DEVELOPMENT OF SPECIES-SPECIFIC MICROSATELLITE MARKERS OF TERMITE *Globitermes sulphurues* Haviland (BLATTODEA: TERMITIDAE) FOR GENETIC POPULATION ANALYSIS BY USING NEXT-GENERATION SEQUENCING (NGS) APPROACH

NUR AIZATUL NATHASHA BINTI KHIZAM

UNIVERSITI SAINS MALAYSIA

2019

DEVELOPMENT OF SPECIES-SPECIFIC MICROSATELLITE MARKERS OF TERMITE *Globitermes sulphurues* Haviland (BLATTODEA: TERMITIDAE) FOR GENETIC POPULATION ANALYSIS BY USING NEXT-GENERATION SEQUENCING (NGS) APPROACH

by

NUR AIZATUL NATHASHA BINTI KHIZAM

Thesis submitted in fulfilment of the requirements

for the degree of

Master Science

June 2019

ACKNOWLEDGEMENT

First and above all, I am grateful to the Almighty God for granting me the strength and capability to complete my Master journey. Alhamdulillah, thank you God for blessing me much more than I deserve. Also, I would like to express my very great appreciation to my supervisor, Associate Professor Dr. Abdul Hafiz Bin Ab Majid for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated. I wish to thank various people for their contribution to this master research project in term of suggestions, guidance, and times. This research work would not have been completed without them.

Special thanks to my beloved parents, Mr. Khizam Hashim and Mrs. Maizatul Akmal for their countless support, understanding and encouragement throughout my entire life. Praise to God, along with their patience and consecutive prayers, I managed to complete my Master studies with ease. I always knew that they believed in me and wanted the best for me. I also really appreciate how all of my family members including my cousins, Muhammad Abid Rosmadi, Muhammad Adib Rosmadi, Aina Khaireena Rozman and Aida Khairunisa Rozman who have been involved in my research studies process.

Also, I am really grateful for having a supportive and thoughtful fiancee, Mr. Alif Amin Bin Abdullah who kept me accompanied during my sampling days just to obtain my samples collection. I thank you for your part through thick and thin in my life journey.

I would like to take a moment to thanks a group of people, Nurul Akmar, Siti Nor Ain, Fadhlina Hazwani, Faezah Syukriah Sabtu, Husna Noordin and Nor Izzaty

Abd Manan who I feel simply haven't been thanked enough, who helped me in giving ideas and helpful assistance in solving my problems through the process of lab work completion. Thanks for the valuable friendship and experience they bring into my days spent in USM.

Next, I have furthermore to extend my sincere gratitude to Universiti Sains Malaysia (USM) for giving me a great opportunity to be one of the postgraduate students studying in their prestige university. Finally, assistance provided by all staff School of Biological Science, either directly or indirectly was greatly appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

%	percent
°C	degree Celsius
/	per
g	gram
mg	microgram
µl	microliter
µM	micromole
rpm	revolution per minute
bp	base pair
PCR	polymerase chain reaction
PIC	polymorphism information content
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
AFLP	amplified fragment length polymorphism
cDNA	complementary DNA
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
EST	expressed sequence tag
EST-SSR	genic SSR
et al	et-alia – and others
TBE	Tris Boric EDTA
Taq	*Thermus aquaticus*
VNTR	variable number tandem repeat

LIST OF APPENDICES

PEMBANGUNAN PENANDA MIKROSATELIT BAGI SPESIES *Globitermes sulphurues* Haviland (BLATTODEA: TERMITIDAE) UNTUK ANALISIS GENETIK POPULASI DENGAN MENGGUNAKAN PENDEKATAN TEKNOLOGI JUJUKAN GENERASI HADAPAN (NGS)

ABSTRAK

Anai-anai bawah tanah *Globitermes sulphureus* (kumpulan anai-anai tertinggi) terdiri daripada keluarga anai-anai yang luas. Maklumat berkaitan fungsi biologi, pembiakan serta kawalan *G. sulphurues* adalah sangat sedikit. Jadi, adanya urutan keseluruhan, rangkaian *G. sulphureus* mempunyai peluang yang tinggi untuk meningkatkan penanda aras mikrosatelit spesies tersebut. Dalam kajian yang dijalankan, sejumlah 243,057 transkrip telah berjaya dikumpul dari dataset transkrip *G. sulphureus*, dengan jumlah dinukleotida yang paling tinggi. Daripada jumlah ini, 30 penanda mikrosatelit baru telah dicipta menggunakan prosedur penjujukan generasi hadapan (NGS). Kesemua 30 penanda berjaya dihuraikan dan menghasilkan polimorfisme yang tinggi. Nilai maklumat polimorfik (PIC) menunjukkan lebih daripada 0.5 di mana ia telah membuktikan bahawa kesemua penanda mempunyai tahap polimorfisme dengan tiga hingga sembilan alel di setiap lokus. Heterosilikat yang dijangkakan adalah diantara 0 hingga 0.83 dan 0 hingga 0.98. Mikrosatelit khusus yang dibangunkan di sini boleh digunakan bagi meramal kepelbagaian genetik, corak pembiakan dan struktur genetik populasi *G. sulphureus*. Di samping itu, corak pembiakan, kepelbagaian genetik dan struktur genetik populasi *G. sulphureus* masih kurang dikaji. Oleh itu, tujuh loci mikrosatelit yang khusus di dalam kajian ini telah dipilih untuk menerangkan corak pembiakan *G. sulphureus* berdasarkan struktur

