POPULATION GENETIC STUDY OF *Tapinoma indicum* (FOREL) IN PENANG ISLAND, MALAYSIA USING MICROSATELLITE MARKER DEVELOPED BY NEXT-GENERATION SEQUENCING

LIM LI YANG

UNIVERSITI SAINS MALAYSIA

2022

POPULATION GENETIC STUDY OF *Tapinoma indicum* (FOREL) IN PENANG ISLAND, MALAYSIA USING MICROSATELLITE MARKER DEVELOPED BY NEXT-GENERATION SEQUENCING

by

LIM LI YANG

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

September 2022

ACKNOWLEDGEMENT

First of all, I would like to express my sincere thanks to my supervisor, Associate Professor Ts. Dr. Abdul Hafiz Ab Majid for his guidance and support throughout the research and gave me the opportunity to gain the knowledge and learn the skills in carrying out the experiments. Besides, I would like to acknowledge my heartfelt gratitude for his patience, concern, encouragement and motivation throughout my research journey.

I am very grateful that my lab mates who are also my friends, Lim Li, Ahmed Ashigar Mohammed, Nurul Akmar Hussin and Abd Hafis Abd Rahim gave me helps and supports physically and morally throughout my master life. I would like to express my deepest appreciation to Kang Siang Yu and Lim Li who is not only my friends but also my guardian in helping to solve various problems that I faced in USM. My appreciation also extended to the staffs of School of Biological Sciences and Institute of Postgraduate Studies who provide countless assistances throughout my research life in USM.

Lastly, I would like to thank my beloved parents, brothers and sister for supporting me both financially and morally to complete my PhD study. Above all, I am very grateful that my past ancestors and grandparents for always blessing me and gives my strength to complete this research.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF PLATES

Page

LIST OF SYMBOLS

bp	Base pairs of sequence
min	Minute
mm	Millimeter
ml	Milliliter
℃	Degree Celsius
×	Magnification times
µl	Microliter
™	Trademark
®	United State trademark
%	Percentage
°	Degree
'	Minute
"	Second

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
AMOVA	Analysis of Molecular Variance
ATP	Adenosine triphosphate
Aug.	August
BOLD	Barcode of life Datasystem
CI	Confidence Interval
CI	Cephalic index
CO1	Cytochrome oxidase subunit 1
CO2	Cytochrome oxidase subunit 2
dNTP	Deoxynucleotide
ddNTP	Dideoxynucleotide
DNA	Deoxyribonucleic acid
E	East
emPCR	Emulsion Polymerase Chain Reaction
EI	Eye index
EL	Eye length
EW	Eye width
F_{null}	Null allele frequency estimation
F_{IS}	Wright's inbreeding coefficient
F_{ST}	Fixation index
F_{IT}	Wright's inbreeding coefficient within total population
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HL	Head length

HW	Head width
HWE	Hardy-Weinberg Equilibrium
Jun.	June
LD	Linkage disequilibrium
Mar.	March
Mbp	Megabase pair
ML	Mesosoma length
MPS	Massive parallel sequencing
N	North
N_A	Number of alleles
NCBI	National Center for Biotechnology Information
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PIC	Polymorphism Information Content
r	Relatedness coefficient
rDNA	Ribosomal deoxyribonucleic acid
s	Second
sp.	Species
spp.	Multiple species
SD	Standard deviation
SE	Standard Error
Sep.	September
SI	Scape index
SL	Scape length
SNPs	Single Nucleotide Polymorphisms
SRA	Sequence Read Archive

SSR	Single Sequence Repeat
USA	United States of America

LIST OF APPENDICES

KAJIAN POPULASI GENETIK *Tapinoma indicum* (FOREL) DI PULAU PINANG, MALAYSIA MENGGUNAKAN PENANDA MIKROSATELIT YANG DIBANGUNKAN OLEH PENJUJUKAN GENERASI SETERUSNYA

ABSTRAK

Tapinoma indicum Forel (Hymenoptera: Formicidae) ialah haiwan perosak isi rumah yang banyak tersebar di negara-negara Asia. Haiwan perosak isi rumah ini juga dikenali sebagai perosak gangguan yang menyebabkan kegusaran dan gangguan dengan membina sarang dan berkumpul dalam bangunan untuk mencari makanan dan air. Walau bagaimanapun, kajian mengenai *T. indicum* adalah terhad terutamanya dalam bidang biologi molekul. Oleh itu, tujuan penyelidikan ini adalah untuk mengkaji struktur genetik populasi *T. indicum* di Pulau Pinang, Malaysia dengan menggunakan penanda novel mikrosatelit yang dihasilkan dari keseluruhan genom *T. indium*. Objektif penyelidikan ini adalah: untuk menentukan taburan populasi *T. indicum* di Pulau Pinang, Malaysia; untuk mengenal pasti hubungan filogenetik *T. indicum* berdasarkan gen mitokondria sitokrom c oksidase subunit I (*CO1*) dan DNA ribosom 16S (16S rDNA); untuk merancang penanda novel mikrosatelit dengan menyaring lokasi mikrosatelit genom *T. indicum* yang diurutkan dari penjujukan generasi masa depan ; untuk mengesahkan polimorfisme penanda mikrosatelit yang dirancang; untuk menganalisis kepelbagaian genetik dan struktur genetik populasi *T. indicum* dengan menggunakan penanda mikrosatelit yang dirancang. Terdapat sejumlah 30 populasi yang dikumpulkan dari Balik Pulau, Gelugor dan George Town. Urutan CO1 dan urutan 16S rDNA *T. indicum* dari setiap populasi disimpan dan diakses melalui pangkalan data Genbank (NCBI). Tidak ada perbezaan genetik yang dapat dikesan antara tiga populasi wilayah yang dikaji berdasarkan urutan *CO1* (F_{ST} <0) tetapi

