GENETIC DIVERSITY OF THE BEETLE, *PAEDERUS FUSCIPES* CURTIS (COLEOPTERA: STAPHYLINIDAE) AND ITS SYMBIOTIC DIVERSITY WITH PEDERIN POLYMORPHISM BASED ON 16S RRNA METAGENOMICS PROFILING

TEOH MIAO CHING

UNIVERSITI SAINS MALAYSIA

2023

GENETIC DIVERSITY OF THE BEETLE, *PAEDERUS FUSCIPES* CURTIS (COLEOPTERA: STAPHYLINIDAE) AND ITS SYMBIOTIC DIVERSITY WITH PEDERIN POLYMORPHISM BASED ON 16S RRNA METAGENOMICS PROFILING

by

TEOH MIAO CHING

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

February 2023

ACKNOWLEDGEMENT

I would like to express my sincere thanks to my supervisor, Dr. Veera Singham for his valuable guidance, assistance and advice throughout this research. I would also like to thank him for helping to get the *P. fuscipes* beetles from other states.

I would like to thank Dr. Go for his guidance, assistance, and advice on the bacterial isolation part. I would also like to thank Prof. Alex and all the staff in Centre of Chemical Biology for providing their assistance throughout this project. A very big thank you to my lab mates in CCB especially Li Shen, Jeremy, Melissa, Diyana, Ka Kei, and Dr. Yasmin for their assistance and support.

I would also like to thank my family and friends for their patience and understandings throughout this study. Their endless support and trust have given me strength to complete this project.

Lastly, I am very grateful to my mother and Li Shen for accompanying me to these locations and in helping me with sampling of *P. fuscipes* beetles. Also, I would like to express my gratitude to the officers from the Seberang Perai Municipal Council and Mr. Patrick Ang for providing the information on the locations infested by the beetles in Penang.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF PLATES

Page

LIST OF ABBREVIATIONS

ACP	Acylcarrier protein
AI	Ayer Itam
AFLP	Amplified fragment length polymorphisms
BLAST	Basic Local Alignment Search Tool
BM	Bukit Mertajam
BP	Bandar Perda
COI	Cytochrome C oxidase I
ddNTPs	Dideoxynucleotides triphosphates
DNA	Deoxyribonucleic acid
FISH	Fluorescent *in situ* hybridization
gDNA	Genomic deoxyribonucleic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
KS	Ketosynthase
KT	Kuala Terengganu, Terengganu
MEGA	Molecular Evolutionary Genetics Analysis
MT	Methyltransferase
mtDNA	Mitochondrial deoxyribonucleic acid
MPCGs	Mitochondrial protein-coding genes
NGS	Next generation sequencing
NSTI	Nearest sequenced taxon index
OTU	Operational taxonomic units
OXY	Oxygenase
PCR	Polymerase chain reaction
PKS/NRPS	Polyketide synthase/nonribosomal peptide synthase

PP	Permatang Pauh
PPB	Pederin-producing bacteria
PIA	*Pseudomonas* isolation agar
PICRUSt	Phylogenetic investigation of communities by reconstruction of unobserved states
RAPD	Random amplified polymorphic DNAs
rpm	Revolutions per minute
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
S. D.	Standard deviation
SNPs	Single-nucleotide polymorphisms
STAMP	Statistical Analysis of Metagenomic Profiles
TAE	Tris-acetate-EDTA
TT	Tanjung Tokong

LIST OF APPENDICES

DIVERSITI GENETIK KUMBANG *PAEDERUS FUSCIPES* CURTIS (COLEOPTERA: STAPHYLINIDAE) DAN DIVERSITI SIMBIOTIK DENGAN POLIMORFISME PEDERIN BERDASARKAN PEMPROFILAN METAGENOMIK 16S RRNA

ABSTRAK

Kumbang *Paederus* adalah pembawa toxin pederin yang menyebabkan linear dermatitis apabila kulit tersentuh dengan toxin ini. Toxin ini dihasilkan oleh bakteria *Pseudomonas* sp. yang masih belum dikulturkan dan kumbang *Paederus* betina adalah pembawa bakteria ini. Polimorfisme pederin telah dilaporkankan di beberapa kumbang spesies *Paederus* dan *Paederidus*, di mana kumbang betina (+) mempunyai toxin pederin manakala betina (−) kekurangan toxin ini. Kumbang betina (−) telah dilaporkan dijangkiti lebih banyak spesies bakteria daripada betina (+), tetapi profil mikrob untuk kedua-dua golongan kumbang betina ini tidak dilaporkan. Selain itu, perkaitan antara polimorfisme pederin dengan profil mikrob dan genetik kumbang masih tidak diketahui. Oleh itu, objektif tesis ini adalah untuk menyiasat profil mikrob kedua-dua golongan kumbang betina *P. fuscipes*, dan untuk mengkaji sitokrom c oksidase I daripada mitokondria kumbang berdasarkan polimorfisme pederin. Gen yang terlibat dalam penghasilan pederin (*ped*) akan dikesan menggunakan teknik tindak balas rantai polymerase (PCR). Kumbang betina juga akan tertakluk kepada penjujukan amplikon 16S rRNA. Sitokrom c oksidase I daripada mitokondria dikaji untuk memahami kepelbagaian genetik kumbang *P. fuscipes* dan perkaitannya dengan polimorfisme pederin. Kajian ini juga tertumpu pada percubaan untuk mengkulturkan dan mengasingkan bakteria yang menghasilkan toxin pederin. Kajian polimorfisme pederin menunjukkan bahawa sebanyak 87% (secara purata) kumbang betina *P.*

