การพัฒนาเครื่องหมายไมโครแซทเทลไลต์ไทป์ 1 และการประยุกต์ใช้ในการสร้างแผนที่จีโนมของกุ้งกุลาดำ *Penaeus monodon*

นายเชิดศักดิ์ มณีรัตนรุ่งโรจน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวเคมี ภาควิชาชีวเคมี
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2547

ISBN 974-53-1929-5

DEVELOPMENT OF TYPE I MICROSATELLITE MARKERS AND APPLICATION IN GENOME MAPPING OF THE BLACK TIGER SHRIMP *Penaeus monodon*

Mr. Cherdsak Maneeruttanarungroj

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Sciences in Biochemistry

Department of Biochemistry

Faculty of Science

Chulalongkorn University

Academic Year 2004

ISBN 974-53-1929-5

Thesis Title	Development of type I microsatellite markers and application in genome mapping of the black tiger shrimp *Penaeus monodon*
By	Mr. Cherdsak Maneeruttanarungroj
Field of Study	Biochemistry
Thesis Advisor	Associate Professor Anchalee Tassanakajon, Ph.D.
Thesis Co-advisor	Prof. Timothy William Flegel, Ph.D., Siriporn Pongsomboon, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

..... Dean of the Faculty of Science

(Professor Piamsak Menasveta, Ph.D.)

Thesis Committee

..... Chairman

(Associate Professor Aran Incharoensakdi, Ph.D.)

..... Thesis Advisor

(Associate Professor Anchalee Tassanakajon, Ph.D.)

..... Thesis Co-advisor

(Siriporn Pongsomboon, Ph.D.)

..... Member

(Associate Professor Siriporn Sittipraneed, Ph.D.)

..... Member

(Teerapong Buaboocha, Ph.D.)

เชิดศักดิ์ มณีรัตนรุ่งโรจน์ : การพัฒนาเครื่องหมายไมโครแซเทลไลต์ไทป์ 1 และการประยุกต์ใช้ในการสร้างแผนที่จีโนมของกุ้งกุลาดำ *Penaeus monodon*. (DEVELOPMENT OF TYPE I MICROSATELLITE MARKERS AND APPLICATION IN GENOME MAPPING OF THE BLACK TIGER SHRIMP *Penaeus monodon*) อ. ที่ปรึกษา : รศ.ดร. อัญชลี ทัศนาขจร, อ. ที่ปรึกษาร่วม : ศ.ดร. Timothy William Flegel, ดร. สิริพร พงษ์สมบูรณ์ 149 หน้า. ISBN 974-53-1929-5.

ไมโครแซเทลไลต์ เป็นเครื่องหมายพันธุกรรมที่มีความหมายสำคัญ และสามารถนำมาประยุกต์ให้เกิดประโยชน์ต่อการเพาะเลี้ยงสัตว์น้ำและการประมง ในการศึกษาครั้งนี้ เครื่องหมายไมโครแซเทลไลต์ไทป์ I ได้ถูกพัฒนาเพื่อนำมาใช้ในสร้างแผนที่จีโนมของกุ้งกุลาดำ *Penaeus monodon* โดยการใช้โปรแกรมทางคอมพิวเตอร์เพื่อค้นหาลำดับเบสชนิดไมโครแซเทลไลต์ จากฐานข้อมูล EST ของกุ้งกุลาดำ (http://pmonodon.biotec.or.th) ได้ทำการค้นหาจากเบสทั้งหมด 10,100 โคลน พบว่า 1,381 โคลน มีไมโครแซเทลไลต์เป็นองค์ประกอบ ผลของการจัด cluster พบว่า 513 โคลน สามารถจัดกลุ่มได้ 129 กลุ่ม ส่วนที่เหลืออีก 868 โคลน เป็นโคลนที่ไม่อยู่ในกลุ่ม จากจำนวนทั้งหมด 2,165 ตำแหน่งของไมโครแซเทลไลต์ ถูกจำแนกประเภทออกเป็น 3 กลุ่ม พบว่ากลุ่ม perfect ไมโครแซเทลไลต์ มีจำนวนตำแหน่งสูงสุด (78.1%) ส่วน imperfect และ compound ไมโครแซเทลไลต์ พบ 16.7 และ 5.2 เปอร์เซ็นต์ตามลำดับ เบสซ้ำสาม AAT พบมากที่สุดในฐานข้อมูลEST รองลงมาได้แก่เบสซ้ำสอง AT ในการเปรียบเทียบ EST clone ที่ประกอบด้วยไมโครแซเทลไลต์ กับข้อมูลใน GenBank โดยใช้โปรแกรม BLASTX พบว่าผลการเปรียบเทียบตรงกับยีนอื่นที่มีรายงานแล้ว (known gene) 11.6% ส่วนที่เป็นยีนของโปรตีนที่ไม่ทราบหน้าที่ (hypothetical protein) 37.3% นอกนั้นเป็น unknown gene 51.1% จากผลการเปรียบเทียบครั้งนี้ทำให้ทราบว่า ไมโครแซเทลไลต์ มักจะอยู่ในบริเวณที่ถูกแปรรหัสเป็นโปรตีน (coding sequence) มากกว่าบริเวณที่ไม่เป็นโปรตีนในช่วง 3'-UTR และ 5'-UTR ทำการออกแบบไพรเมอร์ที่ขนาบข้างไมโครแซเทลไลต์ไป 154 คู่ พบว่าไพรเมอร์ 126 คู่ ให้ผลผลิต PCR และ 50 คู่ จาก 126 คู่นี้พบความหลากหลาย ผลการแสดงลักษณะของความหลากหลายของไมโครแซเทลไลต์ 50 ตำแหน่งนี้ กับกุ้งจำนวน 35-48 ตัว พบว่าอัลลีลเฉลี่ยเท่ากับ 12.6 อัลลีล polymorphic information content (PIC) เฉลี่ยเท่ากับ 0.723 ค่าเฉลี่ยของ observed และexpected heterozygosity เท่ากับ 0.698 และ 0.759 ตามลำดับ นอกจากนี้ได้นำเครื่องหมายพันธุกรรมทั้งสิ้น 50 ตำแหน่งตรวจสอบพันธุกรรมของครอบครัวกุ้งกุลาดำที่ใช้ในการสร้างแผนที่จีโนม ทำการวิเคราะห์ข้อมูลทางสถิติของผลการตรวจสอบทางพันธุกรรมที่ได้ ร่วมกับผลการตรวจสอบพันธุกรรมของเครื่องหมายพันธุกรรม ชนิด AFLPs ไมโครแซเทลไลต์ และเครื่องหมายชนิดอื่น เพื่อวิเคราะห์ตำแหน่งบนแผนที่จีโนมของเครื่องหมายพันธุกรรมทั้ง 50 ตำแหน่งนี้ภายใต้เงื่อนไขค่า LOD ที่ 3.5 และค่า θ ที่ 0.30 จากการวิเคราะห์พบว่าเครื่องหมายพันธุกรรมจำนวน 36 ตำแหน่งสามารถรวมเข้าไปอยู่ในแผนที่จีโนมได้ โดยที่ความยาวของแผนที่ของกุ้งเพศผู้ และเพศเมียคือ 1,101 และ 891.4 เซนติมอร์แกนตามลำดับ ระยะห่างเฉลี่ยระหว่างเครื่องหมายของแผนที่กุ้งเพศผู้และเพศเมียคือ 7 และ 8 เซนติมอร์แกนตามลำดับ