koloni di 8 populasi (semulajadi; n = 4 dan metropolitan; n = 4) di negeri Kedah dan Pulau Pinang. Malah, perbezaan genetik dan jarak genetik lapan populasi di Utara Semenanjung Malaysia telah dianalisis. Dari kajian ini, tujuh lokus polimorfik diuji pada sepuluh individu per populasi untuk mengetahui kepelbagaian genetik antara lapan populasi *G. sulphureus*. Lapan populasi tersebut ialah:(1) Taman Jubli (TJ), (2) Palapes USM (PU), (3) Taman Astana (TA), (4) Arkeologi USM (5) Tikam Batu (TB), (6) Sungai Layar Tengah (SL), (7) Kampung Teluk (KT), dan (8) Taman Negara Pulau Pinang (NP). Hasil kajian ini juga menunjukkan bahawa 6 populasi (TJ, TA, AU, TB, SL, dan KT) mempunyai tahap heterozigositi tertinggi, manakala populasi PU dan NP menunjukkan heterozigositi yang rendah. Oleh itu, saya merumuskan kemungkinan populasi PU dan NP mengalami pengerakan genetik dan pemilihan semulajadi. Sementara itu, 6 populasi yang selebihnya (TJ, TA, AU, TB, SL, dan KT) menunjukkan pengambilan assortatif dengan genotip yang berbeza. Khususnya, populasi PU dan NP mempunyai jarak genetik terendah antara satu sama lain (F_{ST} = 0.108). Secara menyeluruh, jurang genetik dalam kajian ini menunjukkan bahawa wujudnya hubungan positif diantara lokasi geografi dengan lapan populasi di Semenanjung Malaysia. Genetik pembiakan dan perbandingan F-statistik serta tahap hubungkait pekerja anai-anai *G. sulphureus* menunjukkan 60% daripada kesemua koloni adalah koloni yang bercampur, sedangkan koloni yang selebihnya adalah keluarga yang berada di peringkat permulaan. Individu dalam koloni campuran kurang berhubungkait antara satu sama lain (F_{IT} = 0.358) jika dibandingkan dengan koloni yang sederhana. Purata persamaan antara koloni campuran dan permulaan adalah sama (r = 0.121). Nilai positif F_{ST} (F_{ST} = 0.086) menunjukkan semua lapan populasi (> 500 jarak) mempunyai populasi anai-anai dengan perbezaan genetik yang sederhana.

DEVELOPMENT OF SPECIES-SPECIFIC MICROSATELLITE MARKERS OF TERMITE *Globitermes sulphurues* Haviland (BLATTODEA: TERMITIDAE) FOR GENETIC POPULATION ANALYSIS BY USING NEXT-GENERATION SEQUENCING (NGS) APPROACH

ABSTRACT

Subterranean termite *Globitermes sulphureus* (higher termite) comprise a wide distributed family of termites, found throughout the tropical regions of the world. Little is known about *G. sulphureus* role in biology, dispersal range, and their control due to lack of research to gather molecular data information. Hence, from the availability of whole sequences of *G. sulphureus*, a great opportunity to develop species-specific microsatellite markers can be done. In this study, a total of 243,057 transcripts were generated from transcriptome dataset of *G. sulphureus,* with dinucleotide repeats being the most abundant. From these, a total of new 30 species-specific microsatellite markers were designed using next-generation sequencing (NGS) procedures. All of 30 markers were successfully amplified and showed visible scorable polymorphism. The polymorphic information content (PIC) values showed greater than 0.5 in which proved that all markers here had high level of polymorphism, with 3 to 9 alleles per locus were identified. The observed and expected heterozygosities ranged from 0 to 0.83 and 0 to 0.98, respectively. The species-specific microsatellites developed in this study can be used to inferred genetic diversity, breeding pattern and population genetic structure of *G. sulphureus*. In addition, the breeding pattern, genetic diversity and population genetic structure of *G. sulphureus* remain poorly understood. Thus, seven developed of species-specific microsatellite loci in this study were chosen to infer the

breeding pattern of *G. sulphureus* based on colony among eight populations (natural; n=4 and metropolitan; n=4) in Kedah and Penang. Besides that, genetic differentiation and genetic distance of eight populations at Northen part of Peninsular Malaysia were investigated. From the population genetics study, seven polymorphic loci were tested in ten individuals per population to find out genetic diversity between eight populations of *G. sulphureus*. Those eight populations were: (1) Taman Jubli (TJ), (2) Palapes USM (PU), (3) Taman Astana (TA), (4) Arkeologi USM (AU), (5) Tikam Batu (TB), (6) Sungai Layar Tengah (SL), (7) Kampung Teluk (KT), and (8) Penang National Park (NP). The findings showed that all 6 populations (TJ, TA, AU, TB, SL, and KT) had the highest heterozygosity level, whilst both PU and NP populations showed less level of heterozygosity. Thus, I suggest that both PU and NP populations might had experienced genetic drift and natural selection. Meanwhile, those 6 populations (TJ, TA, AU, TB, SL, and KT) suggests that an assortative mating with dissimilar genotypes may had occurred in these populations. In particular, PU and NP populations had the lowest genetic distance from each other (F_{ST} = 0.108). Overall, the acquired genetic distance for all eight populations showed positive relation with geographical location in Peninsular Malaysia. Breeding genetic analysis of family structure and comparisons of estimates F-statistics and also relatedness coefficient (*r*) among *G. sulphureus* workers suggested that 60% of all colonies were mixed families, whereas the remaining colonies were simple families. The mixed family colonies had less significantly inbreeding (F_{IT} = 0.358) compare to simple family colonies. Average relatedness values within simple and mixed family colonies were similar (r = 0.121). Positive F_{ST} values (F_{ST} = 0.086) indicated all eight populations (>500 m apart) had termite populations with significant moderate genetic differentiation.

CHAPTER 1

INTRODUCTION

1.1 History of Study

The wood-feeding subterranean termite *Globitermes sulphureus* Haviland (Blattodea: Termitidae) played a large part and important group among other social insects. They are the most widely distributed family of termites, found throughout the tropical regions of the world such as Myanmar, Thailand, Vietnam, Singapore and Malaysia (Ab Majid et al. 2007; Ab Majid and Ahmad 2011; Neoh et al. 2011; Kuswanto et al. 2015). The soldier termite of the species *G. sulphureus* has a brightly, blazing yellowish-colored abdomen. This unique physical appearance is distinctive, making this termite easily recognized from other termite species by researchers. Also, the soldier termites possess an obvious mandibles or pincer-like hooks to capture predators under threat. Meanwhile, worker termite of this species act as soldiers as they have a suicidal self-defense mechanism, thus, took defensed to even further extreme. By bursting their abdomens and covering their enemies with a sticky secretion, they able to protect themselves from any disturbance and danger (Bordereau et al. 1997). In addition, this termite can develop a large structure nests that extend above ground to dominate local landscapes, thereby, contained colonies with millions of members (Veivers et al. 1983). The walls of the nest are mostly covered with a mixture of soil and excrement, and is quite hard, but thin. As social insects, this subterranean *G. sulphureus* termites live in a caste system composed of royal pairs which are known as king and queen, or supplementary reproductive, soldiers, workers, and nymphs (Dronnet et al. 2015).

