequence by estimating the parameters of the insert size distribution, scans

read for discordant read-pairs, then clusters them to find SV breakpoints. Finally, the run program output was the list of SV breakpoints, predicted SV type, and SV size. (Fan *et al.,* 2014). ANNOVAR further annotated the INS, DEL and INV.

Copy-number variation (CNV) is a form of structural variation characterised by deletions or duplications in the genome (Figure 4.2). Based on the read's depth of the reference genome, CNVnator program version v0.3 was used to detect CNVs of potential deletions and duplications with the following parameter "-call 100". The detected CNVs were also further annotated by ANNOVAR (Fan *et al.,* 2014).

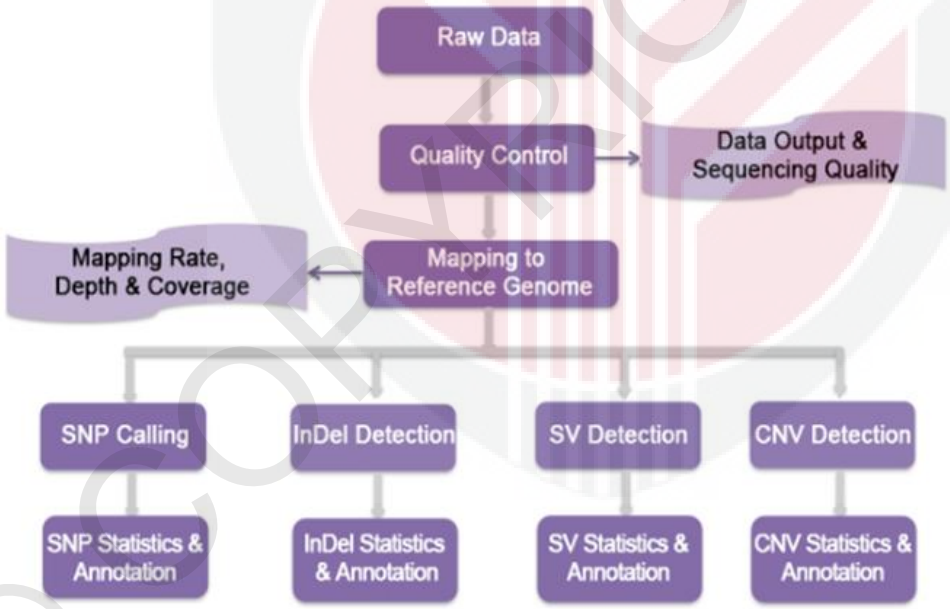

Figure 4.2: Workflow for reference alignment/ mapping analysis of *Burkholderia* species

4.2.5 *De Novo* Assembly Analysis

4.2.5.1 Genome Assembly

Clean data from shotgun whole-genome sequencing genome assembly was carried out using three assembly programs (Abbas *et al.,* 2014). The assembly was first performed using SOAPdenovo v2.04 software with different K-mers (95,107, 119) and the least contigs were chosen as the preliminary assembly result of SOAPdenovo. SPAdes v3.15.4 software was used with different K-mers (99 and 127) and assembly results were chosen based on optimal K-mer and least contigs. Then, AbySS v2.1.5 assembling software was used with K-mer 64 parameter selected, and the assembly result was obtained. Using CISA software to integrate the assembly results from three software and assembly result with the least contigs were selected (Lin and Liao, 2013). The final result was used for further analysis of gene prediction.

4.2.5.2 Genome Component Predictions

Genome component prediction included the prediction of the coding gene and non-coding RNA annotations. The related coding gene predictions were predicted using the GeneMarks v4.28 program (Besemer *et al.,* 2001). Non-coding RNA predictions include small RNA (sRNA), ribosomal RNA (rRNA) and transport RNA (tRNA). The tRNA gene was predicted by tRNAscan-SE, the rRNA gene was analyzed by rRNAmmer and the Rfam database predicted the sRNA (Lowe and Eddy., 1997; Lageson *et al.,* 2007; Gardner *et al.,* 2009).

4.2.5.3 Gene Function Analysis

Sample genome contigs was annotated accordingly to the subsystem coverage, category distribution and feature counts via the Rapid Annotation using Subsystem Technology (RAST) platform. The gene functions were annotated by Gene Ontology (GO), pathways of the annotated gene via Kyoto Encyclopedia of Genes and Genomes (KEGG) databases and phylogenetic classifications of the proteins encoded by gene was annotated via Cluster of Orthologous Groups of Protein Annotation (COG) (Ashburner *et al.,* 2000; Kanehisa *et al.,* 2004; Kanehisa *et al.,* 2006; Galperin *et al.,* 2015; Li *et al.,* 2002). In addition, NCBI non-redundant protein database was annotated based on a protein database and annotation results contain specific information that can be used for species classification (Figure 4.3).

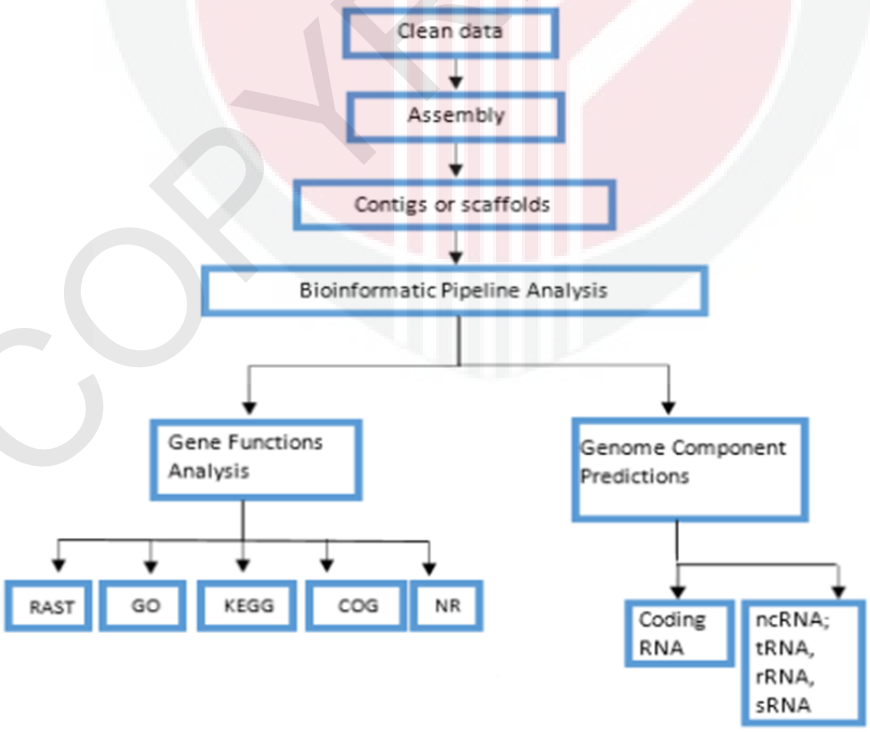

Figure 4.3: Workflow for *De Novo* assemble analysis of *Burkholderia* species.

4.3 Result

4.3.1 Mapping Statistics

After quality control of all samples via FastQC, *B. glumae* strain K6 produced a clean read of 11,343,010, and P3 produced 9,722,426 clean reads. While *B. gladioli* strain UPMBG7 produced 6,702,572 clean reads and UPMBG8 produced 4,886,984 clean reads (Table 4.1). Detailed information for the reference genome used in this study can be obtained from Table 4.2. The clean reads for all samples aligned with the reference sequence according to their species, mapping rate, and coverage. The average depth was counted according to the alignment results (Table 4.3).

Table 4.1: Statistics of Sequencing Data for *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample Name	Raw Reads	Clean Reads	Q20 (%)	Q30 (%)	GC (%)
K6	11,364,102	11,343,010	96.87	92	66.94
P3	9,739,026	9,722,426	97.68	92.32	66.55
UPMBG7	6,718,846	6,702,572	97.81	93.62	67.45
UPMBG8	4,901,304	4,886,984	97.66	93.26	67.52

Table 4.2: Statistics of Reference Genome *B. glumae* BGR1 and *B. gladioli* BSR3

Sample Name	Accession number	Total Length (bp)	N50 Length (bp)	N90 Length (bp)	GC (%)
BGR1 (*B. glumae*)	GCA_000022645.2	7,284,636	3,906,507	2,827,333	67.93
BSR3 (*B. gladioli*)	GCA_000194745.1	9,052,299	3,700,833	403,586	67.4

Table 4.3: Mapping statistics with reference genome for *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample Name	Mapped Reads	Mapping Rate (%)	Average Depth	Coverage at least 1X	Coverage at least 4X
K6	10,383,763	91.28	195.86	86.66	86.52
P3	8,839,866	90.65	171.9	86.61	86.63
UPMBG7	5,748,002	85.65	91	78.86	78.62
UPMBG8	4,181,995	85.46	69.78	78.85	78.57

According to the results above, the mapping rate of each sample toward their reference genome range from 85.46% to 91.28%. Thus, the results are in the qualified normal range and may serve in the subsequent variation detection and related analysis

4.3.2 Structural Variation Analysis of *Burkholderia glumae* and *Burkholderia gladioli*

4.3.2.1 Single Nucleotide Polymorphism Analysis

Single nucleotide polymorphism (SNP) refers to the variation in a single nucleotide that occurs at some specific position in the genome, including transition and transversion of a single nucleotide. SNP variation detected occurs in the upstream region (within 1 kb upstream away from the transcription start site of the gene), Exonic region (Non-synonymous: single nucleotide mutation with changing amino acid sequence; Stop gain/loss: a nonsynonymous SNP that leads to the introduction/removal of a stop codon at the variant site; Synonymous: single nucleotide mutation without changing amino acid sequence), Downstream region (within 1 kb downstream away from transcription termination site of the gene region), Upstream/Downstream region and also Intergenic region (Table 4.4).

Table 4.4: Statistics of SNP Detection & Annotation *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample	Upstream	Exonic				Downstream	Upstream/Downstream	Intergenic	ts[a]	tv[b]	Total
		Stop gain	Stop loss	Synonymous	Non-synonymous						
P3	813	17	3	9912	4513	565	2631	3754	15010	7578	22588
K6	844	17	4	9867	4468	556	2619	3696	14951	7522	22473
UPMBG7	5048	71	3 7	5860 4	2625 2	274 1	1586 5	21	7091 7	3878 3	10970 0
UPMBG8	5044	73	3 8	5855 8	2628 2	272 7	1582 9	21	7082 1	3876 4	10958 5

a Transitions, a point mutation that changes a purine nucleotide to another purine (A ↔ G) or a pyrimidine nucleotide to another pyrimidine (C ↔ T).
b Transversions, the substitution of a (two ring) purine for a (one ring) pyrimidine or vice versa.

4.3.2.2 Insert or Deletion Analysis

Insert or Deletion analysis (InDel) refers to the insertion or deletion of ≤ 50 bp sequences in the DNA. InDel variation detected occur in the upstream region, Exonic region (Stop gain/loss: InDel that leads to the introduction/removal of a stop codon at the variant site; Frameshift deletion/insertion: InDel mutation changing the open reading frame (ORF) with deletion or insertion; Non-Frameshift deletion/insertion: InDel mutation without changing the ORF with deletion or insertion sequences of 3 or multiple of 3 bases), Downstream

region, Upstream/Downstream region and also Intergenic region region (Table 4.5).

Table 4.5: Statistics of InDel Detection & Annotation *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample	Exonic								Downstream	Upstream/Downstream	Intergenic	In/Del[a]
	Upstream	Stop gain	Stop loss	Frameshift deletion	Frameshift insertion	Non-Frameshift deletion	Non-Frameshift insertion					
P3	91	11	1	160	164	105	101	142	515	287	817/828
K6	101	12	0	157	174	104	115	138	510	297	840/838
UPMBG7	588	19	7	652	652	354	366	639	2384	2	2893/2821
UPMBG8	602	20	7	637	642	353	373	615	2387	2	2892/2794

a Total of Insertion and Deletion

4.3.2.3 Structural Variation Analysis

Structural variants (SVs) are genomic variations with mutations of relatively larger size (>50 bp), including detection of insertion (INS), deletion (DEL), inversion (INV), intra-chromosomal translocation (ITX), and inter-chromosomal translocation (CTX) mutations (Table 4.6).

Table 4.6: Statistics of Structural Variants Detection & Annotation *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample	Exonic	INS	DEL	INV	ITX	CTX	Total
P3	98	11	94	21	36	71	233
K6	118	28	103	20	39	73	263
UPMBG7	130	5	120	7	87	65	284
UPMBG8	116	0	109	7	73	60	249

Other regions such as Upstream, Downstream, Upstream/Downstream, Intronic, and Intergenic, have less than 10 structural variants compared to the Exonic region, which indicates most structural variants occur in the coding region gene

4.3.2.4 Copy Number Variation Analysis

Copy-number variation (CNV) is a type of structual variation showing deletions or duplications in the genome. Based on the read's depth of the reference genome (Table 4.7).

Table 4.7: Statistics of Copy Number Variations Detection & Annotation *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample	Exonic	Duplication	Deletion	Duplication length (bp)	Deletion length (bp)	Total
P3	162	63	136	557,100	1,477,600	199
K6	135	40	135	361,100	1,268,500	175
UPMBG7	221	57	167	852,200	2,045,300	224
UPMBG8	218	54	167	781,700	2,045,500	221

The annotation of all SNP, InDel, SV and CNV can be obtained in supplementary information.

4.3.2.5 Visualization of Variations

For proper visualization of the structural variations on the whole-genome, we present according to mutation types with Circos (Krzywinski *et al.,* 2009). The visualization is based on; (1) For for SNP/InDel type, the density distribution is drawn; (2) for SV/CNV type, the location and size are drawn (Figure 4.4 to Figure 4.7)

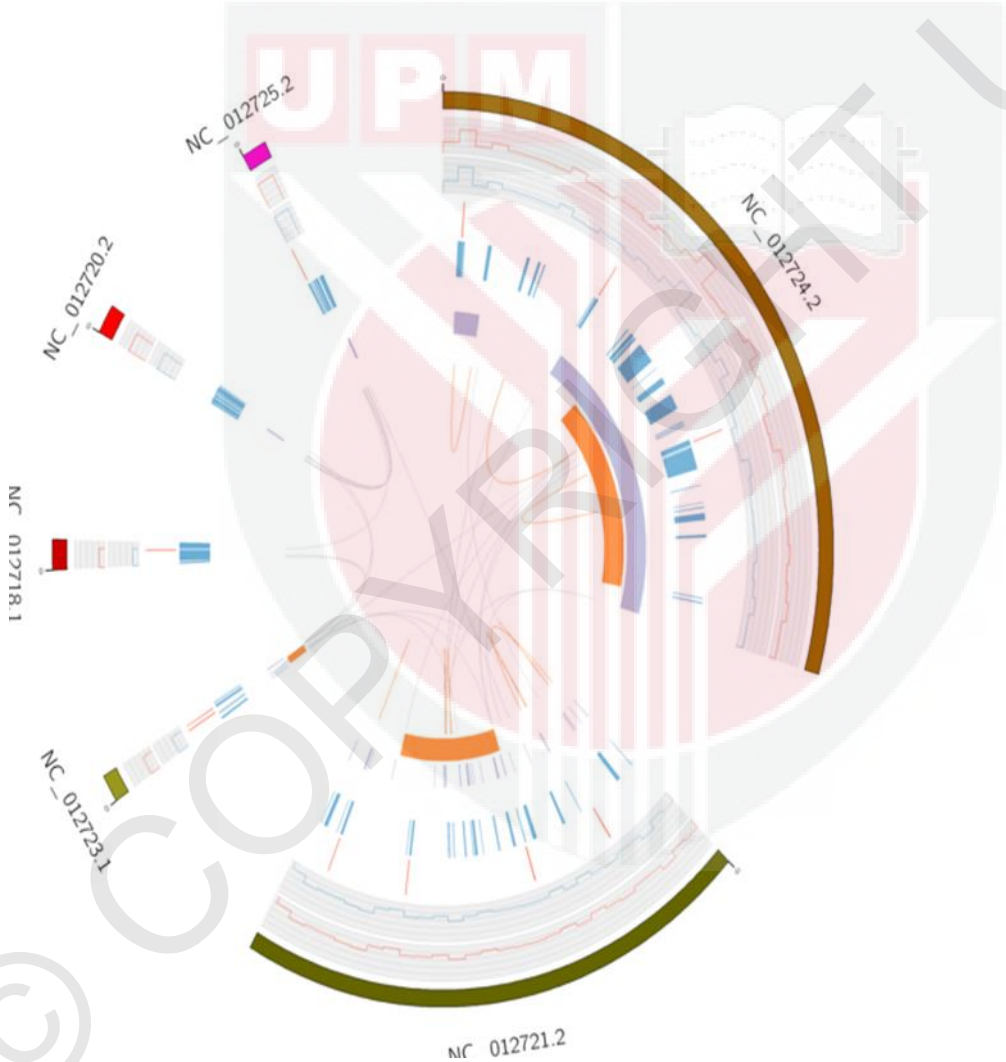

Figure 4.4: Circular presentation of whole genome *B. glumae* K6 variations distributions. From outer to inner of diagram: chromosome, SNP, InDel, CNV duplication, CNV deletion, SV insertion, SV deletion, SV invertion, SV ITX, SV CTX.

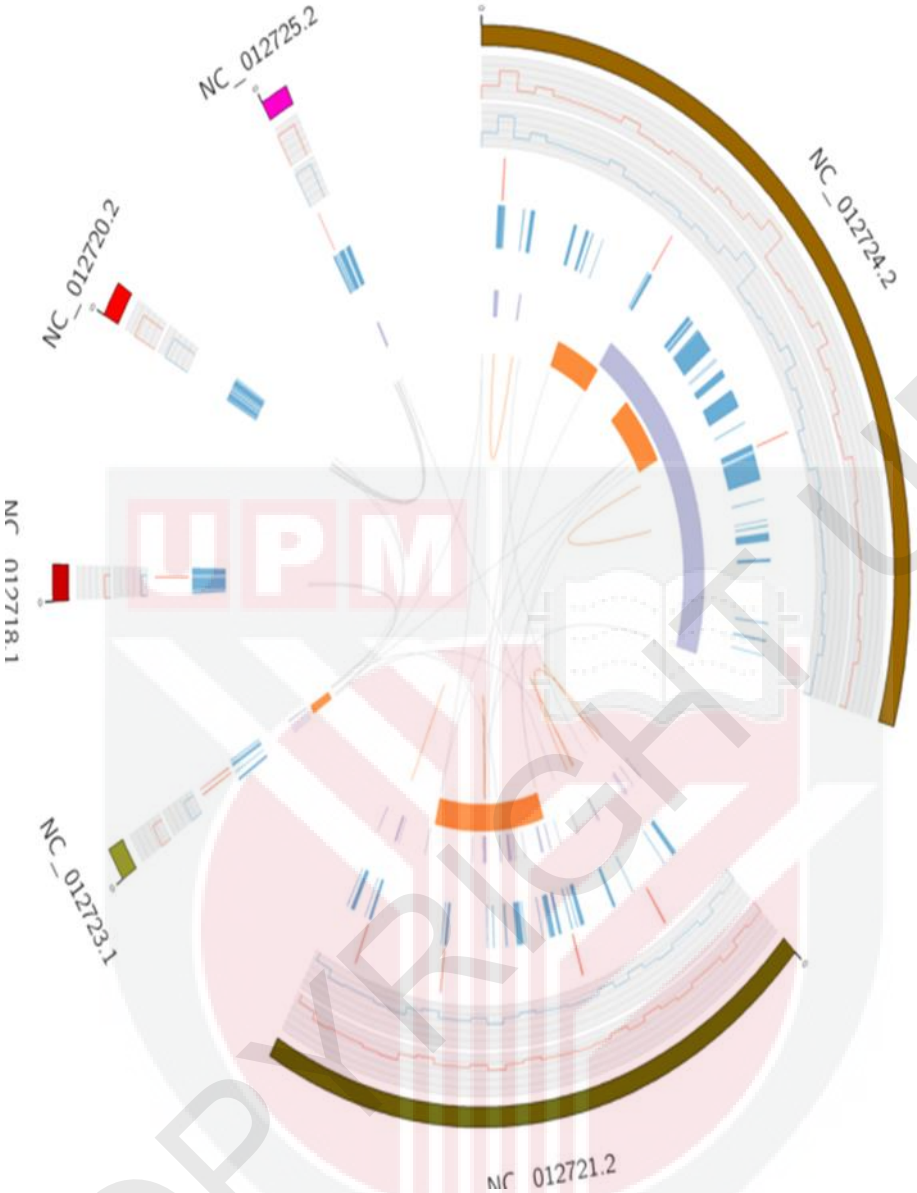

Figure 4.5: Circular presentation of whole genome *B. glumae* P3 variations distributions. From outer to inner of diagram: chromosome, SNP, InDel, CNV duplication, CNV deletion, SV insertion, SV deletion, SV invertion, SV ITX, SV CTX.

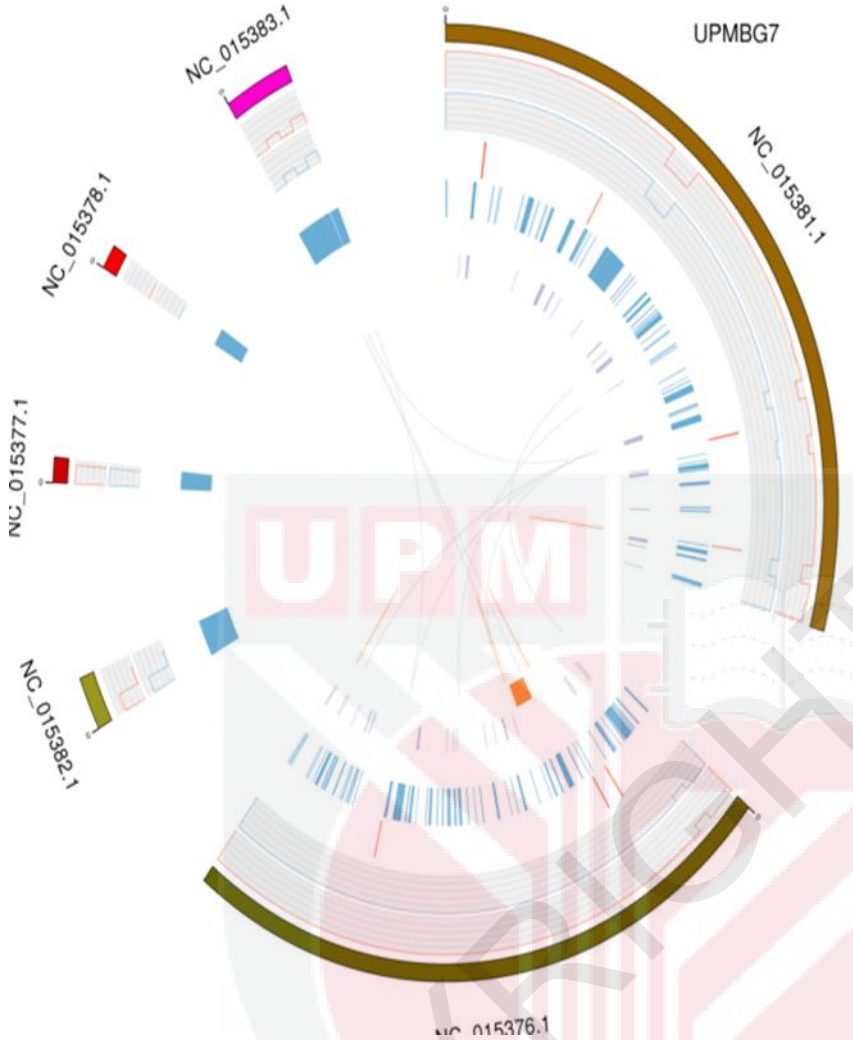

Figure 4.6: Circular presentation of whole genome *B. gladioli* UPMBG7 variations distributions. From outer to inner of diagram: chromosome, SNP, InDel, CNV duplication, CNV deletion, SV insertion, SV deletion, SV invertion, SV ITX, SV CTX

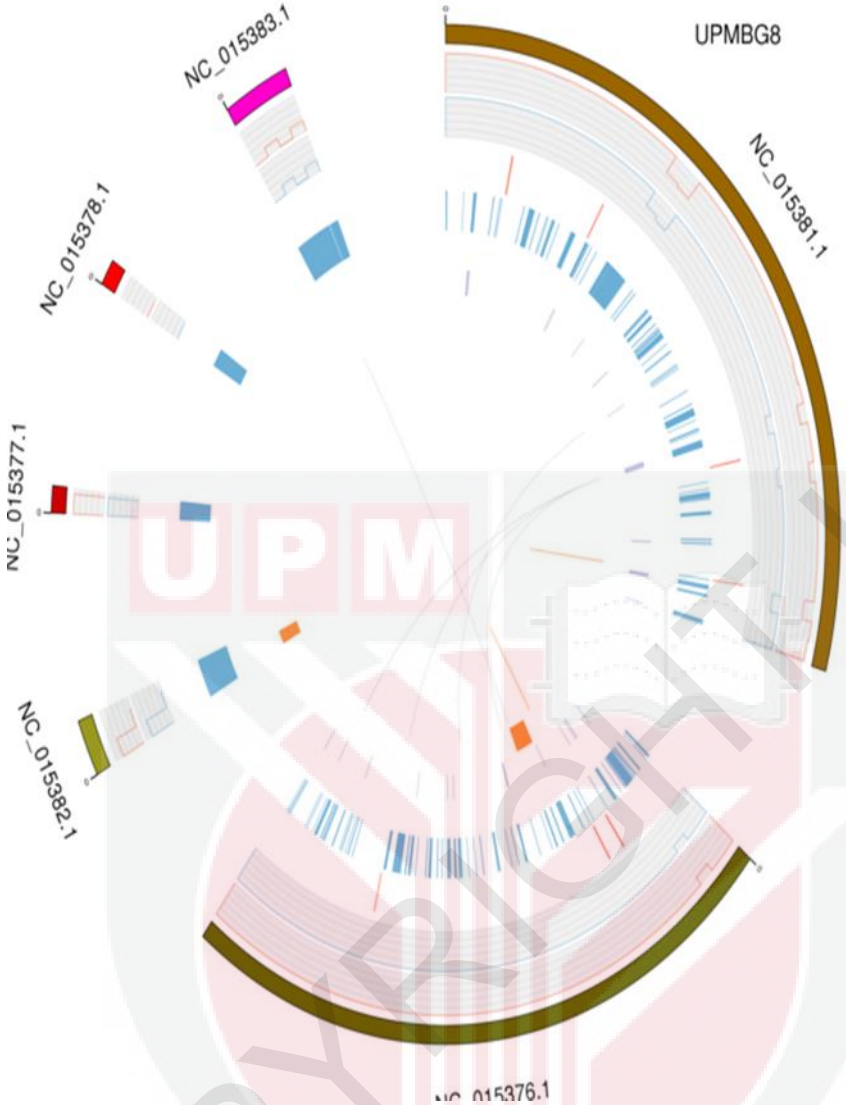

Figure 4.7: Circular presentation of whole genome *B. gladioli* UPMBG8 variations distributions. From outer to inner of diagram: chromosome, SNP, InDel, CNV duplication, CNV deletion, SV insertion, SV deletion, SV invertion, SV ITX, SV

4.3.3 Assembly Result

De novo assembly was performed to produce assembled genome sequences with different contigs. *B. glumae* strain K6 was assembled into 6,572,223 bp genome sized with 210 contigs with an N50 value of 77,057 bp with 68.33% G+C contents, while *B. glumae* P3 strain were assembled into 6,525,024 bp genome size with 499 contigs with N50 value of 25,158 bp with 68.22% G+C contents. As for *B. gladioli* strain, UPMBG7 was assembled into 8,225,388 bp genome size with 124 contigs with an N50 value of 207,968 bp with 67.99% G+C contents, while *B. gladioli* strain UPMBG8 was assembled into 8,056,421 bp genome size with 120 contigs with N50 values of 207,968 bp with 68.01% G+C contents (Table 4.8).

Table 4.8: Statistics of *De Novo* Assembly of *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Strain	Contig	Total Length (bp)	N50 (bp)	N90 (bp)	Max Length (bp)	Min Length (bp)	G+C (%)
K6	210	6,572,223	77,057	20,485	297,634	234	68.33
P3	499	6,525,024	25,158	6,189	97,220	560	68.22
UPMBG7	124	8,225,388	207,968	45,387	446,475	238	67.99
UPMBG8	120	8,056,421	207,968	55,852	343,795	514	68.01

4.3.4 Genome Components Prediction

4.3.4.1 Non-coding RNA Annotation Statistics

Noncoding RNA (ncRNA) is a kind of RNA molecule found in bacteria, archaea, and eukaryotes that serves various biological functions. ncRNA does not convey translation information because it is an end product of RNA. NcRNA contains sRNA, rRNA (5s, 16s, 23s) and tRNA (Table 4.9).