terdapat sedikit perbezaan genetik antara populasi George Town dengan Gelugor, begitu juga populasi antara George Town dengan Balik Pulau telah dapat dikesan berdasarkan urutan 16S rDNA (F_{ST} = 0.0828). Pokok filogenetik menghasilkan tiga klad dengan tiga haplotaip namun struktur genetik tidak dinyatakan. Sebanyak 143,998 penanda satelit mikro dihasilkan berdasarkan mikrosatelit 2.61 × 106 yang diasingkan dari urutan DNA genom *T. indicum*. Tujuh penanda mikrosatelit disahkan sebagai penanda genetik maklumat dan digunakan untuk genotip 300 *T. indicum*, pekerja individu. Kehilangan kepelbagaian genetik telah dijumpai dengan tahap pembiakan yang sederhana dalam populasi (F_{IS} = 0.498). F-statistik Wright dan pekali yang berkaitan menunjukkan *T. indicum* relatif inbred (F_{IS} = 0.490, F_{IT} = 0.529), penurunan kepelbagaian genetik (F_{ST} = 0.075), dan berkaitan secara genetik (r = 0.098). Aliran gen yang tinggi antara populasi menyebabkan struktur genetik populasi tidak dapat dibezakan oleh wilayah termasuk Gelugor, Balik Pulau dan George Town (AMOVA = 1%). Dua kumpulan yang berbeza secara genetik telah dikenal pasti di wilayah yang bertindih. Kesimpulannya, penyelidikan ini berjaya mengkaji hubungan filogenetik, merancang penanda novel mikrosatelit dan mendedahkan struktur genetik populasi *T. indicum* yang telah dikumpulkan di Pulau Pinang, Malaysia.

POPULATION GENETIC STUDY OF *Tapinoma indicum* (FOREL) IN PENANG ISLAND, MALAYSIA USING MICROSATELLITE MARKER DEVELOPED BY NEXT-GENERATION SEQUENCING

ABSTRACT

Tapinoma indicum Forel (Hymenoptera: Formicidae) is a household pest that is widely distributed in Asian countries. It is known as a nuisance pest that causes annoyance and disturbance by constructing nests and foraging in building for food and water. However, studies on *T. indicum* are limited, especially in the field of molecular biology. Hence, this research aims to study the *T. indicum* population genetic structure in Penang Island, Malaysia by using the novel microsatellite markers developed from whole genome of *T. indium*. Objectives of this research include: to determine the *T. indicum* population distribution in Penang island, Malaysia; to identify the phylogenetic relationship of *T. indicum* based on the mitochondrial gene of cytochrome c oxidase subunit I (*CO1*) and 16S ribosomal DNA (16S rDNA); to design novel microsatellite marker by screening microsatellite loci of *T. indicum* genome sequenced from next-generation sequencing; to validate polymorphism of designed microsatellite marker; to analyse genetic diversity and population genetic structure of *T. indicum* using designed microsatellite markers. There was a total of 30 populations collected from Balik Pulau, Gelugor and George Town. The *CO1* sequences and 16S rDNA sequences of *T. indicum* from each population were deposited and accessible via Genbank (NCBI) database. No genetic differences were detected between three area populations studied based on *CO1* sequences ($F_{ST} < 0$) but little genetic difference between the George Town and Gelugor populations, as well as between the George Town and Balik Pulau populations were detected based on 16S rDNA sequences (F_{ST}

= 0.0828). The phylogenetic trees result in three clades with three haplotypes, but the genetic structure is not revealed. A total of 143,998 microsatellite markers were developed based on the 2.61×10^6 microsatellites isolated from *T. indicum* genomic DNA sequences. Seven microsatellite markers were validated as informative genetic markers and used to genotype 300 *T. indicum,* individual workers. Loss of genetic diversity has been found with a moderate level of inbreeding within the populations (F_{IS} = 0.498). Wright's F-statistics and related coefficient reveals the *T. indicum* were relatively inbred (F_{IS} = 0.490, F_{IT} = 0.529), reduced genetic diversity (F_{ST} = 0.075), and genetically related (r = 0.098). A high degree of gene flow in between the populations leads the population genetic structure cannot be distinguished by the areas including Gelugor, Balik Pulau and George Town (AMOVA =1%). Two genetically distinct groups had been identified that overlap the areas. In conclusion, this research had succeeded to study the phylogenetic relationship, designing novel microsatellites markers, and revealing the population genetic structure of *T. indicum* populations collected in Penang Island, Malaysia.

CHAPTER 1

INTRODUCTION

1.1 Background of study

1.1.1 Urbanization

In most of mankind's history, the majority of the human population in the world lived in small settlements which were mainly agrarian and known as rural areas (Berry, 2008). In recent centuries, and notably in the last few decades, there has been a significant change in this perspective. The human population had shifted from small settlements to a more concentrated and relatively compact area, thereby creating the concept of the urban area. This unique process of the mass movement in the human population from rural to urban settings in recent centuries is known as urbanization. Until the present time, over half of the world's population live in urban settings and this phenomenon is expected to be grown every year in the future (Ritchie and Roser, 2018).

