POPULATION STRUCTURE BASED ON MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT I (COI) GENE IN THE BLACK SCAR OYSTER, *Crassostrea iredalei* (FAUSTINO, 1932) IN MALAYSIA

by

DANIAL HARIZ BIN ZAINAL ABIDIN

Dissertation is submitted in fulfillment of the requirements

for the degree of

Master of Science

School of Biological Sciences

Universiti Sains Malaysia

Pulau Pinang, Malaysia

DECEMBER 2013

ACKNOWLEDGEMENTS

In the name of Allah, most Gracious, most Compassionate

Alhamdulillah, with the doa' from Mama and Babah, I have completed my Master's thesis with success.

My deepest gratitude and appreciation to my main supervisor, Dr. Darlina Md. Naim for her guidance, encouragement and concern throughout this project. I'm also indebted to my co-supervisor, Prof. Siti Azizah Mohd. Nor who had provided the opportunity to be a part of this project. Without her guidance and persistent help, this dissertation would not be possible. Special thanks to Prof. Peter Mather from Queensland University of Technology, Australia for his guidance and assistance in developing my microsatellite markers.

I wish to express my sincere thanks to Kak Masa and the Fisheries Research Institute (FRI) Malaysia for samples and information obtained for this study. I take this opportunity to record my heartfelt gratitude to all Lab 308 members, especially Kak Adel, Kak Adib, Su Yin, Kak Mira, Balkhis, Semah, Fong, Marcela, Lim, Kak Ana, Faisal, Jam, and the rest of the lab members, for their advice, assistance, and time given for answering my silly questions sometimes. Not forgetting to all my friends, thank you for your encouragement and inspiration. For me, friends are relatives you make for yourself, and guys.... you are more like brothers and sisters to me.

Last but not least, my profound love to both of my parents, Hj. Zainal Abidin and Hjh. Faridah Osman and families, Abang, Kak Wan, Acu and Kak Azie for their unconditional love, inspiration, and support. Finally, this thesis is completed with so much love and effort from everyone that is important in my life.

TABLE OF CONTENTS

LIST OF TABLES

PAGE

LIST OF PLATES

PAGE

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp	basepair
DNA	Deoxyriboncleic acid
dNTP	Dinucleotide triphosphate
EDTA	Ethylenediamine tetra-acetic acid
EtBr	Ethidium bromide
HCl	Hydrochloric acid
k	Number of groups
kb	kilobase
M	Molar
NaCl	Sodium chloride
SDS	Sodium dodecyl sulphate
TBE	Tris-borate-EDTA
TNES-Urea	Tris-sodium chloride-EDTA-SDS-Urea

LIST OF PUBLICATIONS AND PROCEEDINGS

1.1 Danial Hariz Z. A., Siti Azizah M.N., & Darlina M. N. (2013). Genetic Variation of the Malaysian Oyster: Revealing the Traditional Methods of Harvesting Oyster. In L. R. Norizan Esa, Zuraidah Mohd. Yusoff (Ed.), *Reengineering Local Knowledge - Life, Science and Technology* (pp. 43-50). Malaysia: Penerbit Universiti Sains Malaysia.

1.2 Ahmad Faisal G., Danial Hariz Z.A., Siti Azizah M. N., & Darlina M. N. (2012). Genetic variation of Indian Mackerel (*Rastrelliger kanagurta*) (Cuvier, 1816) of Sabah waters based on mitochondrial D-loop region: A preliminary study. *Asian Journal of Biology and Biotechnology, 1*(1), 1-10.

1.3 Danial Hariz Z. A., Masazurah A. Rahim, & Siti Azizah M. N.. Molecular Approach to Identify Puffer Fish from Malaysian Waters. (Proceeding of the 5th International Barcode of Life Conference, 27th-31st October 2013, Kunming, China). Poster presentation.

1.4 Danial Hariz Z. A., Darlina M. N., Masazurah A. Rahim, & Siti Azizah M. N.. Genetic diversity of the black scar oyster, *Crassostrea iredalei* in Peninsular Malaysia. (Proceeding of the International Fisheries Symposium – IFS 2012, 6th-8th December 2012, Can Tho, Vietnam). Oral presentation.

1.5 Danial Hariz Z. A., Siti Azizah M. N., Masazurah A. Rahim, Suzana Mustaffa, & Darlina M. N.. Phylogenetics and population structure of the black scar oyster, *Crassostrea iredalei* in Peninsular Malaysia as revealed by mitochondrial COI sequences. (Proceeding of the 8th IMT-GT UNINET Biosciences Conference, 22nd-24th November 2012, Darussalam – Banda Aceh, Indonesia). Oral presentation.

1.6 Danial Hariz Z. A., Siti Azizah M. N., & Darlina M. N.. Variasi Genetik Tiram Di Semenanjung Malaysia: Pendedahan Perolehan Tiram Secara Tradisional. (Proceeding of the Regional Conference of Local Knowledge, 15th-16th October 2012, Pulau Pinang, Malaysia). Oral presentation.

STRUKTUR POPULASI BERDASARKAN GEN MITOKONDRIA SITOKROM OKSIDASE SUBUNIT I (COI) DALAM TIRAM, *Crassostrea iredalei* (FAUSTINO, 1932) DI MALAYSIA

ABSTRAK

Gen mitokondria sitokrom oksidase subunit I (COI) telah digunakan untuk menilai genetik populasi dan hubungan filogeografi tiram, *C. iredalei* di kalangan 11 populasi di Semenanjung Malaysia dan Sabah. Secara keseluruhan, populasi *C. iredalei* menunjukkan tahap diversiti nukleotida, π yang rendah (0.000 – 0.004) dan diversiti haplotip, *h* dalam julat nilai rendah ke tinggi (0.000 – 0.795). Penstrukturan genetik telah dikesan antara populasi Semenanjung Malaysia dengan populasi Sabah seperti yang ditunjukkan oleh analisis F_{ST} dan AMOVA. Walaubagaimanapun, analisis gen COI menunjukkan perbezaan yang minimum dan tidak signifikan (P> 0.05) antara populasi di dalam kawasan Semenanjung Malaysia dan Sabah. Kehomogenan secara relatif telah dilaporkan antara populasi pantai masing-masing di Semenanjung Malaysia dan Sabah. Hal ini boleh dikaitkan dengan penyebaran larva yang tinggi di sepanjang pantai barat dan timur dan translokasi benih tiram oleh manusia antara kedua-dua persisiran pantai bagi tujuan pernyemaian benih *C. iredalei*. Analisis neutraliti dan taburan 'mismatch' menyokong hipotesis bahawa *C. iredalei* telah mengalami satu/ beberapa peristiwa pencerutan populasi, diikuti oleh pengembangan populasi. Maklumat molekul yang diperolehi dalam kajian ini boleh digunakan dalam strategi pengurusan pragmatik akuakultur stok induk liar dan ternakan *C. iredalei* di Malaysia. Satu pendekatan penjujukan generasi baru ('Next Generation Sequencing' - NGS), teknologi Ion Torrent telah digunakan dalam pembangunan penanda mikrosatelit baharu untuk *C. iredalei*. Walaupun sejumlah 2424 motif mikrosatelit terulang telah dikenal pasti, hanya 137 layak untuk membentuk primer dalam kajian ini. Walaubagaimanapun, pencirian dan amplifikasi