In particular, *G. sulphureus* species offers great promise as a model in ecology and ecosystem research, which is why knowledge on the genetic background of this species is a pivotal need to select the most appropriate model for particular research studies. Genetic polymorphisms are expected in this study, which affect the outcome of other research field scopes, based on different source of geographical origin. This is due to the fact that *G. sulphureus* colonies used for phylogenetic analysis, foraging population, territory and control management programs are mostly from Indo-Malayan regions (Lee and Ngee et al. 2003; Ab Majid et al. 2011; Ab Majid and Abu Hassan 2011; Neoh et al. 2011).

Over the last century, the use of molecular markers has played an important role in breeding analysis, genetic structure and population genetics. Among different types of molecular markers, microsatellite, also known as short tandem repeat (STR) or simple sequence repeat (SSR) have been utilized most extensively. This is because microsatellite analysis required only small amount of template DNA and simple experimental procedures as they can be readily amplified by polymerase chain reaction (PCR). Despite being highly polymorphic, highly reproducible, high-throughput, tolerance to variation in DNA quality and quantity, widely distributed in genomes and co-dominant inheritance, these microsatellite markers become a famous choice for a variety of genetic studies. Based on the location of microsatellite, these markers can be classified into genomic microsatellite and genic (specific) microsatellite or also known as expressed sequence tag (EST)-microsatellite (Aggarwal et al. 2007; Kalia et al. 2011).

With the advent of bioinformatics, laboratory instruments, sequence databases, sequencing technologies such as Next Generation Sequencing (NGS), advancement and development of species-specific microsatellite marker from transcriptome are

extensively explored. Transcriptome sequencing provide a fast and relatively cost-effective way to generate abundant coding sequence data, offering promising perspectives for the analysis of molecular diversity among species (Hudson 2008; Elmer et al. 2010). In addition, transcriptome serve as a precious source for discovering and identifying polymorphic markers due to information and knowledge obtained in developing microsatellite markers. Because of species-specific microsatellite markers require easy and inexpensive procedures, they can facilitate better cross genome comparisons, a greater probability of being associated with functional genes and greater transferability to related species. Therefore, this study focusing more on identifying and developing species-specific or genic microsatellites compared to genomic microsatellites. By having massive amounts of information derived from polymorphism variation of developed species-specific microsatellites, breeding pattern, genetic structure, and population structure of *G. sulphureus* termites able to explore extensively.

1.2 Problem Statement

The subterranean termite of *Globitermes sulphureus* are group of particular concern in this research because they account for majority of termite damage worldwide (Rust and Su 2012). In a current report, the wood-feeding subterranean *G. sulphureus* termites were mentioned as additional and secondary pests of oil palm and coconut plantations, replacing the primary pest species infestations such as *Coptotermes gestroi* termites. Following the elimination of primary pest species by using baiting system, it is difficult to find species from other genera infesting the same structure or building after the treatment for the time periods. However, in Malaysia, some of the species especially from the genus *Globitermes* (Termitidae) do not respond

well to bait system introduced by several pest control companies. Hence, make this *G. sulphurues* termite species to encounter as secondary pest frequently at both native and grounded areas.

Interestingly, there are also several reported cases of *G. sulphurues* infestations in metropolitan areas showing that this species has successfully occupied outside their primitive ranges, and cause environmental and economic harms to the infested regions through aggressive consumption of wooden or human-made structures and infestations of living trees (Ghaly and Edwards 2011; Perdereau et al. 2011; Rust and Su 2012; Evans et al. 2013). For instance, there was a case reported in Malaysia and Singapore where *G. sulphureus* species are started to infest around urban and suburban areas (Lee and Ngee 2003; Neoh et al. 2011). These conditions become one of the major challenges in managing genera *Globitermes* termite faunas as baiting system usage are generally not effective against this subterranean *G. sulphureus* species. Because of this, several pest control companies in Malaysia are still in the effort to produce suitable and efficient pesticides and bait, particularly for this genus *Globitermes*.

Subterranean *G. sulphureus* termite lives in colonies, yet we still have a poor understanding about number and relatedness among reproductive within colonies and the dynamics of colony-colony interactions, particularly between native and metropolitan regions. Besides that, little is known about *G. sulphureus* role in ecology, biology, dispersal range, and control of G. *sulphureus* due to lack of research to gather molecular data information. Knowledge describing the genetic diversity and genetic structure of this termite species is limited despite majority of *G. sulphureus* termite studies have taken place in surface research only. This research established in order to provide numerous information regarding population genetic analysis which consist of dispersal pattern, breeding pattern, gene flow, genetic distance and genetic

differentiation. By using species-specific microsatellite markers, genetic variation between and within population and the molecular ecology study of *G. sulphureus* can be conducted.

1.3 Significance of Study

This study was conducted to investigate *Globitermes sulphureus* infestation in selected natural and metropolitan region, specifically around Kedah and Penang as they were believed to have resurged from their primitive ranges and invade a new environment as they sometimes attack construction structures, premises and business operation particularly in provincial and metropolitan areas (Ab Majid et al. 2011; Lee and Ngee et al. 2003; Neoh et al. 2011). Also, this study will provide significant amount of information on the availability of species-specific microsatellites related to *G. sulphureus* termites. Besides that, identification of microsatellites related to *G. sulphureus* termites will help to evaluate the disturbance effects on an ecosystem and may be used as a potential indicator to investigate climatic change in a given area. Dispersal ability and breeding system of *G. sulphureus* are among important factors in determining the gene flow between and within population(s). Hence, provide better understanding about genetic structure and population genetic of *G. sulphureus* species.

1.4 Hypothesis

Species-specific microsatellite markers can be developed from genomic sequences of *Globitermes sulphureus* using next generation sequencing (NGS) approach that will give vast information and insight into the breeding system, genetic diversity, genetic structure and population genetic of this *G. sulphureus* species.

1.5 Objectives

In this present study, the objectives were divided into three parts;

Objective 1 (Chapter Three): To develop species-specific microsatellite markers from transcriptome dataset of *Globitermes sulphureus*.

Objective 2 (Chapter Four): To characterize the species-specific microsatellite polymorphism in *Globitermes sulphureus*.

Objective 3 (Chapter Five): To identify and determine population genetic analysis, genetic structure and breeding pattern of *Globitermes sulphureus* from selected native and grounded regions throughout Northen part of Peninsular Malaysia.