Table 4.9: Statistics of Non-Coding RNA statistics of *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample Name	Types	Total number of ncRNA	Average length (bp)	Total length (bp)
P3	tRNA	57	78	4,493
P3	5s rRNA	6	113	678
P3	16s rRNA	1	1,521	1,521
P3	23s rRNA	1	2,879	2,879
P3	sRNA	1	75	75
K6	tRNA	57	75	4,493
K6	5s rRNA	6	78	678
K6	16s rRNA	1	113	113
K6	23s rRNA	1	1,521	1,521
K6	sRNA	1	2,879	2,879
UPMBG7	tRNA	60	78	4,621
UPMBG7	5s rRNA	6	113	678
UPMBG7	16s rRNA	1	1,521	1,521
UPMBG7	23s rRNA	1	2,878	2,878
UPMBG7	sRNA	3	70	211
UPMBG8	tRNA	58	78	4,467
UPMBG8	5s rRNA	6	113	678
UPMBG8	16s rRNA	1	1,521	1,521
UPMBG8	23s rRNA	1	2,878	2,878
UPMBG8	sRNA	2	65	131

4.3.4.2 Coding Gene Annotation Statistics

Protein sequences from other animals, such as model species with high-quality genome annotations and species phylogenetically related to the target species, are the primary sources of proteins employed in this stage. BLAST matched all of the protein sequences to the genome sequence, and GeneWise was used to estimate gene structure-based on credible alignments (Evalue<1e-5). The coding genes were predicted using homologous evidence by GeneMarks (version 4.28). Statistics of coding gene annotation was presented in Table 4.10.

Table 4.10: Statistics of Coding Gene Annotation statistics of *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Sample Name	Genome Sizes (bp)	Total Coding Gene	Total Coding Gene length	Gene Length (%)
K6	6,572,223	5,741	4,961,481	75.49
P3	6,525,024	5,827	5,256,504	80.56
UPMBG7	8,225,388	7,022	6,794,847	82.61
UPMBG8	8,056,421	6,881	6,663,786	82.71

4.3.5 Gene Functions Results

Databases of GO, KEGG, COG, NR, RAST, and NCBI were used for general gene annotation. The table values represent the number of coding genes and proteins annotated in the corresponding database (Table 4.11).

Table 4.11: Statistics of Functional Annotation statistics of *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8

Strain	RAST	NCBI	GO	COG	KEGG	NR
K6	6,688	5,741	3,686	4087	5,271	5,420
P3	6,844	5,827	3,553	3946	5,096	5,243
UPMBG7	7,749	7,022	4,801	5,390	6,637	6,756
UPMBG8	7,594	6,881	4,702	5,275	6,507	6,621

4.3.5.1 Non-Redundant Protein Database Annotation

Results from Non-Redundant Protein Database (NR) annotated species and gene showed that 4,988 (92.03%) gene annotated from K6 strain and 4,820 (91.93%) gene annotated from P3 strain are similar toward *Burkholderia glumae* species. While NR gene annotated for UPMBG7 strain showed 5,700 (84.36%) gene and 5,808 (87.72%) gene annotated from UPMBG8 strain are comparable toward *Burkholderia gladioli* species (Figure 4.8-Figure 4.11).

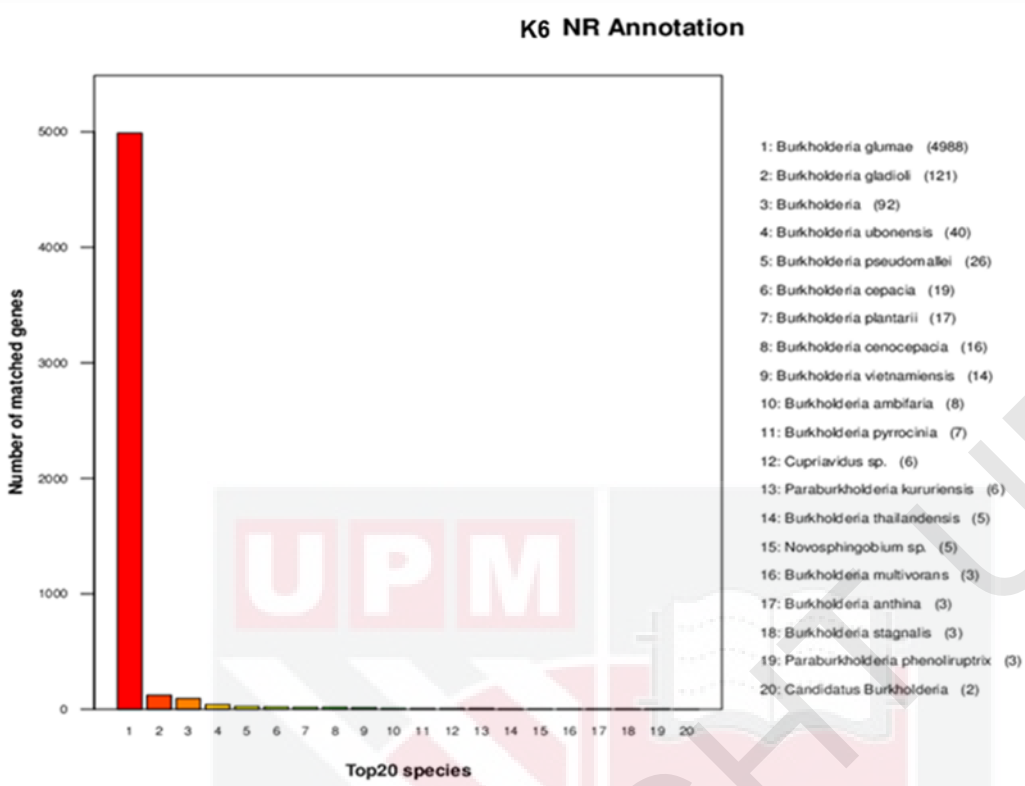

Figure 4.8: Number of matched genes from gene annotation of K6 using NR database

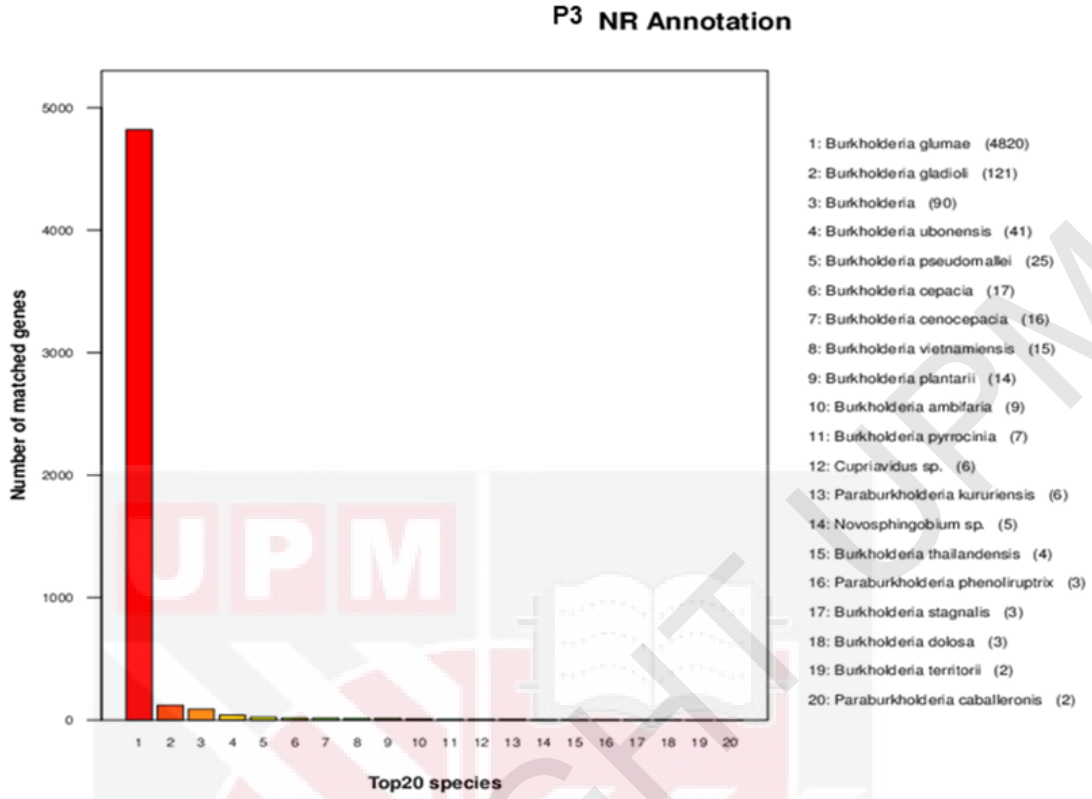

Figure 4.9: Number of matched genes from gene annotation of P3 using NR database

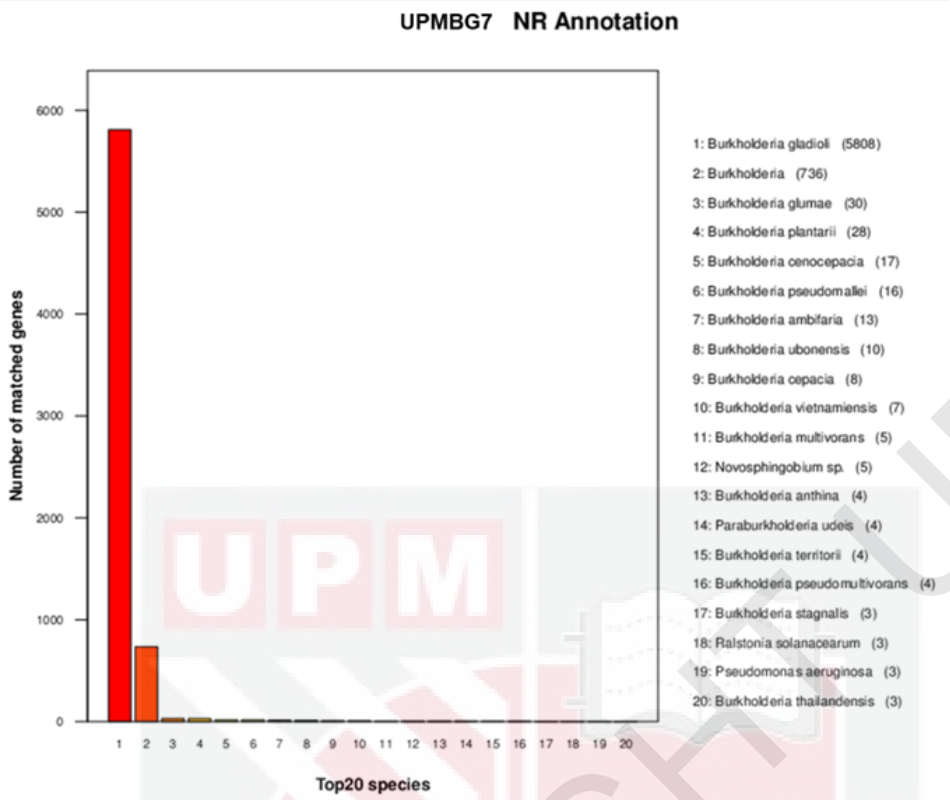

Figure 4.10: Number of matched genes from gene annotation of UPMBG7 using NR database

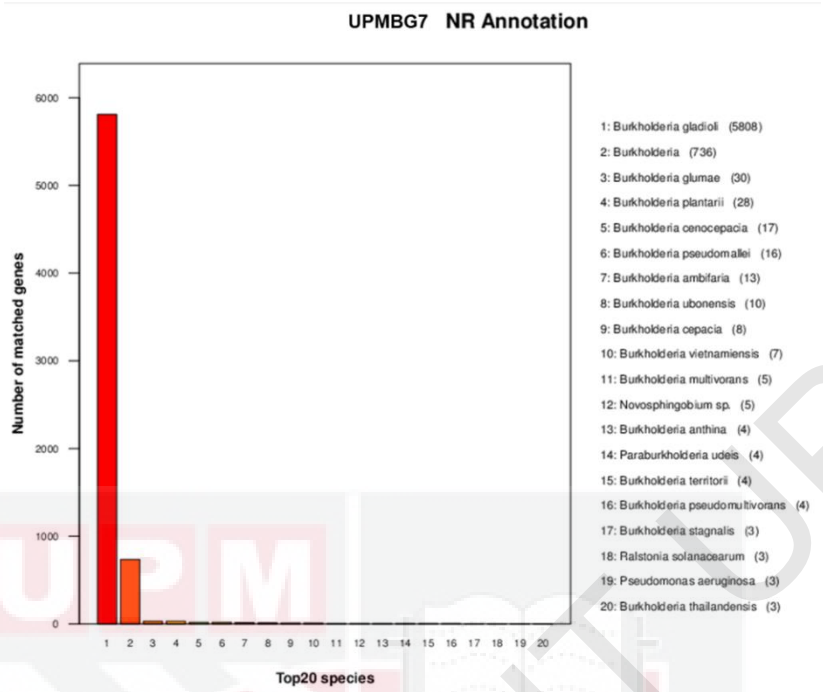

Figure 4.11: Number of matched genes from gene annotation of UPMBG8 using NR database

4.3.5.2 Rapid Annotations Using Subsystem Technology

The draft genome's subsystem coverage, category distribution, and feature counts were revealed by RAST findings. The subsystem coverage in the K6 strain was 22% of the protein with 353 subsystems, whereas the subsystem coverage in the P3, UPMBG7, and UPMBG8 strains was 23% of the protein with 356, 370, and 368 subsystems, respectively (Figure 4.12- Figure 4.15).

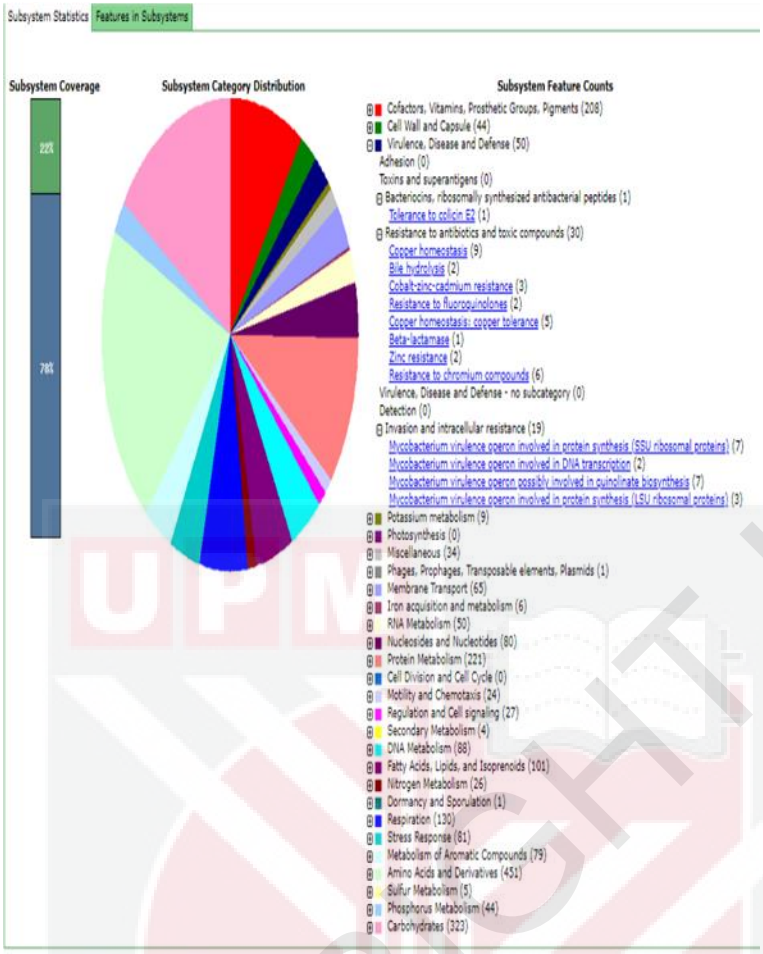

Figure 4.12: Subsystem coverage, category distribution and feature counts of the *B. glumae* strain K6.

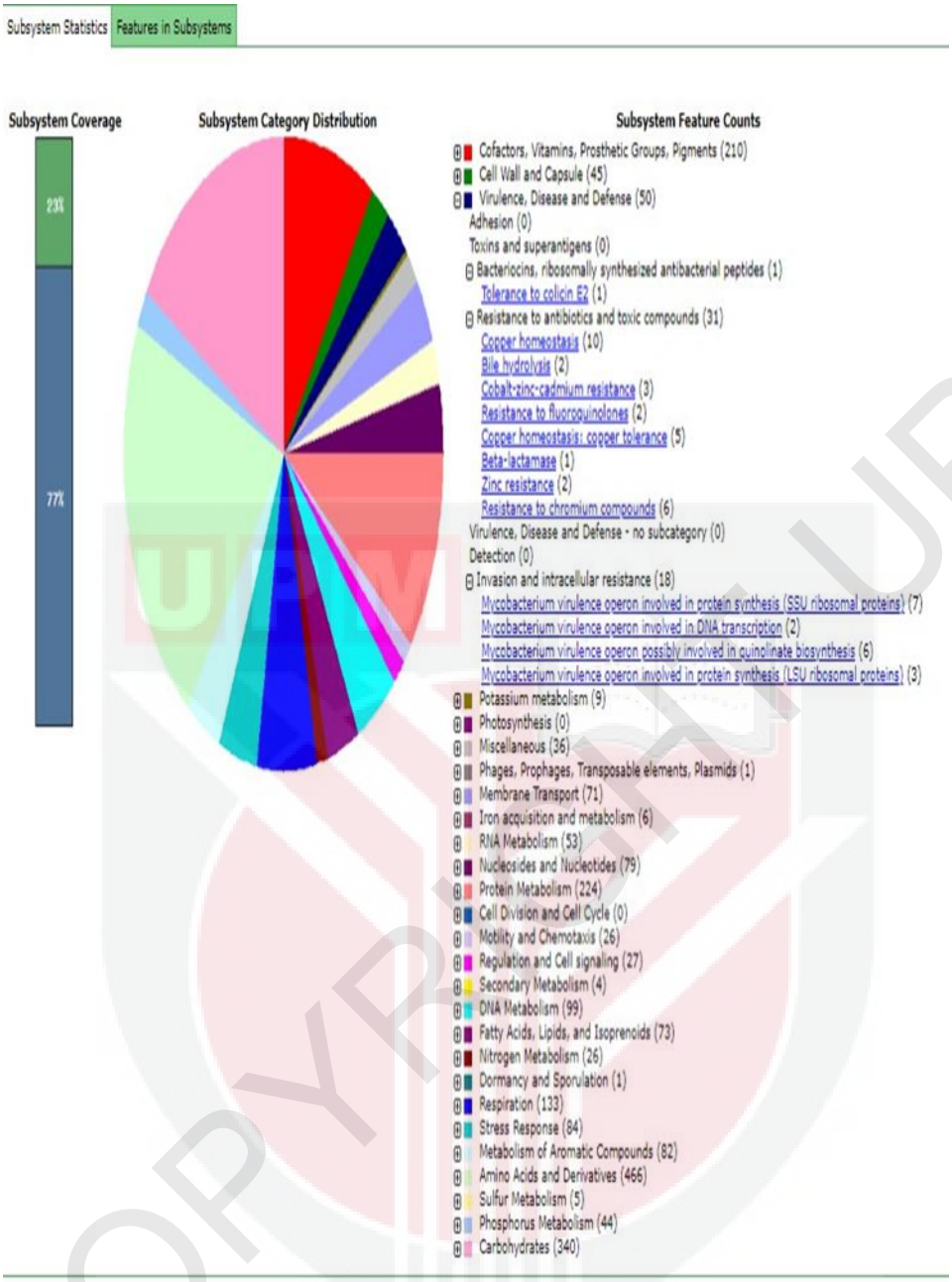

Figure 4.13: Subsystem coverage, category distribution and feature counts of the *B. glumae* strain P3

According to RAST analysis, the subsystem feature counts for the *B. glumae* K6 strain indicated that 50 coding sequences (CDS) related to virulence, disease, and defense mechanism (Figure 4.12). The coding gene annotated showed that the K6 strain is tolerant toward colicin E2, a type of bacteriocin peptide used to kill another same species to reduce competition for nutrient

gaining. Other than that, 30 of the coding sequences of the genome are responsible for the resistance of the *B. glumae* K6 strain toward antibiotic and toxic compounds such as copper, cobalt-zinc-cadmium, fluoroquinolones, zinc, and chromium compounds (Cui *et al.,* 2021). There are 19 coding genes also identified in the mycobacterium virulence operon responsible for the invasion and intracellular resistance of the K6 strain.

From RAST annotated CDS, there are 24 CDS encoded related to flagella motility in the K6 strain. CDS *FliR, FlhB, FlhB and FlhF* found in the K6 strain play a vital role in flagellar biosynthesis protein (Balaban *et al.,* 2009). Next, there are 24 CDS related to the protein secretion system, where 22 of the CDS belong to the type II protein secretion system (T2SS), and 2 of the CDS belong to type VII (T7SS). Seven CDS of T2SS (*Pilin, RcpC*, *RcpA, TadZ, TadA, TadB and TadC*) were annotated in the K6 strain. The *Tad* gene is required for tenacious biofilm development and the creation of bundled *Flp* pili (fibrils) that facilitate adhesion (Perez *et al.,* 2006). T2SS gene cluster encodes a secretion mechanism for fibril export and assembly. T7SS of K6 includes *FimA* gene CDS, essential for synthesizing type 1 pilus. Type 1 pili facilitate receptor structure binding, allowing the bacterium to colonize and invade different host tissues (Korotkov *et al.,* 2012). Thus, the synthesis of type 1 pilus may play crucial role during phytopathogenesis process of *B. glumae* toward rice.

As for subsystem feature counts, the *B. glumae* P3 strain showed the same amount of CDS as the K6 strain related to virulence, disease, and defense mechanism (Figure 4.13). This result indicated that both *B. glumae* strain from

our study resists colicin E2 and resistance to antibiotic and toxic compounds such as copper, Zinc, fluoroquinolones, and other compounds. However, under the subsystem of the mycobacterium virulence operon involved in quinolinate biosynthesis, there is only 4 CDS of *Rv1595* (responsible for quinolinate sythetase) in the P3 strain (Sassetti & Rubin, 2003). In comparison, there are 5 CDS in the K6 strain for *Rv1595*. This is probably due to gapping between contigs, or the P3 strain has less *Rv1595* gene across the genome than the K6 strain (Sassetti & Rubin, 2003).

Under the motility and chemotaxis subsystem, there are 26 CDS pointed out from flagellar motility in the P3 strain. All CDS for flagellar motility in the K6 strain can be found in the P3 strain. However, CDS of CheA (signal transduction histidine kinase CheA) is not found in the genome sequence of the K6 strain. Next, there are 25 CDS related to the protein secretion system, where 23 of the CDS belong to T2SS, and 2 of the CDS belong to T7SS. Compared to the K6 strain, P3 strain have 2 RcpA and 2 TadB protein found on the genome sequences at location 3948, 4888, 2889, and 3945, respectively. *RcpA* protein is secretin from the type II/IV secretion system that is linked with the Flp pilus assembly. As for T7SS, CDS found for the K6 and P3 strains are similar.

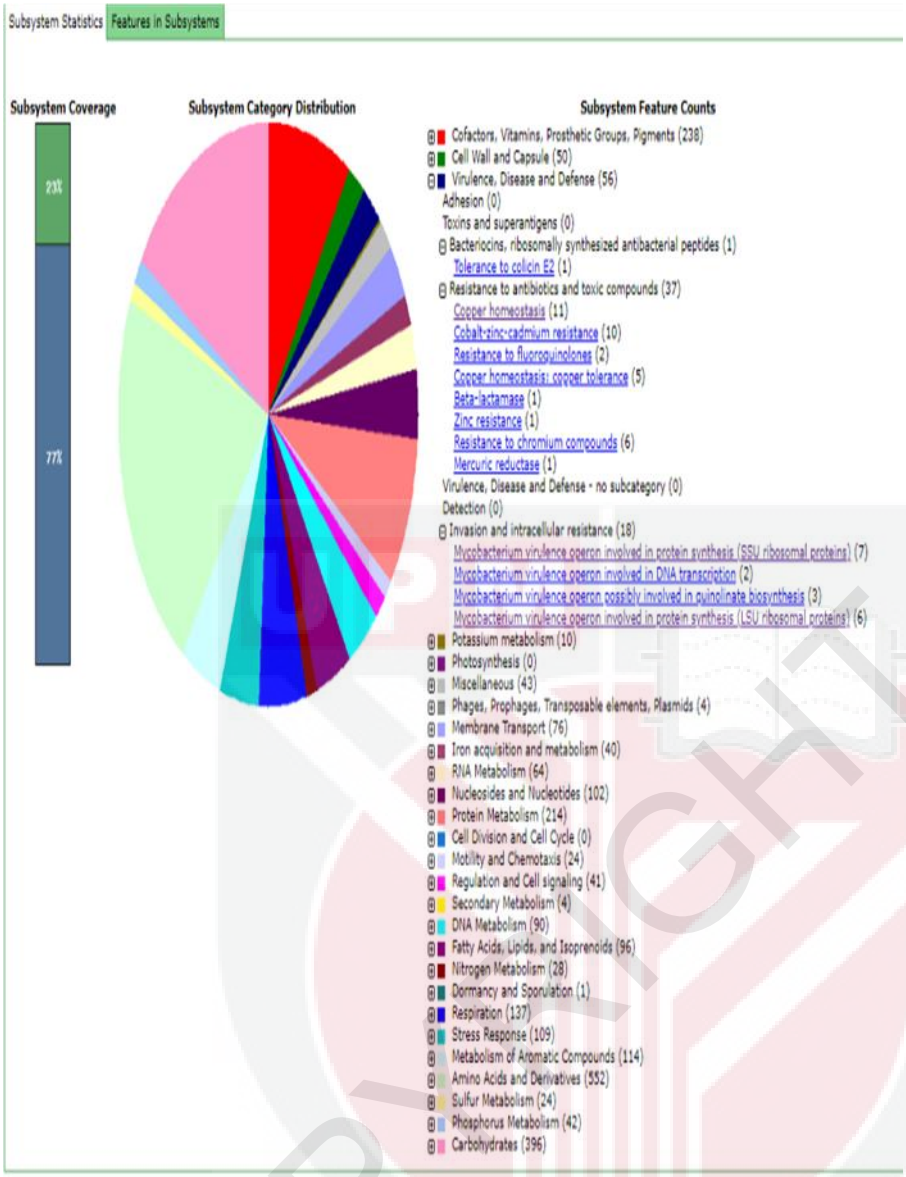

Figure 4.14: Subsystem coverage, category distribution and feature counts of the *B. gladioli* strain UPMBG7

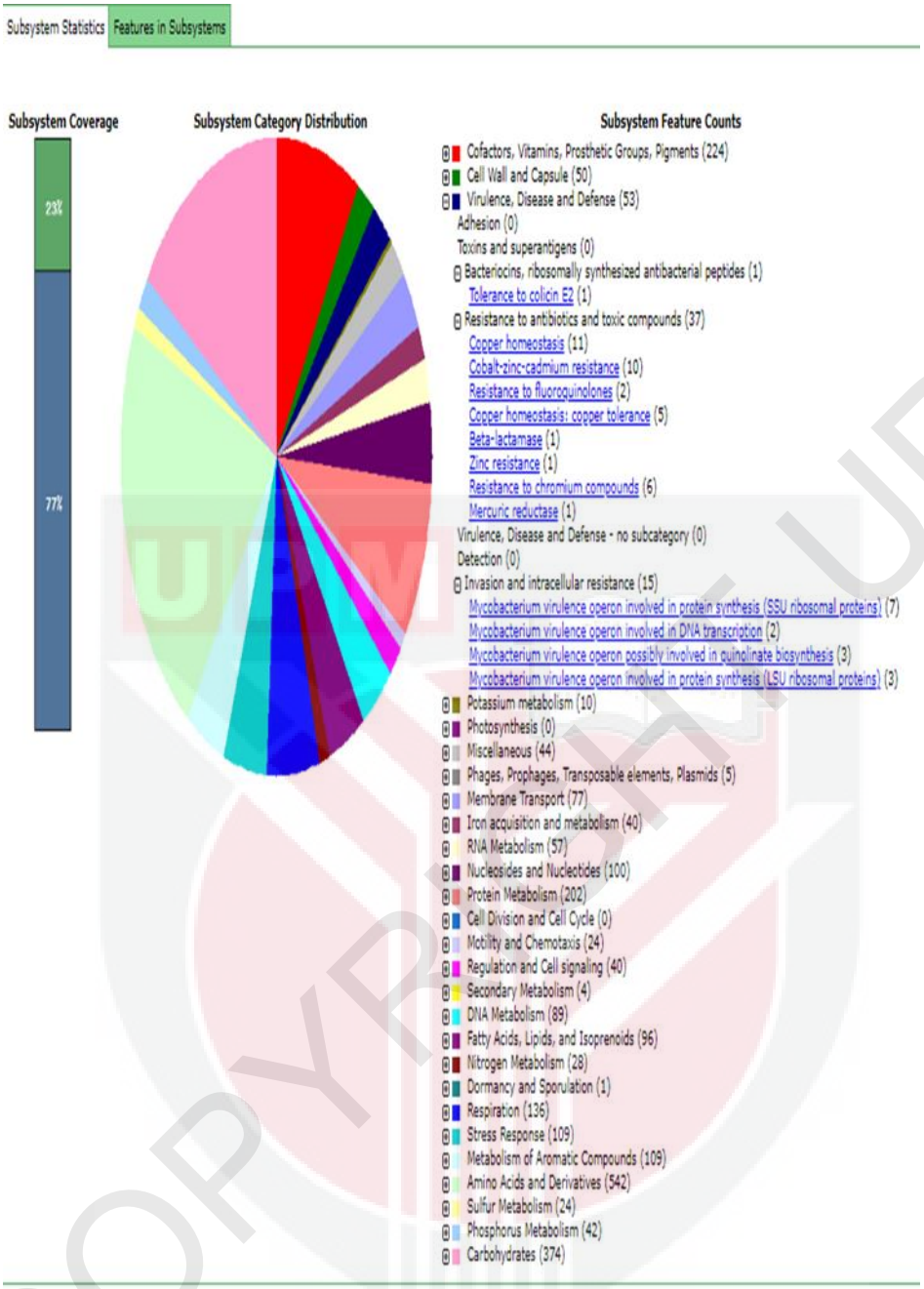

Figure 4.15: Subsystem coverage, category distribution and feature counts of the *B. gladioli* strain UPMBG8

B. gladioli UPMBG7 and UPMBG8 strains subsystem feature counted found that there are 56 and 53 CDS related to virulence, disease and defense mechanism. From the annotated CDS, our study found that *B. gladioli* strain UPMBG7 and UPMBG8 CDS functions as the mercuric reductase (PF00070 family, FAD-dependent NAD(P)-disulphide oxidoreductase), which plays an

essential role in mercury detoxification and biogeochemistry under resistance to antibiotic and toxic compounds subsystem (Figure 4.12- 4.15) (Wang *et al.,* 2009). However, this family gene was not found in *B. glumae* P3 and K6 strains.