Malaysia has been in rapid growth since its independence in 1957. Several areas have succeeded in developing from an agrarian society into an urban society such as Penang. Among Malaysia's states, Penang ranked as the second greatest population density with 3300 people per km^2 of land area (Demographia World Urban Areas, 2021). Penang state is separated into two parts including Penang Island and the mainland, Seberang Perai, as shown in Figure 1.1. The Penang Island is connected to the mainland with only two long bridges. Urbanization of Penang states mainly occurs in the Penang Island with the urban core, George Town.

Figure 1.1 Situation map of Penang state (Brown) in Malaysia (Grey).

As a consequence of urbanization, the increasing human population density has led to the accumulation of food and water resources, as well as a variety of niches serve as habitats (Penick et al., 2015; Sanford et al., 2009). With the ecology changes of Penang Island due to urbanization, there has been an increasing concern regarding the presence of arthropods surrounding the human community (Ridzuan et al., 2022). These arthropods have been recognised as part of urban entomology which impacts both harmful and beneficial to our urban environment. As a result, the study in the household and structural urban entomology has gained in popularity.

Household and structural urban entomology aim to study the insects and pest problems that are potentially associated with urban environments (Rust and Su, 2012). The study includes how they interact with urban settings as well as their impacts on the household and structural damage or human communities as a pest. It provides fundamental knowledge and rationale data for a greater understanding of urban-related insect pests to allow further research towards environmentally friendly control and regulation of these insect pests.

1.1.2 General information of ants

Ants are almost existing all over the planet. There are approximately more than 12000 species of ants that have been identified but most of them are not recognized as a pest. According to National Pest Management Association (2021), there are only about 25 species of ants are known as pests that commonly infest household and structural buildings. Ants are classified as the Formicidae family and belong to the Hymenoptera order with closely related bees and wasps families. They are known as social insects which exhibit eusocial structure in their colony.

1.1.3 Ant as pest

Ant is one of the most common and persistent household pests that may be found around residences and other structures (NPMA, 2012). This pest may survive in human structures, where there is food and water resources and places to hide as a habitat. Ants act as a nuisance pest because they contaminate food, produce ugly mounds on the structure and create annoyance and disturbance to humans. Furthermore, some ant species may sting and trigger hypersensitivity reactions (Potiwat and Sitcharungsi, 2015), while others may hollow out building materials for colony construction and cause structural damage (Hansen et al., 1993; Klotz et al., 2008).

According to the results of three studies on household pest status between 1983-1998 conducted among homeowners and the pest control industry in Malaysia, ants ranked as the third important household pests in all three surveys. Meanwhile, the Malaysian pest control industry has ranked the significance of ants as third among the targeted pests in the survey conducted in 2001, which was improved from the fourth position in the survey carried out in 1995 (Lee, 2002). In comparison to other

countries, ants ranked as the leading annoyance pest in the United States according to the National Pest Management Association (2012). Furthermore, ant control service has contributed averagely 22.2% of industry business sales in 2019 which is greater than every alternative pest management service, according to the 2020 State of the Ant Control Market report (Syngenta, 2020).

Ant is one of the most complicated and requires diligent efforts in pest control management. The most often used and commercially available method in controlling ants as pests is spraying or baiting. Both spraying and baiting methods rely on toxic substances that are lethal to ant species but may become ineffective due to the rise of resistances from genetic mutations (Lim and Ab Majid, 2019). Sometimes, spraying or baiting may not eliminate the entire ant colony but only parts of the ants in the colony that came into contacted with the insecticides (Lim and Ab Majid, 2019). This will only provide a short-term control on the foraging activities of ant infestation. The ant infestation problems may reoccur in a short time as the recovering of the injured colony. Moreover, reinfestation often occurs as spraying and baiting only kills the ant colony and leave no residuals to protect the infestation areas. This will require another effort on the regular implication of the perimeter barrier to prevent the reinfestation. Until the present time, spraying and baiting had been used for many years and only improved through enhancement of the active ingredient that is lethal to ant species. The invention of a novel method in ant control requires an in-depth of understanding of ant species including molecular data information.

1.2 Problem statement

In a recent study conducted in Penang Island, Malaysia, *Tapinoma indicum* was the most numerous species that had been found in household areas, accounting for 17.74 % of the total species abundance (Ab Majid et al., 2016). Unfortunately, studies on *T. indicum* have been quite restricted compared to other species such as *T. melanocephalum* and *T. sessile*. There is no research that has been done in the morphological measurement and molecular information aspect of *T. indicum*. Additionally, there is no published research work on the development of species-specific microsatellite markers for *T. indicum* while researchers have attempted to provide molecular data for various ant species using newly designed species-specific microsatellite genetic markers. Furthermore, the population genetic structure of *T. indicum* is not studied which might be due to lack of suitable genetic markers.

1.3 Significance of study

This research was conducted to study the *T. indicum* infestation in Penang Island, Malaysia. This study provided crucial information including morphological measurements and molecular genetic sequences which may stand as baseline knowledge for *T. indicum* infestation around the world in morphological and molecular aspects.

This study came up with a novel species-specific microsatellite marker from the genomic data of *T. indicum* with next-generation sequencing. These large scales of highly polymorphic genetic markers have provided a tool to explore the genetics and population of the *T. indicum* infestation around the world.

This research could contribute to the understanding of *T. indicum*, as well as contributing to establishing and developing innovative management approaches for this key pest species. The genetic diversity and population genetic structure stand as the starting point for the comparison for further study on others *T. indicum* infestation study sites. The information also provides the setups for predictive framework for prevention of future infestation.