CHAPTER 2

LITERATURE REVIEW

2.1 Higher Group Termites: History, Caste system, Colony dynamics

In general, termites (order Blattodea) can be divided into 280 genera, 7 families and 14 subfamilies (Eggleton et al. 1994; Kambhampati and Eggleton 2000; Ahmed and French 2008). Termites (Termitidae) are eusocial insects that evolved from cockroaches, and that the wood-feeding termite genus *Cryptocercus* is the sister group (Lo et al. 2000; Eggleton 2001; Inward et al. 2007; Korb 2008; Rust and Su 2012; Brune 2014). Termites and cockroaches belong together as they share the same ancestor in which can proved by DNA analysis a few decades ago when the researchers noticed the specialized microbes in the guts of termites are also present in the guts of some cockroaches. However, termites contain more diverse diet and sophicated social system than cockroaches (Beccaloni and Eggleton 2013; Smart News 2018). Furthermore, they also constitute an ecologically and evolutionary diversified group of social insects, approximately more than 2600 termite species that share a common ancestry with cockroaches. Krishna et al. (2013) stated that higher termites are the most diverse of all termite families as they are representing over 85% of all termite genera globally.

The evolution of higher termites become a particular concern because of their diet diversity, social structures, and phenotypes. For termite diets specialization, the gut microbiota of termites plays main roles in the symbiotic digestion of lignocellulose (Watanabe and Tokuda 2010; Brune 2014). Unlike lower termite, higher termites have harbored whole prokaryotic communities in their dilated hindguts, particularly in family Termitidae (Ohkuma and Brune 2011; Bourguignon et al. 2014; Dietrich and

Brune 2014). Besides that, higher termites promote the unique opportunity for researchers to investigate the relationship between co-evolution of gut microbiota and host (Brune and Ohkuma 2011; Brune 2014).

Termitidae are classified into four subfamilies which known as Macrotermitinae, Nasutitermitinae, Apicotermitinae and Termitinae. These group are reported to have a broad variety of genera in Asia, thereby offer an opportunity to examine the evolutionary history of termite features. This is because Asia is reported to be the best-studied region for termites with respect to their taxonomy and distribution among Asian tropics, Afrotropics and Neotropics (Ohkuma et al. 2004; 2008). The family Termitidae (higher termites) contained a large number of species, particularly in tropical ecosystem due to level of latitude. Aside from being economically important insects, another characteristic of termites allowing them to successfully diversify worldwide is their advanced system of social organization. By understanding the reproductive division of labor, the societies and diversities of termites can be characterized. Termites live in a caste system comprised of royal pairs (king and queen) being supported by supplementary reproductive, non-reproductive workers, soldiers and nymphs (Sen et al. 2013). The worker termites placed at the bottom of the termite caste system. Meanwhile, soldier termites protect the colony especially the royal pairs by using their strong and powerful mandibles.

Generally, termite colonies are headed by a pair of primary or winged reproductive and sometimes headed by secondary or neotenic reproductive (Bulmer and Traniello 2002). Colonies that headed by monogamous couples of winged reproductive (alates) and their non-reproductive offspring are called ‘simple families’, or in simple way single queen mated to a single male. Due to the immobility characteristics of neotenics or larvae instars, they unable to fly, stay and mate in the

nest which later engage in inbreeding (Dronnet et al 2005). From the neotenics offspring, ‘extended families’ are built. Additionally, colonies from ‘mixed families’ can be resulted from the fusion between two or more colonies as they are composed of offspring from multiple unrelated reproductive (Lain´e and Wright 2003; DeHeer et al 2005; DeHeer and Vargo 2004; Perdereau et al. 2010).

2.2 Foraging mechanism of subterranean termites

Subterranean termites are unique in that their nests require active contact with the soil which they tunnel through to locate resources and survive (Suiter et al. 2002). In early studies, the foraging behaviour of subterranean termite consists of the extensive underground tunnel system. From the observations, the interactions between termites together with the density and flow rate of individuals through tunnels can formed complete tunnel propagation (Bardunias and Su 2010). This termite tunnels are constructed by a group of workers removing soil particles at tunnel tips which can be used for food searching. Subterranean termites have limited movement as they confined within their gallery systems (Su 2019). Also, they have to travel back to their nest sites using the same tunnels after encountering foods. These factors contributing in the construction and optimization of tunnel pattern for both search and transport efficiency for subterranean termites (Rust et al. 2012).

2.3 Biology of *Globitermes sulphureus* Haviland, 1898

2.3.1 Taxonomy of *Globitermes sulphureus*

The scientific taxonomic classification of *Globitermes sulphureus* according to Catalogue of Life following the Integrated Taxonomy Information System (ITIS) is as below:

Kingdom : Animalia
Phylum : Arthropoda
Class : Insecta
Order : Blattodea
Superfamily : Blattoidea
Family : Termitidae
Genus : *Globitermes*
Species : *Globitermes sulphureus* Haviland, 1898

2.3.2 Morphological characteristics of *Globitermes sulphureus*

Generally, subterranean termite *G. sulphureus* are medium-sized, soft-bodied and light-colored. The soldiers of *G. sulphureus* have bright, blazingly yellowish-colored abdomen. Their soldiers possess long thin mandibles of the reaping type and have a defensive gland extending into the thorax and the forepart of the abdomen. They secreted and released chemical yellow liquid as a weapon of defense and also used their mandibles when they bite or under threat. As for the workers of this species, they commonly have slightly small body size compared to their soldiers. Also, they have black or brownish-coloured at the tip of their abdomen (Bordereau et al. 1997).

2.3.3 Nesting and Foraging territory of *Globitermes sulphureus*

Instead of less efficient, triple mark recapture technique, new single-mark recapture technique had been introduced to modify and identify the foraging territories of *G. sulphurues* species (Lee and Ngee 2003). This finding showed that *G. sulphureus* has foraging distance between 5.7 meter and 10.5 meter (Lee et al. 2002, 2003). Based on the research stated earlier, higher termite such as *G. sulphurues* generally tend to have shorter foraging distance compared to lower termites. This species can build a mound which may reach one to one and half meters in height. A thin bark covering the nest and a wall made of a dense network of fine and large galleries protect the colony.