Both UPMBG7 and UPMBG8 have 24 CDS genes related to flagellar motility, showing that *B. glumae* and *B. gladioli* have flagella structures which may play an important role in virulence toward the host plant. The protein secretion subsystem shows that UPMBG7 have 33 CDS, where 24 CDS belong to T2SS, 1 CDS belongs to T7SS and 8 CDS belong to type V protein secretion system (T5SS), while UPMBG8 has 34 CDS with 9 CDS belong to T5SS compared to UPMBG7. Comparing T2SS of *B. glumae* and *B. gladioli* of our strain show that both strain of *B. gladioli* does not have FimA gene CDS which are encoded for the synthesis of type 1 pilus. Type 1 pilus absence for this gene is probably due to gapping sequence between contigs or *B. gladioli* does not have gene FimA like *Burkholderia multivoran* strain ATCC 17616 (accession number 395019.3).

Both *B. gladioli* strains obtained from this study annotated the CDS for T5SS, but none for *B. glumae* strains (Figure 4.12-4.15). The T5SS issues that *TpsA and TpsB* (TPS) pathways release a variety of major pathogenicity exoproteins, including Ca21-independent cytolysins, an iron acquisition protein, and many adhesins (Jacob-Dubuisson *et al*., 2001).

4.3.5.3 Gene Ontology Annotation

Gene ontology annotated based on three main parts which consists of cellular component level, molecular function level and Biological process level. Under component level, gene that involved in 10 type of activity level were annotated such as transporter activity, antioxidant activity, enzyme regulator activity, catalytic activity and other activity are annotated. While for molecular function level, gene that involved in 10 type of different activity level were annotated such as membrane-enclosed lumen, cell part and other activity are annotated. Lastly for biological process, gene that involved in 23 type of process were annotated such as signaling, locomotion, biological adhesion and others process (Figure 4.16-4.19).

Under biological process annotation of *B. glumae* K6, 22 gene were annotated under biological adhesion, 54 gene were annotated under locomotion process, 242 gene were annotated under signaling process and 3 gene were annotated under immune system process. These biological processes may play an important role during interaction between pathogenic bacteria with host plant.

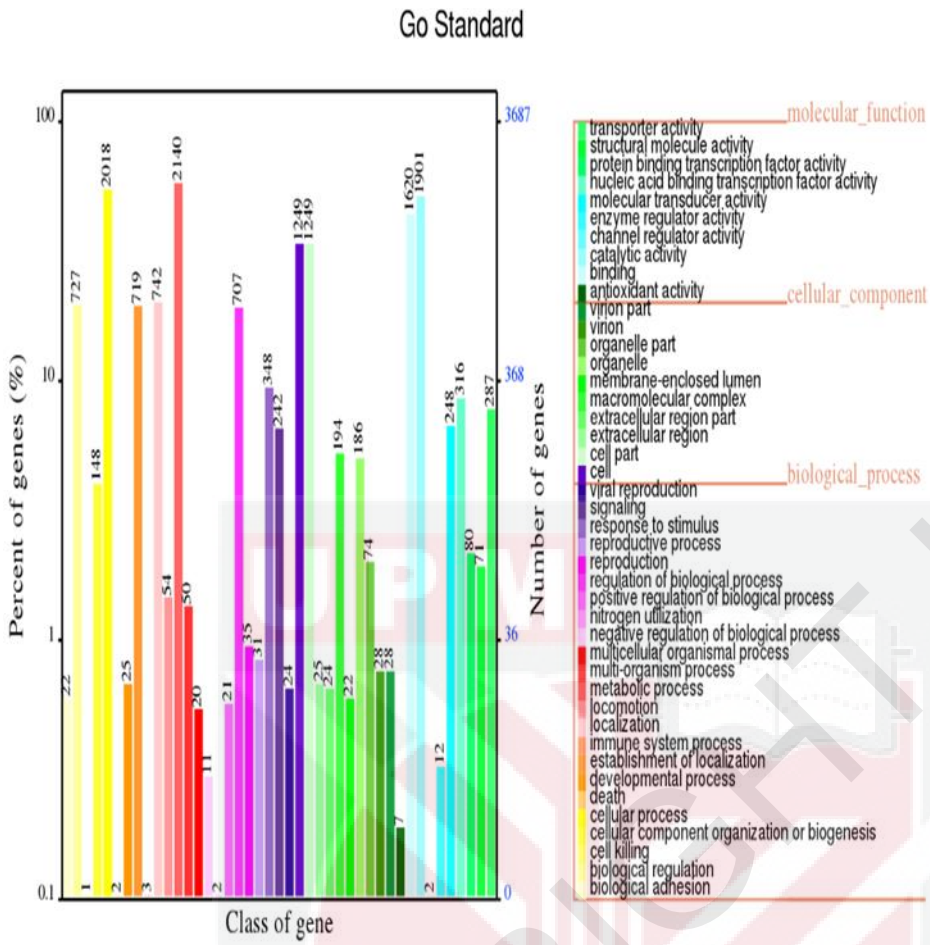

Figure 4.16: GO annotation for *B. glumae* K6 strain

Biological process annotation from GO for *B. glumae* strain P3 showed 21 genes annotated under the biological adhesion process, 50 genes annotated under the locomotion process, 228 genes annotated under the signaling process, and three genes annotated for the immune system process.

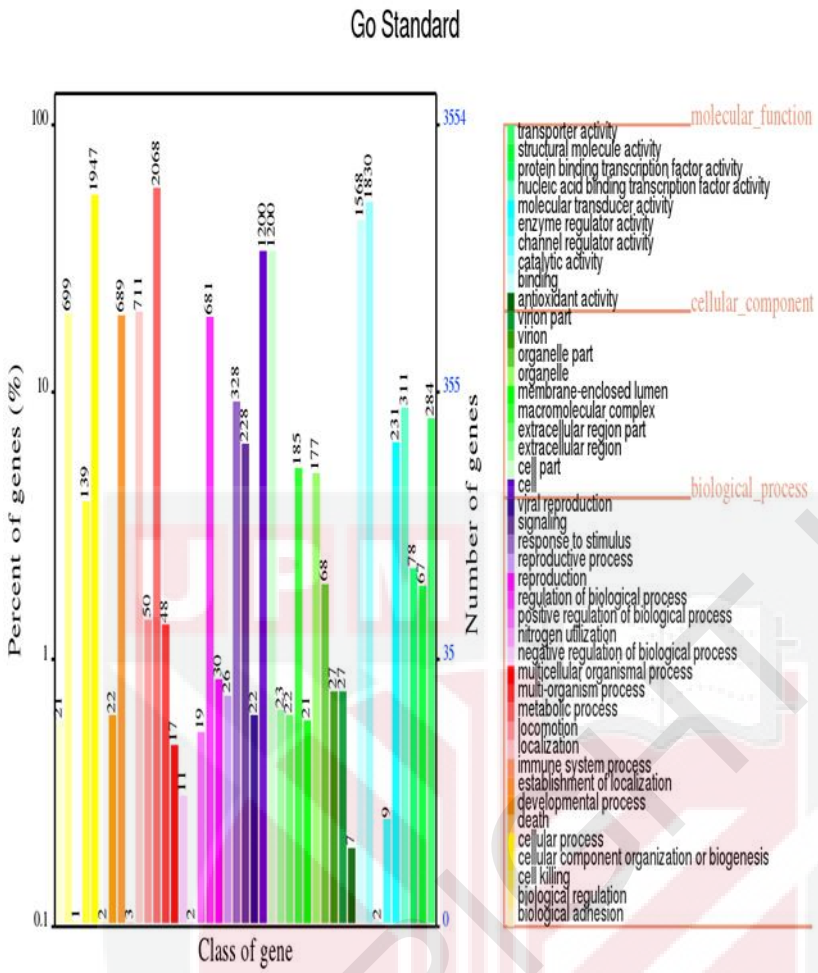

Figure 4.17: GO annotation for *B. glumae* P3 strain

The Gene Ontology (GO) analysis and annotation of the B. glumae K6 and P3 strains' genomes reveal a comprehensive landscape of functional roles and biological processes inherent to these bacterial strains. Within the Biological Process category, it is evident that both strains share commonalities in several aspects. For instance, both strains exhibit genes involved in biological adhesion, biological regulation, cell killing, death processes, negative regulation of biological processes, nitrogen utilization, and immune system processes, suggesting fundamental biological functions conserved across these strains.

However, notable differences emerge. The K6 strain stands out with 22 genes dedicated to biological adhesion, while the P3 strain features 21 genes in this category. Similarly, in the realm of biological regulation, the K6 strain boasts 727 actively annotated genes, while the P3 strain features 699. These disparities may reflect subtle variations in regulatory mechanisms between the two strains, potentially contributing to differences in their biological behaviors.

In cellular processes, both strains exhibit significant genetic involvement, with 2018 genes for K6 and 1947 genes for P3. This suggests a shared commitment to fundamental cellular activities. The presence of genes related to cellular component organization or biogenesis underscores their role in shaping cellular structures, with 148 genes in K6 and 139 in P3 dedicated to this task.

Metabolic processes are notably enriched, with 2140 genes in the K6 strain and 2068 genes in P3. This highlights the strains' metabolic versatility and adaptability to a variety of environmental conditions.

Furthermore, genes associated with multi-organism processes and multicellular organismal processes point to the strains' potential ecological relevance and interactions within complex biological communities.

The presence of genes related to locomotion, immune responses, signaling, and viral production suggests that both strains possess the genetic arsenal for mobility, interactions with their environment, and potential roles in host pathogen interactions.

In conclusion, the GO analysis and annotation offer a comprehensive view of the functional diversity within the genomes of the B. glumae K6 and P3 strains. While both strains share core biological processes, subtle differences in gene counts highlight potential variations in regulatory mechanisms and ecological adaptations. These findings contribute to our understanding of the genetic underpinnings of these strains' biology, with implications for their roles in environmental and host-associated interactions.

As for the *B. gladioli* UPMBG7 strain, biological process annotation showed 22 genes were annotated under the biological adhesion process, 44 genes were annotated under the locomotion process, 254 genes were annotated under the signaling process, and one gene was annotated under the immune system process.

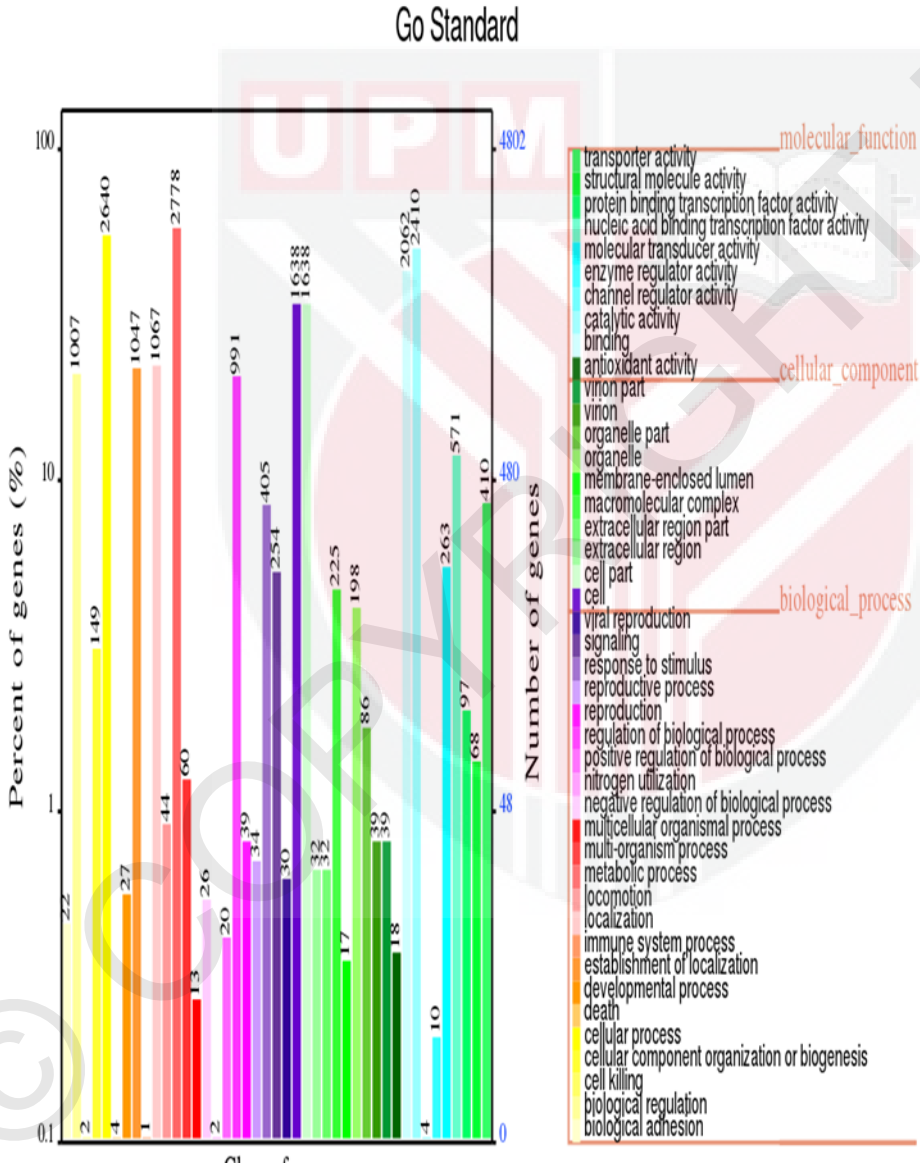

Figure 4.18: GO annotation for *B. gladioli* UPMBG7 strain

For GO annotation of *B. gladioli,* UPMBG8 showed 21 genes annotated under the biological adhesion process, 42 genes annotated under the locomotion process, 247 genes annotated under the signaling process, and one gene was annotated for the immune system process.

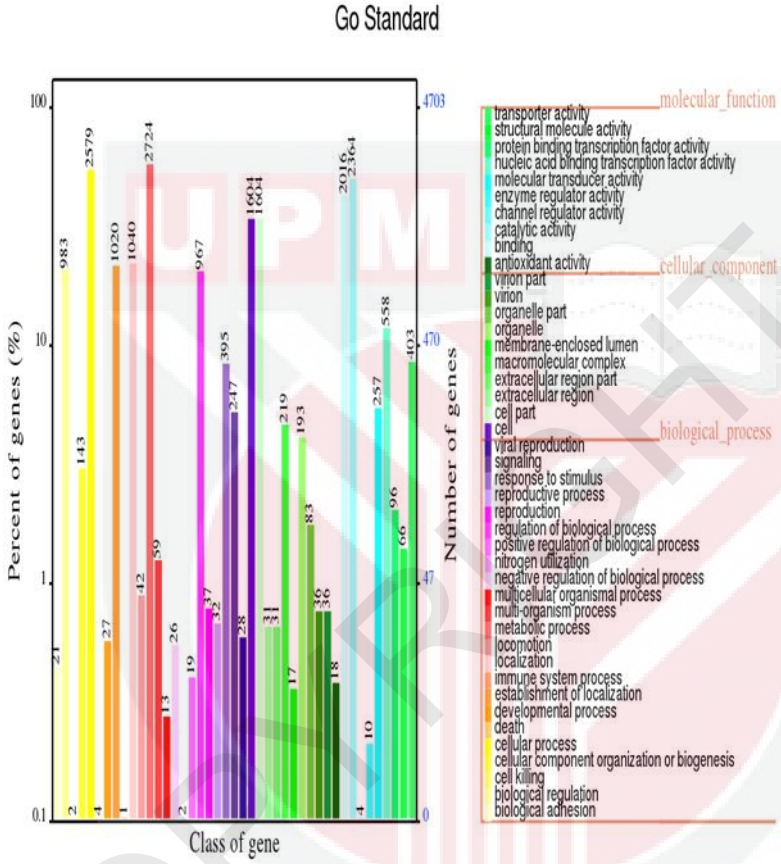

Figure 4.19: GO annotation for B. gladioli UPMBG8 strain.

The Gene Ontology (GO) analysis and annotation of the B. gladioli UPMBG7 and UPMBG8 strains' genomes unveil an intricate landscape of functional attributes and biological processes inherent to these bacterial strains. Within the Biological Process category, these strains share several fundamental processes, indicating core biological functions conserved across the two strains.

Both strains exhibit genes related to biological adhesion, cell killing, death processes, developmental processes, multi-organism processes, multicellular organismal processes, negative regulation of biological processes, nitrogen utilization, and immune system processes, underscoring their shared commitment to essential biological functions.

Notably, differences emerge in certain aspects of biological regulation. The UPMBG7 strain is distinguished by the annotation of 1007 actively regulated genes, while UPMBG8 features 983 genes in this category. These distinctions may suggest nuanced variations in regulatory mechanisms between the two strains, potentially influencing their biological behaviors and adaptability.

In cellular processes, both strains actively engage a substantial number of genes, reflecting their commitment to fundamental cellular activities. Genes related to cellular component organization or biogenesis emphasize their roles in shaping cellular structures, with 149 genes in UPMBG7 and 143 in UPMBG8 dedicated to this function.

Metabolic processes are notably enriched, with UPMBG7 featuring 2778 genes and UPMBG8 featuring 2724 genes. This highlights their metabolic versatility and adaptability to various environmental conditions.

Furthermore, genes related to localization, locomotion, response to stimulus, signaling, and viral production suggest that both strains possess a genetic repertoire for mobility, environmental sensing, and potential roles in host-pathogen interactions.

In conclusion, the GO analysis and annotation provide a comprehensive view of the functional diversity within the genomes of the B. gladioli UPMBG7 and UPMBG8 strains. While they share core biological processes, subtle differences in gene counts hint at potential variations in regulatory mechanisms and ecological adaptations. These findings enhance our understanding of the genetic foundations of these strains' biology, with implications for their roles in environmental and host-associated interactions.

4.3.5.4 Cluster of Orthologous Groups of Protein Annotations

Classifications of the proteins encoded by gene were annotated via Cluster of Orthologous Groups of Protein Annotation (COG) and divided into 24 parts based on their functions (Figure 4.20-4.23).

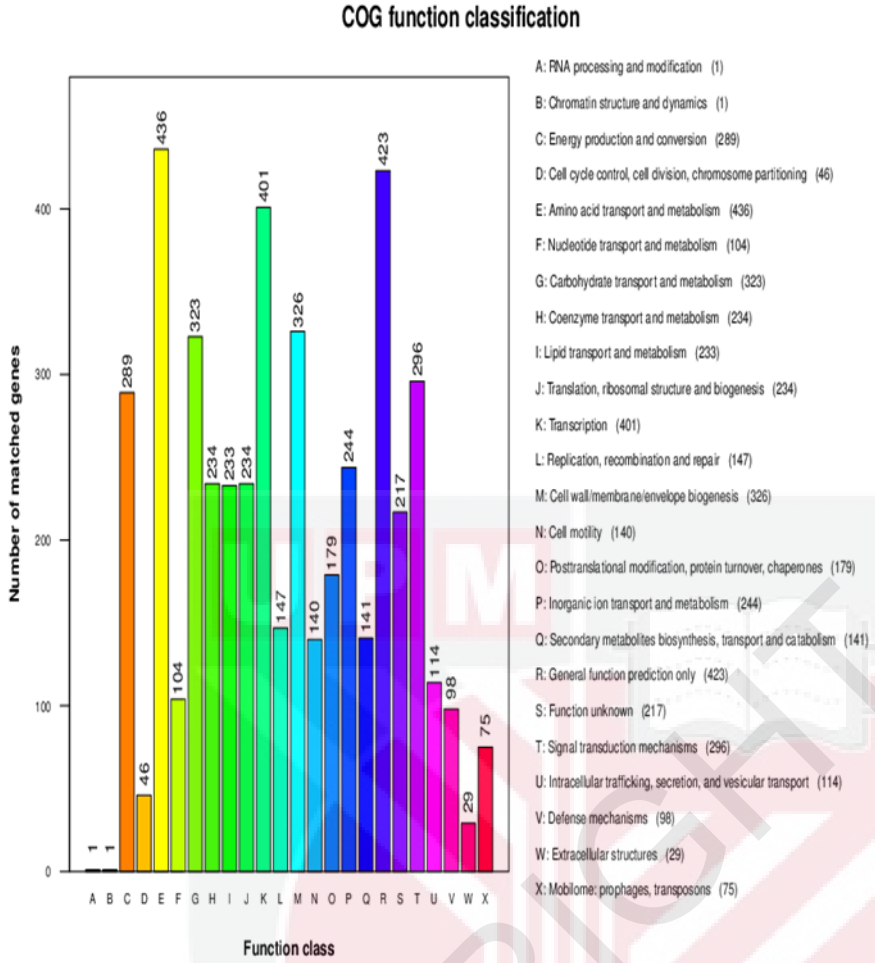

Figure 4.20: The gene functions present in *B. glumae* K6 strain using COG software

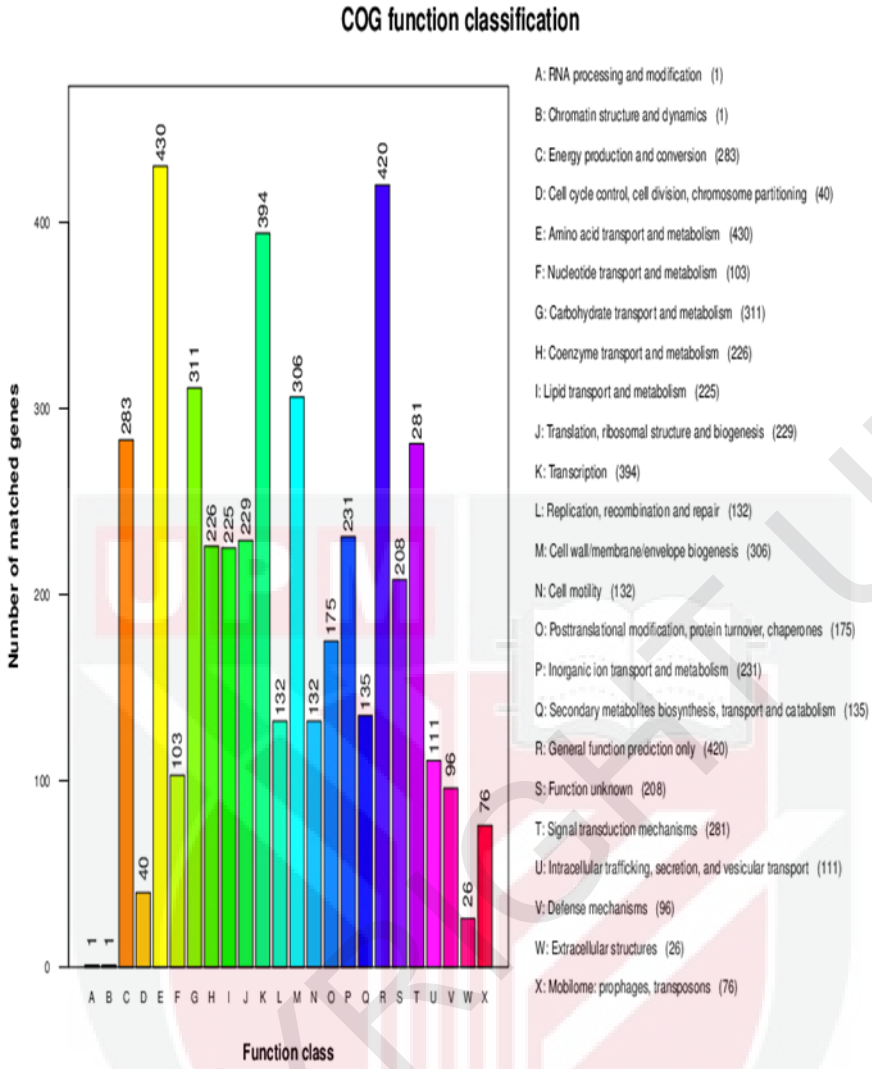

Figure 4.21: The gene functions present in *B. glumae* P3 strain using COG software

The Clusters of Orthologous Groups (COG) annotation analysis provides valuable insights into the functional diversity and potential roles of *B. glumae* K6 and P3 strains in various biological processes. In the category of RNA processing and modification, both strains exhibit limited representation with only one annotated gene each, indicating that they share fundamental mechanisms for RNA processing and modification. Similarly, in the chromatin structure and dynamics category, both strains possess a single annotated

gene, suggesting a commonality in their ability to regulate chromatin organization.

Moving on to energy production and conversion, K6 and P3 strains show slight variations with K6 having 289 annotated genes compared to P3's 283. These differences may reflect subtle distinctions in their energy metabolism strategies, potentially influenced by their respective environmental niches. In the realm of cell cycle control, cell division, and chromosome partitioning, K6 surpasses P3 with 46 annotated genes compared to P3's 40. These genes likely play pivotal roles in regulating cell division and chromosome organization, and the differences in gene counts might indicate diverse strategies in cellular regulation and growth between the two strains.

Amino acid transport and metabolism reveal a wealth of genetic potential in both strains, with K6 featuring 436 annotated genes and P3 having 430. This signifies their adaptability to a broad range of amino acids, essential for their survival in diverse environments. Nucleotide transport and metabolism, although relatively conserved, exhibit a slight difference, with K6 having 104 annotated genes and P3 possessing 103. These genes play roles in nucleotide metabolism, essential for DNA and RNA synthesis.

Carbohydrate transport and metabolism genes are abundant in both strains, with K6 and P3 having 323 and 311 annotated genes, respectively. This reflects their capacity to metabolize a variety of carbohydrates, facilitating their adaptation to different carbon sources. Coenzyme transport and metabolism genes show slight differences, with K6 having 234 genes and P3 having 226.

These genes are vital for enzymatic reactions and metabolic processes, potentially influencing their metabolic versatility.

In lipid transport and metabolism, K6 has 233 annotated genes compared to P3's 225. These genes likely contribute to differences in lipid utilization and membrane composition, impacting their membrane stability and adaptability. Translation, ribosomal structure, and biogenesis-related genes are abundant in both strains, with K6 featuring 234 annotated genes and P3 having 229. This underscores their capacity for protein synthesis and cell growth. Transcription-related genes vary slightly, with K6 possessing 401 annotated genes and P3 having 394. These genes likely influence gene expression patterns and responses to environmental cues, contributing to their distinct biology. Replication, recombination, and repair genes are crucial for maintaining genome stability. K6 has 147 annotated genes, whereas P3 has 132, suggesting potential differences in DNA repair mechanisms.

Cell wall/membrane/envelope biogenesis is essential for interactions with the environment and host organisms. K6 possesses 326 annotated genes, while P3 has 306, indicating variations in cell envelope structure. In cell motility, K6 has 140 annotated genes compared to P3's 132, reflecting the ability of this strain for motility-related strategies and potential ecological niches.

Posttranslational modification, protein turnover, and chaperone-related genes are abundant in both strains, with K6 having 179 and P3 having 175. These genes contribute to protein folding, stability, and cellular responses to stress.

Inorganic ion transport and metabolism show differences, with K6 having 244 annotated genes and P3 having 231. These genes likely influence their ability to acquire and utilize inorganic ions. Secondary metabolites biosynthesis, transport, and catabolism genes exhibit variations, with K6 having 141 and P3 having 135. These genes may be involved in the production of secondary metabolites with potential ecological or pathogenic significance.

General function prediction reveals functional diversity, with K6 featuring 423 genes and P3 having 420. These genes may play multifaceted roles in various cellular processes. Function unknown genes hint at uncharted territories, with K6 having 217 and P3 having 208. Further exploration of these genes could unveil novel functions and mechanisms. Signal transduction mechanisms are diverse, with K6 having 296 genes and P3 having 281. These genes likely mediate responses to environmental cues, contributing to their adaptability.

Intracellular trafficking, secretion, and vesicular transport genes are essential for cellular communication. K6 has 114 annotated genes, while P3 has 111, potentially influencing their interactions with the environment and other organisms. Defense mechanisms are essential for survival, with K6 featuring 98 genes and P3 having 96. These genes may provide protection against external threats.

Extracellular structures genes are involved in surface features, with K6 having 29 and P3 having 26. These structures may play roles in adherence, colonization, or virulence. Mobilome genes related to prophages and

transposons show slight variations, with K6 having 75 genes and P3 having 76. These elements may contribute to genome plasticity and adaptation.