fuscipes adalah pembawa bakteria yang menghasilkan pederin, manakala semua kumbang jantan *P. fuscipes* bukan pembawa bakteria ini. Penjujukan amplikon 16S rRNA menunjukkan betina (+) *P. fuscipes* didominasi oleh bakteria *Pseudomonas* (74.72%) daripada filum Proteobakteria dan betina (−) pula didominasi oleh bakteria *Apibacter* daripada filum Bacteroidetes (48.76%). Kajian pada sitokrom c oksidase I menunjukkan bahawa kumbang *P. fuscipes* mempunyai kepelbagaian genetik yang tinggi. Analisis ini tleah menunjukkan kemungkinan bahawa polimorfisme pederin tidak berkaitan dengan kepelbagaian genetik yang tinggi dalam kumbang *P. fuscipes*. Walau bagaimanapun, gen penanda yang lain seperti gen mikrosatelit atau SNPs perlu disiasatkan untuk mengaitkan perhubungan antara polimorfisme pederin dengan genotip kumbang. Secara keseluruhan, kumbang betina (+) *P. fuscipes* adalah selalu dominan dalam populasi semula jadi, dengan gen *ped* dikesan di sekurang-kurangnya 80% kumbang betina (+) dari semua lokasi. Selain daripada itu, analisis penjujukan amplikon 16S rRNA berdasarkan polimorfisme pederin telah menunjukkan kepelbagain profil mikrob antara kumbang betina (+) dan betina (−). Spesies bakteria yang dominan dalam kedua-dua golongan betina ini juga berbeza. Walaupun gen *ped* telah dikesan dalam kultur-kultur media, bakteria yang menghasilkan toxin pederin tidak berjaya dikulturkan. Selain daripada mengoptimumkan cara mengkulturkan bakteria, strategi yang berlainan seperti penggunaan media axenic dan teknik kultur sel boleh diterokai untuk pengkulturan bakteria ini. Analisis mitokondria menunjukkan bahawa kumbang *P. fuscipes* mempunyai kepelbagaian genetic yang tinggi, tetapi gen penanda yang lain juga harus diselidik untuk memahami perkaitan antara polimorfisme pederin dengan genotip kumbang.

GENETIC DIVERSITY OF THE BEETLE, *PAEDERUS FUSCIPES* CURTIS (COLEOPTERA: STAPHYLINIDAE) AND ITS SYMBIOTIC DIVERSITY WITH PEDERIN POLYMORPHISM BASED ON 16S RRNA METAGENOMICS PROFILING

ABSTRACT

Paederus rove beetles are known to carry the pederin toxin that causes linear dermatitis when crushed on skin. The toxin is produced by an unculturable *Pseudomonas* sp. harboured by the female *Paederus* beetles. Pederin polymorphism has been observed several *Paederus* and *Paederidus* beetle species wherein the (+) female beetles carried the toxin pederin and (−) females lacking this toxic amide. It has been stated that the (−) female *Paederus* beetles are infected with more bacterial species than (+) females, but the microbial profiles between the two female groups were not revealed. Besides, the association of pederin polymorphism with the host beetle's genetic is yet to be investigated. Herein, the objectives of this dissertation are to investigate the microbial profile of the two groups of *P. fuscipes* beetles, and to study the host beetle's mtDNA COI in association with pederin polymorphism. Pederin polymorphism studies on *P. fuscipes* beetles were carried out by PCR detection of pederin (*ped*) biosynthesis genes. The female beetles were then subjected to 16S rRNA amplicon sequencing and mtDNA COI analysis. This study also attempts to culture and isolate the pederin-producing bacteria. The results showed that an average of 87% of female *P. fuscipes* beetles were infected with the pederin-producing bacteria while *ped* genes were not present in all male beetles. The 16S rRNA amplicon sequencing revealed that *Pseudomonas* (74.72%) from the phyla Proteobacteria was dominant in (+) female *P. fuscipes* beetles while (−) female beetles was dominated by

Apibacter (48.76%) from the phyla Bacteroidetes. The investigation on the mtDNA COI region indicated that the *P. fuscipes* beetles were genetically diverse. This analysis showed that pederin polymorphism is unlikely to be associated with the highly diverse mtDNA COI in *P. fuscipes* beetles, as there is no unique haplotypes segregation observed among the sequences. However, further investigations on other marker genes are needed to verify the association of pederin polymorphism with the host genotype. Overall, (+) female *P. fuscipes* beetles are always dominant in natural populations, with *ped* genes detected in more than 80% of females from all locations. Furthermore, the 16S metagenomic sequencing based on pederin polymorphism revealed the differences in the microbial profile and the dominant bacterial groups between the two groups of female *P. fuscipes* beetles. Despite *ped* genes have been detected in agar plates and broth cultures, single colonies of pederin-producing bacteria have not been successfully isolated. Besides optimizing the culture conditions, other strategies such as the use of axenic media and cell culture techniques can be explored to isolate the pederin-producing bacteria. The analysis on partial mtDNA revealed that *P. fuscipes* beetles are genetically diverse, but further investigations on other marker genes, such as 16S and microsatellites are needed to understand the association of pederin polymorphism with host beetle's genotype.