penanda mikrosatelit yang telah dikenal pasti tidak berjaya kerana setelah empat lokus dipilih untuk pencirian populasi, keputusan analisis fragmen yang diperolehi adalah tidak memuaskan. Namun, data ‘Genome Survey Sequence’ (GSS) yang diperolehi dari teknologi Torrent Ion di dalam kajian ini boleh digunakan dalam pelbagai aplikasi masa hadapan untuk penambahbaikan dan pemuliharaan akuakultur *C. iredalei* di Malaysia.

POPULATION STRUCTURE BASED ON MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT I (COI) GENE IN THE BLACK SCAR OYSTER, *Crassostrea iredalei* (FAUSTINO, 1932) IN MALAYSIA

ABSTRACT

Mitochondrial cytochrome oxidase subunit I (COI) gene was utilised to assess the population genetics and phylogeographic relationships of the commercially important black scar oyster, *C. iredalei* among 11 populations throughout Peninsular Malaysia and Sabah. Overall, populations of *C. iredalei* demonstrated low nucleotide diversity π (0.000 – 0.004) and low-to-high haplotype diversity *h* (0.000 – 0.795) levels. Genetic structuring was detected between the Peninsular Malaysia and Sabah populations as revealed by the F_{ST} and AMOVA analyses. However, the COI gene analyses showed minimal and non-significant ($P>0.05$) population differentiation within the Peninsular Malaysia and Sabah regions. Relative genetic homogeneity was reported within the coastal populations of Peninsular Malaysia and Sabah, respectively. These could be attributed to both high larval dispersal along the east and west coasts and human-driven spat translocation between the two coastlines due to *C. iredalei* cultivation practices. The neutrality and mismatch distribution analyses supported the hypothesis that *C. iredalei* had experienced a/several bottleneck event(s), followed by population expansion. The molecular information obtained from this study can be incorporated in a pragmatic aquaculture management strategy of wild broodstock and the hatchery lines of *C. iredalei* in Malaysia. A Next Generation Sequencing (NGS) approach, Ion Torrent technology was utilized in the development of novel microsatellite markers in *C. iredalei*. Although a total number of 2424 microsatellite motif repeats were identified, of which, 137 qualified for primer design in the present study. However, the characterization and amplification of the microsatellite markers was not

successful in this project. The four loci selected for characterization in the population screening did not yield satisfactory fragment analysis results. Notwithstanding these unsatisfactory results, Genome Survey Sequence (GSS) data generated by the Ion Torrent can be utilized in various future applications of *C. iredalei* aquaculture improvement and conservation in Malaysia.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The survival of an organism is greatly dependant on the amount of genetic variation they harbour in their genomes. These variations are fundamental for adaptation to a frequently changing environment. Levels of genetic diversities are determined by the interactions among several evolutionary forces (Medley, 2010). Mutations are the elemental source of variation but do not have much effect on allele frequency changes unless other factors come into play (Medley, 2010). The stochastic force of random drift and the deterministic force of natural selection are two of the major factors that can change allele frequencies (Richard, 2006). Additionally, migration homogenizes populations and recombination regroups alleles found on chromosomes (Morgan *et al.*, 2005; Rose *et al.*, 2006). The interplay of these forces drives evolutionary changes in populations over time and creates the diversity found in all living things.

In order to understand the fundamental of an organism's genetic composition, it is necessary to define the framework in which the evolution took place. The study of population structure and its genetics will provide a better understanding of the temporal and spatial changes in allele frequencies. Connectivity of populations can be inferred by investigating the genetic population structure through molecular markers, both nuclear (RAPDs and microsatellites) and mitochondrial (cytochrome oxidase subunit I-III (COI-III), cytochrome b, ND4, and ND5) genes (Zhang & Hewitt, 1996; O'Connell & Wright, 1997; Chauhan, 2010).

Oysters are bivalves that are of commercial importance globally. In Malaysia, several species are exploited in the fisheries industry. Among them are *C. iredalei*, *C. belcheri*, *S. echinata*, *Saccostrea* sp., *O. folium*, and *H. hyotis* (Ng, 1979).Of these, the black scar oyster, *C. iredalei* is the most important species. *Crassostrea iredalei* (Family Ostredaei) can be naturally found along the east coast of Peninsular Malaysia especially in Kelantan and Terengganu, and along the coast of Sabah (Zulfigar & Aileen, 2000). In addition, the culture programme of *C. iredalei* has been very successful on the west coast of Peninsular Malaysia through spat transplantation from its native habitat in the east coast of Peninsular Malaysia (Devakie & Ali, 2000). Being a highly valued commodity, management of this species is vital.

One of the important aspects for the management of wild and cultured organisms is knowledge on population genetics. Although there have been reports on several aspects of *C. iredalei* (Devakie & Ali, 2000; Bussarawit & Simonsen, 2006; Najiah *et al.*, 2008), documentation on the population genetics of this species in Malaysia is limited. Suzana *et al.* (2011) reported evidence of taxonomic ambiguities in the Malaysian oysters revealed by COI gene sequences. To date, the taxonomy of oysters still remains largely reliant on morphological studies (Lam & Morton, 2004, 2009), but molecular tools utilizing COI gene holds much promise in expediting a more detailed understanding of this challenging group. Numerous studies in assessing the population genetics of bivalves have shown that the mitochondrial COI gene and microsatellites are some of the most efficient and widely used markers for such studies (Luttikhuizen *et al.*, 2003; Shefer *et al.*, 2004; Kochzius & Nuryanto, 2008).