2.3.4 Defensive system of *Globitermes sulphureus*

Most of the termite soldiers used developed defensive strategies such as mechanical and/or chemical weapons to protect termite colonies. As for subterranean *G. sulphureus* termites, the preliminary study indicated that the suicidal behavior by chemical liquid release (autothysis) observed in their soldiers (Brian 2012). The secretion of yellow liquid was believed to originate mainly from hypertrophied salivary glands, thus ejected through the mouth. Some of the reports stated that these secretions can be found among aggressive soldiers showing violent contractions of their whole body or ruptured salivary gland wall and the integument near the basis of the soldier's legs.

However, to date, observations on defensive strategies of this termite soldiers had been carried out intensively. The results revealed that the defensive behavior of this species is not always suicidal and that the secretion is released not from salivary glands, but from a very modified frontal gland in which undergoes an actual metamorphosis during the pre-soldier stage. Under scanning electron microscope, the

location of the defensive gland of *G. sulphureus* corresponds to that of hypertrophied frontal glands observed in some species of Rhinotermitidae. The gummy, rubbery exudation accumulates as a droplet at the inner face of the head capsule.

Apart from autothysis mechanism, the soldier of this termite species used their strong mandibles to fight their enemies without suicidal behavior. Also, they had the capabilities to release an alarm pheromone recruiting congeners on the site of the fights. This multi-defensive strategy developed by *G. sulphureus* appears to be more efficient, thereby becoming important element contributing in the ecological success of their population in Southeast Asia.

2.3.5 Research studies of *Globitermes sulphureus*

In Malaysia, most studies on *G. sulphureus* were conducted to examine the colonies foraging activity, feeding territory and control management by using termiticide applications (Lee et al. 2002, 2003; Ab Majid et al. 2007, 2011). In other Asia countries like Vietnam and Thailand, study on defensive behavior and nitrogen fixation quantification of *G. sulphurues* at the ecosystem scale are done (Bordereau et al. 1997; Yamada et al. 2005). There is a genetic study involving the construction of phylogenetic analysis which based on 16S rRNA (Ab Majid and Husin 2017). Ab Majid and Husin (2017) determined diversity of termite gut bacteria from different *G. sulphureus* mounds using a molecular approach based on16S rRNA genes and found that most of the bacteria from one colony were slightly but distinctly different with bacteria from other colonies. In another study by Ab Majid et al. (2007), where preliminary field observation on *G. sulphurues* towards the efficiency of slow acting termiticide, Premise® 200SC containing imidacloprid was examined. The final external treatment study revealed that Premise® 200SC caused the termite activity to cease about six weeks after the treatment. However, increased efforts are done in

present study to control *G. sulphureus* infestations in which biological control agent such as fungal biocontrol agents proved to be an alternative application (Rath 2000).

2.4 Transcriptome

Transcriptome is a complete set of ribonucleic acid (RNA) molecules or transcripts, that is, at a given moment, present in a given organism, organ, tissue or cell (Tomohiro 2014; Daniel et al. 2018; Anuj et al. 2019). It is a composition of coding and non-coding RNAs, and their degradation products (Susan et al. 2016; Anuj et al. 2019). The level of messenger RNA (mRNA) expression was examines by transcriptome profiling where it uses technology such as DNA microarray. It displays a complete collection of mRNAs present under defined conditions. In particular, the transcriptome is regulated by the process called transcription where it undergoes constant qualitative and quantitative changes. Thus, reflects the changes in expression levels during development and under different conditions.

With the help of several technologies like massively parallel signature sequencing (MPSS), cDNA or oligonucleotide microarrays, serial analysis of gene expression (SAGE), RT-PCR (Reverse transcription PCR) on RNA extract, RNA microarray and construction of EST (Expressed Sequence Tag) collection (Lowe et al. 2017), transcriptome analysis can be done efficiently. The composition of transcriptome analysis able to provide an insight into the way which organisms function. Most researchers can interpret the functional elements of the genome due to the massive amount of genetic information embedded in the transcriptome dataset. Also, researchers can generate a comprehensive insight into the specific cell type constituents, how the cells normally function and genes activity changes which can reflect or contribute to development and disease (Adams and Atkinson 2008).

Ward et al. (2012) states that next generating sequencing (NGS) is one of the most suitable approaches in de novo transcriptome sequencing. This is because next-generating sequencing has emerged as a low-cost, large-scale, fast, and accurate over the past several years. From the transcribed sequences such as ESTs or transcriptome, microsatellite can be derived for a wide variety of applications, including link-age and QTL mapping, comparative genomics and studies of genome evolution (Liu and Cordes 2004). Transcriptome sequencing had been a better alternative for developing genic microsatellite markers over conventional methods which are tedious and only produce few useful markers (Xu et al. 2016). It is particularly favourable for obtaining information on gene expression that can greatly facilitate molecular and genomic studies of species where a sequenced genome is not available (Ryan and Stephen 2014).

2.5 Recognition of Next Generation Sequencing (NGS)

During the past decades, automated Sanger technique is considered a first-generation technology in the 1970s which designed by Sanger and his colleagues to sequence deoxyribonucleic acid (DNA) by chain termination and fragmentation methods, respectively (Chain and Heather 2016). The method commonly referred to as Sanger sequencing and became the prevailing DNA sequencing method for the next several years replacing the pioneering technology introduced by researcher Maxam and Gilbert (2017) which involved chemical degradation of DNA or the chemical cleavage method at specific base (Figure 2.1) (Broadgate 2017; Rowe et al 2017; Ghosh et al. 2018). Limitations from early technologies or the first-generation genome sequencing technologies in term of hazardous toxic chemicals, radioactivity, and acrylamide gels usage (Maxam and Gilbert method), routine genome sequencing for research and clinical usage, average throughput, relatively small size of the genomic areas which could be sequenced in a reasonable amount of time (Sanger method) (Shendure et al. 2017) can caused some difficulties to analyse genome sequencing with high level of accuracy and efficiency with regards to time and cost requirements (Sabatini et al 2016). Therefore, a growing interest and demand for higher throughput with inexpensive technologies has led to the establishment and commercialization of next-generation sequencing (NGS), also known as second-generation technologies (Kulski 2016; Lavezzo et al. 2016; Besser et al. 2018). Massively parallel, next generation sequencing (NGS) or deep sequencing are related terms that describe a DNA sequencing technology which has revolutionised genomic research at an unprecedented speed (Bruger and Marx 2018; Rabbani 2016; Salk et al. 2018). Generally, this well-known technology can be used to sequence whole genomes or constrained to specific regions of interest, including a whole exome or small numbers

of individual genes (Behjati and Tarpey 2013). By using a DNA template, they were able to develop a complementary DNA strand during sequencing (sequencing-by-synthesis or SBS) and optical detection (Chain and Heather 2016; Goodwin et al 2016). Besides that, there are a few studies reported that these new NGS sequencing technique share three major improvements (Mardis 2008;2017). Initially, they depend on the preparation of NGS libraries in a cell free system instead of requiring bacterial cloning of DNA fragments. Next, there are numerous numbers of sequencing reactions can be obtained in parallel instead of hundreds found in principles of Sanger sequencing. Last but not least, sequencing output is directly detected without the need for electrophoresis and base interrogation is performed cyclically and in parallel (Van Dijk et al. 2014).