In summary, the COG annotation results highlight both shared functional aspects and subtle differences between B. glumae K6 and P3 strains. These genomic distinctions could be critical for their ecological niches, pathogenicity, and responses to environmental challenges. Further investigations into specific genes within these categories may reveal the molecular mechanisms underpinning their distinct biological traits and adaptive strategies

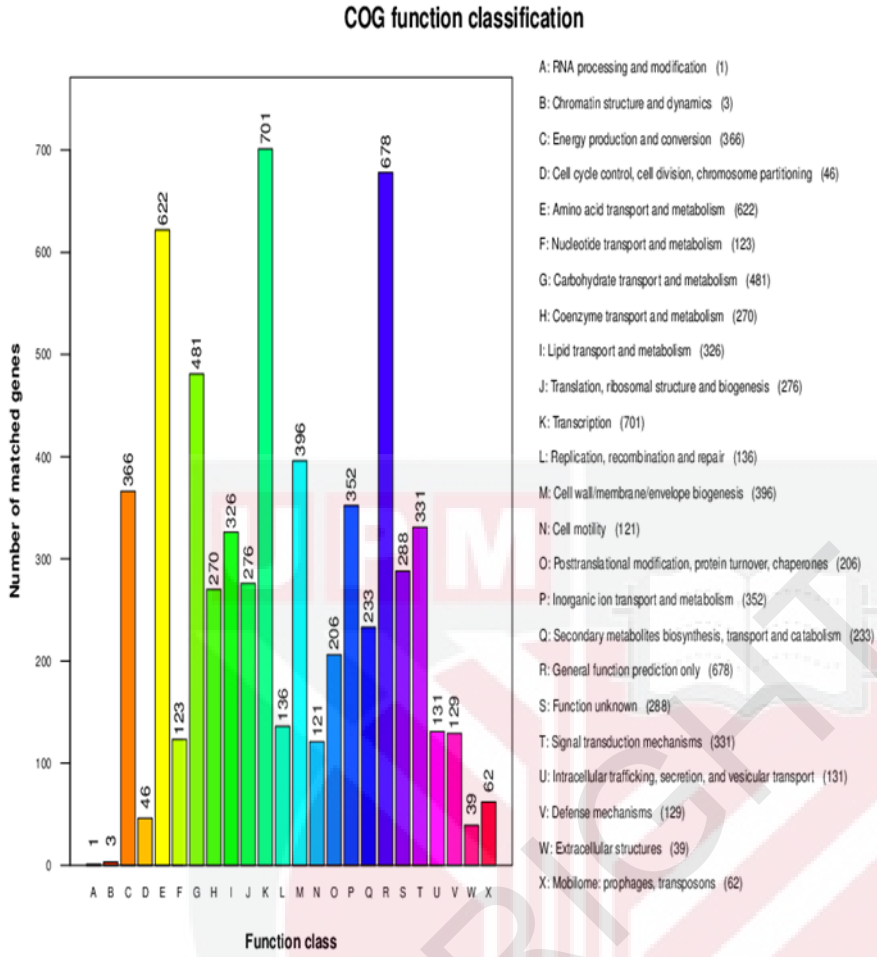

Figure 4.22: The gene functions present in *B. gladioli* UPMBG7 strain using COG software

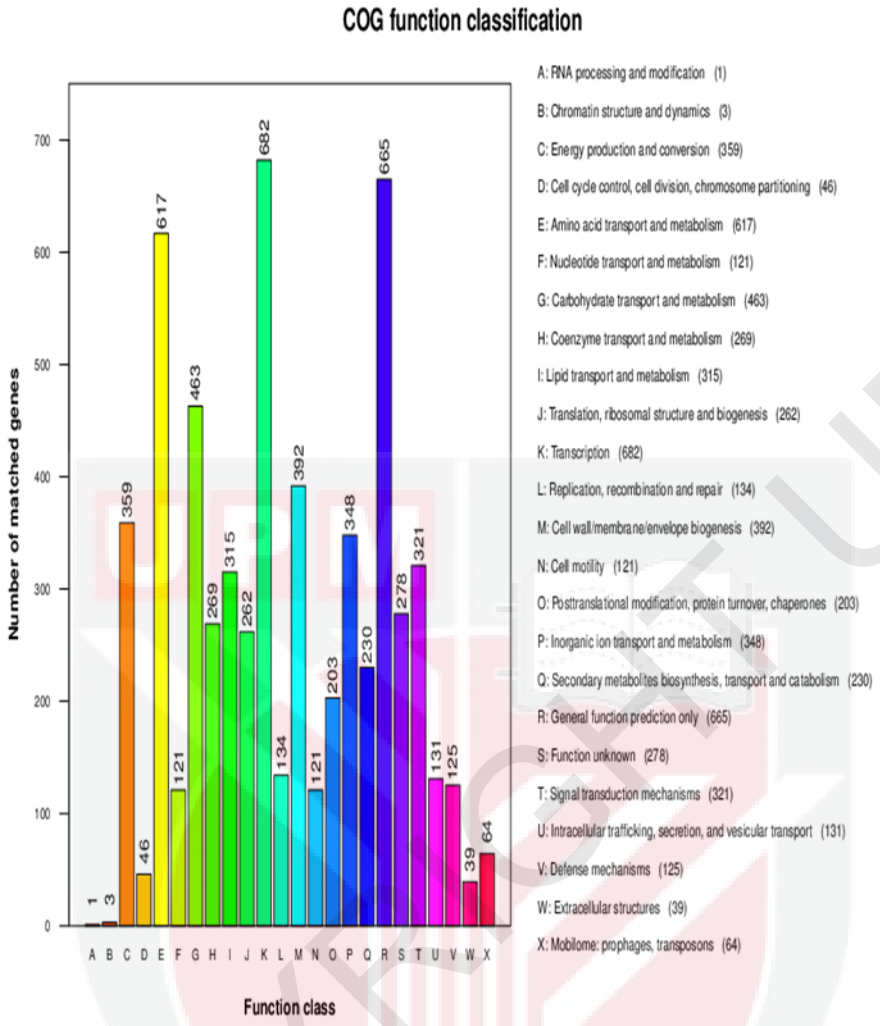

Figure 4.23: The gene functions present in *B. gladioli* UPMBG8 strain using COG software

The Clusters of Orthologous Groups (COG) annotation analysis sheds light on the functional diversity and potential roles of B. glumae UPMBG7 and UPMBG8 strains across various biological processes. In the category of RNA processing and modification, both strains exhibit minimal representation, with just one annotated gene each. This suggests that they share fundamental mechanisms for RNA processing and modification. Chromatin structure and dynamics are also relatively conserved between UPMBG7 and UPMBG8, with three genes annotated in both strains. These genes are likely involved in regulating chromatin organization, reflecting similarities in their epigenetic control mechanisms.

Energy production and conversion are essential for bacterial survival, and while UPMBG7 has 366 annotated genes in this category, UPMBG8 has 359. These differences may indicate slight variations in their energy metabolism strategies, potentially influenced by their respective environmental niches. Cell cycle control, cell division, and chromosome partitioning are critical for genome stability. Both strains exhibit 46 annotated genes in this category, suggesting that they share common regulatory mechanisms for cell division and chromosome segregation.

Amino acid transport and metabolism, a key aspect of bacterial adaptation, show slight differences, with UPMBG7 having 622 annotated genes and UPMBG8 having 617. These genes likely enable them to utilize a wide range of amino acids, reflecting their adaptability to different nutrient sources. Nucleotide transport and metabolism are essential for DNA and RNA

synthesis. UPMBG7 has 123 annotated genes in this category, while UPMBG8 has 121, indicating a conserved capacity for nucleotide metabolism.

Carbohydrate transport and metabolism are vital for energy acquisition. UPMBG7 boasts 481 annotated genes in this category, whereas UPMBG8 has 463, demonstrating their ability to metabolize various carbohydrates. Coenzyme transport and metabolism, crucial for enzymatic reactions, exhibit a minor distinction, with UPMBG7 having 270 annotated genes and UPMBG8 having 269, suggesting functional similarities in this regard.

Lipid transport and metabolism genes, contributing to membrane stability and adaptation, vary slightly, with UPMBG7 having 326 annotated genes and UPMBG8 having 315. Translation, ribosomal structure, and biogenesis-related genes are abundant in both strains, with UPMBG7 featuring 276 annotated genes and UPMBG8 having 262. These genes are vital for protein synthesis and cell growth.

Transcription-related genes, essential for gene expression regulation, are represented with 701 genes in UPMBG7 and 682 in UPMBG8, indicating a complex regulatory network in both strains. Replication, recombination, and repair genes, crucial for genome maintenance, are slightly more numerous in UPMBG7 with 136 annotated genes, compared to 134 in UPMBG8.

Cell wall/membrane/envelope biogenesis is vital for interactions with the environment and host organisms. UPMBG7 possesses 396 annotated genes, whereas UPMBG8 has 392, suggesting potential variations in cell envelope

structure. Both strains exhibit 121 annotated genes related to cell motility, implying similar mechanisms for bacterial movement.

Posttranslational modification, protein turnover, and chaperones are crucial for protein stability and cellular stress responses. UPMBG7 has 206 genes annotated in this category, and UPMBG8 has 203, indicating shared strategies for protein quality control. Inorganic ion transport and metabolism genes are essential for nutrient acquisition. UPMBG7 has 352 annotated genes, while UPMBG8 has 248, potentially reflecting differences in their ion utilization strategies. Secondary metabolites biosynthesis, transport, and catabolism genes, which may contribute to ecological or pathogenic interactions, are abundant in both strains, with UPMBG7 having 233 and UPMBG8 having 230.

General function prediction genes are multifunctional, with UPMBG7 featuring 678 annotated genes and UPMBG8 having 665, suggesting versatile roles in various cellular processes. Function unknown genes hint at uncharted functional territories, with UPMBG7 having 288 and UPMBG8 having 278, warranting further exploration.

Signal transduction mechanisms, crucial for environmental sensing, show 331 genes in UPMBG7 and 321 in UPMBG8, potentially influencing their adaptability. Intracellular trafficking, secretion, and vesicular transport genes are essential for cellular communication, with 131 annotated genes in both strains. Defense mechanisms, crucial for survival, include 129 genes in UPMBG7 and 125 in UPMBG8. Extracellular structures are represented by 39 genes in both strains, possibly contributing to adhesion and colonization.

Lastly, mobilome-related genes, including prophages and transposons, exhibit slight variations, with 62 in UPMBG7 and 64 in UPMBG8, potentially influencing genome plasticity and adaptation.

In summary, the COG annotation results highlight both shared functional aspects and subtle differences between B. glumae UPMBG7 and UPMBG8 strains. These genomic distinctions could be critical for their ecological niches, pathogenicity, and responses to environmental challenges. Further investigations into specific genes within these categories may reveal the molecular mechanisms underpinning their distinct biological traits and adaptive strategies.

4.3.5.5 Kyoto Encyclopedia of Gene and Genome Annotations

KEGG It is a database resource for deducing high-level functions and utility of biological systems, such as the cell, organism, and ecosystem, from molecular-level data, particularly large-scale molecular datasets generated by genome sequencing and other high-throughput experimental methods. From the analysis of the KEGG pathway, results revealed that *B. glumae* K6 and P3 and *B. gladioli* UPMBG7 and UPMBG8 has all the gene necessary for a complete flagellar assembly and quorum sensing system (Figure 4.24-4.26).

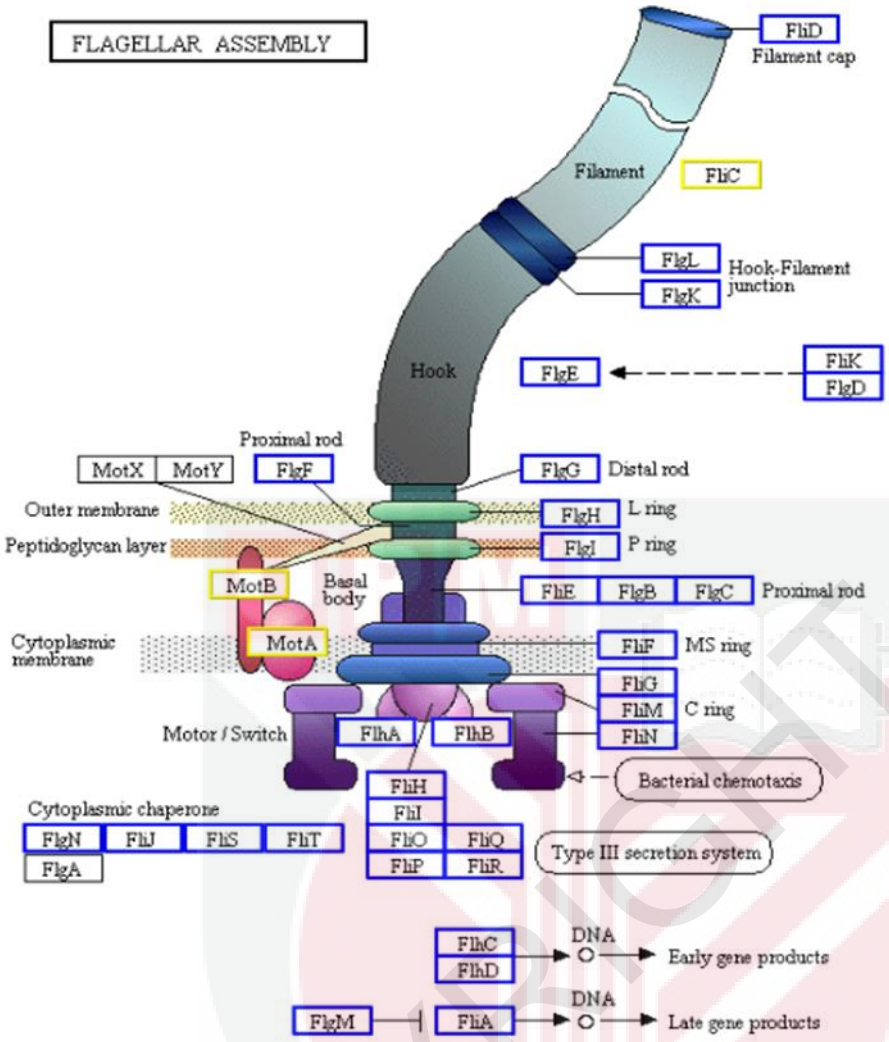

Figure 4.24: The proposed KEGG pathway for flagellar assembly. Highlighted blue and yellow boxes indicate the gene produced by the *B. glumae* K6 and P3 strain and *B. gladioli* UPMBG7 and UPMBG8 strain that are involved in the flagellar assembly

Quorum sensing

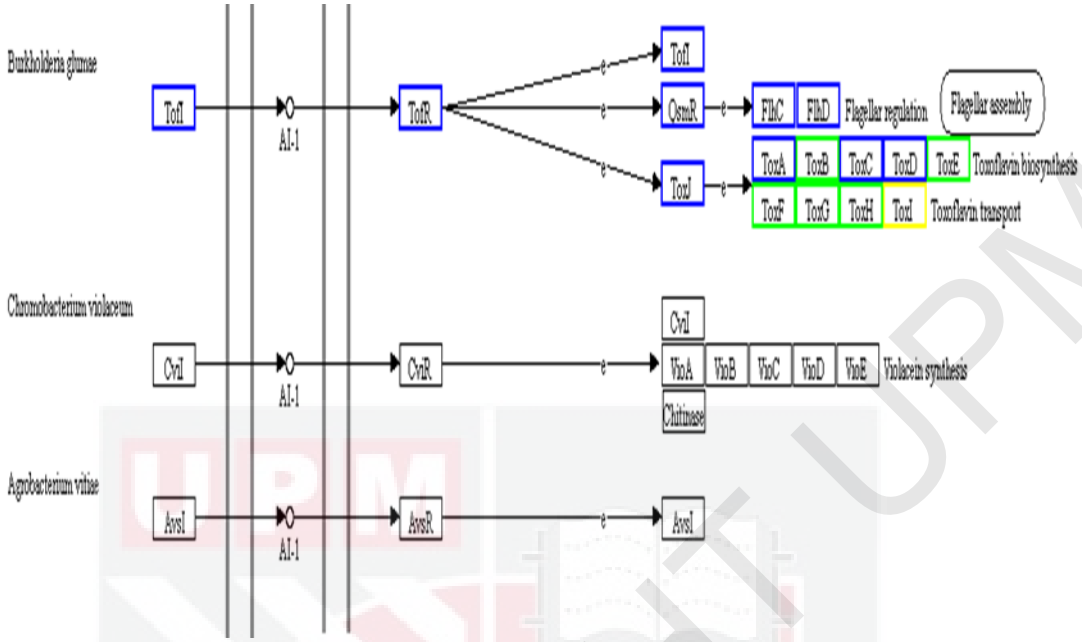

Figure 4.25: The proposed KEGG pathway for the quorum sensing system. Highlighted boxes indicate the gene produced by the *B. glumae* K6 and P3 strains involved in the quorum sensing system
(Toxoflavin and Flagella)

Quorum Sensing

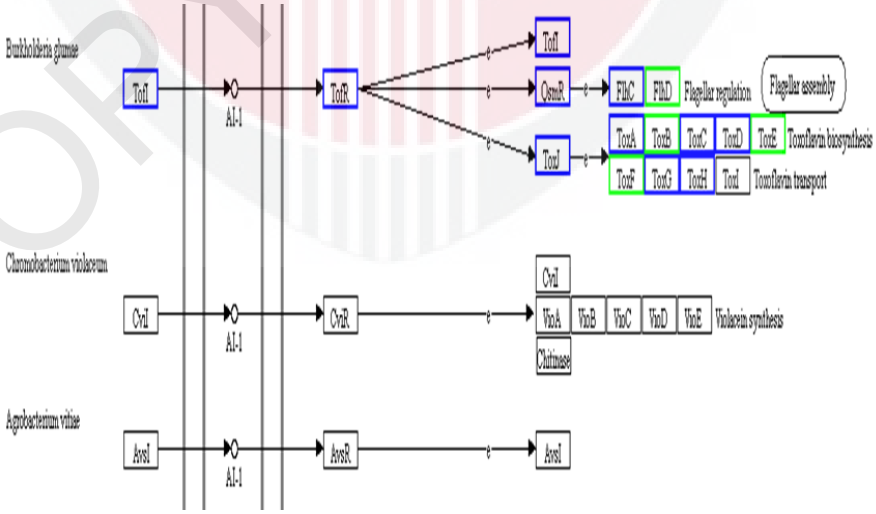

Figure 4.26: The proposed KEGG pathway for the quorum sensing system. Highlighted boxes indicate the gene produced by the *B. gladioli* UPMBG7 and UPMBG8 strains involved in the quorum sensing system
(Toxoflavin and Flagella)

4.3.5.6 Analysis of Virulence Gene Involve in Causing Panicle Blight disease

Putative gene and putative genetic regions expressed by *B. glumae* K6 and P3 and *B. gladioli* UPMBG7 and UPMBG8 present in quorum sensing system (QS), *Hrp* type III secretion system, lipase formation (*LipA and LipB*) and flagella formation were listed in Table 4.12 and Table 4.13 to Table 4.16. There are 19 open reading frames (ORFs) for *B. glumae*, 18 ORFs for *B. gladioli*, and 12 enzymes involved in the virulence factor of *Burkholderia* spp.

Table 4.12: Genes involved in the virulence factor of *Burkholderia* spp

No	System	Gene	Description	Reference
1	Quorum Sensing System	*TofI*	*N-Acyl* homoserine lactone (AHL) synthase for *N-hexanoyl* homoserine lactone (C6-HSL) and C8-HSL	Kim *et al.,* (2004)
2		*TofR*	Cognate receptor for C8-HSL	Kim *et al.,* (2004)
3		*ToxR*	LysR-type transcriptional activator	Ham *et al.,* (2010)
4		*ToxJ*	LuxR C-terminal-related transcriptional regulator	Ham *et al.,* (2010)
5		*toxF, toxG, toxH, toxI*	Toxoflavin tranport	Francis *et al.,* (2013)
6		*ToxA, ToxB, ToxC, ToxD, ToxE*	Toxoflavion biosynthesis	Francis *et al.,* (2013)
7	Flagella Formation	*QsmR*	Ic1R-type transcriptional regulator	Kim *et al.,* (2007)
8		*FlhC*	Flagellar assembly, biosynthesis, chemotaxis	Kim *et al.,* (2007)
9		*FlhD*	Flagellar assembly	Kim *et al.,* (2007)
10	Type III Secretion System	*HrpB*	Regulatory factor for expression of T3SS	Kang et al., (2008)
11	Lipase Formation	*LipA*	Cell-wall degrading enzyme	Zhou-qi (2016)
12		*LipB*	Biosynthesis and activation of *LipA*	Zhou-qi (2016)

Table 4.13: The genetic regions of *B. glumae* P3 that encode an important gene involved as virulence factors.

ORF No.	Start	End	Putative Protein	Gene Name	Bases	Contigs	Accession no.
ORF 1	7036	7646	GNAT family N-acetyltransferase/ AHL synthase	*TofI*	611	95	JAMYCT010000099.1
ORF 2	5211	8882	AHL receptor	*TofR*	3672	95	JAMYCT010000099.1
ORF 3	22985	23984	LysR family transcriptional regulator	*ToxR*	1000	10	JAMYCT010000010.1
ORF 4	14818	15648	LuxR C-terminal-related transcriptional regulator	*ToxJ*	831	10	JAMYCT010000010.1
ORF 5	24476	25213	class I SAM-dependent methyltransferase	*ToxA*	738	10	JAMYCT010000010.1
ORF 6	25335	25913	GTP cyclohydrolase II	*ToxB*	579	10	JAMYCT010000010.1
ORF 7	25970	27661	WD40 repeat domain-containing protein	*ToxC*	1692	10	JAMYCT010000010.1
ORF 8	27769	28749	Formylglycine-generating enzyme family protein/TRP2	*ToxD*	981	10	JAMYCT010000010.1

Table 4.13: Continue

ORF 9	28818	29903	Bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracilreductase *RibD*/deaminase	*ToxE*	1084	10	JAMYCT010000010.1
ORF10	22314	22889	DMT family transporter	*ToxF*	576	10	JAMYCT010000010.1
ORF11	21135	22274	Efflux RND transporter periplasmic adaptor subunit	*ToxG*	1140	10	JAMYCT010000010.1
ORF12	18046	21138	Efflux RND transporter permease subunit	*ToxH*	3093	10	JAMYCT010000010.1
ORF13	16437	17974	Efflux transporter outer membrane subunit	*ToxI*	1538	10	JAMYCT010000010.1
ORF14	11071	11898	IclR family transcriptional regulator	*QsmR*	828	108	JAMYCT010000112.1

Table 4.13: Continue

ORF15	10608 23140	1176224294	Flagellin	*FlhC*	1155	7556	JAMYCT010000078.1 JAMYCT010000058.1
ORF16	4238	4557	Flagellar transcriptional regulator FlhD	*FlhD*	320	75	JAMYCT010000078.1
ORF17	8291	9694	Helix-turn-helix transcription regulator of pathogenicity genes	*HrpB*	1404	153	JAMYCT010000158.1
ORF18	2177	3163	Lipoyl synthase	*LipA*	987	289	JAMYCT010000297.1
ORF19	1432	2184	Lipoyl(octanoyl) transferase *LipB*	*LipB*	753	289	JAMYCT010000297.1

Table 4.14: The genetic regions of *B. glumae* K6 that encode important genes involved as virulence factors.

ORF No.	Start	End	Putative Protein	Gene Name	Bases	Contigs	Accession no.
ORF 1	7035	7646	GNAT family N-acetyltransferase/ AHL synthase	*TofI*	612	84	JAMYCS010000087.1
ORF 2	5211	8882	AHL receptor	*TofR*	3672	84	JAMYCS010000087.1

Table 4.14: Continue

ORF 3	41863	42862	LysR family transcriptional regulator	*ToxR*	1000	6	JAMYCS010000006.1
ORF 4	50199	51029	LuxR C-terminal-related transcriptional regulator	*ToxJ*	831	6	JAMYCS010000006.1
ORF 5	40634	41371	class I SAM-dependent methyltransferase	*ToxA*	738	6	JAMYCS010000006.1
ORF 6	39934	40512	GTP cyclohydrolase II	*ToxB*	579	6	JAMYCS010000006.1
ORF 7	38186	39877	WD40 repeat domain-containing protein	*ToxC*	1692	6	JAMYCS010000006.1
ORF 8	37098	38078	Formylglycine-generating enzyme family protein/TRP2	*ToxD*	981	6	JAMYCS010000006.1

Table 4.15: The genetic regions of *B. gladioli* UPMBG7 that encode important genes involved as virulence factors

ORF No	Start	End	Putative Protein	Gene Name	Bases	Contigs	Accession no.
ORF 1	9949	10527	GNAT family N-acetyltransferase/ AHL synthase	*TofI*	579	43	JANIEE010000046.1
ORF 2	8646	10697	AHL receptor	*TofR*	2052	43	JANIEE010000046.1
ORF 3	29000	29997	LysR family transcriptional regulator	*ToxR*	998	52	JANIEE010000056.1
ORF 4	113847	114661	LuxR C-terminal-related transcriptional regulator	*ToxJ*	815	2	JANIEE010000002.1
ORF 5	27769	28506	class I SAM-dependent methyltransferase	*ToxA*	738	52	JANIEE010000056.1
ORF 6	27070	27648	GTP cyclohydrolase II	*ToxB*	579	52	JANIEE010000056.1
ORF 7	25322	27013	WD40 repeat domain-containing protein	*ToxC*	1692	52	JANIEE010000056.1

Table 4.15: Continue

ORF 8	24226	25202	Formylglycine-generating enzyme family protein/TRP2	*ToxD*	977	52	JANIEE010000056.1

Table 4.16: The genetic regions of *B. gladioli* UPMBG8 that encode important genes involved as virulence factors

ORF No.	Start	End	Putative Protein	Gene Name	Bases	Contigs	Accession no.
ORF 1	136267	136845	GNAT family N-acetyltransferase/ AHL synthase	*TofI*	579	20	JANIEE010000022.1
ORF 2	134964	137015	AHL receptor	*TofR*	2052	20	JANIEF010000022.1
ORF 3	29000	29997	LysR family transcriptional regulator	*ToxR*	998	50	JANIEF010000054.1
ORF 4	249371	250185	LuxR C-terminal-related transcriptional regulator	*ToxJ*	815	2	JANIEF010000003.1

Table 4.16: Continue

ORF 5	27769	28506	class I SAM-dependent methyltransferase	*ToxA*	738	50	JANIEF010000054.1
ORF 6	27070	27648	GTP cyclohydrolase II	*ToxB*	579	50	JANIEF010000054.1
ORF 7	25322	27013	WD40 repeat domain-containing protein	*ToxC*	1692	50	JANIEF010000054.1
ORF 8	24226	25202	Formylglycine-generating enzyme family protein/TRP2	*ToxD*	977	50	JANIEF010000054.1

4.4 Discussion

The results provide evidence of the first shotgun whole-genome sequencing (WGS) of *B. glumae* strain K6 and P3 from Malaysia in the NCBI genome database. As for *B. gladioli* strain, UPMBG7 and UPMBG8 were the first report of *B. gladioli* causing BPB published in Malaysia.

This study provides general genomic analysis using the whole-genome sequencing method for detecting structural variation, gene prediction, gene annotation, and analysis of the virulence gene of *Burkholderia* spp. causing BPB disease. Reference alignment/ structural variation detection refers to aligning sequencing reads from clean reads of whole genome sequencing to a reference genome (*B. glumae* BGR1 and *B. gladioli* BSR3). The reference genome is previously assembled and annotated, typically from a closely related organism, that serves as a guide for aligning the sequencing reads. A reference alignment's outcome can show a genome's quality and completeness. Genome mapping is applied to discover genetic variants and mutations and large-scale chromosome evolution, which can be valuable for understanding the genetic basis of disease and generating novel treatments (Sathya *et al.,* 2014). Based on the mapping rate, statistics show that *B. glumae* K6 and P3 strains showed higher mapping rates to their reference strain (90.65 - 91.28%), while *B. gladioli*'s mapping rate to their reference strain is 85.46 and 85.65%. A genome mapping rate of 100% would indicate that the entire genome was mapped toward the reference genome, which is the ultimate goal of genome sequencing projects. However, achieving a genome mapping rate of 100% is challenging and rare. A genome mapping rate of 90%

or higher is considered high-quality and sufficient for most research and practical applications. While a genome mapping rate of 80% or higher is considered adequate for many applications, lower percentages may result in missing or incomplete information. Several factors can contribute to a lower genome mapping rate. The most common causes include complex genome structure, where genomes with repetitive or highly similar sequences can be challenging to align, resulting in a lower mapping rate (Abbas *et al.,* 2014). Regions with high sequence diversity or regions with structural variation can cause less mapping rate. Limited sequencing coverage is also crucial, meaning short reads were obtained to align clean reads with the reference genome accurately. Sequencing errors and errors during sequencing can lead to gaps or inaccuracies in the genome reads, resulting in a lower mapping rate (Wright *et al.,* 2019).

Single nucleotide polymorphism (SNP) is a type of genetic variation that occurs when a single nucleotide (A, C, G, or T) in the DNA sequence differs from the reference genome sequence. SNPs are the most common type of genetic variation and can occur in any part of the genome, including coding regions, regulatory regions, and non-coding regions (Baird *et al.,* 2008). SNPs can have a wide range of effects on an organism's biology. Some SNPs can cause changes in the amino acid sequence (non-synonymous SNP), leading to altered protein function, while others can affect the regulation of gene expression (synonymous SNP). Some SNPs may have

effect, while others may increase the risk of certain diseases. According to results of SNP for *Burkholderia* species of our strain show that more than 50% of SNPs occur in the exonic region (coding region of a gene) (Hong *et al.,* 2015).

A non-synonymous SNP is a type of SNP that occurs in the coding region of a gene and results in a change in the amino acid sequence of the encoded protein and leads to the introduction or removal of stop and start codon. Non-synonymous SNPs are also known as missense mutations, as they can change the amino acid sequence of a protein and potentially alter its function. Non-synonymous SNPs can have a wide range of effects on an organism's biology, depending on the location and nature of the amino acid change (Robert & Pelletier, 2018). Some non-synonymous SNPs may not affect protein function, while others may lead to loss of function or gain of a new function. Some non-synonymous SNPs may also increase the risk of certain diseases by altering the structure or function of a protein (Dakal *et al.,* 2017). While non-synonymous SNPs can lead to amino acid changes and potentially alter the protein's function, not all amino acid changes will have a negative effect on the protein's function. Some can even increase the protein's efficiency or create a new function. Also, non-synonymous SNPs are not the only type of genetic variation that can alter the protein's function. Insertions, deletions, copy number variations, and structural variations can also have an impact.