1.4 Objective

This study aimed to describe preliminary morphological and molecular genetic data of *T. indicum* particularly on addressing the species genetic diversity and population genetic structure in Penang Island, Malaysia. In order to achieve this primary aim, the study flows were designed to achieve a number of specific objectives, which are listed below:

Objective 1 (Chapter 3): To determine and morpho-molecular characterise the *T. indicum* population in Penang Island, Malaysia.

Objective 2 (Chapter 3): To identify phylogenetic relationship and genetic diversity of *T. indicum* based on the mitochondrial gene of cytochrome c oxidase subunit I (*CO1*) and 16S ribosomal DNA (16S rDNA).

Objective 3 (Chapter 4): To design novel microsatellite marker by screening microsatellite loci of *T. indicum* genome sequenced from next-generation sequencing.

Objective 4 (Chapter 4): To validate the polymorphism of designed microsatellite markers.

Objective 5 (Chapter 5): To analyse the population genetic structure and breeding structure of *T. indicum* using designed microsatellite markers.

CHAPTER 2

LITERATURE REVIEW

2.1 Household ant pest

Generally, an ant colony is divided into three different castes including reproductive female, reproductive male and non-reproductive female which are translated into queen, male and worker (Anderson and McShea, 2001; Oster and Wilson, 1978). Ants may exhibit either monogyny or polygyny in their colony structure, depending on whether the colony contains a single queen or multiple queens (Boomsma et al., 2014; Toennisson et al., 2011). Queen ants have two primary roles including establishing a new colony and laying eggs to produce offspring. Queen ants may produce haploid eggs through parthenogenesis which will eventually develop into male ants at a later stage. On the other hand, the queen ant mates with male ants and produce diploid eggs which will eventually develop into either sterile female (worker) or non-sterile female (queen). The primary roles of male ants are to mate with the queen and fertilize the eggs. A variety of tasks are assigned to workers including taking care of the queen and the developing offspring, building and expanding the colony, foraging for food and water supplies and feeding each other in the colony via trophallaxis or non-trophallaxis behaviour (Meurville and LeBoeuf, 2021).

Ants act as a nuisance pest because they contaminate food, produce ugly mounds on the structure and create annoyance and disturbance to humans. There are several common household ants that are recognized as pest including acrobat ant (*Crematogaster* sp.), Argentine ant (*Linepithema humile*), big-headed ant (*Pheidole* sp.), carpenter ant (*Camponotus* sp.), crazy ant (*Paratrechina* sp. & *Anophlolepis* sp.),

fire ant (*Solenopsis* sp.), pavement ant (*Tetramorium* sp.), pharoah ant (*Monomorium* sp.), tooth Ponera ant (*Odontopenra* sp.) and odorous house ant (*Tapinoma* sp.) (Na and Lee 2001; Smith, 1965). The 2020 State of the Ant Control Market survey had surveyed 248 respondents in United States pest management professionals which ranked odorous house ants as one of the most challenging species to control (Syngenta, 2020).

2.1.1 Medical importance

Some of the ant species may sting and inject ant venoms including acids and alkaloids which may induce allergenic, haemolytic, cytolytic, paralytic, inflammatory and pains (Touchard et al., 2016). Among the most frequent household ants found in urban environments, fire ants are capable of stinging humans and injecting them with ant venom (Wijerathne et al., 2018). The ant venom will cause an immediate, serious burning sensation at the location of the sting, followed by itching that may linger for days (Manuyakorn et al., 2017; Richard and Soto-Aguilar, 1993). Ponera ants are another household ant that plays a significant role in medical entomology, however the protein allergens in their venom have not been well investigated (Potiwat and Sitcharungsi, 2015).

Apart from household ants, there are several ants has been reported as stinging ant that nests around and invades household in urban areas, including Ponera ant, *Pachycondyla striata* (Cantone and de Carvalho Campos, 2015) and arboreal bicoloured ant, *Tetraponera rufonigra* (Sabtu and Ab Majid, 2020). The venom protein composition of both ant species had been studied, and the results showed that they had a high allergenic potential to humans, which might result in pains, inflammation, and a possibly life-threatening allergic response (Naephrai et al., 2021; Santos et al., 2017).

2.1.2 Economic importance

Ants ranked as the leading annoyance pest in the United States according to the National Pest Management Association (2012). Ant pest control is one of the most abundant services provided in the structural pest management sector in the United States, with an estimated annual revenue of $1.7 billion for the pest control industry (Klotz et al., 2008).

A study had found one of the household ants, fire ant, *Solenopsis invicta* may incurs $80.37 per year expenses per household in South Carolina from the pest control treatment expenses, repairing expenses, medical treatment expenses and household pet veterinary treatment expenses (Duke et al., 1999). The study had further estimated the expenses of fire ant impact which was averagely $829 million per year for California residents (Jetter et al., 2002).

According to the results of three studies on household pest status between 1983-1998 conducted among homeowners and the pest control industry in Malaysia, ants ranked as the third important household pests in all three surveys (Lee 2002). Meanwhile, the Malaysian pest control industry has ranked the significance of ants as third among the targeted pests in the survey conducted in 2001, which was improved from the fourth position in the survey carried out in 1995. This has accounts for 10% of business revenue for Malaysian pest control industry which estimated between $18-20 million (Lee et al., 1999).

According to the 2020 State of the Ant Control Market report, ant pest management service has provided averagely 22.2% of industry business earning which is greater than every other pest's management service in 2019 (Syngenta, 2020).