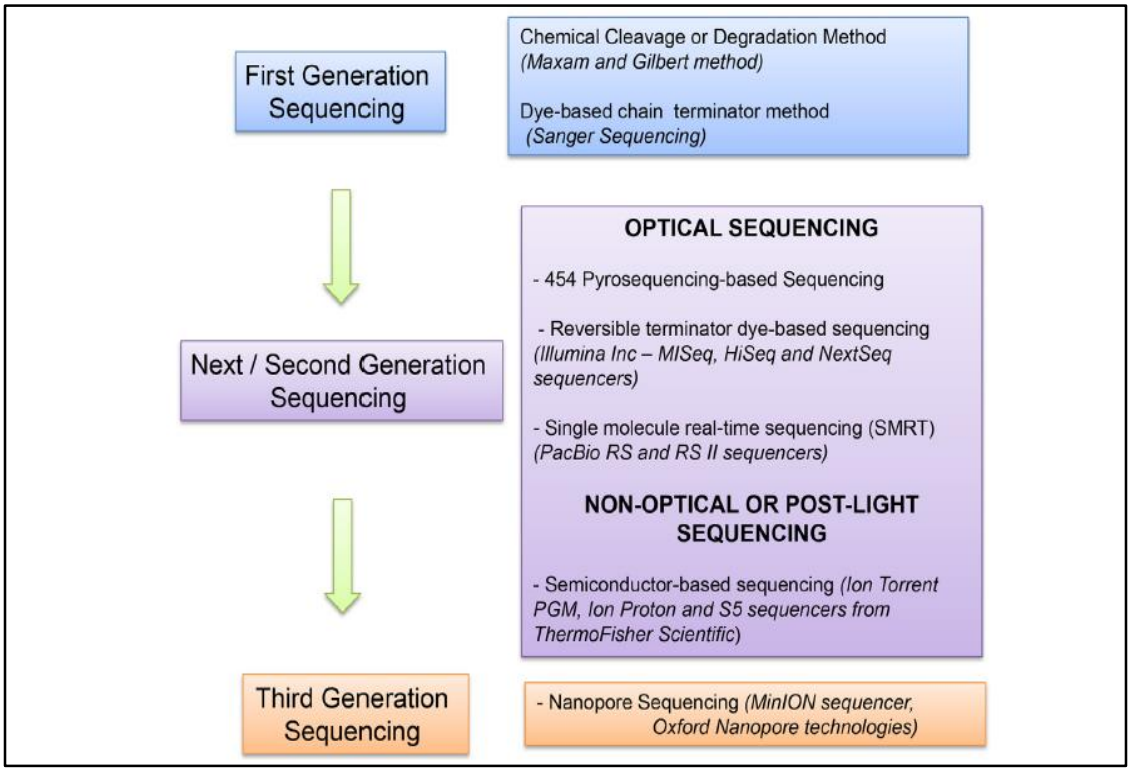

Figure 2.1 Schematic diagram representing the arrival of DNA sequencing technologies which has revolutionised genomic research at an unprecedented speed. (Metzker 2010).

2.5.1 Potential uses and Practises of NGS

The recent approach of next generation sequencing (NGS) as one of the capable instruments to process greater number of sequences reads in parallel in a single run is rapidly altering the prospect of genetics, providing the ability to answer questions in disciplines ranging from evolution and ecology to conservation and agriculture with a peculiar speed (Andrews and Luikart 2014). The enormous increase of NGS sample throughput due to the significant efficiently and advancement in data handling, thus, provided an easier way for researchers especially in bioinformatic analysis and management. In contrast, the previous Sanger sequencing method, used to translate the human genome, took a longer time to deliver the final draft. In addition, the barrier of the expensive cost for genome analysis has recently been broken since the impressive progress in NGS have enabled an immense diversity of novel. For example, there are

numerous studies for investigations of non-model insects by using NGS-based approaches (Ekblom and Galindo 2011; McCormack et al. 2013a, 2013b) which comprised a wide range of insect taxonomy, ecology, their natural history, phylogenetic relationships, phylogeographic interest, genetic diversity patterns, and genetic regions (McCormack et al. 2013; Nadeau and Jiggins 2010; Stapley et al. 2010; Rice et al. 2011). Hence, they provide massive amount of information and knowledges specifically in entomology field.

Concurrently, recent efforts from researchers proved that NGS technology has facilitated the advancement of new genome-assisted advent for correlating genotype and phenotype. Thus, opening up in area of food microbiology by understanding the system level of food microorganisms, able to predict the prevalence of microorganisms in food samples and also control the growth and survival of desirable or undesirable microorganisms in food when we elucidate the molecular basis of how microorganisms respond to different food-associated conditions (Solieri and Giudici 2013). To date, there are limited publications available that describe the application of NGS technology to mitochondrial DNA (mtDNA) testing in the forensic context. With the emergence of Next Generation Sequencing (NGS) technologies, larger mtGenome databases containing mtGenome information able to establish in relatively short terms, thereby facilitating phylogenetic backgrounds and geographic distributions in a forensic genetic case. On the other hand, this advancing NGS technologies had contributed in research and clinical applications where it has potential to revolutionize disease research, forensics, and clinical medicine (Butler et al. 2016; McKiernan and Danielson 2017). In conclusion, NGS technology is believe to become a rapid technique in investigating and measuring any potential therapies

development for a wide variety of cancer and other diseases and also to identify new treatment options after refining patient prognosis.