CHAPTER 1

INTRODUCTION

1.1 Rational of research

Beetles from the Staphylinidae family or most commonly known as rove beetles, have more than 63,657 species described globally (Irmler *et al*., 2018; Frank & Thomas, 2019), which make them the largest beetles' family recorded. Rove beetles are known to be beneficial to the agricultural sector as potential biological control because they feed on other insects. Nevertheless, *Paederus* rove beetles from the Paederinae subfamily have gained medical attentions worldwide. The presence of pederin toxin ($C_{25}H_{45}O_9N$) in these beetles' hemolymph can caused dermatitis linearis or kissing lesions on human skin (Lawrence & Ślipiński, 2013; Beaulieu & Irish, 2016; Krinsky, 2019). These rove beetles are usually found distributed worldwide in moist environments, such as irrigated farmland and edges of freshwater lakes (Manley, 1977; Frank & Kanamitsu, 1987; Bong *et al*., 2012; Beaulieu & Irish, 2016). However, *Paederus* beetles are nocturnal insects with wings, and they are attracted to bright lights from high rise buildings, where infestations occur (Frank & Kanamitsu, 1987; Maryam *et al*., 2016). Besides, Banney *et al*. (2000) stated that climate might play a role in the invasion of *Paederus* rove beetles in residential areas. Nevertheless, Bong *et al*. (2013) reported that climate has a lesser impact on the dispersal of these beetles compared to human activities, such as burning straw, harvesting and cultivation of rice had prompted *Paederus* rove beetles to flee from paddy fields.

Pederin polymorphism in *Paederus riparius* and *Paederus fuscipes* was first reported by Kellner and Dettner (1995). Female beetles that contained pederin are labelled as (+) females, while (–) females indicate that the female beetles are not able to accumulate pederin in their bodies. Kellner and Dettner (1995) also reported that all

male beetles are not able to accumulate pederin. The pederin toxin is passed on to the offspring by (+) females, and the toxin helps to prevent their offspring from spiders' predation (Kellner & Dettner, 1996). At first, pederin was thought to be produced by *Paederus* beetles because the compound can be found throughout the body of the beetle, but further studies have proven that bacteria closely related to *Pseudomonas aeruginosa* are responsible in pederin synthesis (Kellner, 2001; Kellner, 2002b; Piel, 2002). Although the structure of pederin (Cardani *et al*., 1965; Matsumoto *et al*., 1968) and the synthesis pathway (Jewett & Rawal, 2007) had been studied and proposed, the pederin-producing *Pseudomonas* spp. is not well researched because the bacteria has yet to be successfully cultured. Besides, Kellner (2002b) proposed that there are differences in the microbial diversity between (+) and (–) females but lacked the data to support the proposal. Moreover, the population genetics of *Paederus fuscipes* beetles has not been studied, and whether the presence of the pederin-producing bacteria has any effects on the beetle host's genetic or vice versa is also not known.

Though many researchers have been studying *Paederus* rove beetles, not many reported on the prevalence of (+) and (–) females especially in the tropical regions. Besides, the association of pederin polymorphism with the host beetles' microbial diversity as well as the host genetics are still not known. The 16S amplicon sequencing was used to reveal the differences in microbial profile of female *P. fuscipes* beetles in association with the presence or absence of pederin-producing *Pseudomonas*. Moreover, the pederin-producing *Pseudomonas* spp. has yet to be cultured and characterized. Understanding pederin polymorphism at host genetic and microbial level can provide insights into the cost of harbouring the pederin-producing symbiont, how pederin polymorphism affect *P. fuscipes* beetles at host genetic and population level, and may lead to the development of a better control strategy.

1.2 Objectives

The objectives of this study are as follows:

1. To determine the prevalence of pederin polymorphism in natural populations of *P. fuscipes* in tropical country.
2. To investigate the difference in microbial diversity associated with pederin polymorphism between (+) and (–) female *P. fuscipes* using 16S amplicon metagenomics analysis.
3. To investigate the association of host genetic and pederin polymorphism based on mitochondrial DNA diversity in *P. fuscipes*.
4. To culture and isolate pederin-producing *Pseudomonas* spp. from (+) female *P. fuscipes*.

CHAPTER 2

LITERATURE REVIEW

2.1 Rove beetles

Rove beetles are classified under Polyphaga, the largest suborder of order Coleoptera (Krinsky, 2019). 90% of beetle families, including Staphylinidae (commonly known as rove beetles), are classified under this suborder and these polyphagous beetles have various types of feeding habits, such as predation on other insects, feeding on fungi (mycophagous), dead and decaying insects or organic matters as a food source (saprophagous), as well as feeding on plants and algae (phytophagous) (Thayer, 2016). Rove beetles from the subfamily Paederinae are known to contain pederin, a potent vesicant amide that caused *Paederus* or linear dermatitis upon exposure on human skin (Thayer, 2016; Krinsky, 2019). Rove beetles from this subfamily haven been gaining attentions as they are often found in moist environments, such as irrigated crop fields, and the people working at such areas are often affected (Thayer, 2016; Krinsky, 2019). Other than crop fields, Paederinae rove beetles are found distributed at lowlands and mid-elevation of tropical latitudes, northern and southern temperate geographical zones with moist environments, such as leaves debris on the ground, grasslands, freshwater edges and paddy fields (Kellner & Dettner, 1996; Bong *et al*., 2012; Thayer, 2016; Krinsky, 2019). Besides that, Thayer (2016) also pointed out that rove beetles from subfamily Paederinae might be found in nests or burrows of ants, termites, wasps and mammals. Although staphylinids are widely distributed, *Paederus* beetles from the subfamily Paederinae (refer to Table 2.1 for complete taxonomy classification) can be found across Asia, Europe, Africa, Australasia, and America (North, South and Central), but not in Antarctica (Frank & Kanamitsu, 1987; Krinsky, 2019; Thayer, 2016).