Moreover, the survey had identified odorous house ants as the largest percentage of pest control service calls in 2019 which accounted to 24% of survey responds.

2.2 Overview of household ant control and management

Ant is one of the most complicated and requires diligent efforts in pest control management. Despite the fact that there are several ant control managements and strategies, most of them involved insecticide treatments which are formulated as liquid sprays, dust, fogs and baits (Drees and Summerlin, 1998). Although it is possible to dust or fog with insecticide to control household ants, this is generally not suggested since it might spread chemical residues throughout the house, which can be harmful to humans if they inhale them in (Meftaul et al., 2020; Pogoda and Preston-Martin, 1997).

When used in conjunction with spraying and baiting, a fast-acting insecticide will only give short-term control over the foraging activities of an ant infestation since it does not eliminate the whole ant colony but only the ants in the colony that have interacted with the insecticides. Long-term control of ants requires the use of a slow-acting insecticide because the transmission of toxic substance within the colony is desired (Hainze, 1993). Therefore, baiting method with slow-acting insecticides is often preferable.

The application of barrier sprays with ant repellents around the perimeter of a building to repel ants is another frequent strategy for household ant prevention and control (Klotz et al., 1997). This approach does not eliminate ants, but it creates a barrier around the structure to repel ants from entering the structure. The use of a non-repellent pesticide to treat the perimeter of the building is also a common practise to manage the home ant (Field et al., 2007). This approach involves creating a barrier

around the structure that prevents ants from entering when they pass through the barrier. Furthermore, granular bait in conjunction with baiting stations around the structure represents a sustainable control method (Klotz, 2000). This strategy may help to reduce the possibility of human exposure to the toxicants as the insecticide baits is in a discrete area. It also shields the baits from direct exposure to rain and sunshine, which allows the granular baits to remain effective for a longer period.

2.3 *Tapinoma indicum* Forel (Hymenoptera: Formicidae)

2.3.1 Tapinoma indicum morphology

Tapinoma indicum Forel (Hymenoptera: Formicidae) has been known as nuisance pest that is found across Asia countries (Plate 2.1). It was originally identified in 1895 (Forel, 1895). *Tapinoma indicum* is referred to as the household odorous ants due to its rotting coconut odour emission when it is crushed. In addition, it is also commonly known as a ghost ant because of its smaller size and faster speed of movement when compared to other ant species. Way et al. (2002) have reported *T. indicum* as a highly aggressive predator species.

Tapinoma spp. can be distinguished from other genus ants by several morphological characteristics (Mallis and Moreland, 2011). One of the most unique morphological characteristics is the workers of *Tapinoma* spp. have a flattened petiole that is virtually obscured by the front edge of the abdomen. Furthermore, *Tapinoma* spp. workers have no prominent club on the 12-segmented antennae, no spine on the uneven thorax and no hairs surrounding the abdomen's tip.

The colour appearances of *T. indicum* can be distinguished from other well-known species such as *T. melanocephalum* and *T. sessile* (Lee and Tan, 2004). The

Plate 2.1 Image of *T. indicum*.

head and thorax of *T. indicum* workers are black in colour, whilst the legs and abdomen are brown in colour. In contrast, *T. melanocephalum* workers have dark brown colour on the head and thorax, whilst the abdomen and legs are a milky white in colour., *T. sessile* workers are uniformly dark coloured for the entire body. Unfortunately, there are no morphological measurements on the *T. indicum* had been reported to date.

2.3.2 *Tapinoma indicum* molecular genetic

Tapinoma indicum has not been the main subject of a molecular investigation to date. There is only a series of mitochondrial cytochrome c oxidase subunit 1 (*CO1*) sequences for *T. indicum* had been published in the National Center for Biotechnology Information (NCBI). Unfortunately, the research article published did not address or discuss these sequences (Wang et al. 2018). The only study that involved the molecular genetic aspect of *T. indicum* was focused on the ant biodiversity by building the species-level phylogenetic tree with 18S ribosomal RNA gene, 28S ribosomal RNA gene and wingless (wnt-1) gene (Zhao et al., 2020).

2.3.3 Other published studies on *T. indicum*

Despite its great prevalence as a nuisance, *T. indicum* has received less attention. The first study of *T. indicum* focused on the foraging choice and activity (Chong & Lee, 2006). The study found that *T. indicum* prefer both carbohydrate and proteinaceous food and a peak foraging activity was at 0730 due to low temperature and high relative humidity. The majority of research has been on *T. indicum*'s bait preferences. Lee (2008) reported that *T. indicum* workers as active foragers who found the baits within a very short period, and they prefer gel and liquid baits compared to paste and granule baits. A study on *Anoplolepis gracilipes* preference on brown cane sugar-based liquid bait containing fipronil, indoxacarb, and boric acid noticed that *T. indicum* was found to be the most often encountered species that show up after the

elimination of long-legged ant (Chong & Lee, 2009). Ab Majid et al. (2018) have developed a coffee-based bait that significantly attracted *T. indicum* workers with 0.01 % Arabica extract as coffee-induced odour stimuli. The impact of Arabica extract as insecticide had resulted in high mortality on *T. indicum* (Yeoh et al., 2018). The most recent work with *T. indicum* is the creation of an insecticide generated from plants including *Citrus hystrix*, *Mentha piperita* and *Ocimum basilicum* (Lim & Ab Majid, 2019).

2.4 *Tapinoma indicum* infestation in Penang Island, Malaysia.