Our study shows that high frameshift indel occur in exonic regions compared to non-frameshift indel. A frameshift indel is a type of insertion or deletion

(indel) that occurs in a gene's coding region and changes the DNA sequence's open reading frame. A frameshift indel can add or remove one or more nucleotides from the DNA sequence, leading to a shift in the codon boundaries and altering the amino acid sequence of the encoded protein. Frameshift indels can have severe consequences on protein function and may lead to loss of function or altered protein function (Pagel *et al.,* 2019). On the other hand, a non-frameshift indel is a type of indel that occurs in the coding region of a gene but does not change the reading frame of the DNA sequence. A non-frameshift indel can add or remove one or more nucleotides, but the codon boundaries remain intact. Non-frameshift indels can have minimal effect on protein function and may have no effect or lead to a silent mutation (Pagel *et al.,* 2019).

In addition, frameshift indels are considered high-impact mutations as they can change the entire protein sequence downstream from the point of insertion or deletion. Non-frameshift indels are considered low-impact mutations as they usually do not affect the protein sequence or only cause silent mutations. Other than that, frameshift indels can change the amino acid sequence of the encoded protein, which can alter its function or lead to loss of function. Next, frameshift indels can also provide insights into a gene's evolution, as they can alter the encoded protein and create new functions. Lastly, from frameshift indel, we can use the variation as genetic markers for association studies, as they can help identify genetic variations associated with a particular virulence factor.

Structural variation analysis (SVA) is the process of identifying and characterizing large-scale genomic variations such as insertions (INS), deletions (DEL), inversions (INV), intra-chromosomal translocations (ITX), and inter-chromosomal translocations (CTX), which can affect the structure and function of the genome (Fan *et al.,* 2014). The results show that 11, 28, 5, and 0 insertion sites with more than 50 kb occur in P3, K6, UPMBG7, and UPMBG8, respectively. This result shows that insertion with large size occurs more in the *B. glumae* strain than in the *B. gladioli* strains. However, deletion site with more than 50kb appears more in *B. gladioli* strain than in the *B. glumae* strains, 94, 103, 120, and 109 sites for P3, K6, UPMBG7, and UPMBG8, respectively. Insertion and deletion can have severe consequences on the protein function and gene expression as they can change the entire gene's function downstream from the point of insertion and deletion. However, some of the reason why insertion and deletion occur in SVA is probably due to replication errors, transposable elements, replication slippage, and homologous recombination (Bourque *et al.,* 2018; Carvalho and Lupski, 2016). Replication errors occur during DNA replication when the replication machinery fails to copy the DNA sequence accurately. These errors can result in extra nucleotides added or removed from the genome (Carvalho and Lupski, 2016). Transposable elements can insert or remove themselves in the genome, creating new copies or removing themselves as they move around (Bourque *et al.,* 2018). Next, homologous recombination occurs due to homologous recombination, a process by which similar or identical sequences in the genome are exchanged. Replication slippage occurs when an error

occurs during replication, leading to a loop in the DNA strand and resulting in adding or removing nucleotides to the sequences (Viguera *et al.,* 2001).

Translocation is a type of structural variation that occurs when piece of DNA sequence with more than 50kb is moved from one location in the genome to another. There are two main types of translocation which are ITX and CTX. CTX is when more than 50kb of DNA sequence is moved from one chromosome to another chromosome. This can result in the exchange of genetic material between chromosomes and can lead to changes in the structure and function of the affected genes. In contrast, ITX occurs when more than 50kb of DNA is moved to a new location within the same chromosome. It can lead to changes in the affected genes' structure and function and the formation of new genetic elements, such as inversions and duplications. Both types of translocation can have severe consequences on the structure and function of the genome, including the disruption of gene function, the formation of new genetic elements, and the creation of new genetic disorders (Weckselblatt *et al.,* 2015).

Copy number variation (CNV) analysis identifies and characterizes changes in the number of copies of a specific DNA sequence or gene within the genome with the reference genome (Pös *et al.,* 2021). CNV can occur in any part of the genome, including coding, regulatory, and non-coding regions. CNV can infer evolutionary relationships between different strains by comparing the CNV patterns in their genomes.

According to the results, most of the CNV deletion and duplication occurs in the exonic region of our strain. CNVs in the exonic regions, the genome regions that code for proteins, are more likely to have a functional impact on the organism. One reason CNVs are more likely to occur in exonic regions since the regions tend to have a higher density of genes. Another reason is exonic regions are more conserved across different species, which means that the sequences in these regions are more similar across other organisms (Rigau *et al.,* 2019). CNVs that occur in exonic regions are more likely to disrupt the function of the encoded protein and therefore have a more significant impact on the species in terms of adaptability and evolution (Pös *et al.,* 2021).

The scope of our study focused on general resequencing analysis, including SNP analysis, SV analysis, In/Del analysis, and CNV analysis, comparing our four strains with their respective reference genomes. However, it's crucial to acknowledge that our analysis did not include the specific identification of regions where variations occurred within the genomes of our four strains. Consequently, we were unable to conduct a detailed comparative analysis with other published genomic sequences of the same species to determine whether the observed variations were also present in those strains.

This limitation underscores the need for more comprehensive and region-specific comparative analyses in future research. Pinpointing the exact locations and nature of genetic variations within the genomes can provide deeper insights into the genetic diversity of the species and its potential

implications for pathogenicity and virulence. Such analyses may involve targeted sequencing, whole-genome alignment, and functional characterization of specific regions of interest. Therefore, while our study provides a foundational understanding of genomic variations, it should serve as a starting point for more detailed investigations into the genetic diversity and pathogenic potential of these *Burkholderia* strains.

In this research venture, a comprehensive bioinformatic pipeline was meticulously drafted and evaluated for the purpose of annotating and identifying virulence genes from raw sequencing data. The initial phase involved rigorous quality control and data cleansing using FastQC, ensuring that only high-quality reads were included in subsequent analyses. Subsequently, multiple genome assembly tools, including SPAdes, SOAPdenovo, and ABySS, were employed to construct contiguous sequences (contigs) from the initially fragmented data. The pipeline reached its pinnacle in the annotation phase, where the extensive resources of NCBI, KEGG, RAST and many more databases were harnessed to illuminate the genetic intricacies of virulence genes. The rigorous evaluation process serves as evidence of the pipeline's effectiveness and its crucial role in unraveling the genetic basis of virulence, establishing it as an indispensable tool in genomics and pathogenomics research.

In our dataset, we noticed that the GC content and genome size of our *Burkholderia* strain are similar. To better understand this, we compared our strain's genome size and GC content to reference genomes and previously

published *Burkholderia* genomes.Our analysis reveals a close match between our Burkholderia glumae strain and the reference sequences in terms of both genome size and GC content. According to the NCBI database, the genome size of *Burkholderia glumae* ranges from 5.8 to 7.89 Mbp, with a GC content of approximately 68%. These findings are consistent with our *B. glumae* strain, which has a genome size of around 6.5 Mbp and a GC content of 68%. This alignment underscores the accuracy of our genomic analysis. The NCBI database indicates that the genome size of *Burkholderia gladioli* ranges from 4.1 to 9 Mbp, with a GC content of approximately 67-68%. These findings are in alignment with our *B. gladioli* strain, which has a genome size of around 8 to 8.25 Mbp and a GC content of 68%. This indicates a high level of similarity with reference genomes and other *Burkholderia* strains, confirming its classification and relevance in the broader context of *Burkholderia* genetics.

In our comparison with the reference genome and other published Burkholderia genomes, we have found that the genome annotation of our strain bears strong similarities. It is noteworthy that even though our strain lacks a complete assembly due to the limitations of second-generation sequencing, resulting in contigs and some missing gene and repeated sequences, we were still able to successfully annotate key components. Notably, our Burkholderia strain's virulence genes have been accurately annotated, and their structures closely resemble those in the reference genome. This robust annotation confirms the presence of virulence genes in our strain, showcasing its potential significance in terms of virulence factors.

In our analysis, we compared the annotation and genomic features of our *Burkholderia* strain with both the reference genome and other published Burkholderia genomes of the same species through Genbank NCBI database. Despite the inherent limitations of second-generation sequencing, resulting in contigs and occasional gaps in the assembly, our annotation efforts were remarkably successful in capturing key genetic components.

One noteworthy achievement was the precise annotation of virulence genes within our *Burkholderia* strain. These genes play a pivotal role in the pathogenicity and virulence of the bacterium. Remarkably, our annotation results revealed that the structures of these virulence genes closely resemble those found in the reference genome. This robust annotation confirms not only the presence of these virulence genes in our strain but also highlights their integrity and potential functionality. This finding carries significant implications, suggesting that our strain possesses a repertoire of virulence factors that could contribute to its pathogenicity.

Furthermore, when we compared the annotation and analysis of our four strains and reference genome to other published genomic sequences of the same species, we found striking similarities in the genomic content and virulence gene profiles. This alignment of genetic features across different strains of the same species underscores the conserved nature of key virulence factors within the species.

Such conservation suggests that these virulence genes are integral to the species' pathogenicity, and their presence across multiple strains highlights their importance in the overall pathogenic strategy of the species.

Despite the challenges posed by the limitations of second-generation sequencing, our comprehensive annotation efforts and comparisons with other published genomes have shed light on the genetic composition and potential virulence of our Burkholderia strain. These findings not only contribute to a deeper understanding of the species but also emphasize the importance of these virulence genes in the pathogenicity and virulence of Burkholderia. Moreover, they provide valuable insights that can guide future research into the mechanisms underlying pathogenicity in this bacterial species, ultimately enhancing our ability to combat associated diseases and infections.

B. glumae K6, P3 strain, and *B. gladioli* UPMBG7, UPMBG8 were also subjected to assembly for gene prediction, gene annotation, and gene virulence analysis. *B. glumae* isolates' draft genome ranged from 6.52 to 6,57 Mbs with 68.22 to 68.33% G+C content, while *B. gladioli* isolates range from 8.05 to 8.22 Mbs with 67.99 to 68.01%. The assembly results showed that *B. gladioli* strain have a larger genome size than *B. glumae*. One of the reasons for this difference in genome size is that *B. gladioli* have a higher number of genes and additional genetic material than *B. glumae.* For instance, additional genetic material, such as transposable elements, can act as a source of new genes.

Other than that, *B. gladioli* and *B. glumae* may have different genetic variations and mutations that could contribute to their different sizes. These genetic variations could result in different growth rates, metabolic pathways, and overall physiology (Kang *et al.,* 2022). *B. gladioli* and *B. glumae* are known to occupy different ecological niches. *B. gladioli* is found in soil, rhizosphere, and plant tissues, whereas *B. glumae* is found in plant tissues. *B. gladioli* may have evolved to be larger to better adapt to its environment (Pedraza *et al.,* 2018). However, the genome size of a microorganism may not necessarily indicate its complexity, pathogenicity, or virulence. It can impact the organism's metabolic pathways, resistance mechanisms, and adaptation to different environments (Kang *et al.,* 2022). Based on genome components prediction for non-coding annotation of *B. glumae* and *B. gladioli* strain, the number of ribosomal RNA (5s rRNA, 16s rRNA, 23s rRNA) for all strains are the same with almost similar size. Ribosomal RNA plays a critical role in protein synthesis by binding to the mRNA and tRNA molecules and catalyzing the formation of peptide bonds between amino acids (Moore and Steitz, 2011). As for transfer RNA (tRNA), both *B. glumae* strains consist of 57 tRNA, while *B. gladioli* UPMBG7 strain consists of 60 tRNA, and the UPMBG8 strain consists of 58 strains. In general, organisms have a set of tRNA genes in their genome, which are transcribed into pre-tRNA molecules and then processed by a series of enzymes to generate mature tRNA molecules. The number of tRNA genes varies depending on the organism and the number of codons in its genetic code (Moore and Steitz, 2011).

In addition, tRNA molecules are crucial for the translation of the genetic code into functional proteins, and the presence of more than one type of tRNA for a given amino acid allows for flexibility in the decoding of the genetic code. It can help to minimize errors during translation (Reynolds *et al.,* 2017).

The result from coding gene annotation showed that *B. gladioli* strain has a higher number of coding gene than the *B. glumae* strain. However, the number of coding gene annotated between the same species is probably due to gaps between the contigs sequence. A genome with many exons may have a more complex gene structure, with more alternative splicing events, which could lead to a greater diversity of proteins (Kandul & Noor,2009). A high number of exons can also indicate a higher number of functional domains within the proteins, which can be important for the organism's adaptation to different environments or for the organism's diverse functions, such as in the case of resistance cells or tissues (Reynolds *et al.,* 2017). On the other hand, a low number of exons could indicate a more streamlined gene structure, with fewer alternative splicing events, which could be a sign of a more efficient use of genetic resources (Kandul & Noor, 2009).

The data provided through RAST analysis offers valuable insights into the genomic features and functional capabilities of various *Burkholderia* strains, including *B. glumae* K6, *B. glumae* P3, and *B. gladioli* UPMBG7 and UPMBG8. These findings shed light on the potential roles of these bacteria in virulence, disease, and defense mechanisms.

Firstly, in the *B. glumae* K6 strain, the presence of 50 coding sequences related to virulence, disease, and defense mechanisms underscores its potential as a pathogen. Notably, its tolerance to colicin E2, a bacteriocin peptide used to combat competition, suggests an adaptive advantage in competitive environments. Additionally, its resistance to various antibiotics and toxic compounds reveals its ability to thrive in challenging conditions, which is particularly important in understanding its ecological and agricultural impact.

Furthermore, the identification of 19 coding genes in the mycobacterium virulence operon implies the K6 strain's capacity for invasion and intracellular resistance. The presence of 24 coding sequences associated with flagella motility highlights its mobility, potentially aiding in host colonization and disease initiation. Additionally, the 24 coding sequences related to the protein secretion system, particularly the type II protein secretion system (T2SS) and type VII (T7SS), suggest an arsenal of tools for the bacterium to interact with its environment and hosts, including the synthesis of type 1 pili for colonization and invasion.

Comparatively, the *B. glumae* P3 strain exhibits a similar genomic profile regarding virulence, disease, and defense mechanisms, indicating its ability to resist colicin E2 and various toxic compounds.

The discrepancy in the mycobacterium virulence operon could be attributed to genome gaps or genetic variations. Interestingly, the P3 strain lacks *CheA* in its genome sequence, potentially influencing its motility and chemotaxis mechanisms.

Intriguingly, *B. gladioli* strains UPMBG7 and UPMBG8 exhibit a higher number of coding sequences related to virulence, disease, and defense mechanisms. Their unique function as mercuric reductases highlights their capability for mercury detoxification and its potential relevance in biogeochemical cycles. Both strains maintain flagellar motility systems, suggesting their importance in virulence toward host plants. The presence of the type V protein secretion system (T5SS) in *B. gladioli* strains, absent in *B. glumae* strains, implies different strategies for pathogenesis. However, the absence of *FimA* gene CDS in T7SS suggests variability in colonization mechanisms.

In conclusion, this data reveals intriguing differences and similarities in the genomic features of Burkholderia strains. The presence of virulence-related genes, antibiotic resistance, and distinct secretion systems suggest varied pathogenic potential and adaptive strategies among these strains. Further research is warranted to explore the functional significance of these genomic characteristics and their implications for plant-pathogen interactions, environmental dynamics, and potential applications in agriculture and biotechnology.

The Gene Ontology (GO) annotation analysis of the four bacterial strains, *B. glumae* K6, P3, *B. gladioli* UPMBG7, and UPMBG8, reveals intriguing insights into their genomic functional diversity. There is shared Biological Processes between All four strains exhibit including cell killing, death processes, developmental processes, multi-organism processes, multicellular organismal processes, negative regulation of biological processes, nitrogen utilization, and

immune system processes. These shared processes suggest fundamental biological functions conserved across the strains, potentially vital for their ecological niches. Biological Regulation, notably, differences emerge in biological regulation. *B. glumae* K6 and P3 strains display subtle disparities, with K6 featuring 727 annotated genes and P3 with 699. Likewise, UPMBG7 and UPMBG8 show variation in this category, with UPMBG7 having 1007 genes and UPMBG8 featuring 983. These distinctions hint at nuanced regulatory mechanisms that may influence their adaptability and responses to environmental cues. Next, Cellular Processes, all strains actively engage in cellular processes, reflecting their commitment to fundamental cellular activities. K6 features 2018 genes in this category, while P3 has 1947, UPMBG7 boasts 2640, and UPMBG8 includes 2579 genes related to cellular processes. This shared emphasis underscores the strains' dynamic nature and adaptability.

Other than that, Metabolic Versatility, metabolic processes are significantly enriched in all strains, highlighting their metabolic versatility and adaptability to diverse environmental conditions. K6 leads with 2140 genes, followed by UPMBG7 with 2778, UPMBG8 with 2724, and P3 with 2068 genes in this category. This underlines their capacity to utilize various nutrients for growth and survival. Specific Biological Functions, variations are observed in the number of genes related to biological adhesion, establishment of localization, and immune system processes among the strains. These differences suggest unique ecological niches and interaction strategies. Response to Stimulus and Signaling, showed that all strains possess genes related to response to

stimulus and signaling, indicating their ability to sense and respond to environmental changes. Viral Production, genes associated with viral production are found in all strains, suggesting potential interactions with viruses, possibly for defense or other biological interactions.

In summary, under GO annotation analysis provides a comprehensive view of the functional diversity within these bacterial genomes. While common biological processes indicate shared functions, differences in regulatory mechanisms and ecological adaptations are hinted at by variations in gene counts. The enrichment of metabolic processes underlines their metabolic adaptability, and the variations observed suggest that each strain has evolved unique genetic attributes to thrive in its specific environment. These findings enhance our understanding of these strains' biology and their potential roles in various ecological contexts.

The Clusters of Orthologous Groups (COG) annotation analysis of the four *Burkholderia* strains provides valuable insights into their functional profiles and potential adaptations. In the category of RNA processing and modification, it's notable that all four strains exhibit limited representation, with only one annotated gene each. This suggests a fundamental and conserved machinery for RNA-related processes across these strains. This functional constraint indicates that these strains may not heavily rely on post-transcriptional RNA modifications or have evolved streamlined mechanisms for RNA regulation. Chromatin structure and dynamics, which play a pivotal role in gene regulation, show a consistent pattern with three annotated genes in all strains. This

commonality suggests that these Burkholderia strains share similar epigenetic regulatory mechanisms, ensuring proper gene expression control and chromosomal organization.

Energy production and conversion annotations reveal subtle differences among the strains, with K6 and UPMBG7 having slightly more annotated genes than P3 and UPMBG8. These differences could be indicative of variations in energy metabolism strategies, potentially influenced by their respective environmental niches. Further investigations are warranted to explore these distinctions and their ecological implications. In terms of cell cycle control, cell division, and chromosome partitioning, all strains exhibit the same number of annotated genes (46). This suggests a conserved machinery governing cell division and genome stability, crucial for bacterial survival and reproduction.

The category of amino acid transport and metabolism, vital for nutrient acquisition, indicates that all strains possess a comprehensive suite of genes involved in amino acid utilization. The minor variations likely reflect their adaptability to diverse nutrient sources, which is a common feature among versatile environmental bacteria like Burkholderia. Nucleotide transport and metabolism annotations are quite similar, emphasizing the importance of nucleotide biosynthesis and utilization for DNA and RNA synthesis in these strains.

Carbohydrate transport and metabolism annotations exhibit a shared ability to access a wide range of carbohydrate sources, indicating their metabolic

versatility. Annotations related to coenzyme transport and metabolism are remarkably consistent, suggesting that these strains possess comparable mechanisms for coenzyme utilization and cofactor-mediated enzymatic reactions. Lipid transport and metabolism annotations reveal functional similarities among the strains, emphasizing the importance of lipid-related processes in their biology.

Translation, ribosomal structure, and biogenesis-related annotations are abundant in all strains, indicating a shared capability for protein synthesis and cell growth, essential for their survival and proliferation. Annotations related to transcription, a key aspect of gene regulation, are well-represented in all strains, reflecting complex regulatory networks in each strain's genome. Annotations associated with replication, recombination, and repair underscore the importance of genome maintenance and integrity in these *Burkholderia* strains. Cell wall/membrane/envelope biogenesis annotations suggest variations in cell envelope structure among the strains, which could relate to their interactions with different environmental or host factors. Annotations in the cell motility category indicate a shared capability for bacterial movement among the strains. Posttranslational modification, protein turnover, and chaperone annotations emphasize the significance of protein stability and stress responses across these strains. Inorganic ion transport and metabolism annotations suggest their ability to acquire essential ions from the environment, with minor variations likely related to specific ion requirements. Annotations related to secondary metabolite biosynthesis, transport, and

catabolism highlight the potential for ecological interactions or pathogenicity among these strains.

General function prediction annotations indicate multifunctionality in various cellular processes, reflecting the versatility of these *Burkholderia* strains. Annotations of unknown function signify uncharted functional territories within the genomes, emphasizing the need for future research to elucidate these roles.

Annotations related to signal transduction mechanisms demonstrate shared traits, suggesting that these strains employ similar strategies for environmental sensing and response. Intracellular trafficking, secretion, and vesicular transport annotations reveal essential processes for cellular communication, which are conserved among the strains.

Defense mechanisms are well-represented in all strains, indicating their ability to defend against various challenges, potentially including host immune responses or competition with other microorganisms. Annotations related to extracellular structures suggest shared features that may contribute to adhesion, colonization, or biofilm formation. Annotations associated with mobilome elements, such as prophages and transposons, point to genome plasticity and potential adaptations in response to changing environments.

In conclusion, the COG annotation analysis underscores both the common functional attributes and nuanced distinctions among these *Burkholderia* strains. These variations in specific functional categories likely correspond to

unique ecological niches, pathogenic adaptations, and responses to environmental cues. Further investigations and comparative genomics can shed more light on the functional diversity and evolutionary strategies of these bacteria, contributing to a deeper understanding of Burkholderia biology and ecology.

However, annotating coding sequences using GO, COG and RAST database does not categorize coding sequences related to quorum sensing (toxoflavin biosynthesis and transport) and flagellar assembly, which are crucial in the pathogenicity factor of *Burkholderia* spp. According to the KEGG pathway, the color box (yellow, green, and blue) indicated a specific gene that has been experimentally verified to be involved in the pathway. While for white box gene indicates a gene that has not been experimentally verified or inferred to be involved in the pathway in question. This means that the gene's role in the pathway is unknown or uncertain and not included in the pathway diagram.

Under the quorum sensing pathway, *toxI* genes were indicated with a white box for both *B. gladioli* strains, indicating that this gene was not present in toxoflavin transport for *B. gladioli*. It is unclear why the *ToxI* gene is absent in *B. gladioli*. It could be that the *ToxI* protein is not necessary for the pathogenesis of *B. gladioli*, or that the gene has been lost through evolution. Another possibility is that the *ToxI* gene is present in some strains of *B. gladioli* but has not yet been identified or characterized. Further research is needed to understand why the *ToxI* gene is absent in Burkholderia gladioli. On the other

hand, genes that are involved in toxoflavin biosynthesis, transport, and flagella assembly were found in both *B. glumae* and *B. gladioli*.

Quorum sensing (QS) is a process of cell-to-cell communication used by bacteria to coordinate gene expression in response to changes in cell population density. For *Burkholderia* spp., the QS is mediated by N-acyl homoserine lactones (AHLs) and N-octanoyl homoserine lactone (C8-HSL) (Chun *et al.,* 2009). *TofI* and *TofR* genes encode the AHL synthesis for *N-hexanoyl* homoserine lactone (C6-HSL) and C8-HSL, as well as the cognate receptor for C8-HSL, respectively. TofI/TofR promotes the expression of the *toxJ* gene, which in turn activates the transcription of the *toxR* (LysR-type transcriptional activator), and both of these genes

activate the expression of the *toxABCDE* (toxoflavin biosynthesis) and *toxFGHI* genes (toxoflavin transport) (Ham *et al.,* 2010; Kim *et al.,* 2011).

The QsmR is an IcIR-type transcriptional regulator that binds to the promoter regions of target genes and regulates their expression in response to changes in cell population density and homoserine lactone level. It attaches to the promoter regions of the *flhC* and *flhD* genes and activates their expression, producing the *FlhC* and *FlhD* proteins (Chun *et al.,* 2009). *FlhC* and *flhD* genes are important in the virulence factor of *Burkholderia* spp. since they regulate the expression of virulence factors (Moule *et al.,* 2015). The *flhC* and *flhD* genes code for transcriptional activators that control the expression of many genes in the flagellar regulon, including those involved in motility, quorum sensing, and type III secretion system. *Burkholderia* spp. use motility and

secretion systems to infect and colonize host tissues. By regulating the expression of genes involved in these processes, the *flhC* and *flhD* genes help to control the virulence of *Burkholderia* spp. *FlhC* and *FlhD* proteins also regulate the genes encoding the type III secretion system, which inject virulence factors directly into host cells. By regulating the expression of the type III secretion system, the *flhC* and *flhD* genes control the ability of *Burkholderia* spp. to cause disease (Vander Broek and Stevens, 2017). Thus, the *flhC* and *flhD* genes play a crucial role in the virulence of *Burkholderia* spp. by regulating the expression of genes involved in motility, quorum sensing, and type III secretion system essential for the pathogenesis of this pathogen.

Bacteria inject virulence factors directly into host cells using the type III secretion system (T3SS). The *HrpB* protein plays a role in the pathogenesis of these bacteria by promoting bacterial attachment and invasion of host cells, as well as by modulating host cell signaling pathways (Green & Mecsas, 2016). The T3SS comprises several components, including a needle complex, responsible for injecting virulence factors into host cells, and a translocon, accountable for transporting the virulence factors across the host cell membrane (Lipscomb & Schell, 2011). The *HrpB* protein interacts with other needle complex components, such as *HrpF*, to form a stable needle-like structure that can inject virulence factors into host cells (Mannaa *et al.,* 2018). Once the virulence factors are injected into host cells, they can modulate host cell signaling pathways and cause cell death (Lipscomb & Schell, 2011). This results in bacterial attachment, invasion, and multiplication within the host cell, leading to infections and diseases caused by bacteria.

The *lipA* and *lipB* genes found in some strains of the bacterial genus *Burkholderia are* involved in the pathogenesis of these bacteria by promoting bacterial attachment and invasion of host cells (Tribble and Lamont, 2010).

The *lipA* and *lipB* genes encode proteins involved in the biosynthesis of lipopolysaccharides (LPS), which are significant components of the outer membrane of Gram-negative bacteria (Ham *et al.,* 2010). LPS comprises a lipid A component, a core oligosaccharide, and an O-antigen repeat. *LipA* is responsible for the biosynthesis of lipid A, which is the toxic portion of LPS, and it is responsible for the majority of the toxicity of LPS (Bertani & Ruiz, 2018). In comparison, *LipB* is responsible for the biosynthesis of the core oligosaccharide, an essential component of LPS. The core oligosaccharide is involved in the outer membrane's stability and integrity and the recognition and binding of LPS by host cells (Bertani & Ruiz, 2018). In addition, these LPS molecules on *Burkholderia* spp. can promote bacterial attachment and invasion of host cells by binding to host cell receptors (Lee *et al.,* 2019). Once the bacteria can attach to host cells, they can invade the host cells and cause infection. The LPS also plays a role in the evasion of the host's immune system by masking the bacteria from the host's immune cells and modulating host cell signaling pathways, leading to the development of infections and diseases caused by *Burkholderia* spp. (Vellasamy *et al.,* 2016).

In our analysis, where we examined the predicted sequences of virulence genes against those archived in the NCBI database, a striking result emerged: the predicted virulence genes exhibited a remarkable similarity of over 98%.

This high degree of sequence similarity is an intriguing finding and warrants careful consideration.

One plausible explanation for the observed high similarity is the functional indispensability of virulence genes. These genes typically encode factors crucial for an organism's pathogenicity, such as adhesion proteins, toxins, and immune evasion mechanisms. Given their pivotal role in the virulence process, these genes may be subjected to strong purifying selection pressure, constraining sequence variations. Additionally, the conservation of virulence genes across strains and isolates might be a reflection of their critical contribution to the organism's survival and infectious potential. Such high conservation ensures the effective execution of pathogenic strategies, which, if altered significantly, could compromise the organism's ability to infect and cause disease. Furthermore, the high similarity underscores the reliability of our gene prediction methods, affirming their accuracy. This result not only solidifies the credibility of our data but also suggests that the virulence genes identified are indeed integral components of our Burkholderia strain's pathogenic arsenal.

In conclusion, the exceptional similarity observed in the predicted virulence gene sequences, exceeding 98%, emphasizes the functional importance and evolutionary conservation of these genes. This finding underscores their vital role in pathogenicity, highlights the accuracy of our predictive methods, and reaffirms the significance of our *Burkholderia* strain in the context of virulence gene research.

The inability to directly correlate the pathogenicity/virulence of the strains with the results of the pathogenicity test can be attributed to practical challenges and timing disparities inherent in the research process. One of the primary reasons for this disconnect lies in the substantial time lag between conducting pathogenicity testing and completing whole-genome sequencing. While pathogenicity testing can yield rapid results, the comprehensive process of whole-genome sequencing, from sample submission to data analysis, typically spans several months.