Table 2.1 Taxonomy classification of Paederus rove beetles (Karthikeyan & Kumar, 2017)

Kingdom	Animalia
Phylum	Arthropoda
Subphylum	Hexapoda
Class	Insecta
Order	Coleoptera
Suborder	Polyphaga
Superfamily	Staphylinoidea
Family	Staphylinidae
Subfamily	Paederinae
Tribe	Paederini
Subtribe	Paederina
Genus	*Paederus* Fabricius, 1775

2.1.1 *Paederus* rove beetles

Paederus beetles are found in moist environments, such as irrigated farmlands, marshes, plantations and edges of freshwater lakes (Manley, 1977; Bong *et al*., 2012). Although *Paederus* beetles can be active during daylight, they are known to be more active at night and are attracted to lights (Kurosa, 1958; Kellner & Dettner, 1996; Silva *et al*., 2015; Maryam *et al*., 2016). In Malaysia, they are known to have flown to high-rise buildings near paddy fields during harvesting season (Bong *et al*., 2012). Besides harvesting season, Banney *et al*. (2000) mentioned that the weather also plays a role in the dispersal of these beetles. The flight activity of *Paederus* beetles always peaked during the rainy season, and during the hot and humid weather during spring (Banney *et al*., 2000; Nikbakhtzadeh & Tirgari, 2008; Nasir *et al*., 2015). Nikhita *et al*. (2014) reported that the highest number of *P. fuscipes* beetles recorded is usually during the summer months (February to July) in Tamilnadu, India.

2.1.2 Adult *P. fuscipes* morphology and life cycle

Paederus fuscipes beetles, commonly known as "Charlie ants" in Malaysia, are often confused as ants by the locals because of the beetles' physical features. *P. fuscipes* beetles are usually 7 to 10 mm long and 0.5 to 1 mm wide (Kurosa, 1958; Krinsky, 2019). In comparison to the dull appearance of most of the rove beetles, *P. fuscipes* beetles have orange thorax and the 1st to 4th segments of the abdomen are also orange in colour (Plate 2.1) (Thayer, 2016; Krinsky, 2019). Besides that, the legs and antenna of these beetles are black and orange in colour. The segments near to the head are orange in colour, while the rest of the antenna is black. Then, black colouration can be seen on the joints on the legs. In addition, *P. fuscipes* have wings that are covered by metallic blue elytra, whereas the head and 5th to 6th abdominal segments are black. Male and female beetles can be differentiated by observing the presence of median excision in the

sixth sternite (abdomen) under a microscope, where males will have the excision, but it is absent in the females (Plate 2.2).

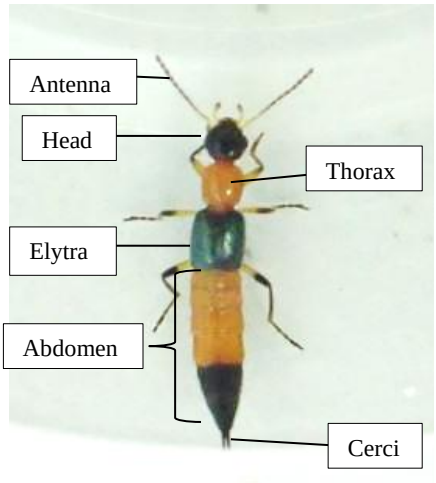

Plate 2.1 ***Paederus fuscipes*** **(dorsal view)**

Plate 2.2 Ventral view on the sixth sternite of male (left) and female (right) *P. fuscipes*

P. fuscipes beetles undergo complete metamorphosis with 2 larvae instars in their life cycle (Kurosa, 1958; Kellner, 1998; Bong *et al*., 2012). *Paederus* females lay eggs singly (Kellner & Dettner, 1996; Bong *et al*., 2012; Nikhita *et al*., 2014), where the females can lay up to almost 150 eggs in their lifetime (Bong *et al*., 2012). The duration of immature stages of *P. fuscipes* is around 16-18 days (Devi *et al*., 2002; Bong *et al*., 2012). According to Bong *et al*. (2012), the eggs will incubate for an average of 4-6 days before they hatch. The 1st instars (L1) will then take 3-5 days to emerge into 2nd instars (L2). Then, they will take 5-7 days for the 2nd instars to grow into pupae. Lastly, the pupae will use 3-4 days until they fully enclosed into adults. Devi *et al*. (2002) documented that the longevity for males and females were 54 days and 72 days, respectively, while Bong *et al*. (2012) reported that the beetles' mean longevity was 43-48 days regardless of gender. Manley (1977) mentioned that *P. fuscipes* beetles can live for up to 3 months, while Kellner and Dettner (1995) had documented that *P. fuscipes* can live for more than 150 days. The differences in the reported longevity of *P. fuscipes* beetles are likely to be influenced by the diet provided in each laboratory. Bong *et al*. (2014) reported that *P. fuscipes* beetles provided with high carbohydrate diet can live up to 70 days, while beetles fed with protein-rich and lipid-rich diets live up to 40 days and 12 days, respectively.