During the mid-1990s, the first research on the household ant species composition in Penang Island, Malaysia found no evidence for the presence of *T. indicum* (Yap and Lee, 1994). The research surveyed four locations including Balik Pulau, Gelugor, George Town and Universiti Sains Malaysia Minden campus had discovered the pharaoh ant, *Monomorium pharaonis* as the predominant species which followed by the odorous house ant, *T. sessile* and the crazy ant, *Paratrechina longicornis*. The first study that detected the occurrence of *T. indicum* was conducted in 2001 (Lee et al., 2002). The study covered Air Itam, George Town, Gelugor and University Sains Malaysia Minden campus had found the presence of *T. indicum* in low abundance with only 4.2 % of species-abundance, whilst the Singapore ant, *M. destructor* being the most prevalent species, followed by odorous house ant, *T. melanocephalum*, big-headed ant, *Pheidole* sp. and bicolored trailing ant, *M. floricola*.

A total of 13 ant species were discovered in residential areas on Penang Island, Malaysia, according to the most recent study (Abdul Hafiz et al., 2016). The research indicates *P. longicornis* (22.09 %) as the most abundant ant species found in residential areas of Balik Pulau, Relau, Sungai Ara and George Town, followed by *T. indicum*

(17.74 %), *T. melanocephalum* (17.0 %), *M. pharaonis* (10.61 %) and *Pheidole* spp. (5.22 %).

The increasing infestation of *T. indicum* from non-detected in the mid-1990s, to 4.2 % of species abundance in 2001 and 17.74 % of species abundance in 2016 is observed. However, this result may not indicate the current population of ants in Penang Island because of the survivability of native species and the threat of invasive species. Invasive species usually represent stronger competitiveness compared with native species in adaptation to the new climatic condition due to highly competitive biological traits, widely spread geographic ranges and stronger toleration to a different environment (Hellmann et al., 2008). *Tapinoma* species have been commonly recognised as a prevalent tropical tramp species that is defined as ant species with small and sterile workers which widely dispersed throughout the world by human traffic and lives in close contact with human communities (Passera, 1994). Consequently, the increasing infestation of *T. indicum* in Penang Island, Malaysia is presumed introduced from outside of the natural surrounding sea barrier via human traffic and distributed widely by human activities inside Penang Island, Malaysia after being introduced.

2.5 DNA barcoding

2.5.1 Background

DNA barcoding is a technique that involves sequencing a short fragment gene which referred to as the "DNA barcode" from a taxonomically unidentified organism and comparing the sequencing results to a reference barcodes library such as the Barcode of life Datasystem (BOLD) in order to determine the taxonomic species identification (Wilson, 2012). This technique has been employed in several genetic

applications such as phylogenetic study owning to the capabilities in detecting Single Nucleotide Polymorphisms (SNPs), insertions and deletions.

In genetics, SNP is a variation in the sequence of a single nucleotide between two or more organisms in the same DNA area. A SNP may induce gene mutation when it occurs in the gene area. A SNP that does not modify the protein acid sequence is featured as synonymous SNP while SNP creates alteration in the protein amino acid sequence is recognised as non-synonymous SNP. In non-synonymous SNP, SNP induced different amino acid is classified as missense and SNP causes premature stop codon is defined as nonsense.

A gene area should always fulfil the following three requirements in order to be relevant as a DNA barcode: it must contain conserved flanking areas enabling the development of universal PCR primers for broad taxonomic application; it must possess sufficient species-level genetic diversity and variability; it must only include a short sequence length in order to sustain existing practices of DNA extraction and amplification (Dong et al., 2021; Kress and Erickson, 2008).

2.5.2 Cytochrome oxidase c subunit 1

Cytochrome oxidase c subunit 1 (*CO1*) mitochondrial gene had been proposed to be implemented as the DNA barcode and served as the core of species identification system for animals (Hebert et al., 2003). The *CO1* gene is a small mitochondrial gene with a length of around 650 bp that is present in most of the eukaryotes (Dong et al., 2021; Simon et al., 1994). This gene has been shown to be highly conserved across species as it involved the conversion of food into useful energy for the organism in the mitochondrial complex. Moreover, there are numerous copies of mitochondria complex present in each cell, facilitating the DNA extraction and DNA amplification

in practical situations with a restricted sample to be accomplished. Folmer et al., (1994) have developed the universal primer pair for the *CO1* gene including L*CO1*490 (5′-GGTCAACAAACATAAAGATATTGG-3′) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′).

The *CO1* gene sequences had been used in phylogenetic analysis as genetic evidence to support the taxonomic separation of the *Tapinoma* genus. Escárraga et al., (2021) used the *CO1* gene sequences concatenated with three other sequences to construct a phylogenetic tree that included *Tapinoma atriceps* and *Tapinoma breviscapum*, which were described as two *Tapinoma* ants that were morphologically similar. The results showed that *T. atriceps* was recovered as a monophyletic group that was distinct from the *T. breviscapum* lineage as well as other outgroups including *Tapinoma opacum* and *T. melanocephalum*.

2.5.3 Limitations

DNA barcoding using the *CO1* gene has been well developed and actively utilised in the past decade. However, several drawbacks have been recognised when conducting DNA barcoding with insects such as incorrect classification of searches that are reflected in the reference database and incorrect classification of searches due to the limited insect references in a database (Virgilio 2010). One of the major problems encountered is the presence of nuclear pseudogenes derived from the mitochondrial DNA (Song et al., 2008). A nuclear mitochondrial DNA pseudogene is a copy of a mitochondrial sequence that has been integrated into the nuclear genome.