2.6 Molecular or Genetic marker

Molecular or genetic marker is a sequence of DNA or protein that can be screened to reveal key attributes of its state or composition and thus used to reveal genetic variation (Sunnucks 2000; Burange et al. 2015). Indeed, DNA sequences determine the diversity of organisms, and therefore, the methods used to evaluate DNA polymorphisms directly measure the genetic diversity. Due to the Mendelian inheritance in molecular markers, it is possible to trace the fingerprint of each organism and determine the evolutionary history of the species by phylogenetic analysis, studies of genetic relationship, population genetic structures and genetic mapping. There are two types of molecular marker which recognised as DNA markers and protein markers. First, DNA markers produced due to the mutation arises which causes genetic variation at DNA level. Subsequently, genetic variation at DNA aroused due to mutation will cause variation in protein (protein markers). Hence, can cause differences in basic and acidic amino acid composition without disturbing their enzyme function. Besides that, there is also little changes in primary structure which can alter secondary and quaternary structure of the protein. In particular, these genomic DNA (gDNA) mutations are heritable and become a source of new variation (polymorphism). They can be classified into: (a) base substitution, (b) deletion, (c) insertion, (d) duplication and (e) inversion (Figure 2.2). Recently, a variety of molecular markers, including restriction fragment length polymorphisms (RFLPs), random amplification of polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), minisatellites and microsatellites or simple sequence repeats (SSRs), have been developed in different crop plants and animals. The rapid

development of molecular techniques offers a palette of technical approaches for population biologists to better understanding on how variations in survivorship, fertility, and gene flow contribute to changes in allele frequencies within and among populations. Besides that, molecular markers also allow precise and rapid varietal identification, which has been proved to be an efficient tool for crop germplasm characterization, collection and management. Furthermore, variation found in molecular markers contained characteristic biological properties that can be detected and measured in any parts of the body such as the blood or tissue at any stage (Haley and Koning 2006; Singh et al. 2014) and they are not confounded by environment, pleiotropic, or epistatic effects (Agarwal et al. 2008).

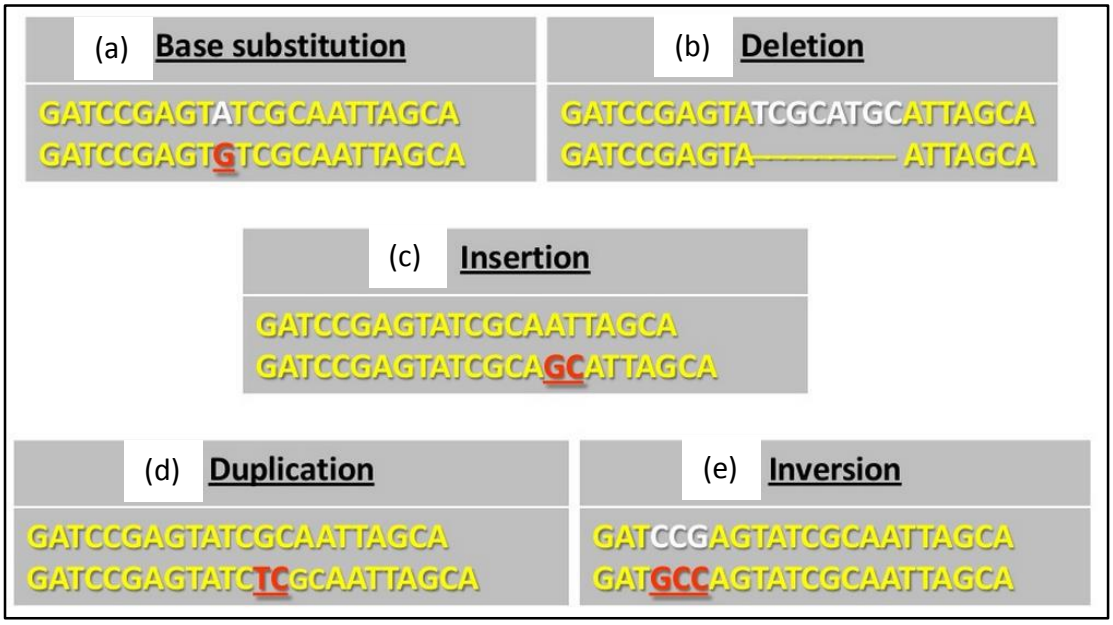

Figure 2.2 Classification of genomic DNA mutation phenomena.
(Wright et al. 1995).

2.7 Microsatellite marker

Simple sequences (Tautz 1989), Simple Sequence Repeats (SSRs) (Jacob et al. 1991) and Short Tandem Repeats (STRs), (Edwards et al. 1991) are several other names for the microsatellite marker (Lee et al. 2004). Microsatellites are stretches of DNA sequences consisting of short repeats of highly variable number of 1 to 7 base pair (bp) in length, repeating itself a number of times that can be found in the organism's DNA (Thiel et al. 2003; McDonald 2005; Garrido 2017). They also distributed evenly throughout eukaryotic nuclear, chloroplasts, and mitochondrial genomes. In addition, microsatellite markers are ubiquitous in the coding and noncoding regions (Tautz and Renz 1984a; Toth et al. 2000) with a higher density of simple sequence motifs in the noncoding regions of eukaryotes (Katti et al. 2001; Li et al. 2002; Zhang et al. 2004; Cruz et al. 2005). In particular, they referred to a class of highly mutable genomic sequences known as variable number of tandem repeat (VNTR) elements (Buschiazzo and Gemmell 2006; Chambers and MacAvoy 2000) and also simple sequence length polymorphisms (SSLPs) (Jarne and Lagoda 1996). This is because they contain multiple redundant tandem repeats, for instance in human genome in which 50, 000 short tandem repeats are found and approximately 5000 had been characterized as molecular markers.

Also, this marker can be divided into genomic microsatellite marker and genic (specific) microsatellite marker which depend on the location of microsatellite in the genome, thereby, determines its functional role (Lawson and Zhang 2006). These have the potential to affect all aspects of genetic functions including gene regulation, development and evolution (Kashi and King 2006; Lawson and Zhang 2006). A microsatellite located in a coding region can affect the activation of a gene and therefore, the expression of a protein. If located in a no coding or genic region, the

microsatellite may impact gene regulation or gene transcription (Lawson and Zhang 2006; Vieira et al. 2016). However, in contrast with study of insects, microsatellite length and frequency were suggested to have correlation with the size of genome (Hancock 2002; Toth et al. 2000).