The eggs of *P. fuscipes* beetles are spherical and yellowish-white in colour, with a diameter of 0.5-1.0 mm (Plate 2.3). L1 have 10 segments of the abdomen and are approximately 2.0 mm long. On the 9th abdominal segment, a pair of cerci can be observed, with a shorter 10th segment in between them. L2 are around 5.0 mm with well-developed legs and thorax. Pupa of *P. fuscipes* are approximately 4.0 mm in size, with a white to yellowish appearance. Unlike butterflies and flies' pupa, *P. fuscipes*' pupa are unsclerotized and exarate. The pupa is not encapsulated in a cocoon and can

have limited movements. Besides the changes of colour and appearance, after the adults emerge, the abdominal segments are reduced from 10 to just 6 segments, with the cerci attached on the 6^{th} abdominal segments (Plate 2.1; Plate 2.4).

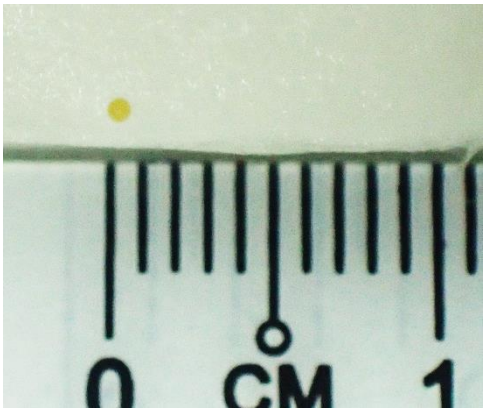

Plate 2.3 *Paederus fuscipes* beetle's egg

Plate 2.4 Different developmental stages of *Paederus fuscipes* (Nikbakhtzadeh & Tirgari, 2007)

2.1.3 Agricultural importance of *Paederus* rove beetles

The predaceous feeding habit of *Paederus* rove beetles is known to be beneficial to the agricultural sector as a potential biological control agent. As mentioned in the previous section, *Paederus* beetles are habitually found in paddy fields, as well as maize (corn), cotton and other vegetations (Nikbakhtzadeh & Tirgari, 2008; Rana *et al*., 2013; Thayer, 2016; Krinsky, 2019). Paddy, also known as rice, is the staple food throughout Asia, and the productions of rice are often affected by agricultural pests. Examples of a few common pests found in paddy fields are *Cnaphalocrosis medinalis* (rice leaf roller), *Marasmia patnalis* (rice leaf folders), *Nilaparvata lugens* (brown planthopper), *Scirpophaga incertulus* (rice yellow stem borer) and *Sogatela furcifera* (white-backed planthopper) (de Kraker *et al*., 2000; Ul Ane & Hussain, 2015). Manley (1977) reported that the adult *Paederus fuscipes* and their larvae actively preyed and fed on leafhoppers, aphids, small Orthopteran and Dipteran, except for the larvae of *Tryporyza incertulas* (stem borer). Jadhao and Shukla (2016) studied the population dynamics of *P. fuscipes* in paddy fields also showed that *P. fuscipes* beetles are one of the dominant predators in the rice ecosystem in Bhandara, India. Besides preying and feeding on rice pests, *Paederus* beetles (adults and larvae), also known as generalist predators, can feed on *Drosophila melanogaster* (fruit flies), cockroaches (lobster, German, American, etc.), *Aedes aegypti* (mosquitoes), *Tenebrio molito*r (mealworm) larva, *Calliphora* (blowflies) pupa, as well as decayed vegetables when reared in the lab (Kellner, 1998; Bong *et al*., 2012; Nasir & Akram, 2012).

2.1.4 Medical importance of *Paederus* beetles

Paederus beetles first gained concerns in medical field because they were identified as the causative agent of linear dermatitis (Frank & Kanamitsu, 1987; Nikbakhtzadeh & Tirgari, 2008; Nasir *et al*., 2015; Srihari *et al*., 2017). Linear

dermatitis is also known as *Paederus* dermatitis, dermatitis linearis, spider lick and whiplash dermatitis (Beaulieu & Irish, 2016; Krinsky, 2019). According to Somerset (1961), the first published case on *Paederus* dermatitis was described by Vorderman in 1901 where an outbreak of dermatitis lesions caused by semoet-kalong occur in Jawa. The causative agent was later identified as *Paederus* rove beetles (Somerset, 1961). However, according to Krinsky (2019), this type of skin lesion had been documented in China at around AD 739 and this skin reaction was known as Ch'ing yao ch'ung. The common names of *Paederus* beetles vary in different countries and they were listed in Table 2.2. Up until now, there have been at least 600 species were described under the genus *Paederus,* with approximately 4% of the described species are known to have caused dermatitis blisters on human (Frank & Kanamitsu, 1987; Krinsky, 2019). The occurrences of *Paederus* rove beetles and linear dermatitis have been reported in Russia, India, Australia, China, Italy, countries in Africa and South America (Frank & Kanamitsu, 1987; Krinsky 2019). The *Paederus* species documented in these regions include *P. melampus, P. riparius, P. sabaeus, P. ilsae,* and *P. fuscipes,* with *P. fuscipes* was being the most reported species in Italy, Russia, and countries across Asia (Frank & Kanamitsu, 1987; Krinsky 2019). Besides *Paederus* rove beetles, beetles from the genus *Paederidus* can also cause linear dermatitis (Kellner, 2002a; Piel, 2002).