Microsatellites are repetitive DNA sequences of several nucleotides, usually two to six base pairs present in eukaryotic genomes (Tautz & Renz, 1984; Tautz, 1989). This marker system has been employed in different fields of studies ranging from ancient and forensic DNA studies, to population genetics and conservation of biological resources (Jarne & Lagoda, 1996; Chambers & MacAvoy, 2000; Adibah *et al.*, 2012). Due to their high levels of allelic variation, codominant inheritance, and ease of analysis, microsatellite markers are very useful in evaluating variation and population genetics of target species. However, the utilization of microsatellite markers have some drawbacks which include a lengthy and costly development phase (by conventional approach) and a relatively low throughput because of difficulties for automation and data management (Guichoux *et al.*, 2011). However, advanced technology in the Next Generation Sequencing (NGS) approach has revolutionised genomics studies through its high-throughput and immense bioinformatics data generation. Thus, in this study, an NGS platform, Ion Torrent technology was utilized in the development of novel microsatellite markers in *C. iredalei*. The data output also generates partial Genome Survey Sequence (GSS) for identification of putative genes in biological processes of the species.

Knowledge of the genetic variation of this species would allow a more systematic programme for management of cultured and wild populations. For instance, selection of broodstocks can be based on genetically rich populations as identified by this study. Furthermore, once established, cultured populations can be monitored to maintain their genetic diversity so as to prevent inbreeding, through these markers. Similarly, these data will also permit identification of closely related populations should the need arise for a translocation programme.

1.2 Objectives

Thus, considering the status of the existing genetic knowledge of the black scar oyster, *C. iredalei*, the main objectives of this dissertation are to:

1) Characterize the population genetics and phylogeographic relationships of *C. iredalei* inferred from the mtDNA COI gene analyses.
2) Develop novel microsatellite markers for *C. iredalei* utilizing the Next Generation Sequencing (NGS) approach, Ion Torrent technology.

CHAPTER 2

LITERATURE REVIEW

2.1 Biology of *Crassostrea iredalei*

2.1.1 Nomenclature and taxonomy

A total of more than 790 species of bivalves are currently collected globally, mostly for the craft and food industries (Berthou *et al.*, 2009). In Malaysia, *C. iredalei* has been recognized as the commercially potential species which is favoured for its sweet flavour and creamy coloured meat (Devakie *et al.*, 1993). However, the culture of *C. iredalei* is not limited just as a luxurious cuisine, but there are also projects utilizing this species (the shells) as substitutes for silicon dioxide (silica sand) in ceramic tile manufacturing (Agbayani & Espinosa, 2006). The oyster shells which are of economic value for their aesthetic appeal are mostly found in tropical countries, such as Malaysia. The ornamental shells are also a leading export item used for buttons and jewellery as the development of the tourism and leisure industries (Berthou *et al.*, 2009).

The focus of this study, the slipper cupped oyster (FAO, 2010), *C. iredalei* is a commercially important species in Malaysia. It is commonly known as the black scar oyster (Bussarawit & Simonsen, 2006) because of the presence of a distinct black adductor muscle scar. It is also known as the Philippine cupped oyster (Berthou *et al.*, 2009) due to its abundance in the country. According to Devakie and Ali (2000), the local coastal inhabitants traditionally consume oysters collected from the wild. Oyster culture in Malaysia has shown great potential in the aquaculture perspective as for its commercial value and great demands (Devakie & Ali, 2002).

Crassostrea iredalei usually has an asymmetrical, medium sized shell of approximately 8 cm, but this could reach 15 cm high (FAO, 2010). The anterior-posterior height is higher than the dorsal-ventral distances, giving the shell a flattened appearance (Poutiers, 1998). The shell shape is roughly rounded, slanted triangular or elongated ovate in outline. The lower valve of the species forms a deep, cupped-shape, while the upper valve is opercular and sometimes a row of denticles is present along the outer edge laterally from the hinge (Thomson, 1954). The upper and lower valves are connected by a small, but sometimes large attachment area. The upper valve is smaller than the lower valve which is more rounded and lighter. There is a clear adductor muscle scar which is the attachment area for the flesh in the internal part of the shells (Lam & Morton, 2003). The adductor muscle scar is black, roundish or bean-shaped and located dorsally, nearer to the ventral margin than to the hinge (FAO, 2010). Visootiviseth *et al.* (1998) noted that the colour of the adductor muscle scar is the primary characteristic used to differentiate between *Crassostrea* oysters, which are observed to be black in *C. iredalei* (Plate 2.1).

The morphological descriptions provide a very good guideline for basic identification of the species. However, the taxonomy of the genus *Crassostrea* Sacco (1897) on the whole has been problematic because of its phenotypic shell plasticity (Lam & Morton, 2003). In general, shell morphology and anatomy of several *Crassostrea* species often overlap (i.e. *C. iredalei* with *C. madrasensis; C. gigas* with *C. angulata*) (Littlewood, 1994; Boudry *et al.*, 2003; Suzana *et al.*, 2011). On the other hand, intraspecies variation has also been documented in many species (Littlewood, 1994). Suzana (2011) observed that the elongated shells of *C. iredalei* can appear in variable shapes in an investigation of the Malaysian

populations. Furthermore, molecular studies have revealed taxonomic ambiguities of the Malaysian oysters initially identified by morphological characteristics (Suzana *et al.*, 2011)