In 1983, the first discovery of nuclear mitochondrial DNA pseudogene in insect genome was reported on the grasshopper, *Locusta migratoria* (Gellissen et al., 1983). Similar findings were published subsequently in which the pseudogenes found in the

nuclear genome of grasshopper, *Schistocerca gregaria* was amplified in conjunction with the mitochondrial gene sequences (Zhang and Hewitt, 1996). It was also discovered in the same year that aphids, *Sitobion* spp., had several nuclear mitochondrial DNA pseudogenes matching to the mitochondrial *CO1* and cytochrome c oxidase subunit 2 (*CO2*) areas were present in their genome (Sunnucks and Hales, 1996).

There have been numerous insects had been reported as a result of the discovery of the nuclear mitochondrial DNA pseudogene including fruit fly, African malaria mosquito, flour beetle, honeybee (Pamilo et al., 2007), yellow fever mosquito (Hlaing et al., 2009), Hawaiian bee (Magnacca and Brown, 2010) and corixid aquatic insect (Yano et al., 2020). Aside from that, the presence of nuclear mitochondrial DNA pseudogenes had also been found in the ants (Maykon et al., 2014).

2.5.4 16S ribosomal DNA

The inclusion of another relevant gene in the DNA barcoding applications may be beneficial in order to minimize the possibility of bias that might arise from the use of solely *CO1* genes in the DNA barcoding applications. In fact, 16S ribosomal DNA (16S rDNA) is widely used in bacteria identification as a primary tool for DNA barcoding. 16S rDNA is a highly conservative gene with a length of around 1500 bp that encodes the 16S small subunits of the ribosomes, which plays an important role in assisting with the catalysis of protein synthesis.

The use of 16S rDNA genes for DNA barcoding has been shown to meet the criteria for a universal DNA barcoding marker in amphibians (Vences et al., 2005). Yang et al., (2013) had validated the 16S rDNA as a rapid and reliable DNA barcoding marker for insects which identified the unknown psocid species of under genus

Liposcelis. Another study conducted on 84 tick specimens belonging to Ixodida order had recommended 16S rDNA to be utilized as the DNA barcode in cases where *CO1* does not produce reliable results (Lv et al., 2014). Yoshizawa and Johnson (2003) had successfully employed 16S rDNA to construct a phylogenetic study to analyse the phylogenetic relationship between Phthiraptera (chewing and sucking lice) and Liposcelididae (booklice and barklice). Moreover, the application of 16S rDNA also can be found in the study of ant species, where 16S rDNA was used to determine the molecular identity and phylogeny of the pharoah ant, *M. pharaonic* (Alajmi et al., 2020).

2.6 Next-generation sequencing

2.6.1 Development of DNA sequencing technology

In 1977, Frederick Sanger developed a method of DNA sequencing that uses the chain-termination method (Sanger et al., 1977). The chain-termination method adds dideoxyribonucleotides (ddNTPs) in the DNA synthesis solution. DNA synthesis will stop whenever a ddNTP is added into the extension strand due to lack of 3'-OH group for bond formation with a 5'-phosphate group of next deoxyribonucleotide (dNTP). Therefore, different lengths of DNA will be produced, and the sequence can be identified by gel electrophoresis. Sanger sequencing becomes the gold standard and is widely used from the 1980s to the 2000s.

In the following years, many improvements have been applied to Sanger sequencing, the most significant of which was the substitution of phospho- or tritium-based radiolabelling with fluorometric-based detection (Heather and Chain, 2016). It improved the sequencing to be occurring on single vessel rather than 4 different vessels and allowed reading method carried out by capillary-based electrophoresis instead of

gel-based electrophoresis. These improvements eventually contributed to the first-generation sequencing platforms. The first fully automated DNA sequencing platform is developed and commercialized by Applied Biosystems in 1987s. The first-generation sequencing platforms had pushed DNA sequencing into a new era especially helped the Human Genome Project to complete years ahead of schedule (Hood and Rowen, 2013)

In 1993, a new DNA sequencing method was invented known as pyrosequencing (Nyrén et al., 1993; Ronaghi et al., 1998). Pyrosequencing utilized the luminescent technique to estimate synthesis of pyrophosphate that converted into ATP and produced illumination proportionate to the quantity of pyrophosphate by sulfurylase and luciferase, respectively. By monitoring pyrophosphate formation, this method was utilised to determine sequence. It allowed the DNA sequencing can be performed using natural nucleotides and observed in real-time which reduces the cost and time-consuming in DNA sequencing.

2.6.2 Next-generation sequencing platforms

The first major commercially successful next-generation sequencing platform was developed by 454 Life Science in 2004 (later acquired by Roche) based on the pyrosequencing technique (Margulies et al., 2005). It uses massive parallel sequencing (MPS) which clonally amplified DNA template in a single bead by emPCR. These beads coated with amplified DNA sequences are now transferred onto a reaction plate containing multiple wells that can be filled with one bead. Pyrosequencing is now being performed on the reaction plate. Wheeler et al., (2008) used a 454-sequencing platform for human genome sequencing and resulting in a significant reduction in time consumed and cost compared to first-generation sequencing.

In 2006, another important next-generation sequencing platform was introduced by Solexa (later acquired by Illumina) (Bennett, 2004). Illumina platforms use MPS by linking different adapters on each end of the sequencing library DNA, then binding it to a slide coated with the complementary sequences for each primer (Meyer and Kircher, 2010). This allows "bridge PCR", producing spots of amplified DNA on the slide. These spots are then sequenced based on the principle of Sanger sequencing by using fluorescent-tagged ddNTPs.