Table 2.2 Common names of *Paederus* rove beetles by the locals from different countries (Banney *et al*., 2000; Karthikeyan & Kumar, 2017; Krinsky, 2019)

Location	Common name
Australia	Whiplash rove beetle
Brazil	Poto
Egypt	El-Rawagha / the escaper
Indonesia	Semoet-kalong
Iran	Balaloos / onion fly / dracula
Iraq	Phosphorus insect
Malaysia	Semut semai / semut kayap / Charlie ants
Venezuela	Puri-puri / tar-tari
Africa	Nairobi fly / champion fly

Dermatitis linear will only occur when the beetle is accidentally crushed on the skin because the pederin toxin is found in the hemolymph of *Paederus* rove beetles. The symptoms of *Paederus* dermatitis are often confused with blister beetle dermatitis, acute allergic, liquid burns, fungal infections and similar dermatitis disorder (Nasir *et al*., 2015; Beaulieu & Irish, 2016; Karthikeyan & Kumar, 2017). In *Paederus* dermatitis, linear lesions usually can be observed after 24 to 72 hours (vary between patients) of exposure. Sometimes, kissing lesions or mirror lesions can also occur if the affected area is near the flexures (Banney *et al*., 2000; Beaulieu & Irish, 2016; Karthikeyan & Kumar, 2017). Erythema with small blisters is commonly seen on the affected skin and patients described that the affected area often felt itchy with burning sensations. However, the mechanisms of formation of blisters caused by exposure to pederin toxin are still not known. Banney *et al*. (2000) also noted that in many cases of this dermatitis condition, bullae with a size up to 1 cm in diameter are also observed along with the minute blisters and erythema. The lesions usually do not affect regions other than the exposed skin (usually uncovered areas, such as the face, hands, feet and neck). However, if the bullae burst and other unexposed areas are smeared with the fluid, inflammation and dermatitis linearis will then spread to other uninfected areas. Conjunctivitis might be developed if the pederin toxin is accidentally transferred and exposed to the periorbital area (Beaulieu & Irish, 2016). Fortunately, the infection in the eyes is not permanent and can be cured. There are also some severe cases where the patient experienced fever, vomiting, arthralgia (joint pain) and neuralgia (intense intermittent pain along a nerve) (Nasir *et al*., 2015; Karthikeyan & Kumar, 2017). Besides that, the affected area is prone to secondary infection if the exposed skin is excessively scratched. If appropriate care is taken, the lesions and symptoms will eventually subside and the exposed skin can be healed. Nonetheless, after healed, post-

inflammatory hyperpigmentation (scarring) might be visible on the affected skin area as documented by Banney *et al*. (2000), Beaulieu & Irish (2016), and Karthikeyan & Kumar (2017).

2.2 Pederin

Pederin is the toxic amide that caused linear dermatitis in humans and is reported to be found in the hemolymph of *Paederus* rove beetles. Pederin was detected at high intensity in female *Paederus* beetles' fourth abdominal segment (Bhandari *et al.*, 2015). It was also reported that pederin-producing bacteria were detected in the whole body of *P. fuscipes* beetles, including digestive tracts and reproductive organs, where the bacteria were highly abundant in the ovaries (Song *et al*., 2022). Besides beetles, researchers have found out that pederin-family compounds can be also found produced by marine sponges and cyanobacterium (Piel, 2002; Kust *et al*., 2018). Other than being the causative agent for *Paederus* dermatitis, this toxin has gained a lot of interest from researchers because the amide has the potential to be developed as antitumor agent due to its structure and chemical properties (Piel *et al*., 2005). This amide compound is catalogued in National Center for Biotechnology Information (NCBI, PubChem ID: 5381287) and European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI, ChEBI ID: CHEBI:78591).

According to Piel (2002) and Kellner (2001b, 2002b), pederin is a polyketide and a bacterial metabolite synthesized by *Pseudomonas* spp. in *Paederus* rove beetles. Pederin toxin is a crystalline amide that can be solubilized in water and alcohol (Frank & Kanamitsu, 1987). According to the NCBI and EMBL-EBI databases, the IUPAC (International Union of Pure and Applied Chemistry) name for pederin is (2*S*)-*N*-[(*S*)-[(2*S*,4*R*,6*R*)-6-[(2*S*)-2,3-dimethoxypropyl]-4-hydroxy-5,5-dimethyloxan-2-yl]-

methoxymethyl]-2-hydroxy-2-[(2*R*,5*R*,6*R*)-2-methoxy-5,6-dimethyl-4-methylideneoxan-2-yl]acetamide, and the amide is also known as paederine, pederine, (+)-pederin or (+)-pederine. The molecular formula of pederin is $C_{25}H_{45}O_9N$, with a molecular weight of 503.6 g/mol. The structure of pederin is as shown in Figure 2.1.

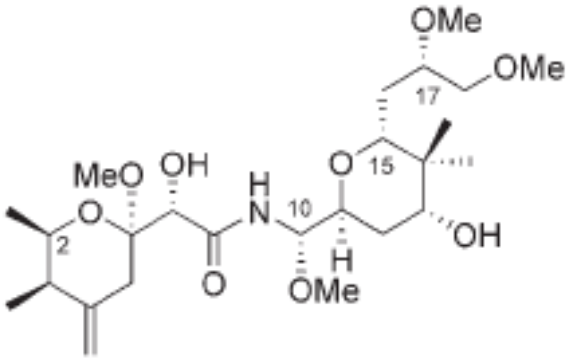

Figure 2.1 Structure of pederin (Jewett & Rawal, 2007; Piel, 2002).