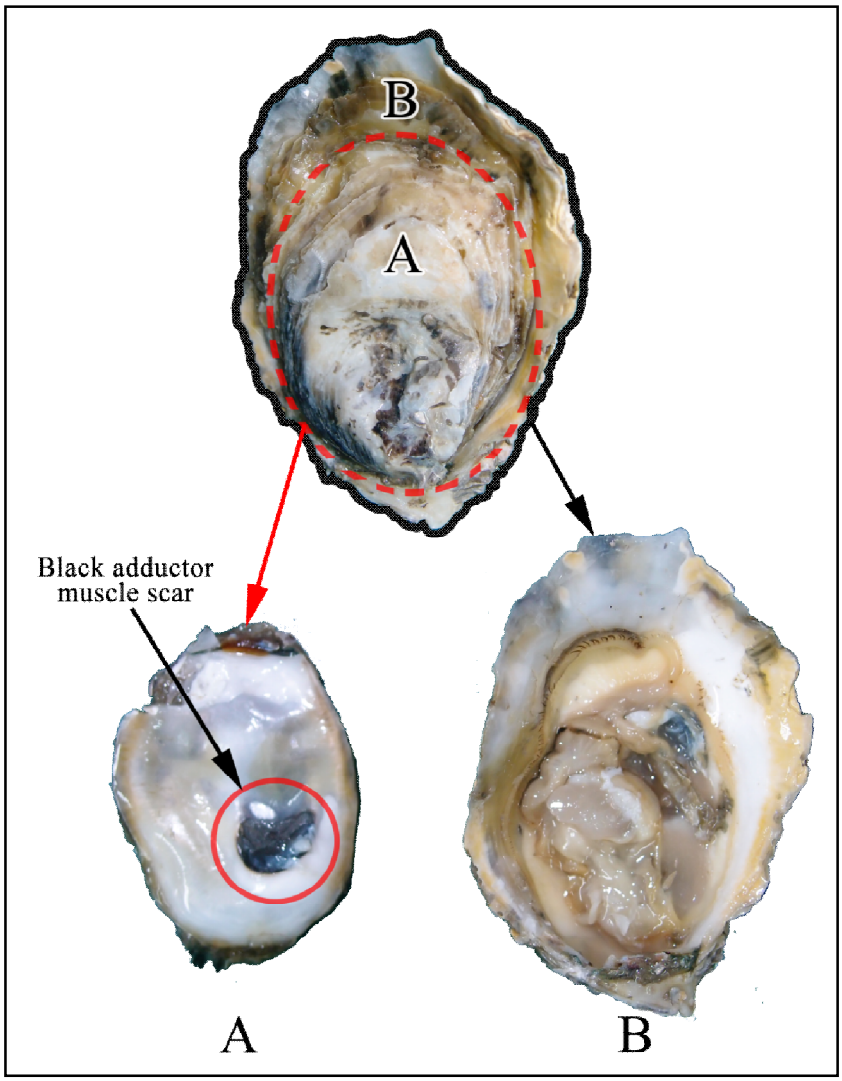

Plate 2.1: Dorsal view of *Crassostrea iredalei*. Circled in red is the black adductor muscle scar.

The taxonomic classification of *C. iredalei* according to the Integrated Taxonomy Information System (ITIS) is as follows:

Kingdom : Animalia
Subkingdom : Bilateria
Branch : Protostomia
Infrakingdom : Lophotrochozoa
Superphylum : Eutrochozoa
Phylum : Mollusca
Class : Bivalvia
Subclass : Pteriomorpha
Order : Ostreoida
Suborder : Ostreoidea
Family : Ostreidae
Genus : *Crassostrea*
Species : *Crassostrea iredalei* (Faustino, 1932)

2.1.2 Reproduction and life history

Crassostrea spp. is oviparous organism; both sperm and eggs are released into the water column with fertilization occurring outside the organism (Galtsoff, 1964). This species is a dioecious alternate hermaphrodite (Angell, 1992) where it exhibits sex reversal. Oysters are usually male when they mature and become female as they grow older (Medley, 2010). Thompson *et al.* (1996) showed that the sex change process is reversible in subsequent years. The adaptive significance of sex reversals in oysters and the factors influencing them are not clear, however there are a lot of related studies that indicated the influence of food limitation and environmental stresses in a more male dominated population (Thompson *et al.*, 1996).

Reproduction occurs in response to temperature cues during the warmer periods of the year (Medley, 2010). The spawning temperature varies among populations in the range of 20˚C to 25˚C (Bahr & Lanier, 1981). The fertilization

process usually takes place within a few minutes following the release of the eggs (Galtsoff, 1964) and hatching occurs approximately six hours after fertilization (Loosanoff, 1965) producing free-swimming planktonic larvae within 48 hours (Buroker, 1983). The free-swimming larvae are passively transported by currents in the 2-3 week period (Medley, 2010). The larvae than moves towards the headwaters of the estuary by remaining at the bottom during ebb tide and moving into the water column during floodtide in the last few days of the period (Babb, 2005).

Oyster larvae conceivably travel hundreds of kilometres as planktons for two to three weeks before settlement resulting in its high gene flow (Kennedy, 1996). The long-range gene flow provides demographic connections between areas with localized restoration activities (Rose, 2008). Palumbi (1994) discussed that the prcesses of natural selection and variance in reproductive success determine the magnitude of effective gene flow by the presettlement movement of the larvae. High fecundity and stochasticity of the larval viability can result in extreme variance in reproductive success (sweepstakes event) in marine organism (Hedgecock, 1994). Such events could potentially create genetic heterogeneity among cohorts, transiently among localities in breeding populations (Rose, 2008).

2.1.3 Habitat and distribution

Oysters generally live in dense aggregations referred to as reefs or beds in shallow well-mixed estuaries, oceanic bays, and lagoons (Andrews, 1979). Menzel *et al.* (1966) reported that oysters also inhabit mud flats and offshore sand bars where it tolerates widely fluctuating water temperatures and salinities (Andrews, 1979). All of these habitats should provide adequate amount of food, with suitable substrate, non-turbulent water currents, moderate pH levels (6.75 to 8.75), and moderate dissolved

oxygen levels (Stanley & Sellers, 1986). *Crassostrea iredalei* was reported to inhabit enclosed brackish water lagoons (Devakie *et al.*, 1993), estuaries, intertidal zone, shallow subtidal water (FAO, 2010) and even rocky shores which have lower salinity than the open sea (Young & Serna, 1982).

Crassostrea iredalei is one of the most widely distributed species with a native range covering South China Sea, Andaman Sea, and Gulf of Thailand (Yoosukh & Duangdee, 1999). It has been noted by Berthou *et al.* (2009) that *C. iredalei* production is mainly restricted to Malaysia and the Philippines. In Malaysia, *C. iredalei* can be naturally found along the east coast of Peninsular Malaysia especially in Kelantan and Terengganu, and along the coast of Sabah (Zulfigar & Aileen, 2000). However, cultured populations can be found in the west coast, a result of a transplantation programme for aquaculture activities initiated by the Malaysian Fisheries Research Institute (Devakie & Ali, 2000).