Additionally, the random selection of strains for sequencing, undertaken before the pathogenicity test results were available, was essential to maintain the study's unbiased nature. However, this approach meant that the specific virulence levels of the sequenced strains remained unknown at the time of selection.

Despite these logistical constraints and the inability to synchronize the timing of pathogenicity testing and sequencing, the study still holds substantial scientific value. It provides crucial insights into the genomic characteristics of these *Burkholderia* strains, offering a foundation for understanding their potential virulence factors and genetic diversity. While the direct correlation between genome data and pathogenicity results may not be immediately achievable, the research contributes to a comprehensive comprehension of these strains' capabilities and their potential as pathogens. Moreover, the data generated through genome sequencing can serve as a valuable resource for guiding future investigations into the pathogenicity and virulence mechanisms

of these strains, complementing the insights obtained from the pathogenicity tests.

4.5 Conclusion

In conclusion, the structural variation and gene annotation in bacterial plant pathogens is vital for understanding these organisms' evolution, genetic makeup and mechanisms of pathogenesis. Mapped reads showed that the *B. glumae* strain has a mapping rate of 90-91%, while *B. gladioli* have a mapping rate of 85%. The mapping rate could not achieve 100% due to the structural variation between samples sequenced with the reference genome. Results from structural variation through SNP, InDel, SV, and CNV showed that *B. glumae* strain K6, P3, and *B. gladioli* strain UPMBG7 and UPMBG8 have most of the variation occurred in their exonic region (coding region). Variations in these regions can lead to changes in the amino acid sequence of a protein, which can impact its function or stability. Thus, further study should be conducted to understand their evolution and impact in causing disease in paddy from Peninsular Malaysia.

On the other hand, gene annotation in bacterial plant pathogens can help identify genes involved in pathogenesis, such as those encoding for virulence factors or the production of toxins. Non-redundant protein identification, gene pathway, gene function, gene class, and gene-related subsystem categories have been annotated using NR, KEGG, GO, COG, and RAST. Genome deposition on NCBI can be retrieved using accession numbers of JAMYCS000000000 (K6 strains), JAMYCT000000000 (P3 strains), JANIEE000000000 (UPMBG7 strains), and JANIEF000000000 (UPMBG8 strains).

Lastly, a gene related to quorum sensing, flagellar assembly, lipase production, and *HrpB* T3SS has been annotated and detected, which is associated with the virulence of *Burkholderia* spp. in their hosts.

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATION FOR FUTURE RESEARCH

In the first objective, we proposed a hypothesis that phenotypic and molecular characterization may distinguish *Burkholderia* spp. causing BPB disease in rice. Panicles with dark grey florets and a reddish-brown line (border of the lesion) across the floret between the darker and straw-colored sections are symptoms of panicle blight. The leaf sheath has a grey, necrotic core lesion with a reddish-brown border that can be several centimeters long, and the leaves have small circular tan lesions with brown borders ranging from 1 to 5 mm (Nandakumar *et al.,* 2009). A range of systemic identification and characterization approaches were performed to isolate and discover *Burkholderia* spp. associated with BPB disease. Procedures including field observation and sample collection of panicle blight disease of rice, morphological selection of *Burkholderia* spp. isolates on King's B agar medium, biochemical test, molecular characterization, and phylogenetic analysis were done during the phenotypic and molecular characterization of *Burkholderia* spp. isolates causing BPB.

In our study, three states of rice fields in Peninsular Malaysia (Kedah, Perak, and Selangor) were investigated, each having significant damage to the panicles and the possibility of BPB disease outbreaks. Isolation of pure *Burkholderia* spp. on King's B medium revealed that isolates' morphology is

cream-white, with circular features, flat colony elevation, and smooth margin on King's B agar (Ramachandran *et al.,* 2021).

Burkholderia spp. produced diffusible yellow pigment on King's B agar. However, *B. gladioli* produced intense yellow pigment on King's B agar compared to *B. glumae*. Biochemical tests on 27 isolates showed that they were Gram-negative, negative for oxidase, sulfur production, and indole production tests, and positive for catalase, KOH, motility, and carbohydrate fermentation tests. These are the major features of *Burkholderia* species (Mulaw *et al.,* 2018; Zhou-qi *et al.,* 2016). Results from 16S rRNA sequencing and phylogenetic analysis showed that 22 strains were clustered to *B. glumae* and 5 clusters related to *B. gladioli*. Identification using *gyrB* primer showed higher genetic variation compared to 16s rRNA primer. While *toxB* gene primer also has low genetic diversity, probably due to restricted nucleotide sequences.

In the second objective, the pathogenicity test showed that all isolates were classified into three functional groups: highly virulent, moderately virulent, and weakly virulent. Based on the statistical results from pathogenicity assays, isolate UPMBG7, UPMBG8, and BGS.AS5, BGP.A4, BGS.AS4, BGP.A6, BGK.AZ7 and BGS.AS1 exhibited the highest degree of severity of panicle blight disease to MR219, followed by isolates BGS.AS3, BGK.AZ4, BGP.A2, which were moderately virulent, and BGK.AZ1 as the weak virulent toward MR219. Pathogenicity analysis showed that *B. glumae* BGS.AS1 and *B. gladioli* UPMBG7 were the most virulent strain for both cultivars of MR219. In

conclusion, our results revealed that *Burkholderia* spp. strains associated with the BPB disease of rice from Peninsular Malaysia are pathologically diverse and differ in virulence. Our first two objectives show that *B. glumae* and *B. gladioli* are the causative agents of BPB disease in Peninsular Malaysia.

These findings are critical for a better understanding and identification of *Burkholderia* spp. strains, since rice has been regarded as Malaysia's most important commodity crop with a high economic value (Bray, 2014).

In the third objective, the whole-genome sequencing assembly of *B. glumae* K6, P3 and *B. gladioli* UPMBG7 and UPMBG8 revealed that the size of *B. glumae* was ~6.5 Mb with 68% G+C contents, while *B. gladioli* strain size was ~8-8.2 Mb with 68% G+C contents. The results showed that the *B. gladioli* genome size is larger than *B. glumae*. The mapping rate for *B. glumae* strains was 90-91% toward *B. glumae* BGR1, while *B. gladioli* strains have a mapping rate of 85% toward *B. gladioli* BSR3. These results indicated that the *B. gladioli* strain has more structural variation to its reference genome. The SNP, InDel, SV, and CNV analysis also showed that most variation occurs in both species' exonic regions. These variations can contribute to the evolution and adaptation of these pathogens to different host plants and the emergence of new pathogenic strains.

Structural variations in bacteria can have a variety of effects on the organism. Single nucleotide polymorphism, deletions, or insertions can disrupt the function of genes or change the regulation of gene expression. It can lead to

changes in the phenotype of the bacteria, such as changes in antibiotic resistance or virulence.

Other types of structural variations, such as duplications or inversions, can lead to the formation of multiple copies of a gene or a change in the orientation of a gene (Kang *et al.,* 2022), resulting to increase expression that can benefit the bacteria, e.g., increased antibiotic resistance or ability to adapt to new environments. Structural variations can also lead to the formation of novel genetic elements, such as plasmids or transposable elements, which can facilitate the spread of antibiotic resistance or other beneficial traits among different bacterial populations.

In summary, structural variations in bacteria can have a wide range of effects, depending on the type of variation and its location within the genome. Some structural variations may be beneficial, while others may be detrimental. Understanding the impact of structural variations in bacteria is essential for understanding the evolution and ecology of bacterial populations and developing new strategies for controlling bacterial infections. The identity of P3, K6, UPMBG7, and UPMBG8 as *Burkholderia* spp. was reflected with the NR protein database annotation. P3 and K6 have 91-92% of their protein matched to *B. glumae,* while UPMBG7 and UPMBG8 have 84-87% of their protein annotated to *B. gladioli.* It was apparent that all strains could perform flagellar assembly, toxoflavin biosynthesis, and transport due to a complete functional pathway based on KEGG analysis. Based on RAST analysis, we

could identify the coding sequence of our strain related to virulence, disease and defense mechanism, flagella motility, T2SS, T5SS, and T7SS.

Gene annotation in bacterial plant pathogens can help identify genes involved in pathogenesis, such as those encoding for virulence factors or for the production of toxins. This information can be used to develop new strategies for controlling bacterial plant pathogens, such as identifying new drug targets or developing new vaccines. In conclusion, structural variation and gene annotation in bacterial plant pathogens are critical for understanding these organisms' evolution, adaptations, genetic makeup and mechanisms of pathogenesis and for developing new strategies for controlling bacterial plant diseases. Further research in this field will continue to provide insights into the genetic basis of bacterial plant pathogens and the mechanisms of their evolution and host interactions.

For future prospects, we recommend further research as the following:

1. Pathogenicity test on several other cultivars to determine and evaluate the sensitivity of different cultivars toward *Burkholderia* spp. under field conditions, especially in deciding whether *B. gladioli* cause more severity to compare to *B. glumae*.

2. Study on *Burkholderia* spp. infection by analysis of gene expression in the host, which will help evaluate the virulence-associated genes expressed by this pathogen in rice hosts.

3. To conduct an in-depth comparative genomic analysis of diverse *Burkholderia* strains within the same species, with a focus on detecting and characterizing genetic variations, in order to provide deeper insights into the genetic diversity and variation of the species and its potential implications for pathogenicity and virulence.

REFERENCES

Ab Razak, S., Nor Azman, N. H. E., Kamaruzaman, R., Saidon, S. A., Mohd Yusof, M. F., Ismail, S. N., Jaafar, M. A., & Abdullah, N. (2020). Genetic diversity of released Malaysian rice varieties based on single nucleotide polymorphism markers. *Czech Journal of Genetics and Plant Breeding*, *56*(2), 62–70.

Abbas, M. M., Malluhi, Q. M., & Balakrishnan, P. (2014). Assessment of de novo assemblers for draft genomes: a case study with fungal genomes. *BMC genomics*, *15* (Suppl 9).

Abdurrahman, A., & Wahid, N. A. (2019). The relationship between government spending and economic growth in Malaysia. *International Journal of Economics, Commerce and Management*, *7*(8), 1-11.

Alaloul, W. S., Musarat, M. A., Rabbani, M. B. A., Iqbal, Q., Maqsoom, A., & Farooq, W. (2021). Construction Sector Contribution to Economic Stability: Malaysian GDP Distribution. *Sustainability*, *13*(9), 5012.

Alam, M. M., Zainal Abidin, M. A., Ismail, M. R., & Awang, Y. (2016). Pathogenicity and genetic diversity of *Burkholderia* spp. associated with rice sheath rot and bacterial panicle blight of paddy in Malaysia. *Plant Pathology Journal*, *32*(2), 124-136.

Ali, S. A., Rukunnudin, I. H., Kasim, K. F., Said, W. & Hosni, H. (2018). Evaluation of the Effects of Milling Time on the Physical Quality of Sarawak's Mamut Rice Traditional Variety. *ASM Science Journal*, *13*(4).

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian journal of microbiology*, *56*(4), 394–404.

An, J. H., Goo, E., Kim, H., Seo, Y. S., & Hwang, I. (2014). Bacterial quorum sensing and metabolic slowing in a cooperative population. *Proceedings of the National Academy of Sciences*, *111*(41), 14912–14917.

Ari, Ş., & Arikan, M. (2016). Next-Generation Sequencing: Advantages, Disadvantages, and Future. *Plant Omics: Trends and Applications*, 109–135.

Aryal, S. (2019, June 14). Citrate Utilization Test- Principle, Media, Procedure and Result. *Microbiology Info.Com*.

Ashburner, M., Ball, C., Blake, J. et al. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Asselin, J. A. E., Bonasera, J. M., & Beer, S. V. (2016). PCR Primers for Detection of *Pantoea ananatis, Burkholderia* spp., and *Enterobacter* sp. from Onion. *Plant Disease*, *100*(4), 836–846.

Baharudin, M. A. H., Ismail, A. M., & Abdullah, R. (2019). The impact of foreign direct investment on Malaysia's economic growth: Evidence from the manufacturing sector. *Journal of Asian Finance, Economics and Business*, *6*(2), 43-52.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*(10), e3376.

Balaban, M., Joslin, S. N., & Hendrixson, D. R. (2009). *FlhF* and its GTPase activity are required for distinct processes in flagellar gene regulation and biosynthesis in *Campylobacter jejuni*. *Journal of bacteriology*, *191*(21).

Baltrus, D. A., & Clark, M. (2020). Complete Genome Sequence of *Pseudomonas coronafaciens* pv. oryzae 1_6. *Microbiology resource announcements*, *9*(3), e01564-19.

Bangratz, M., Wonni, I., Kini, K., Sondo, M., Brugidou, C., Béna, G., Gnacko, F., Barro, M., Koebnik, R., Silué, D., & Tollenaere, C. (2020). Design of a new multiplex PCR assay for rice pathogenic bacteria detection and its application to infer disease incidence and detect co-infection in rice fields in Burkina Faso. *PLOS ONE*, *15*(4), e0232115.

Bertani, B., & Ruiz, N. (2018). Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus*, *8*(1).

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic acids research*, *29*(12), 2607–2618.

Bishwajit, G., Sarker, S., Kpoghomou, M. A., Gao, H., Jun, L., Yin, D., & Ghosh, S. (2013). Self-sufficiency in rice and food security: A South Asian perspective. *Agriculture & Food Security*, *2*(1), 10.

Bourque, G., Burns, K.H., Gehring, M. et al. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*, 199.

Bowden, R., Davies, R.W., Heger, A., Pagnamenta, A. T., de Cesare, M., Oikkonen, L. E., Parkes, D., Freeman, C., Dhalla, F., Patel, S. Y., Popitsch, N., Ip, C. L. C., Roberts, H. E., Salatino, S., Lockstone, H, Lunter, G., Taylor, J. C., Buck, D., Simpson, M. A., & Donnelly, P. (2019). Sequencing of human genomes with nanopore technology. *Nature Communications*, *10*(1), 1869.

Bray, F. (2014). Rice in Malaya: A Study in Historical Geography. *The AAG Review of Books*, *2*(4), 127–129.

Carvalho, C. M., & Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nature reviews*. Genetics, *17*(4), 224–238.

Chun, H., Choi, O., Goo, E., Kim, N., Kim, H., Kang, Y., Kim, J., Moon, J. S., & Hwang, I. (2009). The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *Journal of bacteriology*, *191*(13), 4152–4157.

Coenye, T., & Vandamme, P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environmental Microbiology*, 5(9), 719-729.

Cui, Z., Ojaghian, M., Tao, Z., Kakar, K., Zeng, J., Zhao, W., Duan, Y., Vera Cruz, C., Li, B., Zhu, B., & Xie, G. (2016). Multiplex PCR assay for simultaneous detection of six major bacterial pathogens of rice. *Journal of Applied Microbiology*, *120*(5), 1357–1367.

Cui, Z., Wang, S., Kakar, K. U., Xie, G., Li, B., Chen, G., & Zhu, B. (2021). Genome Sequence and Adaptation Analysis of the Human and Rice Pathogenic Strain Burkholderia glumae AU6208. *Pathogens* (Basel, Switzerland), 10(2), 87.

Dakal, T. C., Kala, D., Dhiman, G., Yadav, V., Krokhotin, A., Dokholyan, N. V. (2017). Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. *Scientific Reports*, 7(1), 6525.

Del Vecchio, F., Mastroiaco, V., Di Marco, A. et al. (2017). Next-generation sequencing: recent applications to the analysis of colorectal cancer. *Journal of Translational Medicine*, 15, 246.

Department of Agriculture. (2021). Annual Gross Domestic Product 2015-2020. https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=491&bul_id=YnhhZ2g5QlpZWG9RcVNwTGhLaHE4UT09&menu_id=TE5CRUZCblh4ZTZMODZIbmk2aWRRQT09

Department of Statistics Malaysia. (2020). Agriculture census 2019: Preliminary report. https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=a2FRVFBWbEpoeXZVNFdvN1lRZGgzdz09

Devescovi, G., Bigirimana, J., Degrassi, G., Cabrio, L., LiPuma, J. J., Kim, J., Hwang, I., & Venturi, V. (2007). Involvement of a Quorum-Sensing-Regulated Lipase Secreted by a Clinical Isolate of *Burkholderia glumae* in Severe Disease Symptoms in Rice. *Applied and Environmental Microbiology*, *73*(15), 4950–4958.

El-Tarabily, K. A., Nassar, A. H., Hardy, G. E. S. J., & Sivasithamparam, K. (2009). Plant growth promotion and biological control of *Pythium aphanidermatum*, a pathogen of cucumber, by endophytic actinomycetes. *Journal of Applied Microbiology*, *106*(1), 13-26.

Fahmi, Z., Samah, B. A., & Abdullah, H. (2013). Paddy industry and paddy farmers well-being: A success recipe for agriculture industry in Malaysia. *Asian Social Science*, *9*(3), 177.

Fan, X., Abbott, T. E., Larson, D., & Chen, K. (2014). BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping. *Current Protocols in Bioinformatics*, *45*(1).

FAO. (2021). GIEWS Country Brief on Malaysia. https://www.fao.org/giews/countrybrief/country.jsp?code=MYS

FAOSTAT. (2021). Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL

Firdaus, R. B. R., Leong Tan, M., Rahmat, S. R., & Senevi Gunaratne, M. (2020). Paddy, rice and food security in Malaysia: A review of climate change impacts. *Cogent Social Sciences*, *6*(1), 1818373.

Fory, P. A., Triplett, L., Ballen, C., Abello, J. F., Duitama, J., Aricapa, M. G., Prado, G. A., Correa, F., Hamilton, J., Leach, J. E., Tohme, J., & Mosquera, G. M. (2014). Comparative Analysis of Two Emerging Rice Seed Bacterial Pathogens. *Phytopathology*, *104*(5), 436–444.

Francis, F., Kim, J., Ramaraj, T., Farmer, A., Rush, M. C., & Ham, J. H. (2013). Comparative genomic analysis of two *Burkholderia glumae* strains from different geographic origins reveals a high degree of plasticity in genome structure associated with genomic islands. *Molecular Genetics and Genomics*, *288*(3–4), 195–203.

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic acids research*, *43*, D261–D269.

Gao, R., Krysciak, D., Petersen, K., Utpatel, C., Knapp, A., Schmeisser, C., Daniel, R., Voget, S., Jaeger, K. E., & Streit, W. R. (2015). Genome-Wide RNA Sequencing Analysis of Quorum Sensing-Controlled Regulons in the Plant-Associated *Burkholderia glumae* PG1 Strain. *Applied and Environmental Microbiology*, *81*(23), 7993–8007.

Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., Wilkinson, A. C., Finn, R. D., Griffiths-Jones, S., Eddy, S. R., & Bateman, A. (2009). Rfam: updates to the *RNA* families database. *Nucleic acids research*, 37(Database issue), D136–D140.

Goo, E., An, J. H., Kang, Y., & Hwang, I. (2015). Control of bacterial metabolism by quorum sensing. *Trends in Microbiology*, *23*(9), 567–576.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351.

Green, E. R., & Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology spectrum*, 4(1).

Gupta, A. K., & Gupta, U. D. (2014). Next Generation Sequencing and Its Applications. *Animal Biotechnology*, 345–367.

Ham, J. H., Melanson, R. A., & Rush, M. C. (2010). *Burkholderia glumae*: next major pathogen of rice? Molecular Plant Pathology, *12*(4), 329–339.

Hamid, F. S., Kamarudin, S. K., & Sulaiman, M. R. (2021). The Impact of Rice Productivity on Economic Growth in Malaysia: An Empirical Study. *International Journal of Academic Research in Business and Social Sciences*, *11*(2), 294-307.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1–8.

Hindmarsh, P. C., & Geertsma, K. (2017). Biochemical Tests Used for Diagnosis. *Congenital Adrenal Hyperplasia*, 95–110.

Hong, K. W., Koh, Y. S., Kim, S. J., Lee, H. J., Shin, J. H., Lee, J. Y., ... & Park, K. U. (2015). SNP-based molecular characterization and genotyping of Burkholderia pseudomallei from sputum samples of patients with

suspected melioidosis in Korea. *PLoS Neglected Tropical Diseases*, *9*(4), e0003694.

Hou, Y., Xu, Y., Zhang, Y., Yu, L., Liu, L., Wang, L., & Huang, S. (2020). First Report of Bacterial Panicle Blight of Rice Caused by *Burkholderia glumae* in Southern China. *Plant Disease*, *104*(4), 1252.

Ishihara, H., & Kiba, A. (2015). Effects of the environment on the virulence of *Burkholderia glumae*. *Journal of General Plant Pathology*, *81*(4), 315-320.

Ismail, A. M., Baharudin, M. A. H., Abdullah, R., & Muhammad, F. (2019). The Contribution of the Rice Sector to the Malaysian Economy: A Computable General Equilibrium Analysis. *Journal of Agribusiness and Rural Development*, *53*(4), 421-434.

Jacob-Dubuisson, F., Locht, C., & Antoine, R. (2001). Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Molecular microbiology*, *40*(2), 306–313.

Jamil, M. & Anwar, F. (2016). Properties, Health Benefits and Medicinal Uses of *Oryza sativa*. *European Journal of Biological Sciences*, 8(4).

Jonghan, K., Kang, S., & Tenhunen, J. (2013). Impacts of climate change on paddy rice yield in a temperate climate. *Global Change Biology*, *19*(2).

Jung, B., Lee, S., Ha, J., Park, J. C., Han, S. S., Hwang, I., Lee, Y. W., & Lee, J. (2013). Development of a Selective Medium for the Fungal Pathogen *Fusarium graminearum* Using Toxoflavin Produced by the Bacterial Pathogen *Burkholderia glumae. The Plant Pathology Journal*, *29*(4), 446–450.

Jung, Y., Han, D. (2022). BWA-MEME: BWA-MEM emulated with a machine learning approach, *Bioinformatics*. *38*(9), 2404–2413.

Kandul, N.P., & Noor, M.A. (2009). Large introns in relation to alternative splicing and gene evolution: a case study of Drosophila bruno-3. *BMC Genetics*, *10*, 67.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic acids research*, *32*, D277–D280.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., & Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research*, *34*, D354–D357.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, *45*, D353-D361.

Kang, Y., Kim, J., Kim, S., Kim, H., Lim, J. Y., Kim, M., Kwak, J., Moon, J. S., & Hwang, I. (2008). Proteomic analysis of the proteins regulated by *HrpB* from the plant pathogenic bacterium *Burkholderia glumae*. *PROTEOMICS*, *8*(1), 106–121.

Kang, M., Lim, J. Y., Kim, J., Hwang, I., & Goo, E. (2022). Influence of genomic structural variations and nutritional conditions on the emergence of

quorum sensing-dependent gene regulation defects in *Burkholderia glumae*. *Frontiers in microbiology*, *13*, 950600.

Kanzi, A. M., San, J. E., Chimukangara, B., Wilkinson, E., Fish, M., Ramsuran, V., & de Oliveira, T. (2020). Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Frontiers in Genetics*, *11*.

Kchouk, M Gibrat, J. F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*, *9*(3).

Kim, B. K., Cho, M. S., Kim, M. H., Choi, H. J., Kang, M. J., Shim, H. S., Ahn, T. Y., Kim, J., & Park, D. S. (2012). Rapid and Specific Detection of *Burkholderia glumae* in Rice Seed by Real-Time Bio-PCR Using Species-Specific Primers Based on an *rhs* Family Gene. *Plant Disease*, *96*(4), 577–580.

Kim, H., Goo, E., Kang, Y., Kim, J., & Hwang, I. (2011). Regulation of Universal Stress Protein Genes by Quorum Sensing and RpoS in Burkholderia glumae. Journal of Bacteriology, *194*(5), 982–992.

Kim, J., Kim, J. G., Kang, Y., Jang, J. Y., Jog, G. J., Lim, J. Y., Kim, S., Suga, H., Nagamatsu, T., & Hwang, I. (2004). Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in Burkholderia glumae. Molecular microbiology, *54*(4), 921–934.

Kim, J., Kang, Y., Choi, O., Jeong, Y., Jeong, J. E., Lim, J. Y., Kim, M., Moon, J. S., Suga, H., & Hwang, I. (2007). Regulation of polar flagellum genes is mediated by quorum sensing and *FlhDC* in Burkholderia glumae. Molecular microbiology, *64*(1), 165–179.

Kim, S., Park, J., Kim, J. H., Lee, J., Bang, B., Hwang, I., & Seo, Y. S. (2013). RNAseq-based Transcriptome Analysis of *Burkholderia glumae* Quorum Sensing. *The Plant Pathology Journal*, *29*(3), 249–259.

Kim, S., Park, J., Lee, J., Shin, D., Park, D. S., Lim, J. S., Choi, I. Y., & Seo, Y. S. (2014). Understanding pathogenic *Burkholderia glumae* metabolic and signaling pathways within rice tissues through in vivo transcriptome analyses. *Gene*, *547*(1), 77–85.

Korotkov, K. V., Sandkvist, M., & Hol, W. G. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. Nature reviews. *Microbiology*, *10*(5), 336–351.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, *19*(9), 1639–1645.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mole*cular Biology and Evolution*, *35*(6), 1547–1549.

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, *35*(9), 3100–3108.

Lalithya, K. A., Manjunatha, G., Raju, B., Kulkarni, M. S., & Lokesh, V. (2017). Plant growth regulators and signal molecules enhance resistance against bacterial blight disease of pomegranate. *Journal of Phytopathology*, *165*(11-12), 727-736.

Lee, C., Mannaa, M., Kim, N., Kim, J., Choi, Y., Kim, S. H., Jung, B., Lee, H. H., Lee, J., & Seo, Y. S. (2019). Stress Tolerance and Virulence-Related Roles of Lipopolysaccharide in *Burkholderia glumae*. *The plant pathology journal*, *35*(5), 445–458.

Lee, J., Park, J., Kim, S., Park, I., & Seo, Y. S. (2016). Differential regulation of toxoflavin production and its role in the enhanced virulence of *Burkholderia gladioli*. *Molecular Plant Pathology*, 17(1), 65–76.

Lee, H. H., Lelis, T., Ontoy, J., Bruno, J., Ham, J. H., & Seo, Y. S. (2021). Complete Genome Sequence Data of Four Burkholderia glumae Strains Isolated from Rice Fields in the United States. *Molecular plant-microbe interactions: MPMI*, *34*(11), 1324–1327.

Li, J., Huang, Y., Li, Y., Yang, Z., & Wei, H. (2020). Comparative genomic analysis of six *Burkholderia gladioli* strains reveals core and pan-genome evolution. *Frontiers in Microbiology*, *11*, 61.

Li, L., Wang, L., Liu, L. M., Hou, Y. X., Huang, S. W., & Li, Q. Q. (2016a). Infection Process of Burkholderia glumae in Rice Spikelets. *Journal of Phytopathology*, *165*(2), 123–130.

Li, L., Wang, L., Liu, L. M., Hou, Y. X., Li, Q. Q., & Huang, S. W. (2016b). Infection Process of *Burkholderia glumae* Before Booting Stage of Rice. *Journal of Phytopathology*, *164*(10), 825–832.

Li, W., Jaroszewski, L., & Godzik, A. (2002). Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics*, *18*(1), 77–82.

Lim, J., Lee, T. H., Nahm, B. H., Choi, Y. D., Kim, M., & Hwang, I. (2009). Complete genome sequence of *Burkholderia glumae* BGR1. *Journal of bacteriology*, *191*(11), 3758–3759.

Lin, K., Smit, S., Bonnema, G., Sanchez-Perez, G., & de Ridder, D. (2014). Making the difference: integrating structural variation detection tools. *Briefings in Bioinformatics*, *16*(5), 852–864.

Lin S-H, Liao Y-C (2013) CISA: Contig Integrator for Sequence Assembly of Bacterial Genomes. *PLoS ONE, 8*(3), e60843.

Lipscomb, L., & Schell, M. A. (2011). Elucidation of the regulon and cis-acting regulatory element of *HrpB*, the AraC-type regulator of a plant pathogen-like type III secretion system in *Burkholderia pseudomallei*. *Journal of bacteriology*, *193*(8), 1991–2001.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research*, *25*(5), 955–964.

Magbanua, Z. V., Arick, M., Buza, T., Hsu, C. Y., Showmaker, K. C., Chouvarine, P., Deng, P., Peterson, D. G., & Lu, S. (2014). Transcriptomic

dissection of the rice – *Burkholderia glumae* interaction. *BMC Genomics*, *15*(1).

Mannaa, M., Park, I., & Seo, Y.-S. (2018). Genomic Features and Insights into the Taxonomy, Virulence, and Benevolence of Plant-Associated *Burkholderia* Species. *International Journal of Molecular Sciences*, *20*(1), 121.

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, *470*, 198–203.

McCann, H. C., Li, L., Liu, Y., Li, D., Pan, H., Zhong, C., Rikkerink, E. H. A., Templeton, M. D., Straub, C., Colombi, E., Rainey, P. B., & Huang, H. (2017). Origin and Evolution of the Kiwifruit Canker Pandemic. *Genome biology and evolution*, *9*(4), 932–944.

McCoy, R.C., Taylor, R.W., Blauwkamp, T.A., et al. (2014). Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS One*, *9*(9), e106689.

McGinnis, J., Laplante, J., Shudt, M., & George, K. S. (2016). Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. *Journal of Clinical Virology*, *79*, 44–50.

Md. Zali, A., & Iberahim, R. (2021). Relationship between Routine Handling Practices with Potential Pathogenic Bacteria Isolated from Contact Lenses Among Students in UiTM Negeri Sembilan. *Journal of Academia*, *9*(1), 106-116.

Md-Yassin, F. (2018). Growth drivers of the Malaysian economy: A comparative analysis between pre and post financial crisis. *Journal of Applied Economic Sciences*, *13*(4), 966-978.