The isolation and studies on the *Paederus* beetles' putative pederin biosynthesis pathway by Piel (2002) and Piel *et al*. (2004a, b, c) revealed that *ped* genes are classified as bacterial type I polyketide synthase (PKS). PKSs are modular structures of multifunctional enzymes mainly consist of ketosynthase (KS) domain that catalyses the polyketide chain elongation, acyltransferase (AT) domain that selects the correct acyl-CoA unit, acyl carrier protein (ACP) domain that binds the intermediates to the polyketide chain, and other domains involved in further modifications (methylations, dehydrations, and reductions of double bonds or keto groups) (Piel, 2002; Piel *et al*., 2004a, b, c). Unlike most PKSs genes, the unique feature of pederin biosynthesis (*ped* genes) are separated into three clusters distributed on the symbiont's genome due to the tranposase pseudogenes located at both sides of the gene clusters (Piel, 2002; Piel *et al*., 2004b, c). The pederin biosynthesis gene clusters contained 18 genes *pedA* to *pedR*, where the assembling of pederin molecules occur in PKS modules, *pedF* and *pedI* (Piel, 2002; Piel *et al*., 2004b, c). *pedH* was reported to have similar modular structure as *pedF* and *pedI*, and might be involved in further elongation of the intermediates generated by *pedF* (Piel *et al*., 2005). However, the roles of this gene in pederin biosynthesis require further studies, as *pedH* might be a residue from the evolutionary process (Piel *et al*., 2005). Further studies on *ped* genes revealed that the *ped* system's PKS modules are categorized as *trans*-AT PKSs (Piel, 2002; Piel *et al*., 2004c, 2005). Instead of located in *pedF* and *pedI* along with KS and ACP domains, the AT domains of *ped* PKS modules are located in different genes, *pedC* and *pedD* (Piel, 2002). Besides that, the detection of domain related to the proteins from GCN5-related *N*-acetyltransferase (GNAT) family at *pedI* indicate that loading module in pederin biosynthesis is primed with acetate, which is not reported in other known *trans*-AT PKS systems (Piel *et al*., 2004c).

Researchers had been investigating its cytotoxicity against in vitro human, mammal, and insect cell lines (Hisada & Emura, 1965; Soldati *et al*., 1966; Brega *et al*., 1968; Yamada *et al*., 2019). Pederin is an amide compound known as an inhibitor of protein synthesis and DNA synthesis, but it does not affect RNA synthesis (Soldati *et al*., 1966; Brega *et al*., 1968; Mosey & Floreancig, 2012). The cytotoxicity of pederin on eukaryotic cells might be due to the ability of pederin to bind to eukaryotic ribosomes, and potentially inhibit protein and DNA synthesis (Nishimura *et al*., 2005; Wan *et al*., 2011). Mammalian cell lines, including human, mouse, baby hamster and dog kidney cells, exposed to pederin at a concentration as low as 1 ng/mL showed severed cytological modifications, with the documentation of a decrease in the rate of DNA replication and protein synthesis (Soldati *et al*., 1966; Brega *et al*., 1968). The mitosis process of pederin-treated HeLa cells was halted at metaphase and eventually led to the inhibition of mitosis (Soldati *et al*., 1966; Brega *et al*., 1968). Decreased cell culture growth rate, cellular degeneration, cytoplasmic alterations (cellular busting and/or vacuolization), and fragmentation of nuclear chromatin were also observed (Soldati *et al*., 1966; Brega *et al*., 1968). Complete lysis of HeLa cells was documented at 24 hours to 4 days after exposure to 1 to 1.5 ng/mL of pederin (Soldati *et al*., 1966; Brega *et al*., 1968). Other than high cytotoxicity against mammalian cells, pederin is shown to be slightly toxic to insects (i.e., aphids and ladybeetles), insect cell lines (i.e., Sf9 cell lines), and plants (Yamada *et al*., 2019).

2.3 Insect microbial communities and interactions

Insects have been regularly infected and colonized by microorganisms through their food source (horizontal transfer) or from their parents (vertical transfer), and the microbiota formed inside insect hosts play important roles in shaping the hosts'

evolutionary history (Douglas, 2015). The microbial composition in insects include archea, bacteria, fungi, viruses, and several unicellular eukaryotic microorganisms, but not all microorganisms are associated with insects (Sachs *et al*., 2011; Douglas, 2015). The associations of bacteria with insect hosts are widely researched compared to other microbial groups as bacteria are often found colonizing the insect's gut, hemocoel, on the exoskeleton, and within the cells of insect host (Douglas, 2015). The exposure of insects to these bacteria can be mutually beneficial to both organisms (mutualism), commensal or detrimental to the insects (pathogenic or parasitism) (Douglas, 2011; Mushegian & Ebert, 2015; López-García *et al*., 2017; Teoh *et al*., 2021).

One of the most common interactions between insect hosts and bacteria is endosymbiosis, where the bacteria live inside the insect host are known as endosymbionts. Obligate endosymbionts (primary endosymbionts) and their insect hosts are highly dependent on each other for survival (provision of essential nutrients and protections against biotic and/or abiotic factors), and these primary symbionts are strictly vertically transmitted, such as reported in aphids-*Buchnera*, whiteflies-*Portiera*, tsetse flies-*Wigglesworthia*, and carpenter ants-*Blochmannia* (Baumann, 2005; Kikuchi, 2009; Feldhaar & Gross, 2009). Besides obligate endosymbionts, insects are also infected with facultative endosymbionts (secondary endosymbionts). Multiple infections of facultative endosymbionts along with obligate endosymbionts are reported in insect hosts, such as in aphids and whiteflies, and their infection rate is variable among population from different geographical regions and host species (Marubayashi *et al*., 2014; Guo *et al*., 2017). Some of these facultative endosymbionts may provide some benefits to their insect host such as protections against parasitoids and microbial infections, but can be costly to the host fitness. The costs of harbouring facultative endosymbionts include the decline in the insect host's behaviour, decrease in fecundity,