2.1.4 Aquaculture status of *C. iredalei* in Malaysia

There are four main genera of oysters that have been identified in Malaysia; *Crassostrea* (*C. iredalei* and *C. belcheri*), *Saccostrea* (*S. echinata* and *Saccostrea* sp.), *Ostrea* (*O. folium*), and *Hyotissa* (*H. hyotis)* but only the first three genera are commercially cultured in Malaysia (Ng, 1979). Devakie and Ali (2000) reported that the culture programme of *C. iredalei* has been very successful on the west coast of Peninsular Malaysia through spat transplantation from its native habitat in the east coast of Peninsular Malaysia. Even though the transplanted spat grow relatively slower than the other local oyster, *C. belcheri*, the former has successfully adapted to the west coast environment with low mortality rate (Devakie *et al.*, 1993). However, due to the monsoon seasons occurring yearly from November to January, oyster spats

are regularly imported from Thailand to supplement the scarce spat supply (Devakie & Ali, 2000).

Crassostrea iredalei has a very high potential in the aquaculture industry due its high demand among consumers compared to other oyster species such as *S. cucullata* and *O. folium* (Davy & Graham, 1982; Suzana, 2011). Recognizing the success of Thailand and Philippines in the oyster industry (Suzana, 2011), the Malaysian government initiated a systematic aquaculture programme for the development of the local oyster industry in the late 1980's. Through financial and technical assistance from the Danish International Development Assistant (DANIDA) and the Swedish International Development Authority (SIDA) to the Fisheries Research Institute (FRI) Malaysia under the Bay of Bengal Programme, a systematic culture programme was initiated in 1988 with focus on *C. iredalei* and *C. belcheri* (Devakie *et al.*, 1993). However, the local oyster industry has not reached the same success as the neighbouring countries, perhaps due to inadequate management practices and poor knowledge of existing resources, for example in accurate species identification and genetic diversity (Devakie & Ali, 2002; Lam & Morton, 2009; Suzana, 2011). Thus, to stimulate further growth of the local oyster aquaculture industry, species identification and proper stock management is required (Liu *et al.*, 2011).

The total oyster production in Malaysia increased from 373.72 metric tonne (mt) in 2005 to 869.72 mt in 2007 with increased profit of 74.02% (RM 3.4 billion) (DOF, 2008). Unfortunately, production decreased to 275.47 mt in 2008 with a total value of only RM 1.25 billion (DOF, 2008). Shellfish culture has been established primarily in coastal areas and estuaries that are environmentally sensitive (Berthou *et al.*, 2009). Therefore, overall landings are correlated with appropriate environmental

management as well as cultural practices, production cost, and market demands (Berthou *et al.*, 2009). The drastic decline in the one year period could have been due to one or several factors; destructive activities at the cultural sites, overfishing, sedimentation, unstable freshwater inflow, disease, intrusion of new species, and unbalanced nutrients and pollutants (Beck *et al.*, 2009).

2.1.5 Molecular markers in phylogenetics and population genetics studies

2.1.5.1 Mitochondrial DNA

Mitochondrial DNA (mtDNA) has been widely utilized in assessing phylogenetics (Yu, 2006), population genetic studies (Medley, 2010; Suzana *et al.*, 2011; Rahim *et al.*, 2012; Tan *et al.*, 2012), evolutionary history (Brown *et al.*, 1979; Helgason *et al.*, 2000) and species identification (Klinbunga *et al.*, 2003; Lam & Morton, 2003; Hebert *et al.*, 2004). Mitochondrial DNA is widely used in aquaculture and fishery for instance the investigation of stock structure (Liu & Cordes, 2004), conservation of population (Ferguson *et al.*, 1995) and management in aquaculture and fisheries (Iguchi *et al.*, 1999).

The majority of the invertebrate mitochondrial genome is made up of nearly the same number of genes as in vertebrates (Pereira, 2000). It normally encodes 37 genes in most animals of which 13 are protein coding, 22 are for the relatively conserved transfer RNA (tRNA), and two are for the large and small subunits of ribosomal RNA (rRNA). In addition it has a non-coding gene known as the control region or D-loop (Boore, 1999; Pereira, 2000). While the typical mtDNA genome ranges in size between 16 000 – 20 000 base pairs (Boore, 1999; Kucuktas & Liu, 2007), the mitochondrial genome composition of *C. iredalei* is documented to be much longer than in all other oysters (22 446 bp in length) (Wu *et al.*, 2010).

Mitochondrial DNA is documented to be inherited only through the maternal line (Birky, 2001) and hence no recombination occurs (Gyllensten *et al.*, 1985; Marshall *et al.*, 2009). However, there are a few exceptions in the mtDNA inheritance pattern as recorded in certain organism such as in mussels (Hoeh *et al.*, 1991), flies (Kondo *et al.*, 1990), anchovies (Magoulas & Zouros, 1993), honeybees (Meusel & Moritz, 1993), mice (Gyllensten *et al.*, 1991) and recently also found in humans (Schwartz & Vissing, 2002). Seven bivalve lineages namely Mytilidae, Unionidae, Margaritiferidae, Hyriidae, Donacidae, Solanidae, and Veneridae have been found to have an unusual system termed doubly uniparental inheritance of mtDNA (DUI) (Theologidis *et al.*, 2008; Doucet-Beaupré *et al.*, 2010) and deviate from the typical maternal inheritance where the mtDNA transmission route is governed by the female and male lineage (Xu, 2005).

The evolution rates of mtDNA gene sequences is higher (10 – 30 times) than nuclear genome (Lane, 2009), as mtDNA does not harbour the repair enzymes to correct errors occurring during the replication process (Clayton, 1982). As a result, mtDNA possess a high level of transitions and transversion with high rate of small length mutation leading to high amount of variation in the mtDNA sequences (Cann & Wilson, 1983). Johnson (2000) and Mahidol *et al.* (2007) revealed that variations in mtDNA are more sensitive to the effects of isolation of populations, hence increased sensitivity to genetic drift and bottleneck effect. Therefore, understanding the evolutionary behaviour of molecular characters such as nucleotide substitution pattern could reveal the effects of more recent historical subdivision and provide useful information in resolving phylogenetic relationship of the studied organism (Johnson, 2000).

The 650 bp segment at the 5' end of the mitochondrial cytochrome oxidase subunit I (COI) gene or widely known as the DNA barcoding marker has been recognised as the standard mitochondrial fragment in identifying species taxa as proposed by Hebert *et al.* (2003). Numerous studies in a variety of taxa have been carried out to examine its efficacy and accordingly have proven its validity i.e. oysters - Liu *et al.* (2011); birds - Hebert *et al.* (2004); fish - Ward *et al.* (2005). It is very useful in the assignments of cases where morphological features are missing or misleading (Schindel & Miller, 2005; Liu *et al.*, 2011) as well as in species identification for taxa with the high levels of morphological plasticity such as the oysters. Although the sequence gathered for DNA barcoding is insufficient in assessing population-level genetics of the studied organism, the COI gene can provide an early insight into the patterning of genomic diversity within a species (Hajibabaei *et al.*, 2007). This particular sequence will provide a first signal in the degree of nature of population divergence and could facilitate comparative studies of population diversities in many species (Hajibabaei *et al.*, 2007). To date, the taxonomy of oysters still remains largely reliant on morphological studies (Lam & Morton, 2004, 2009), but molecular tools utilizing COI gene holds much promise in expediting a more detailed understanding of this challenging group.