Ministry of Agriculture and Agro-Based Industry. (2020). Agricultural trade statistic. Retrieved from https://www.moa.gov.my/muat-turun/perangkaan-perdagangan-pertanian

Minoche, A. E., Dohm, J. C., & Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, *12*(11), R112.

Misman, S. N., & Zakaria, L. (2019). Pathotype Identification of Rice Blast Pathogen, *Pyricularia oryzae* Using Differential Varieties in Peninsular Malaysia. *Tropical Life Sciences Research*, *30*(2), 181–190.

Mizobuchi, R., Fukuoka, S., Tsuiki, C., Tsushima, S., & Sato, H. (2018). Evaluation of major Japanese rice cultivars for resistance to bacterial grain rot caused by *Burkholderia glumae* and identification of standard cultivars for resistance. *Breeding Science*, *68*(4), 413–419.

Mizobuchi, R., Fukuoka, S., Tsushima, S., Yano, M., & Sato, H. (2016). QTLs for Resistance to Major Rice Diseases Exacerbated by Global Warming: Brown Spot, Bacterial Seedling Rot, and Bacterial Grain Rot. *Rice*, *9*(1).

Mohamad, N. H., Idris, A. B., & Amin, M. S. M. (2017). The role of rice production in ensuring food security in Malaysia. *Journal of Tropical Agriculture and Food Science*, *45*(1), 69-78.

Mohd Kasim, N. A., Teh, C. Y., Namasivayam, P., Mohd. Yusoff, N. F., & Ho, C. L. (2020). Analyses of yield related agronomic traits of Malaysian rice varieties. *Asia Pacific Journal of Molecular Biology and Biotechnology*, 59–74.

Moon, H., Park, H. J., Jeong, A., Han S. W., & Park C. J. (2017). Isolation and identification of *Burkholderia gladioli* on *Cymbidium* orchids in Korea. *Biotechnology & Biotechnological Equipment*, *31*(2), 280-288.

Moore, P. B., & Steitz, T. A. (2011). The roles of RNA in the synthesis of protein. *Cold Spring Harbor perspectives in biology*. 3(11), a003780.

Mostafa, E. M., Aly, S. & Bolbol, D. (2015). Overviews of "next-generation sequencing". *Research and Reports in Forensic Medical Science*, 1-5.

Moule, M. G., Spink, N., Willcocks, S., Lim, J., Guerra-Assunção, J. A., Cia, F., Champion, O. L., Senior, N. J., Atkins, H. S., Clark, T., Bancroft, G. J., Cuccui, J., & Wren, B. W. (2015). Characterization of New Virulence Factors Involved in the Intracellular Growth and Survival of Burkholderia pseudomallei. *Infection and immunity*, *84*(3), 701–710.

Mulaw, T., Wamishe, Y., & Jia, Y. (2018). Characterization and in plant detection of bacteria that cause bacterial panicle blight of rice. *American Journal of Plant Sciences*, *9*(04), 667.

Nandakumar, R., Shahjahan, A. K. M., Yuan, X. L., Dickstein, E. R., Groth, D. E., Clark, C. A., Cartwright, R. D., & Rush, M. C. (2009). *Burkholderia glumae* and *B. gladioli* Cause Bacterial Panicle Blight in Rice in the Southern United States. *Plant Disease*, *93*(9), 896–905.

Nilsson, S. (2020). Rice in Malaysia | OEC. OEC - The Observatory of Economic Complexity. https://oec.world/en/profile/bilateral-product/rice/reporter/mys

OECD-FAO Agricultural Outlook. (2017). OECD-FAO Agricultural Outlook 2019-2028. https://stats.oecd.org/Index.aspx?DataSetCode=HIGH_AGLINK_2019

Omar, S. C., Shaharudin, A., & Tumin, S. A. (2019). The status of the paddy and rice industry in Malaysia. *Khazanah Research Institute*.

Ortega, L., & Rojas, C. M. (2021). Bacterial Panicle Blight and *Burkholderia glumae*: From Pathogen Biology to Disease Control. *Phytopathology, 111*(5), 772–778.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., ... & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*, *42*, D206-D214.

Pagel, Kymberleigh A.; Antaki, Danny; Lian, AoJie; Mort, Matthew; Cooper, David N.; Sebat, Jonathan; Iakoucheva, Lilia M.; Mooney, Sean D.; Radivojac, Predrag; Kann, Maricel G (2019). Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome. *PLOS Computational Biology*, *15*(6), e1007112.

Park, J. Y., Kim, Y. C., Kim, C. J., Choi, I. Y., Lee, Y. H., & Cho, K. Y. (2006). Toxoflavin produced by *Burkholderia glumae* causing rice grain rot is responsible for inducing bacterial wilt in many field crops. *Plant Disease*, *90*(7), 856-861

Pasquali, F., do Valle, I., Palma, F., Remondini, D., Manfreda, G., Castellani, G., Hendriksen, R., & de Cesare, A. (2019). Application of different DNA extraction procedures, library preparation protocols and sequencing platforms: impact on sequencing results. *Heliyon*, *5*(10), e02745.

Pedraza, L. A., Bautista, J., & Uribe-Vélez, D. (2018). Seed-born Burkholderia glumae Infects Rice Seedling and Maintains Bacterial Population during Vegetative and Reproductive Growth Stage. *The Plant Pathology Journal*, *34*(5), 393–402.

Peng, J. (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. *LSU Master's Theses*. 2722.

Perez, B. A., Planet, P. J., Kachlany, S. C., Tomich, M., Fine, D. H., & Figurski, D. H. (2006). Genetic analysis of the requirement for flp-2, tadV, and rcpB in *Actinobacillus actinomycetemcomitans* biofilm formation. *Journal of bacteriology*. *188*(17), 6361–6375.

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., & Szemes, T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical journal*, *44*(5), 548–559.

Rahmat, S. R., Firdaus, R. R., Mohamad Shaharudin, S., & Yee Ling, L. (2019). Leading key players and support system in Malaysian paddy production chain. *Cogent Food & Agriculture*, *5*(1), 1708682.

Rajamoorthy, Y., Rahim, K. B. A., & Munusamy, S. (2015). Rice industry in Malaysia: Challenges, policies and implications. *Procedia Economics and Finance*, *31*(2015), 861–867.

Ramachandran, K., Vijaya, S. I., Ahmad, F. N., Amzah, B., & Zakaria, L. (2021). Characterization and identification of *Burkholderia glumae* as the causal pathogen of bacterial panicle blight of rice (*Oryza sativa* L.) in Malaysian rice granaries. *Journal of General Plant Pathology*, *87*(3), 164–169.

Remenant, B., Babujee, L., Lajus, A., Médigue, C., Prior, P., & Allen, C. (2012). Sequencing of K60, type strain of the major plant pathogen *Ralstonia solanacearum*. *Journal of bacteriolog. 194*(10), 2742–2743.

Reynolds, N. M., Vargas-Rodriguez, O., Söll, D., & Crnković, A. (2017). The central role of tRNA in genetic code expansion. *Biochimica et biophysica acta*. *1861*(11), 3001–3008.

Riera-Ruiz, C., Vargas, J., Cedeño, C., Quirola, P., Escobar, M., Cevallos-Cevallos, J. M., Ratti, M., & Peralta, E. L. (2014). First Report of *Burkholderia glumae* Causing Bacterial Panicle Blight on Rice in Ecuador. *Plant Disease*, *98*(7), 988.

Rigau, M., Juan, D., Valencia, A., & Rico, D. (2019). Intronic CNVs and gene expression variation in human populations. *PLoS genetics*. *15*(1), e1007902.

Robert, Francis., & Pelletier, Jerry (2018). Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Frontiers in Genetics*, *9*, 507.

Sabran, S. H., & Abas, A. (2021). Knowledge and Awareness on the Risks of Pesticide Use Among Farmers at Pulau Pinang, Malaysia. *SAGE Open*, *11*(4), 215824402110648.

Sadek, S. M., Yusoff, S. A., & Omar, M. A. (2019). The Contribution of Rice to the Malaysian Economy: An Input-Output Analysis. *Journal of Agribusiness and Rural Development*, *53*(2), 161-173.

Saikia, S. P., Bhuyan, S. I., Deka, M., & Gogoi, D. K. (2013). Production and characterization of toxoflavin from *Burkholderia gladioli* isolated from soil. *BioMed Research International*, 548635.

Sassetti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences*, *100*(22), 12989-12994.

Sathya, B., Dharshini, A. P., & Kumar, G. R. (2014). NGS meta data analysis for identification of SNP and INDEL patterns in human airway transcriptome: A preliminary indicator for lung cancer. *Applied & translational genomics*, *4*, 4–9.

Sayler, R. J., Cartwright, R. D., & Yang, Y. (2006). Genetic Characterization and Real-Time PCR Detection of *Burkholderia glumae*, a Newly Emerging Bacterial Pathogen of Rice in the United States. *Plant Disease*, *90*(5), 603–610.

Seo, Y. S., Lim, J., Choi, B. S., Kim, H., Goo, E., Lee, B., Lim, J. S., Choi, I. Y., Moon, J. S., Kim, J., & Hwang, I. (2011). Complete genome sequence of *Burkholderia gladioli* BSR3. *Journal of bacteriology*. *193*(12), 3149.

Shaner, G. and Finney, R.E. (1977). The effect of nitrogen fertilization in the expression of slow-mildewing resistance in Knox wheat. *Phytopathology*, *67*, 1051-1056.

Shew, A. M., Durand-Morat, A., Nalley, L. L., Zhou, X.-G., Rojas, C., & Thoma, G. (2019). Warming increases Bacterial Panicle Blight (*Burkholderia glumae*) occurrences and impacts on USA rice production. *PLOS ONE*, *14*(7), e0219199.

Siwar, C., Diana Mohd Idris, N., Yasar, M., & Morshed, G. (2014). Issues and Challenges Facing Rice Production and Food Security in the Granary Areas in the East Coast Economic Region (ECER), Malaysia. *Research Journal of Applied Sciences, Engineering and Technology,* 7*(*4), 711–722.

Slatko, B. E.; Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59.

Son, M. S., & Taylor, R. K. (2011). Preparing DNA libraries for multiplexed paired-end deep sequencing for Illumina GA sequencers. *Current protocols in microbiology*, Chapter 1, Unit1E.4–1E.4.

Sundaramoorthy, S., Karthikeyan, G., Suresh, S., & Raguchander, T. (2017). Virulence evaluation of Burkholderia glumae strains in rice cultivars under greenhouse condition. *Biotechnology & Biotechnological Equipment*, *31*(5), 897-902.

Sunian, Elixon & Ramli, Asfaliza & Omar, Othman & Misman, Siti Norsuha & Mohamad Saad, Maisarah & Jack, Allicia & Hashim, Shahida. (2017). Evaluation on yield, yield component and physico-chemicals of advanced rice lines. *Journal of Tropical Agriculture and Food Science*. *45*, 131-143.

Takenaka, Y., Takeda, K., Yoshii, T., Hashimoto, M., & Inohara, H. (2012). Gram Staining for the Treatment of Peritonsillar Abscess. *International Journal of Otolaryngology*, *2012*, 1–5.

Tan, B. D. (2019). Where does Malaysia's paddy and rice industry stand? *The Star*.

Tan, L. W., A. Rahman, Z., Goh, H. H., Hwang, D. J., Ismail, I., & Zainal, Z. (2017). Production of Transgenic Rice (indica 1 cv. MR219) Overexpressing *Abp57* Gene through Agrobacterium-Mediated Transformation. *Sains Malaysiana*, *46*(5), 703–711.

Tankeshwar, A. (2012). oxidase test: Principle Procedure and oxidase positive organisms. *Microbeonline.com*.

Tayeb, L. A., Lefevre, M., Passet, V., Diancourt, L., Brisse, S., & Grimont, P. A. (2008). Comparative phylogenies of *Burkholderia, Ralstonia, Comamonas, Brevundimonas* and related organisms derived from *rpoB, gyrB* and *rrs* gene sequences. *Research in microbiology*, *159*(3), 169–177.

Thomas, M., Fenske, G. J., Antony, L., Ghimire, S., Welsh, R., Ramachandran, A., & Scaria, J. (2017). Whole genome sequencing-based detection of antimicrobial resistance and virulence in non-typhoidal *Salmonella enterica* isolated from wildlife. *Gut Pathogens*, *9*(1).

Thompson, J.F., & Milos, P.M. (2011). The properties and applications of single-molecule DNA sequencing. *Genome Biology*, (12), 217.

Toschka, H. Y., Höpfl, P., Ludwig, W., Schleifer, K. H., Ulbrich, N., & Erdmann, V. A. (1988). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Pseudomonas aeruginosa. Nucleic acids research*. *16*(5), 2348.

Tribble, G. D., & Lamont, R. J. (2010). Bacterial invasion of epithelial cells and spreading in periodontal tissue. *Periodontology 2000*, *52*(1), 68–83.

United States Department of Agriculture. (2021). Grain: World Markets and Trade. Retrieved from https://apps.fas.usda.gov/psdonline/circulars/grain.pdf

Vaghefi, N. & Shamsudin, Mad Nasir & Radam, Alias & Rahim, Khalid. (2015). Impact of climate change on food security in Malaysia: economic and policy adjustments for rice industry. *Journal of Integrative Environmental Sciences*. 1-17.

Vandamme, P., Peeters, C., de Smet, B., Price, E. P., Sarovich, D. S., Henry, D. A., Hird, T. J., Zlosnik, J. E. A., Mayo, M., Warner, J., Baker, A., Currie, B. J., & Carlier, A. (2017). Comparative Genomics of *Burkholderia singularis* sp. nov., a Low G+C Content, Free-Living Bacterium That Defies Taxonomic Dissection of the Genus *Burkholderia. Frontiers in Microbiology*, *8*.

Vander Broek, C. W., & Stevens, J. M. (2017). Type III Secretion in the Melioidosis Pathogen *Burkholderia pseudomallei*. *Frontiers in cellular and infection microbiology*. *7*, 255.

Vellasamy, K. M., Mariappan, V., Shankar, E. M., & Vadivelu, J. (2016). *Burkholderia pseudomallei* Differentially Regulates Host Innate Immune Response Genes for Intracellular Survival in Lung Epithelial Cells. *PLOS Neglected Tropical Diseases*, *10*(7), e0004730.

Viallard, V., Poirier, I., Cournoyer, B., Haurat, J., Wiebkin, S., Ophel-Keller, K., & Balandreau, J. (1998). *Burkholderia graminis* sp. nov., a rhizospheric *Burkholderia species*, and reassessment of [*Pseudomonas*] *phenazinium*, [*Pseudomonas*] *pyrrocinia* and [*Pseudomonas*] *glathei* as *Burkholderia*. *International journal of systematic bacteriology*, *48*, 549–563.

Viguera, E., Canceill, D., & Ehrlich, S. D. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *The EMBO journal*, *20*(10), 2587–2595.

Wamishe, Y.A., Belmar, S.B., Kelsey, C.D., Gebremariam, T.A., McCarty, D.L. (2014). Effects of excessive nitrogen fertilizer on rice diseases with emphasis to bacterial panicle blight. *Proceedings of Rice Technical Working Group (RTWG).* 86-87.

Wang, Y., Freedman, Z., Lu-Irving, P., Kaletsky, R., & Barkay, T. (2009). An initial characterization of the mercury resistance (*mer*) system of the thermophilic bacterium *Thermus thermophilus* HB27. *FEMS microbiology ecology*, *67*(1), 118–129.

Wang, S., Nie, W., Gong, Q., Lee, Y., Shui, H., Chen, G., & Zhu, B. (2020). Complete Genomic Data of *Burkholderia glumae* Strain GX Associated with Bacterial Panicle Blight of Rice in China. *Plant Disease*, *104*(6), 1578–1580.

Weckselblatt, B., Hermetz, K. E., & Rudd, M. K. (2015). Unbalanced translocations arise from diverse mutational mechanisms including chromothripsis. *Genome research*, *25*(7), 937–947.

Weinberg, J. B., Alexander, B. D., Majure, J. M., Williams, L. W., Kim, J. Y., Vandamme, P., & LiPuma, J. J. (2007). *Burkholderia glumae* infection in an infant with chronic granulomatous disease. *Journal of clinical microbiology*, *45*(2), 662–665.

Weny, Safni, I., Lisnawita, & Lubis, K. (2019). Screening for disease resistance in rice varieties against bacterial panicle blight disease (*Burkholderia glumae*) in Northern Sumatra of Indonesia. *IOP Conference Series: Earth and Environmental Science*, *260*(1), 012118.

Wright, B., Farquharson, K.A., McLennan, E.A. (2019). From reference genomes to population genomics: comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics*. *20*, 453.

Xiao, T., & Zhou, W. (2020). The third-generation sequencing: the advanced approach to genetic diseases. *Translational pediatrics, 9*(2), 163–173.

Yasar, M. & Siwar, C. (2016). Paddy field conversion in Malaysia: issues and challenges. *Jurnal Rona Teknik Pertanian*, *9*(2).

Zhou, Q., Sun, W., Liu, X., Wang, X., Xiao, Y., Bi, D., & Shi, D. (2016). Third-Generation Sequencing and Analysis of Four Complete Pig Liver Esterase Gene Sequences in Clones Identified by Screening BAC Library. *PLOS ONE*, *11*(10), e0163295.

Zhou, X. G. (2019). Sustainable Strategies for Managing Bacterial Panicle Blight in Rice. *Protecting Rice Grains in the Post-Genomic Era*.

Zhou, X. G. (2014). First Report of Bacterial Panicle Blight of Rice Caused by *Burkholderia glumae* in South Africa. *Plant Disease*, *98*(4), 566.

Zhou-qi, C., Bo, Z., Guan-lin, X., Bin, L., & Shi-wen, H. (2016). Research Status and Prospect of Burkholderia glumae, the Pathogen Causing Bacterial Panicle Blight. *Rice Science*, *23*(3), 111–118.

APPENDICES

Appendix 1

1X TBE buffer

To prepare 1X TBE buffer, 100 mL of 10X TBE buffer is diluted with 900 mL distilled water.

Agarose gel

To prepare agarose gel, dissolve 1 g Agarose Biotechnology Grade in 100 mL 1X TBE buffer.

FloroSafe DNA stain

2 µL FluroSafe DNA Stain, 100 mL agarose gel solution

Appendix 2: Supplementary information in Google Drive link

Structural Variation analysis

Burkholderia glumae

https://drive.google.com/drive/folders/1ilsQSYyFo5YqvnQViyb5xz2r5nIYBl_7?usp=share_link

Burkholderia gladioli

https://drive.google.com/drive/folders/1LACDXz6VU1dUJtyO8tFtz9oKfXrcmG9c?usp=share_link

***De novo* annotation analysis**

Burkholderia glumae

https://drive.google.com/drive/folders/11wlRJWadpPT4W7loc2mWek-n7IbyK1Z6?usp=share_link

Burkholderia gladioli

https://drive.google.com/drive/folders/13u7M7FZ3yPd2VvlOOqYyl_WDVA0La4Ru?usp=share_link

Appendix 3: Analysis of variance (ANOVA) of BPB of rice

The ANOVA Procedure

Dependent Variable: disease severity (MR219)

The ANOVA Procedure

Dependent Variable: disease severity (MR219)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	6457.088057	496.699081	13.40	<.0001
Error	22	815.345002	37.061136		
Corrected Total	35	7272.433058			

R-Square	Coeff Var	Root MSE	diseaseseverity Mean
0.887886	10.29288	6.087786	59.14559

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	2	343.803996	171.901998	4.64	0.0209
trt	11	6113.284060	555.753096	15.00	<.0001

Appendix 4 : Nucleotide sequence of all 27 bacterial isolates using 16s rRNA

BGK-AZ1

TGCAAGTCGAACGGCAGCACGGGTGCTTGCACCTGCTGGCGAGTGGC
GAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGATAGCC
CGGCGAAAGCCGGATTAATACCGCATACGATCTACGGAAGAAAGCGGG
GGACCTTCGGGCCTCGCGCTATAGGGTTGGCCGATGGCTGATTAGCTA
GTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCAGTAGCTGGTCTGA
GAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCA
GCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGT
CCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGGGATGACGGTACCG
GAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGT
AGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGC
GGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGAACTGC
ATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGGGGGTAGAATTCCAC
GTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGGCGAAG
GCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAAGCGTGGGGA
GCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAAC
TAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCGTGAAGTT
GACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGAC
GGGGACCCGCACAAGCGGTGGATGATGTGGATTAATTCGATGCAACGC

GAAAAACCTTACCTACCCTTGACATGGTCGGAATCCTGAGGAGACTCGG
GAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
TGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAGACTGCCGGTGACAA
ACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTATGGGT
AGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTTGCCAACCCG
CGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTCCGGATTGCACTC
TGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCGGATCAGC
ATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACA
CCATGAGAGTGGGTAACACCAGAAGTGGCTAGTCTAACCGCAAGGAGG
CCCAGCTCGCCCAC

BGK.AZ2

CACGGGTGCTTGCACCTGCTGGCGAGTGGCGAACGGGTGAGTAATACA
TCTGGAACATGTCCTGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTA
ATACCGCATACGATCTTCGGAAGAAAGCGGGGGACCTTCGGGCCTCGC
GCTATAGGGGTGGCCTATGGCTGATTACCTAGTTGGTGGGCTAATTGCC
TACCAAGGCGACCATCAGTACCTGGACTGATAGGACAACCAGCCAGAC
TGGGACTGACACTCGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGA
ATTTTGGACAATGGGCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGA
AGAAGGCCTTCATGTTGTATATCACTTTTGTCCGGAATGAAATCCTGAAG
GCTAATATCCTTCGGGGATGACGGTACCGGAAGA
ATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGT
GCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTTT
GCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGCAACTGCATTGG
GGACTGGCAGGCTAGAGTATGGCACAGGGGTGTAGAATTCCACGTGTA
GCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGGCGAAGGCAGC
CCCCTGGGCCAATACTGACGCTCATGCACGAAAGCGTGGGGAGCAAAC
AGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATCTCCGCTAATTG
TTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCCTGAAGTTGACCGC
CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTCGCGAGGAA
TTAAACACTCCACAGAGGACTTGAATTAATTCGAACCCAAATCAAAAACC
TTACCTACCCTTGACATGGTCGGAATCCTGAGGAGACTCGGGAGTGCT
CGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCGTCAGCTCGTG
TCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGT

BGK.AZ3

CTGCCTTAAAACTGGCAAGTCGAAGCGGACTGCATGTGGTACTTGCATC
TGTGTGGAGTGTGGCGAACGGGTGAGTAATACATCTGAACATGTCCTGT
ATTGGAGGATAACCCGGCCGAAAACCGGATTACTACCGCATACTATCTT
CTGAAAAAAGCTGGGTACCTTCTGACCTCCCTGCTATAGGTGTGAATTA
TGCCATGACTACAGTATAGCAGAGTGAGCTAAAGGCCTACCAAGGCGA
CGATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGA
CACGGCCCAGACTCCTACGGGAGATAGCTGAAGGAAATTTTGGACAAT
GGGCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCG
GGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCT
TCGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCA
GCAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGC
CGTAAAGCGTGCGCAGGCGGTTTGCTAAGACCGATCTGAAATCACTGA
GCTCAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAG
AGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGA
GGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCAT

GCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC
GCCCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGT
AGCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAA
AACTCAAAGGAATTGACGACCCGCACAAGCGGTGGATGATGTGGATTAA
TTCGATACAAGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCCT
GAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCA
TAGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
AGCCCAACTCTTGTCCTTAGTTGATACGCAAGAGCATTCTAAGGAGACT
GCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATG
GCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAGG
GTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTC
CGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAAT
CGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACAC
CGCCCGTCACACCATGGGAGTGGGTAACACCAGAAGTGGCTAGTCTAA
CCGCAAGGAGG

BGK-AZ4

TACACATGCAAGTCGAACGGCAGCACGGGTGCTTGCACCTGGTGGCGA
GTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGA
TAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTACGGAAGAAA
GCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCGATGGCTGAT
TAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCAGTAGCTG
GTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACT
CCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTG
ATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCAC
TTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGGGATGACGG
TACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCG
CAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGA
ACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGGGGGTAGAAT
TCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGG
CGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAAGCGTG
GGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATG
TCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCGTGA
AGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATT
GACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAATTCGATGCAA
CGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCCTGAGGAGACT
CGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCG
TCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
CCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAGACTGCCGGTGA
CAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTATG
GGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTTGCCAAC
CCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTCCGGATTGCA
CTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCGGATCA
GCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCA
CACCATGGGAGTGGGT

BGK.AZ5

TGCCTTACCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACCTGGTG
GCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGG
GGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTACGGAA
GAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCGATGGC

TGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCAGTA
GCTGGTCTGAGAAGACGACCAGCCACACTGGGACTGAGACACGGCCCA
GACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAG
CCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAA
GCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGGGATG
ACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCG
GTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCG
TGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCT
GGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGGGGGT
AGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACC
GATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAA
GCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAA
CGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACG
CGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAA
GGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAATTCG
ATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCCTGAG
GAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGG
CTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAG
CGCAACCCTTGTCCTTAGTTGCTACGCAAGACCACTCTAAGGAGACTGC
CGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCC
CTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTT
GCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTCCGG
ATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGC
GGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGC
CCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAACCGC
AAGGAGGACGGTCACCCACCGGTAGATCGC

BGK.AZ6

AGGGGATGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCA
CCTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCT
GTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATC
TACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGC
CGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACG
ATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACA
CGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGG
GCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGG
TTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTC
GGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGC
AGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCG
TAAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGC
TCAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAG
GGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGG
AATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGC
ACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGC
CCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAG
CTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAA
CTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATT
AATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAAT
CCTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCT
GCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGC
AACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGA

GACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTC
ATGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAG
TCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTA
ATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACAC
ACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTA
ACCGCAAGGAGGACGTCACCCACGGTAGATCAGGTA

BGK.AZ7

TTGGGGGTGCTTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGC
ACCTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCC
TGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGAT
CTACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGG
CCGATGGCTGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGAC
GATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGAC
ACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATGCACAATGG
GCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGG
TTGTAAAGTACTTTCAGCCGGAAAGAAATCCTGAGGGCTAATATCCTTC
GGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGC
AGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCG
TAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGC
TCAACCTGGGAACTGCATTCGAAACTGGCAGGCTAGAGTCTTGTAGAG
GGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGG
AATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGT
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGC
CGTAAACGATGTCGACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTA
GCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA
ACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTT
TAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATGGTCGGAA
TCCTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGC
TGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGC
AACGAGCGCAACCCTTATCCTTTGTTGCCACGCAAGAGCACTCTAAGGA
GACTGCCGGTGACAAACCGGGAGAAAGGTGGGGATGACGTCAAGTCAT
CATGGCCCTTACGAGTAGGGCTACACACGTCATACAATGGTCGGAACA
GAGGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGT
AGTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTA
GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTA
CACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTGGCTAG
TCTAACCGCAAGGAGGACGGTCACCACGGTAGTTCAGGTT

BGK.AZ8

AATGGGGGTGCTTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTG
CACCTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTC
CTGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGA
TCTACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGG
CCGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGAC
GATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGAC
ACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATG
GGCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGG
GTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTT
CGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAG
CAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGC

GTAAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGG
CTCAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGA
GGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAG
GAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATG
CACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACG
CCCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTA
GCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAA
ACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGAT
TAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAA
TCCTGAGGAGACTCGG
GAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
TGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAGACTGCCGGTGACAA
ACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTATGGGT
AGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTTGCCAACCCG
CGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTCCGGATTGCACTC
TGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCGGATCAGC
ATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACA
CCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAACCGCAAGGAGGA
CGGTCACCACGGTAGTTCAGGTT

BGS.AS1

ACATGCAAGTCGAACGGCAGCACGGGTGCTTGCACCTGGTGGCGAGTG
GCGAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGATAG
CCCGGCGAAAGCCGGATTAATACCGCATACGATCTACGGAAGAAAGCG
GGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCGATGGCTGATTAGC
TAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCAGTAGCTGGTCT
GAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTA
CGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTGATCC
AGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTG
TCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGGGATGACGGTACC
GGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACG
TAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGG
CGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGAACTG
CATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGGGGGTAGAATTCCA
CGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGGCGAA
GGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAAGCGTGGGG
AGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAA
CTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCGTGAAGTT
GACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGAC
GGGGACCCGCACAAGCGGTGGATGATGTGGATTAATTCGATGCAACGC
GAAAAACCTTACCTACCCTTGACATGGTCGGAATCCTGAGGAGACTCGG
GAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
TGTCCTTAGTTGCTACGCAAGACCACTCTAAGGAGACTGCCGGTGACAA
ACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTATGGGT
AGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTTGCCAACCCG
CGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTCCGGATTGCACTC
TGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCGGATCAGC

ATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACA
CCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAACCGCAAGGAGGA
CGGTTCACCCACGGTAG

BGS.AS2

ACATGCAAGTCGAACGGCAGCACGGGTGCTTGCACCTGGGTGGCGAGT
GGCGAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGATA
GCCCGGCGAAAGCCGGATTAATACCGCATACGATCTACGGAAGAAAGC
GGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCGATGGCTGATTAG
CTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCAGTAGCTGGTC
TGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCCAGACTC
CTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTGA
TCCAGCAATGCCGCGTGTGTTGAAGAAGGCCTTCGGGTTGTAAAGCAC
TTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGGGATGACGG
TACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCG
CAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGA
ACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGGGGGTAGAAT
TCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGG
CGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAAGCGTG
GGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATG
TCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCGTGA
AGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATT
GACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAATTCGATGCAA
CGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCCTGAGGAGACT
CGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCG
TCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
CCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAGACTGCCGGTGA
CAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTATG
GGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTTGCCAAC
CCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTCCGGATTGCA
CTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCGGATCA
GCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCA
CACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAACCGCAAGGAG
GACGGTCACCCACGTAG