The generation of microsatellite are facilitated with the availability of microsatellite mining tools such as TROLL (Castelo et al. 2002), MISA (Thiel et al. 2003), SciRoKo (Kofler et al. 2007), and MSATCOMMANDER (Faircloth 2008). These software able to search and identify microsatellite repeat motifs in the expresses sequence tag sequences or databases. Hence, exploit the possibility of converting it into polymorphic microsatellite markers.

2.7.1 Types and Classifications of microsatellite

Microsatellite can be classified into four categories which based on: (a) occurrence and source for development, (b) type of repeat sequence, (c) length of repeat motif and (d) arrangement of nucleotides in the repeat motif (Figure 2.3).

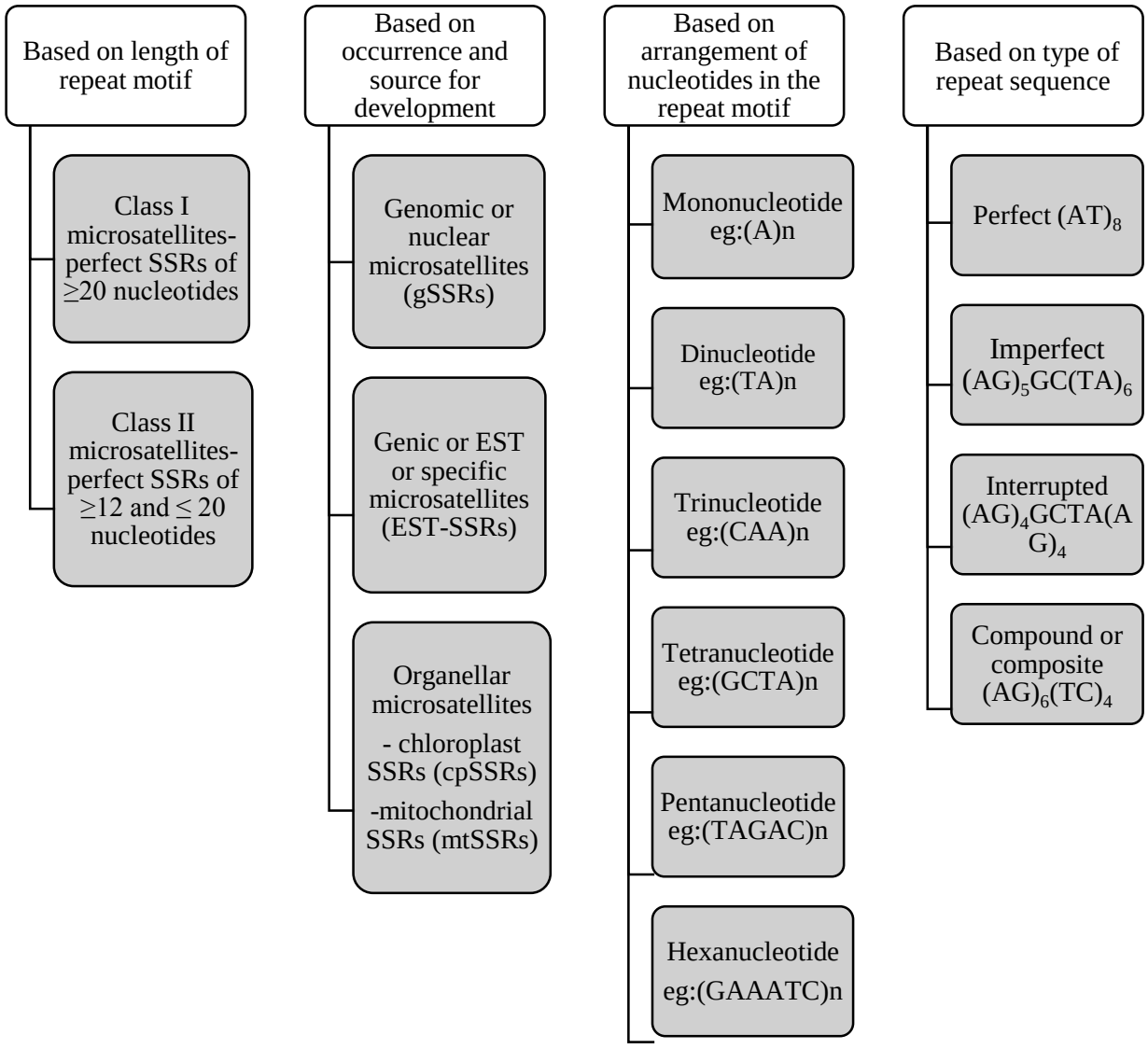

Figure 2.3 Types and classifications of microsatellite marker.
(Charlesworth 2013).

With respect to the length of repeat motif, microsatellites can be classified as Class I microsatellite where the perfect microsatellites are more than 20 nucleotides in length or Class II microsatellite where the perfect microsatellites are less than 20 nucleotides in length but more than 12 nucleotides in length (Temnykh et al. 2001; Takezaki 2017). Based on occurrence and source for development, microsatellites isolated from the nuclear genome in which the genomic DNA of an organism with or without the construction of genomic DNA library will produced genomic or nuclear microsatellite (g-SSRs) (Abdelkrim 2009; Santana et al. 2009). Meanwhile, by exploiting or data-mining expressed sequence tag (EST) stored in public databases, genic or EST microsatellites (EST-SSRs) are obtained (Aggarwal et al. 2007; Yasodha 2008). With respect to the arrangement of nucleotides in the repeat motif, microsatellite can be classified as mono-, di-, tri-, tetra-, penta- or hexanucleotide repeat motifs (Li et al. 2002; Lai and Sun 2003; Kofler et al. 2007; Cavagnaro et al. 2010). According to the arrangement of nucleotides within the repeat motifs, type of microsatellite sequences also can be determined: (a) perfect repeats, when showing only perfect repetition, (b) imperfect repeats, when a pair of bases that are not repeated interrupted the repeated sequences, (c) interrupted repeats, when a small fragment of sequence that are not repeated interrupted the repeated sequences, and (d) compound or composite, when there are two or more distinctive sequence-repeats (Toth et al. 2000; Ferguson et al. 2004; Kalia et al. 2011).