2.1.4.2 Microsatellite DNA

Microsatellites are simple sequence repeat regions (SSRs) that are distributed throughout the genomes of eukaryotes (Messier *et al.*, 1996; Mulvey *et al.*, 1998). Since microsatellite sequences were first discovered in eukaryotic genomes 20 years ago (Tautz & Renz, 1984; Tautz, 1989), they have been utilized as a powerful genetic marker in genome mapping, disease studies, population genetics, and for

evaluation of patterns of evolution and mutations in the repeat region themselves (Jarne & Lagoda, 1996; Chambers & MacAvoy, 2000). These non-coding, co-dominant and neutral DNA repeating motifs (di-, tri-, tetra-, and penta- nucleotides) are abundant throughout the genome of an organism (O'reilly & Wright, 1995). However, some microsatellite markers have also been found to be associated with the coding region (i.e. EST-linked), typically in introns or flanking untranslated regions (Vasemaegi *et al.*, 2005; Carlsson & Reece, 2007). Li *et al.* (2002) stated that microsatellite mutation rates are generally high (10^{-2}-10^{-6} events per locus per generation) and appears as changes in the number of repeat units (Xiao, 2009). Putative mutation mechanisms include slippage during DNA replication and recombination between DNA strands (Li *et al.*, 2002). The rapid evolution rate in microsatellite markers allows the detection of recent demographic events (Morin *et al.*, 2004). Undoubtedly, microsatellite markers have become the preferred tool for application in fisheries stock management, population analyses and biodiversity preservation programme due to their ability to detect differences between closely related populations and their efficiency for revealing extensive allelic variation (O'Connell & Wright, 1997; Romana-Eguia *et al.*, 2004; Xiao *et al.*, 2010).

These neutral markers are highly polymorphic within and among populations with the variation arising from changes in the number of repeated motifs (Mulvey *et al.*, 1998). Microsatellites could provide more detailed information on population structure based on levels of heterozygosity and number of alleles at each locus as it is highly variable in nature (Banks *et al.*, 1999). The microsatellite markers are efficient markers for determination of population structuring and genetic variation within and between populations (Charrier *et al.*, 2006) but are less informative for interspecific differentiation due to high level of polymorphism (Shaw *et al.*, 1999). The isolation

of such hypervariable DNA loci has led to more accurate assessment of population differentiation and intrapopulation variation, and has provided a means of identifying populations that are vulnerable due to lack of genetic diversity (Mulvey *et al.*, 1998; Kikuchi & Isagi, 2002; Richard, 2006).

Microsatellite markers are recognized to have more advantages compared to the other common nuclear markers, and have been utilized in many marine fish population genetic studies, parentage and kinship analyses and genome mapping of organism (Liu & Cordes, 2004). Since Clabby *et al.* (1996) carried out the first attempt to clone and study satellite DNA in *C. gigas*, microsatellite markers have been developed for a number of commercially important *Crassostrea* species over the past ten years (Xiao, 2009). These markers have been successfully applied in assessing the genetic variation of wild and hatchery stocks of oysters (Xiao, 2009), examining the geographic structure of wild oyster populations (Launey *et al.*, 2002; Rose *et al.*, 2006), performing parentage analyses (Huvet *et al.*, 2001; Boudry *et al.*, 2002), studying the inbreeding effects (Bierne *et al.*, 1998) and in linkage mapping (Hubert & Hedgecock, 2004). Although there are microsatellite markers developed for various *Crassostrea* species, including *C. gigas* (Huvet *et al.*, 2000a; Sekino *et al.*, 2003; Hedgecock *et al.*, 2004), *C. ariakensis* (Xiao *et al.*, 2008), *C. virginica* (Reece *et al.*, 2004; Carlsson *et al.*, 2006; Rose *et al.*, 2006), they are of limited application in *C. iredalei* due to the low cross-species transferability (Hedgecock *et al.*, 2004).

One of the drawbacks of utilizing microsatellites markers in assessing the genetic variabilty of an organism is because of the high frequency of null alleles which could complicate data analysis (Launey & Hedgecock, 2001). However, to

date, they are still capable of providing the best resolution (Richard, 2006). Dewoody *et al.* (2006) noted that allelic dropout, null alleles and stuttering are the most common problems encountered in accurate allelic scoring. Allelic dropout and null alleles would influence the scoring as it will be biased towards excess homozygosity thus underestimating polymorphism in a particular population (Dewoody *et al.*, 2006; Tan, 2011). Stutters or artefact peaks will impair the interpretation of genotypes (Arif *et al.*, 2010). The variability and low quality of DNA template (below 0.05ng/10 µL) has been shown to result in allele dropout error (Gagneux *et al.*, 1997). Hence, in mitigating this error, a standardized DNA concentration should be applied for higher accuracy (Tan, 2011). Notwithstanding all the limitations in utilizing microsatellite markers in molecular study, further development and use of these neutral markers will prove to be valuable in defining population structure of the studied organism.

2.2 Development and characterization novel microsatellite markers of *C. iredalei* utilizing the Next Generation Sequencing approach, Ion Torrent technology

DNA sequencing is a powerful approach in understanding the biology of living organisms. Over the past few years, the advent of next-generation sequencing (NGS) technologies had been recognized as a superior option for obtaining global genomic information (Su *et al.*, 2011). NGS or also known as massively parallel sequencing technologies has revolutionized genomic and genetic research (Quail *et al.*, 2012). Over the past decade, the cost of NGS has decreased significantly, making whole genome sequencing a possible way of investigating emerging environmental issues (Mehinto *et al.*, 2012). The NGS approach enables researchers to obtain large amount of raw data in a short period of time. Additionally, the advancement in bioinformatics to assemble the sequences and annotate the genes, helps in facilitating

the management of these large datasets (Mehinto *et al.*, 2012). One of the most recent applications of NGS is the assembly of complete genome sequence of *C. gigas* (Zhang *et al.*, 2012). The oyster genome sequence provides valuable resources for studying molluscan biology and for genetic improvement of oysters and other important aquaculture species.