growth and/or longevity were observed, such as reported in aphids infected with defensive symbiont *Hamiltonella* or *Spiroplasma* (Guo *et al*., 2017). Besides that, there are certain groups of the insect-associated bacteria are pathogenic to the insect hosts (Ruiu, 2015; Glare *et al*., 2017), while some are reproductive manipulators that can affect the insect hosts' population dynamic (Ferrari and Vavre, 2011; Correa and Ballard, 2016). Some of the commonly identified entomopathogenic bacteria are from the family Bacillaceae (*Bacillus* spp., *Lysinibacillus* spp., *Paenibacillus* spp.), Enterobacteriaceae (*Serratia* spp., *Yersinia* spp., *Proteus* spp.), Pseudomonadaceae (*Pseudomonas* spp., *Photorhabdus* sp., *Xenorhabdus* sp.), and Rickettsiaceae (*Rickettsia* and *Wolbachia*) (Ruiu, 2015; Glare *et al*., 2017; Teoh *et al*., 2021). Some of the bacterial groups commonly identified as reproduction manipulators in insects are *Wolbachia*, *Spiroplasma*, *Cardinium*, and *Rickettsia*, and these symbionts can be found in insects from different orders (Ferrari and Vavre, 2011).

2.3.1 Next-generation sequencing (NGS) in the study of insect microbial profiles

Recently, the study on the microbial profile in insects has been gaining popularity, as studies have revealed that insect-associated bacteria can influence the insect's survival, aid in insect hosts' adaptations to the ecological changes, insect hosts' resistance against insecticides, as well as the therapeutic potential in the secondary metabolites produced by insect-associated bacteria (Bode, 2011; Kikuchi *et al*., 2012; Douglas, 2016). Moreover, the process of infections and colonization of bacteria in insect hosts are often complex and involve multiple mechanisms to overcome the insects' immune defence mechanisms (Nielsen-LeRoux *et al*., 2012; Glare *et al*., 2017). Culturable bacteria allow researchers to study the physiology and genetics of the bacteria that can lead to better understandings in the infection route and the interactions

of bacteria and their infected insect hosts (Kikuchi, 2009; Masson & Lemaitre, 2020). However, most of the insect-associated microorganisms are not cultivable outside of their insect hosts, and hence, researchers are utilizing next-generation sequencing (NGS) in order to investigate the microbial profiles in insects.

Previously, the use of clone libraries of microbial 16S rRNA marker gene (~1500 bp), followed by Sanger sequencing are commonly used to investigate the insect host's microbial diversity (Pidiyar *et al*., 2004; Rani *et al*., 2009; Russell *et al*., 2012). The construction and detection of the clone library have been a "gold standard" method to identify microbial profiles, but it can be laborious and costly (DeSantis *et al*., 2007; Bailón-Salas *et al*., 2017). Currently, researchers are relying NGS technologies, such as pyrosequencing and 16S rRNA metagenomics approach, along with bioinformatics analysis to study the microbial profile in insects (DeSantis *et al*., 2007; Kolasa *et al*., 2019; Xu *et al*., 2019). Unlike Sanger sequencing, these metagenomics sequencings produce short amplicons (~250 bp to 600 bp, depending on sequencing platforms) targeting the hypervariable regions (V1-V9). Studies have shown that regardless of host species (e.g.: human, shrimps, insects, and environmental samples), amplicons spanning both hypervariable V3 and V4 (V3V4) regions produced higher number of bacterial OTUs, species richness and diversity compared to other regions (Thijs *et al*., 2017; Rausch *et al*., 2019; García-López *et al*., 2020). Besides that, NGS technologies, depending on the platform and depth of sequencing, also allow researchers to have an overview on the functional profiles on both the insect hosts and their microbial communities, which will provide better insights on the interactions between insect hosts and their microbial diversity (DeSantis *et al*., 2007; Kolasa *et al*., 2019; Xu *et al*., 2019).

2.4 Insect population genetic diversity

Genetic diversity, also known as genetic polymorphism, is a central concept of evolutionary biology, where the magnitude of inherited traits within a population is linked to the ability of a species to adapt to environmental changes (Bazin *et al*., 2006; Hughes *et al*., 2008; Ellegren & Galtier, 2016). The variations in DNA sequences among the individuals of the same species are important in the study of a species evolution history and have important implications for the conservation of the species (Hughes *et al*., 2008; Ellegren & Galtier, 2016). Genetic polymorphism often occurred across chromosomes and loci with substantial differences, and it can be affected by factors such as the life cycle of the organism studied, mating systems, natural selection, the structure of the population, and population bottlenecks (Amos & Harwood, 1998; Ellegren & Galtier, 2016).

In the early 1970s, the allozyme technique was used to analyse species-level genetic variation in plants and animals (Ellegren & Gartier, 2016; Hoy, 2019). This technique, also known as protein electrophoresis, is based on the electrophoretic properties of protein to detect the allelic diversity in a population (Hubby & Lewontin, 1966). By using 10-30 loci, extensive allozyme diversity has been estimated in hundreds of plant and animal species (Ellegren & Gartier, 2016). Although this technique is cost-effective and easy to execute, there are several limitations that may lead to the underestimation of the genetic variation (Ellegren & Gartier, 2016; Hoy, 2019). These limitations include gene duplication and the effects of post-translational modifications of proteins (Ellegren & Gartier, 2016).

With the discovery of genetic markers, researchers have been relying on these molecular data to study the biology, cryptic species identity, genetic diversity, host