BGK.AS3

AGGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCAC
CTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTG
TAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCT
ACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCC
GATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGA
TCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGG
CGAAAGCCTGATCCAGCAATGCCGCGTGTGGGAAGAAGGCCTTCGGGT
TGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCG
GGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCA
GCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGT
AAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCT
CAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGG
GGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGA
ATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCA

CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
CTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGC
TAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAAC
TCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTA
ATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATC
CTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTG
CATGGCTGTCGTCAGCTCCTGTCGTGAAAATGTTGGGTTAAATCCCGCA
ACGAGCGCAACCCTTGTCCTTAATTGCTACCCAAGAGCACTCTAAGGAG
ACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCA
TGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAG
TCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTA
ATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACAC
ACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTA
ACCGCAAGGAGGACGTCACCCACGGTAGTTCGGGT

BGK.AS4

TGCGGTGCCTTACCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGAT
CAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGC
GAAAGCCTGATCCAGCAATGCCGCGTGNTGTGAAGAAGGCCTTCGGGT
TGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCG
GGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCA
GCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGT
AAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCT
CAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGG
GGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGA
ATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCA
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
CTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGC
TAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAAC
TCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTA
ATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATC
CTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTG
CATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCA
ACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAG
ACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCA
TGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAG
TCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTA
ATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACAC
ACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTA
ACCGCAAGGAGGACGGTCACCCACGGTAGTTC

BGS.AS5

AAGGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCA
CCTGGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCC
TGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGAT

CTACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGG
CCGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGAC
GATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGAC
ACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATG
GGCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGG
GTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTT
CGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAG
CAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGC
GTAAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGG
CTCAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGA
GGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAG
GAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATG
CACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACG
CCCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTA
GCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAA
ACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGAT
TAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAA
TCCTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGC
TGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
CAACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGG
AGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCT
CATGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACA
GAGGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGT
AGTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTA
GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTA
CACACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGT
CTAACCGCAAGGAGGACGGTTCACCCACGGTAGTTC

BGS.AS6

CGGGGTGCCTTAACCATGCAAGTCGAACGGCAGCACGGGTGCTTGCAC
CTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTG
TAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCT
ACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCC
GATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGA
TCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGG
CGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGT
TGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCG
GGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCA
GCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGT
AAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCT
CAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGG
GGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGA
ATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCA
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
CTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGC
TAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAAC
TCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTA
ATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATC
CTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTG
CATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCA

ACGAGCGCAACCCTTGTCCTTAGTTGCTACCCAAGACCACTCTAAGGAG
ACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCA
TGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAG
TCCGGATT
GCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCGG
ATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCC
GTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAACCGCAA
GGAGGACGGTCACCCACGGTTAGTTC

BGS.AS7
GGGGTGCCTTAACCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGAT
CAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGC
GAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTT
GTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGG
GGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTC
AACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGG
GGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAA
TACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCAC
GAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCC
TAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCT
AACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACT
CAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAA
TTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCC
TGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAGA
CTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCAT
GGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAG
GGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGT
CCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAA
TCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACA
CCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAA
CCGCAAGGAGGACGGTCACCCACGGTAGAT
BGP.A1
GGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGAT
CAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG

GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGC
GAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTT
GTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGG
GGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGCAGGCGGTTTGCTAAGACCGATG
TGAAATCCCCGGGCTCAACCTGGGAACTGCATTGGTGACTGGCAGGCT
AGAGTATGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCG
TAGAGATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAAT
ACTGACGCTCATGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACC
CTGGTAGTCCACGCCCTAAACGATGTCAACTAGTTGTTGGGGATTCATT
TCCTTAGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGGGAGTACG
GTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGG
TGGATGATGTGGATTAATTCGATGCAACGCGAAAAACCTTACCTACCCT
TGACATGGTCGGAATCCTGAGGAGACTCGGGAGTGCTCGAAAGAGAAC
CGATACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGT
TGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTTAGTTGCTACGCA
AGAGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGGGGGG
ATGACGTCAAGTCCTCATGGCCCTTATGGGTAGGGCTTCACACGTCATA
CAATGGTCGGAACAGAGGGTTGCCAACCCGCGAGGGGGAGCTAATCC
CAGAAAACCGATCGTAGTCCGGATTGCACTCTGCAACTCGAGTGCATGA
AGCTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGT
TCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTTAC
CAGAAGTGGCTAGTCTAACCGCAAGGAGGACGGTCACCCACGGTAGAT
CAGGT

BGP.A2

TGGGGCTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCA
CCTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCT
GTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATC
TACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGC
CGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACG
ATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACA
CGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGG
GCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTTCGG
GTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTT
CGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAG
CAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGC
GTAAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGG
CTCAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGA
GGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAG
GAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATG
CACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACG
CCCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTA
GCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAA
ACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGAT
TAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAA
TCCTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGC
TGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
CAACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGG
AGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCT

CATGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACA
GAGGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGT
AGTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTA
GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTA
CACACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGT
CTAACCGCAAGGAGGACGGTCACCCACGGTAGTTCAGGT

BGP.A3

GGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGAT
CAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGC
GAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTT
GTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGG
GGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTC
AACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGG
GGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAA
TACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCAC
GAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCC
TAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCT
AACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACT
CAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAA
TTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCC
TGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAACACTCTAAGGAGA
CTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCAT
GGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAG
GGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGT
CCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAA
TCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACA
CCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAA
CCGCAAGGAGGACGGTCACCCACGGTAG

BGP.A4

AGGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCAC
CTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTG
TAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCT
ACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCC
GATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGA
TCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGG
CGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGT
TGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCG
GGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCA
GCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGT

AAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCT
CAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGG
GGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGA
ATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCA
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
CTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGC
TAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAAC
TCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTA
ATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATC
CTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTG
CATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCA
ACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAACACTCTAAGGAA
ACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCA
TGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAG
TCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTA
ATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACAC
ACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTA
ACCGCAAGGAGGACGTCACCCACGGTAG

BGP.A5

GGGTGCCTTAACCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACCT
GGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGTA
GTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTAC
GGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCGA
TGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATC
AGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCG
AAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTG
TAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGG
GATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGC
CGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAA
AGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCA
ACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGGG
GGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATA
CCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGA
AAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTA
AACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAA
CGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCA
AAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAATT
CGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAATCCTG
AGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCTGCAT
GGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
AGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAAGGAGACT
GCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATG
GCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAGG
GTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTAGTC
CGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAAT
CGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACAC
CGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCTAAC
CGCAAGGAGGACGGTCACCCACGGTAGTTC

BGP.A6

GGCGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCAC
CTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTG
TAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCT
ACGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCC
GATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGA
TCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACAC
GGCCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGG
GCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGG
TTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTC
GGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGC
AGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCG
TAAAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGC
TCAACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAG
GGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGG
AATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGC
ACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGC
CCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAG
CTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAA
CTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATT
AATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAAT
CCTGAGGAGACTCGGGAGTGCTCGAAAGAGAACCGATACACAGGTGCT
GCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGC
AACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAACACTCTAAGGA
GACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTC
ATGGCCCTTATGGGTAGGGCTTCCACACGTCATAAAATGGTCGGAAAAA
AAGGGTTGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGT
AGTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTA
GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTA
CACACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGT
CTAACCGCAAGGAGGACGGTCACCCACGGTAGTT

BGP.A7

GGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGAAGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGAT
CAGTAGCTGGTCTGAAAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGC
GAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTT
GTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGG
GGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGCAGGCGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTC
AACCTGGGAACTGCATTGGTGACTGGCAGGCTAGAGTATGGCAGAGGG
GGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAA
TACCGATGGCGAAGGCAGCCCCCTGGGC
CAATACTGACGCTCATGCACGAAAGCGTGGGGAGCAAACAGGATTAGA
TACCCTGGTAGTCCACGCCCTAAACGATGTCAACTAGTTGTTGGGGATT
CATTTCCTTAGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGGGAGT

ACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAG
CGGTGGATGATGTGGATTAATTCGATGCAACGCGAAAAACCTTACCTAC
CCTTGACATGGTCGGAATCCTGAGGAGACTCGGGAGTGCTCGAAAGAG
AACCGATACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTTAGTTGCTAC
GCAAGAGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTG
GGGATGACGTCAAGTCCTCATGGCCCTTATGGGTAGGGCTTCACACGT
CATACAATGGTCGGAACAGAAGGTTGCCAACCCGCGAGGGGGAACTAA
TCCCAGAAAACCGATCGTAGTCCGGATTGCACTCTGCAACTCGAGTGCA
TGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATA
CGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTT
TACCAGAAGTGGCTAGTCTAACCGCAAGGAGGACGGTCACCCACGGTA
GATCAGGT

UPMBG7

CCGGGGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTT
GCACCTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGT
CCTGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACG
ATCTACGGATGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTG
GCCGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGA
CGATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGA
CACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAAT
GGGCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCG
GGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCT
TCGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCA
GCAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGG
CGTAAAGCGTGCGCATGCGGTTTGTTAAGACCGATGTGAAGTCCCCGG
GCTCTCCCTGGGAGGTGCGTTGGTGACTGGCAAGCTAGAGTATGGCAG
AGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGA
GGAATACCGATGGCGAAGGCAGCCGCCTGGGCCAATACTGACGCTCAT
GCATGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC
GCCCTAAACGATGTCAAGTAGTTGTTGGGGATTCATTTCCTTAGTAACGT
AGCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAA
AACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGA
TTAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGA
ACCTTGGAGAGATCTGAGGTTGCTCGAAGGAGAACCGATACACAGGTG
CTGCATGGCTGTCGTCAGCTCGTGTCAGTGAGATGTTGGGTTAAGTCCC
GCAACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAG
GGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGT
CCTCATGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAA
CAGAGGGTCGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGAT
CGTAGTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCG
CTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTT
GTACACACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCT
AGTCTAACCGCAAGGAGGACGTCACCCACGGTAGTTCAGG

UPMBG8

CCGGGGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTT
GCACCTGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGT
CCTGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACG
ATCTACGGATGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTG
GCCGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGA

CGATCAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGA
CACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAAT
GAGCGAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCG
GGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCT
TCGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCA
GCAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGG
CGTAAAGCGTGCGCAGGCGGTTTGTTAAGACCGATGTGAAATCCCCGG
GCTCAACCTGGGAACTGCATTGGTGACTGGCAAGCTAGAGTATGGCAG
AGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGA
GGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCAT
GCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC
GCCCTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGT
AGCTAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAA
AACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGA
TTAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGA
ACCTTGAAGAGATCTGAGGGTGCTCGAAAGAGAACCGATACACAGGTG
CTGCATGGCTGTCGTCACCTCGTGTCAGTGAAAAGTTGGGTTAAGTCCC
GCAACGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAG
GGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGT
CCTCATGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAA
CAGAGGGTCGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGAT
CGTAGTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCG
CTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTT
GTACACACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCT
AGTCTAACCGCAAGGAGGACGTCACCCACGGTAGTTCAGG

UPMBG9

GGGGAGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGATGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGAT
CAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGC
GAAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTT
GTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCGG
GGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGCAGGCGGTTTGTTAAGACCGATGTGAAGTCCCCGGGCTC
AACCTGGGAACTGCATTGGTGACTGGCAAGCTATAGTATGGCAGAGGG
GGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAA
TACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCAC
GAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCC
TAAACGATGTCAAGTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCT
AACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAACT
CAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTAA
TTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAACCT
TGGAGAGATCTGAGGTTGCTCGAAGGAGAACCGATACACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGTCAGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAGGGAG
ACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCA

TGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTCGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTA
GTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGT
AATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACA
CACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCT
AACCGCAAGGAGGACGTCACCCACGGTGGTTCTGG

UPMBG15

GGGGTGCCTTACCCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACC
TGGTGGCGAGTGGCGAACGGGTGAGTAATACATCGGAACATGTCCTGT
AGTGGGGGATAGCCCGGCGAAAGCCGGATTAATACCGCATACGATCTA
CGGATGAAAGCGGGGGACCTTCGGGCCTCGCGCTATAGGGTTGGCCG
ATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGAT
CAGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAAGGGG
CGAAAGCCTGATCCAGCAATGCCGCGTGTGGGAAGAAGGCCTTCGGGT
TGTAAAGCACTTTTGTCCGGAAAGAAATCCTGAGGGCTAATATCCTTCG
GGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCA
GCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGT
AAAGCGTGCGCAGGCGGTTTGTTAAGACCGATGTGAAATCCCCGGGCT
CAACCTGGGAACTGCATTGGTGACTGGCAAGCTAGAGTATGGCAGAGG
GGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGA
ATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCA
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
CTAAACGATGTCAACTAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGC
TAACGCGTGAAGTTGACCGCCTGGGGAGTACGGTCGCAAGATTAAAAC
TCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGATGTGGATTA
ATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGGTCGGAACC
TTGGAAAGATCTGAGGGTGCTCAAAAAAAAACCGATACACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTTGTCCTTAGTTGCTACGCAAGAGCACTCTAGGGAG
ACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCA
TGGCCCTTATGGGTAGGGCTTCACACGTCATACAATGGTCGGAACAGA
GGGTCGCCAACCCGCGAGGGGGAGCTAATCCCAGAAAACCGATCGTA
GTCCGGATTGCACTCTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGT
AATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACA
CACCGCCCGTCACACCATGGGAGTGGGTTTACCAGAAGTGGCTAGTCT
AACCGCAAGGAGGACGTCACCCACGGTAGTTCAGG

UPMBG17

GTCGAACGGCAGCACGGGTGCTTGCACCTGGTGGCGAGTGGCGAACG
GGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGATAGCCCGGCG
AAAGCCGGATTAATACCGCATACGATCTACGGATGAAAGCGGGGGG
ACCTTCGGGCTCTCGCGCTAATAGGGTTGGCCGATGGCTGATTAGCTA
GTCGGTGGGGTAAAGGCCCACCAAGGCGACGATCAGTAGCTGGTCTGA
GAGGACGACCAGCCACACTGGGACTGAGAACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCA
GCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGT
CCGGAAAGAAATCCTGAGGGCTAATATCCTTCGGGGATGACGGTACCG
GAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGT
AGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGC
GGTTTGTTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGAACTGC

ATTGGTGACTGGCAAGCTAGAGTATGGCAGAGGGGGGTAGAATTCCAC
GTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGGCGAAG
GCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAAGCGTGGGGA
GCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAAC
TAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCGTGAAGTT
GACCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGAC
GGGGACCCGCACAAGCGGTGGATGATGTGGATTAATTCGATGCAACGC
GAAAAACCTTACCTACCCTTGACATGGTCGGAACCTTGGAGAGATCTGA
GGGGTGCTCGAAAGAGAACCGATACACAGGTGCTGCATGGCTGTCGTC
AGCTCGTGTCGTGAAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACC
CTTGTCCTTAGTTGCTACGCAAGAGCACTCTAGGGAGACTGCCGGTGA
CAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTATG
GGTAGGGCTTCACACGTCATACAATGGTCGGAACAGAGGGTCGCCAAC
CCGCAAGGGAGAGCTAATCCCAGAAAACCGAATCGTAATTCCGGAATTG
CACTCTGTGCAACTCGAGTGCATGAAGCTGGAATCGCTAGTAATCGCG
GATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCC
CGTCACACCATGGGAGTGGGTTTTACCAGAAGTGGCTAGTCTAACCGC
AAGGAGGACGGTCACCACG

Nucleotide of all 27 isolates using *gyrB* region

BGK.AZ1
CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCT CAACAACGGC
GTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTT
CGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGAC
CAAGAGCGTGCTGCACCCGACCATCTTCCACATCA TCGGCGAGAA
GGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACA
ACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCG
GCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAAC
AAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACG
TCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTG
BGK.AZ2
CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGC
GTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTT
CGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGAC
CAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAA
GGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACA
ACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCG
GCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAAC
AAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACG
TCGGGCGACGACATGCGCGAAGGGCTCTCA TGCGTGCTG
BGK.AZ3
CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGTGC
GCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCGCG

TTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCAAG
AGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACGGC
GTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAGAA
CGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACCCA
CCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTACAT
CGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGGCG
ACGACATGCGCGAAGGGCTCTCATGCGTGCTG

BGK.AZ4

CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGC
GTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTT
CGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGAC
CAAGAGCGTGCTGCACCCGACCATCTTCCACATCA TCGGCGAGAA
GGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACA
ACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCG
GCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAAC
AAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACG
TCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTG

BGK.AZ5

CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGTGC
GCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCGCG
TTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCAGA
GCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACGGCG
TGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAGAAC
GTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACCCAC
CTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTACATC
GTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGGCGA
CGACATGCGCGAAGGGCTCTCATGCGTGCTG

BGK.AZ6

CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGTGC
GCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCGCG
TTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCAAG
AGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACGGC
GTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAGAA
CGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACCCA
CCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTACAT
CGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGGCG
ACGACATGCGCGAAGGGCTCTCATGCGTGCTG

BGK.AZ7

CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGTGC
GCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCGCG
TTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCAAG
AGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACGGC

GTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAGAA
CGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACCCA
CCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTACAT
CGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGGCG
ACGACATGCGCGAAGGGCTCTCATGCGTGCTG

BGK.AZ8

CGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCAT
TTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGACA
TCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGTGC
GCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCGCG
TTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCAAG
AGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACGGC
GTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAGAA
CGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACCCA
CCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTACAT
CGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGGCG
ACGACATGCGCGAAGGGCTCTCA TGCGTGCTG

BGS.AS1

GAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCAC
CGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATAC
CACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACA
ACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGAC
GACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAAC
AAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAG
AAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTA
CAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGG
CGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCA
ACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAA
CGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTCGGTG
AAGG

BGS.AS2

GCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGC
ATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGA
CATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGT
GCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCG
CGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCA
AGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACG
GCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAG
AACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACC
CACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTA
CATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGG
CGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTCGGTGAAG

BGS.AS3

CCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGAAGTGCA
TTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCACTACGAC
ATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAACGGCGTG
CGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGACTTCGC
GTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAAGACCAA
GAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAAGGACGG
CGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACAACGAGA

ACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCGGCACCC
ACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAACAAGTACA
TCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACGTCGGGC
GACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGS.AS4

GTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCACCGA
AGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATACCAC
TACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAACAAC
GGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAGACGA
CTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCAACAA
GACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCGAGAA
GGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGCTACA
ACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGACGGCG
GCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGATCAAC
AAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGAAACG
TCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGS.AS5

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTCGGT
GAAG

BGS.AS6

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTCGGT
GAAG

BGS.AS7

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGATTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG

AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
GACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGP.A1

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGP.A2

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGP.A3

GTGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGC
ACCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAAT
ACCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGATTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
GACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGP.A4

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC

GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTCGGT
GAAGG

BGP.A5

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTCGGT
GAAGG

BGP.A6

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

BGP.A7

TGAAGTGTCGCCGATGGAGGTCGTCGGCGAGACCGAGAACCGCGGCA
CCGAAGTGCATTTCATGGCCGATCCGACCATCTTCGGCACCGTCGAATA
CCACTACGACATCCTCGCCAAGCGGATTCGCGAGCTGTCGTTCCTCAA
CAACGGCGTGCGCATCCGCCTGACCGACCAGCGCAGCGGCAAGGAAG
ACGACTTCGCGTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCACCCGACCATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGGGTGGAAGTGGCGATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAACATCCCGCAGCGCGAC
GGCGGCACCCACCTGACCGGCCTGCGCGCGGCGATGACGCGCGTGAT
CAACAAGTACATCGTCGACAACGAGATCGCCAAGAAGGCCAAGGTGGA
AACGTCGGGCGACGACATGCGCGAAGGGCTCTCATGCGTGCTGTC

UPMBG7

CTGCGCCTGGTGGTGAAGCGCAACGGCAAGAAGCACTTCATGGAGTTC
CATCGCGGCGTGGCGCAGGACCGCGTGCTGGAAACGCGCGACGGCGT
GGAAGTGTCGCCGATGGCAGTCACCGGCGACACCGAGAACCGCGGCA
CCGAAGTGCACTTCATGGCCGATCCGACTATCTTCGGCACGGTCGAGT
ATCACTACGACATCCTGGCCAAGCGGATTCGCGAACTGTCGTTCCTGAA
TAACGGCGTGCGAATCCGCCTGACCGACCAGCGCTCGGGCAAGGAAG
ACGATTTCGCCTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCATCCGACGATCTTCCACATCATCGGCG

AGAAGGATGGCGTGGGCGTGGAAGTGGCCATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAATATTCCGCAGCGGGACG
GA

UPMBG8

CTGCGCCTGGTGGTGAAGCGCAACGGCAAGAAGCACTTCATGGAGTTC
CATCGCGGCGTGGCGCAGGACCGCGTGCTGGAAACGCGCGACGGCGT
GGAAGTGTCGCCGATGGCAGTCACCGGCGACACCGAGAACCGCGGCA
CCGAAGTGCACTTCATGGCCGATCCGACTATCTTCGGCACGGTCGAGT
ATCACTACGACATCCTGGCCAAGCGGATTCGCGAACTGTCGTTCCTGAA
TAACGGCGTGCGAATCCGCCTGACCGACCAGCGCTCGGGCAAGGAAG
ACGATTTCGCCTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCATCCGACGATCTTCCACATCATCGGCG
AGAAGGATGGCGTGGGCGTGGAAGTGGCCATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAATATTCCGCAGCGGGACG
G

UPMBG9

CGTCCCGCTGCGGAATATTGTTGGTGAAACACAGCACGTTCTCGTTGTA
GCTGTCGTTCCACTGCATGGCCACTTCCACGCCCACGCCATCCTTCTCG
CCGATGATGTGGAAGATCGTCGGATGCAGCACGGTCTTGGTCTTGTTG
ATGTACTCGACGAAGCCCTTCACGCCGCCCACGAAGGCGAAATCGTCT
TCCTTGCCCGAGCGCTGGTCGGTCAGGCGGATCCGCACGCCGTTATTC
AGGAACGACAGTTCGCGAATCCGCTTGGCCAGGATGTCGTAGTGATAC
TCGACCGTGCCGAAGATGGTCGGATCGGCCATGAAGTGCACTTCGGTG
CCGCGGTTCTCGGTGTCGCCGGTGACTGCCATCGGCGACACTTCCACG
CCGTCGCGCGTCTCCAGCACGCGGTCCTGCGCCA

UPMBG15

CGTCCCGCTGCGGAATATTGTTGGTGAAACACAGCACGTTCTCGTTGTA
GCTGTCGTTCCACTGCATGGCCACTTCCACGCCCACGCCATCCTTCTCG
CCGATGATGTGGAAGATCGTCGGATGCAGCACGGTCTTGGTCTTGTTG
ATGTACTCGACGAAGCCCTTCACGCCGCCCACGAAGGCGAAATCGTCT
TCCTTGCCCGAGCGCTGGTCGGTCAGGCGGATCCGCACGCCGTTATTC
AGGAACGACAGTTCGCGAATCCGCTTGGCCAGGATGTCGTAGTGATAC
TCGACCGTGCCGAAGATGGTCGGATCGGCCATGAAGTGCACTTCGGTG
CCGCGGTTCTCGGTGTCGCCGGTGACTGCCATCGGCGACACTTCCACG
CCGTCGCGCGTCTCCAGCACGCGGTCCTGCGCCA

UPMBG17

CTGCGCCTGGTGGTGAAGCGCAACGGCAAGAAGCACTTCATGGAAGTT
CATCGCGGCGTGGCGCAGGACCGCGTGCTGGAGACGCGCGACGGCGT
GGAAGTGTCGCCGATGGCGATCACCGGCGACACCGAGAACCGCGGCA
CCGAAGTGCACTTCATGGCCGATCCGACCATCTTCGGCACGGTCGAGT
ATCACTACGACATCCTGGCCAAGCGGATTCGCGAACTGTCGTTCCTGAA
TAACGGCGTGCGGATCCGCCTGACCGACCAGCGCTCGGGCAAGGAAG
ACGATTTCGCCTTCGTGGGCGGCGTGAAGGGCTTCGTCGAGTACATCA
ACAAGACCAAGAGCGTGCTGCATCCGACGATCTTCCACATCATCGGCG
AGAAGGACGGCGTGGGCGTGGAAGTGGCCATGCAGTGGAACGACAGC
TACAACGAGAACGTGCTGTGTTTCACCAACAATATTCCGCAGCGGGACG
G

Nucleotide Sequence for 22 strain of *B. glumae* using *toxB* region

BGK.AZ1

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTCGGTTATCTG

BGK.AZ2

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTTGGTTATCTG

BGK.AZ3

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTCGGTTATCTG

BGK.AZ4

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT

CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTCGGTTATCTG

BGK.AZ5

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTCGGTTATCTG

BGK.AZ6

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTCGGTTATCTT

BGK.AZ7

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTCGGTTATCTG

BGK.AZ8

GAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCCGAACATATCGCG
CTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGTGCGCCTGCATTC
CGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGCGCTGCGACTGCG
GCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGCGAGCGCGGCGGC
ATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCATCGGGCTCTACAAC
AAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGGACTCGACACCTTC
GCTGCCAATCGCGCGCTCAACTTCCCCGACGACCTGCGCGACTTCCGC
GTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCGGCGAGGTCTCGCT
GGTCACCAACAACCCCGACAAGACCGCGCAACTGACCCGTCACGGCAT
CAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTCGCCAACAGCGTCAA
CTTTGGTTATCTG

BGS.AS1

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGS.AS2

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGS.AS3

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGS.AS4

GGCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGC
CGAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGS.AS5

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGATTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAAGGC

BGS.AS6

GGCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGC
CGAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAAGGC

BGS.AS7

GGCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGC
CGAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGP.A1

GGCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGC
CGAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGP.A2

GGCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGC
CGAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT

GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGGTCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGP.A3

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGP.A4

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAAGG

BGP.A5

GGCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGC
CGAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAAGG

BGP.A6

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC

GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BGP.A7

GCATTTGAAACCGAGATGGTGACCTTCAACGGTCTTTGCGACAGCGCC
GAACATATCGCGCTGGTGTTCGGCAAGATCGGCGAGGCGCCGCTGGT
GCGCCTGCATTCCGAGTGCCTGACGGGCGACGTGTTCGGCTCCGCGC
GCTGCGACTGCGGCGAGCAGCTCGACGAGTCGATCGCGCTGTTCGGC
GAGCGCGGCGGCATCCTGCTCTACCTGCGCCAGGAAGGGCGCGGCAT
CGGGCTCTACAACAAGCTCGACGCCTATCGCCTGCAGATCTCGCAGGG
ACTCGACACCTTCGCTGCCAATCGCGCGCTCAACTTCCCCGACGACCT
GCGCGACTTCCGCGTGGCGGCCCAGATGCTCAAGGCGCTCGGCGTCG
GCGAGGTCTCGCTGGTCACCAACAACCCCGACAAGACCGCGCAACTGA
CCCGTCACGGCATCAAGGTGCGCCAGGTGAAGTCGACCGGCGTGTTC
GCCAACAGCGTCAACTTTGGTTATCTGCATGCGAA

BIODATA OF STUDENT

Adam Zafdri Md Zali was born on May 30th, 1998 in Skudai, Johor. He completed her formal education with Sijil Pelajaran Malaysia in Sekolah Menengah Sains Sembrong, Johor. In 2016, he pursued his study in the Science Matriculation program at Kolej Matrikulasi Negeri Sembilan. A year later, he continued her undergraduate degree in Bachelor of Science with honours (Biologi) at Universiti Teknologi MARA (UiTM) and graduated in July 2020.

Since March 2020, he has enrolled as a full-time candidate in the Department of Plant Protection, Faculty of Agriculture, Universiti Putra Malaysia, to pursuing a Master of Science (Plant Pathology) under the supervision of Dr Dzarifah Mohamed Zulperi from March 2020 until the present.

LIST OF PUBLICATIONS

Journal

Md-Zali, A. Z., Ja'afar, Y., Paramisparan, K., Ismail, S. I., Saad, N., Hata, E. M., ... & Yusof, M. T. (2023). First report of *Burkholderia gladioli* causing bacterial panicle blight of rice in Malaysia. *Plant Disease*, *107*(2), 551.

Poster and Conference

1. First LRGS Climate Ready Rice Colloquium (2022): Characterization and Genomic analysis of *Burkholderia glumae* causing Panicle Blight (BPB) Disease in Malaysia.

2. 11th International Conference on Plant Protection in the Tropics (2022): Molecular Identification of *Burkholderia glumae* from Infected Rice Samples in Malaysia.

Award

1. UPM Graduate Research Funding for Post-Graduate student.

Genbank

1. Deposited sequences:

2. 16S rRNA gene (Genbank accesion nos. OK632514-OK632516, OK631741, OL347392-OK34795, ON062124-ON062137, OM869953-OM869957)

3. *GyrB* specific gene (Genbank accesion nos. OL461777-OL461784, ON086704-ON86717, OM824438-OM24442)

4. *ToxB* specific gene (Genbank accesion nos. ON075583-ON075590, ON086719-ON86732)

5. Draft genome sequence (Genbank accesion nos. JAMYCS000000000, JAMYCT000000000, JANIEE000000000, JANIEF000000000)