In February 2010, the Ion Torrent was launched as a sequencer that differs from the conventional sequencer as it is based not on dye-labeled oligonucleotides and expensive optics but on ion detection. Hence, the machine has been casually referred to as a 'pH meter that sequences' (Rusk, 2010). The Ion Torrent is a modified version of the 454 pyrosequencing approach and operates on the same principle, except it makes use of the hydrogen ion (H^+) that is released with every nucleotide incorporated, instead of the pyrophosphate (Rothberg *et al.*, 2011). The fundamentals of Ion Torrent are quite straightforward as discussed by Rusk (2010) and Rothberg *et al.* (2011). Firstly, the DNA to be sequenced is captured in a microwell and unmodified nucleotides are floated across the wells, one at a time. Thereupon, the ion sensor detects the pH change in the solution as hydrogen ion is released when the polymerase incorporates appropriate oligonucleotide into the growing strand. In order to detect the hydrogen ion being released, the picotiter plate sits on top of massively parallel semiconductor-sensing devise on an ion chip. The integrated circuits take advantage of metal-oxide semiconductor technology, which significantly reduces the cost of sequencing since luciferase and other costly enzymes and scanners are not needed in the technology (Mehinto *et al.*, 2012). The mechanism of Ion Torrent also allows real time sequencing (Rusk, 2010). Despite having shorter read length compared to other NGS platforms, expanding the size of

the semiconductor chips utilized in the Ion Torrent technology will increase the throughput (Rusk, 2010).

Microsatellite marker development can be problematic in many non-model species because of the high cost and time required in utilizing the traditional approach of library construction (Guichoux *et al.*, 2011; Wang *et al.*, 2012). Nevertheless, this issue can be resolved using NGS approach of Ion Torrent sequencing. Significant random genomic DNA fragment (genome survey sequence or GSS) and large expressed sequence tag (EST) data sets can be generated by the NGS (i.e. Ion Torrent and Roche 454) (Iranawati *et al.*, 2012) and can be applied to identify new genes and data resources including for microsatellite marker and single nucleotide polymorphism (SNP) marker development (Strong & Nelson, 2000; Salem *et al.*, 2010; Jung *et al.*, 2011; Wang *et al.*, 2012). As reported by Strong and Nelson (2000), the GSS approach is more productive and efficient than an EST approach for gene identification because redundant mRNA sequences from highly expressed genes can be avoided. Since microsatellite markers are neutral markers, mainly from non-coding regions, some of the markers developed in functional genes are useful in evaluating functional diversity (Shikano *et al.*, 2010; Guichoux *et al.*, 2011; Iranawati *et al.*, 2012).

2.3 Population genetics study of *C. iredalei*

Marine invertebrates provide a unique challenge in interpreting their data as their life history is typically characterized by high fecundity, early mortality and planktonic larval stage that allows long distance dispersal (Flowers *et al.*, 2002; Xiao *et al.*, 2010). Having such life history strategy helps marine invertebrates namely oyster, to make them successful as exemplified by their worldwide distribution.

Marine invertebrates are a diverse group of organisms, so generalisations are difficult to be applied across taxa. In spite of that, understanding patterns of genetic variation of the oyster will assist the determination of different aspects of life history that influence the evolution of species (Xiao *et al.*, 2010).

The low level of population structure observed in marine species is very common as a result of long distance larval dispersal (Bohonak, 1999). The substantial variation in average effective dispersal distances among species is usually concomitant to the larval duration in the water column prior to its settlement (Shanks *et al.*, 2003; Siegel *et al.*, 2003). Hence, genetic panmixia over a small population scale is a reasonable null hypothesis of long larval periods in marine species (Xiao *et al.*, 2010).

Despite the abundance, ecological and economic significance of *C. iredalei* in Southeast Asia, there is a scarcity of population genetics and population structure information on the species. To date, there are only few studies on population genetics of the species. Thailand is one of the most active countries that had focused much attention to their local oysters. Molecular genetics keys have been developed and applied in the taxonomy and resolving genetics issue of oysters within the genera *Crassostrea*, *Saccostrea* and *Striostrea* in Thailand (Klinbunga *et al.*, 2000, 2001, 2003 & 2005). A study on *C. iredalei* and *C. belcheri* was also conducted to reveal population differentiation along the coast of Thailand including a small number of specimens from Penang, Malaysia by means of isozyme analysis (Bussarawit & Simonsen, 2006). Suzana *et al.* (2011) did a study on the local oysters in Malaysia and showed evidence of taxonomic ambiguities and biological invasion. Accurate identification of the local oysters will be vital for effective conservation of wild

populations, management of invasive species and development of future aquaculture breeding programmes in Malaysia.

2.4 Management and conservation of *C. iredalei*

Asia Pacific has long been considered to be the epicentre of oyster speciation (Ren *et al.*, 2010). Nevertheless, much of the oyster classification remains a challenge partly due to the lack of well-defined morphological characters and has resulted in considerable errors and confusion. Without proper identification and biological information on the genetics of the local oyster, the management and conservation of *C. iredalei* will be hindered. The population genetics and phylogenetics study would facilitate the elucidation of the extent of genetic variation within species from distinct populations and identify management units. Such study provides important evidence on natural population structure that can be utilised as a parameter to protect genetically depauperate populations and also long term management for fisheries resources (Ciftci & Okumus, 2002; Gum *et al.*, 2009). From the perspective of aquaculture, information on levels of genetic variability within and among broodstock populations allows the breeders to avoid potential detrimental effects of inbreeding and other genetic changes from one generation to another (Gjedrem, 1992). Furthermore, genetic markers can be used to assist establishment of base-line stock, including family lines for selection purposes (Cross *et al.*, 2000). In addition, detection of low levels of genetic variation in wild stocks is evidence of overexploitation and anthropogenic perturbation and requires *in-situ* monitoring and habitat preservation (Rodrigues *et al.*, 2008).