Applied Biosystems, leader of first-generation sequencing platforms commercially released a new next-generation sequencing platform in 2006 known as SOLiD (branded as Life Technologies). SOLiD uses MPS which clonally amplified DNA template in a single bead by emPCR. The beads are then covalently attached to a glass slide. SOLiD uses ligation-based sequencing with fluorescence labelled Di-base probes. After detection, the fluorescence part in probes is cleaved and the ligation cycle is repeated until the sequencing is completed.

A new next-generation sequencing platform, Ion torrent owned by Life Technologies is developed in 2010 which sequencing based on the detection of hydrogen ions (Rothberg et al., 2011). Ion torrent uses MPS which clonally amplified DNA template in a single bead by emPCR. The beads are then fit into microwells on a semiconductor chip. The sequences are then inferred by semiconductor chip detect the production of hydrogen ion by DNA synthesis biochemical reaction.

2.6.3 Genome sequencing of ants

Genome sequencing has become more popular, thanks to the usage of next-generation sequencing. However, only a small number of studies have been done on the genome sequencing of ant species. To present, genome sequencing of ants

conducted using next-generation sequencing includes the following species: Argentine ant, *Linepithema humile* (Smith et al., 2011); clonal raider ant, *Cerapachys biroi* (Oxley et al., 2014); Florida carpenter ant, *Camponotus floridanus* (Bonasio et al., 2010); Indian jumping ant, *Harpegnathos saltator* (Bonasio et al., 2010); Alpine silver ant, *Formica selysi* (Purcell et al., 2014); black garden ant, *Lasius niger* (Konorov et al., 2017); leaf-cutting ant, *Acromyrmex echinatior* (Nygaard et al., 2011); leaf-cutting ant, *Atta cephalotes* (Suen et al., 2011); leaf-cutting ant, *Atta colombica* (Nygaard et al., 2016); tiny myrmicine ant, *Cardiocondyla obscurior* (Schrader et al., 2014); fungus-growing ant, *Cyphomyrmex costatus* (Nygaard et al., 2016); pharaoh ant, *Monomorium pharaonis* (Mikheyev and Linksvayer, 2015; Warner et al., 2017; Gao et al., 2020); red harvester ant, *Pogonomyrmex barbatus* (Smith et al., 2011); workerless inquiline ant, *Pogonomyrmex Colei* (Smith et al., 2015); desert harvester ant, *Pogonomyrmex rugosus* (Smith et al., 2015); slender twig ant, *Pseudomyrmex gracilis* (Rubin and Moreau, 2016); fire ant, *Solenopsis invicta* (Wurm et al., 2011); fungus-growing ant, *Trachymyrmex cornetzi* (Nygaard et al., 2016); fungus-gardening ant, *Trachymyrmex septentrionalis* (Nygaard et al., 2016); fungus-growing ant, *Trachymyrmex zeteki* (Nygaard et al., 2016); Japanese ant, *Vollenhovia emeryi* (Smith et al., 2015); social parasitic ant, *Vollenhovia nipponica* (Smith et al., 2015); South American giant ant, *Dinoponera quadriceps* (Patalano et al., 2015); neotropical ants *Crematogaster levior* (Hartke et al., 2019); wood ant, *Formica exsecta* (Dhaygude et al., 2019)

2.6.4 Microsatellite marker design using genome sequences

According to Park et al., (2019), they analysed 41 published ant genomes and assessed the genome size which resulted in a range of 177 Mbp to 429 Mbp with an average of 297 Mbp. Microsatellite analysis was carried out on the 41 ant genomes

and a total of 2,526,854 microsatellite loci were detected with an average of 61,631 microsatellite loci.

In contrast to previous conventional design strategies, only a restricted number of microsatellite markers for Formicidae was produced. Pearcy et al. (2004) constructed eight microsatellite markers by ligating fragmented DNA into a cloning vector and screening for microsatellites using synthetic tagged oligonucleotides $(TC)_{10}$ and $(TG)_{10}$. Lemos et al. (2020) developed 42 microsatellite markers for *Odontomachus chelifer* by employing hybridization-based capture using $(CT)_8$ and $(GT)_8$ biotin-linked probes to create a microsatellite-enriched library for cloning. Similarly, Kronauer and Gadau (2002) only designed five polymorphic microsatellite markers for honey ant *Myrmecocystus mimicus* by using the enrichment procedure with biotin-labelled $(CA)_{10}$ probe and Arthofer et al. (2007) only produced 20 microsatellite markers for *Lasius austriacus* by using the library enrichment protocol with $(AC)_8$ and $(GA)_8$ biotinylated oligoprobe.

Tsutsui et al. (2008) had estimated the genome size of an ant species by comparing it to the genomes of 40 distinct ant species. The results show that the average genome size across the ant species studied is 361.8 Mbp, with the smallest genome size (*Cerapachys edentata*) being 210.7 Mbp and the biggest genome size (*Ectatomma tuberculatum*) being 690.4 Mbp. Following another whole-genome sequencing effort using Illumina next-generation sequencing platform including six *Aphaenogaster* spp., the mean genome size was determined to be 370.5 Mbp, with the smallest genome size being 310.3 Mb and the largest being 429.7 Mbp (Lau et al., 2018). These findings may be used to estimate the genome size of the undiscovered ant species.