CHAPTER 3

MATERIALS AND METHODS

3.1 Sample localities

Wild and cultured oysters were collected from 11 locations over a range of 2,885 km coastline along the coastal waters of west Peninsular Malaysia (facing the Straits of Malacca), east Peninsular Malaysia, and Malaysian Borneo (Sabah) (the latter two are separated by the South China Sea) (Table 3.1 and Figure 3.1). However, samples from Pantai Chenang, Kedah (PCE), Sg. Dedap, Kedah (SD), Penaga, Pulau Pinang (P), and Sg. Pasir Hitam, Perak (ST) were originated from Setiu, Terengganu as the samples were translocated for culture purposes. Up till now, there is no documentation of *C. iredalei* in the Sarawak waters (Devakie & Ali, 2002). The oysters were collected by dredge or diver from natural subtidal reefs and supplied by Fisheries Research Institute (FRI). Wild oysters originated from natural settling substrates such as oyster shells, mangrove roots, submerged rocks and mostly covered in silt. Most oysters were found as individuals; however, some small clusters were present. The cultured populations (PCE, SD, P, ST) were obtained from grow-out individuals collected from hatchery sites (estuaries). All specimens were identified based on traditional keys described in Visootiviseth *et al.* (1998); Siddiqui and Ahmed (2002). Oysters were kept on ice during transit and subsequently stored at -20°C prior to analysis.

3.2 Sample storage and tissue preservation

Oysters were shucked and 10 g of freshly collected mantle tissue specimens were preserved in replicates of 1000 µL TNES-Urea (100 mM Tris-HCL pH 7.5, 125

mM NaCl, 10 mM EDTA pH 7.5, 1% SDS, 3 M Urea) (Tan, 2011) with 95% ethanol and stored at room temperature (~25°C). Tissues in TNES-Urea buffer were kept at least two weeks for complete digestion before extraction. Fresh samples in ethanol were ready to be extracted after 24 hours.

Table 3.1: Sampling localities, site abbreviations (abbrev.), coordinates (latitude and longitude), sample sources and collection date.

Localities	Abbrev.	Latitude (N)	Longitude (E)	Sample source	Collection date
Peninsular Malaysia					
Setiu, Terengganu	S	05°31'	102°56'	Wild	30/06/2008
Merchang, Terengganu	MG	05°01'	103°17'	Wild	25/10/2009
Pantai Seri Tujuh, Kelantan	L	06°13'	102°14'	Wild	25/10/2009
Semerak, Kelantan	SM	05°51'	102°30'	Wild	23/03/2012
Sungai Dedap, Kedah	SD	05°39'	100°22'	Cultured	03/11/2008
Pantai Chenang, Kedah	PCE	06°17'	99°43'	Cultured	21/08/2009
Penaga, Pulau Pinang	P	05°31'	100°22'	Cultured	26/02/2012
Sg. Pasir Hitam, Perak	ST	04°39'	100°36'	Cultured	26/02/2012
Malaysian Borneo					
Serusup, Sabah	R	06°13'	116°16'	Wild	24/09/2008
Kuala Penyu, Sabah	KP	05°33'	115°35'	Wild	07/11/2008
Trayong, Sabah	T	06°08'	116°08'	Wild	23/09/2008

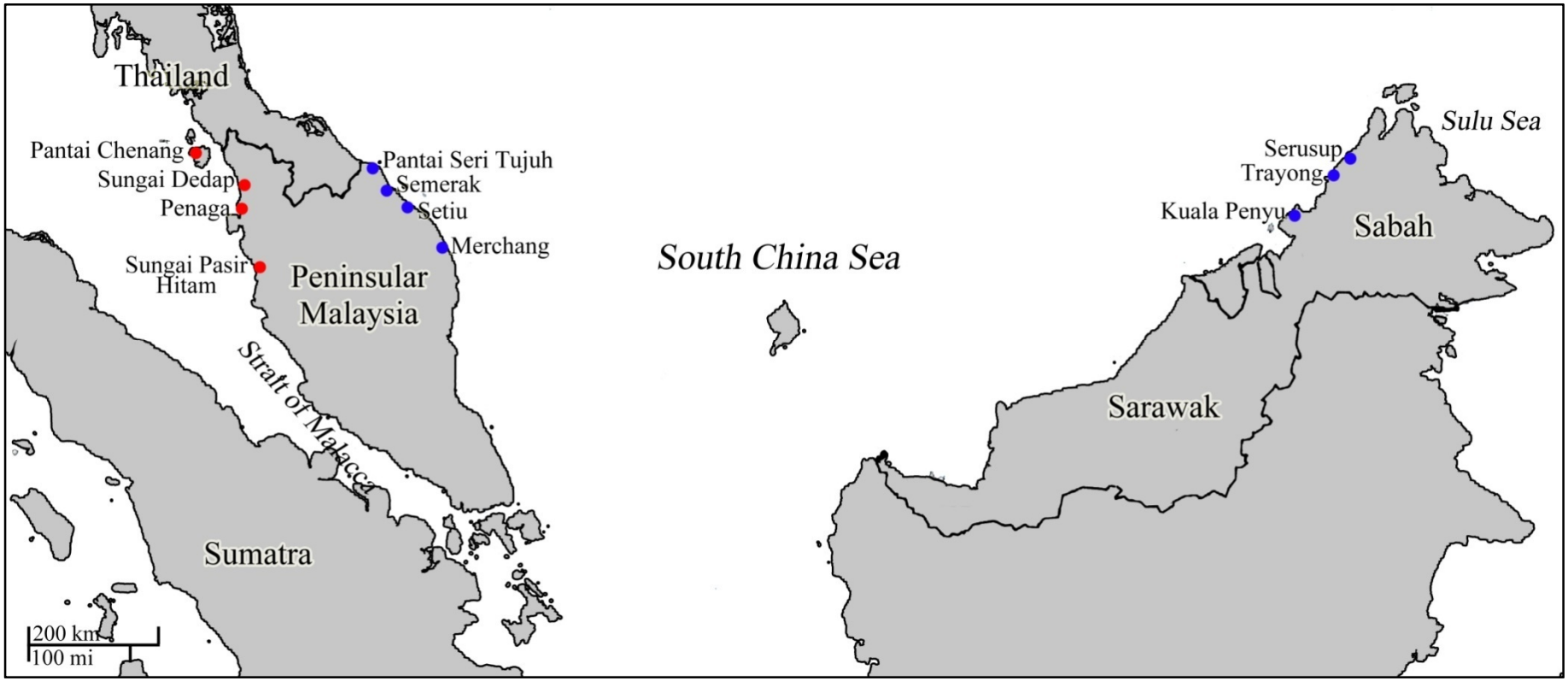

Figure 3.1: Sampling localities of eleven *C. iredalei* populations analyzed in the present study. The populations marked with blue are the wild samples while red indicates the translocated samples.