ภาควิชา ชีวเคมี ลายมือชื่อนิสิต.....

สาขาวิชา ชีวเคมี ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา 2547 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

##4672252323: MAJOR BIOCHEMISTRY

KEY WORD: black tiger shrimp/ *Penaeus monodon*/ type I microsatellite/ bioinformatics/ genetic linkage mapping

CHERDSAK MANEERUTTANARUNGROJ : DEVELOPMENT OF TYPE I MICROSATELLITE MARKERS AND APPLICATION IN GENOME MAPPING OF THE BLACK TIGER SHRIMP *Penaeus monodon*. THESIS ADVISOR: ASSOC. PROF. ANCHALEE TASSANAKAJON, Ph.D., THESIS CO-ADVISOR : PROF. TIMOTHY WILLIAM FLEGEL, Ph.D., SIRIPORN PONGSOMBOON, Ph.D., 149 pp. ISBN 974-53-1929-5

Microsatellites are useful markers for numerous applications in aquaculture and fisheries research. In this study, type I microsatellite markers were developed for use in constructing a genetic linkage of *Penaeus monodon*. A software tool was applied for the identification of microsatellite repeats in the black tiger shrimp (*P. monodon*) ESTs database (http://pmonodon.biotec.or.th). A bioinformatics analysis of 10,100 ESTs identified 1,381 ESTs containing microsatellites. Clustering analysis indicated that 513 of these ESTs fell into 129 contigs, and the remaining 868 ESTs were singletons. A total of 2,165 microsatellite were identified. Perfect microsatellites were predominant in this study whereas imperfect and compound microsatellites were found at a much lesser extent (16.7% and 5.2%, respectively). Trinucleotide AAT repeat type appeared to be the most abundant type distributed in *P. monodon* database, followed by dinucleotide AT repeat type. Homology searching by the BLASTX program revealed that the microsatellite containing clones represent 11.6% know gene products, 37.3% hypothetical protein and 51.1% unknown gene. Microsatellite in ESTs were mainly located in coding region, followed by 3'-UTR and 5'-UTR region, respectively. One hundred and fifty-four primer pairs flanking microsatellite loci have been used for screening polymorphism. As results, 126 primer pairs produced PCR products and 50 pairs were polymorphic. Characterization of the 50 new microsatellite markers on a panel of 35-48 unrelated shrimps showed high levels of genetic polymorphism with the average of 12.6 alleles per locus and the average polymorphic information content (PIC) of 0.723. The average observed and expected heterozygosities were 0.698 and 0.759, respectively. These 50 microsatellite loci were used to genotype the reference family for international genetic mapping of *P. monodon*. The genotyping data was analyzed with AFLP primer combination, microsatellites, and other markers with LOD score of 3.5 and maximum θ of 0.30. Thirty-six microsatellite markers were integrated into the previously shrimp genetic linkage map. The total lengths of linkage groups covering the male and female maps were 1,101 and 891.4 cM, respectively. The average spacing between 2 markers of male and female maps were 7 and 8 cM, respectively.

Department Biochemistry Student's signature Chat. M.

Filed of study Biochemistry Advisor's signature A. Tassanakajon

Academic year 2004 Co-advisor's signature

Co-advisor's signature Siriporn Pongsomboon

Acknowledgements

I would like to express my deepest gratitude to my advisor, Associate Professor Dr. Anchalee Tassanakajon for her guidances, supervision, encouragement and supports throughout my study. I am very grateful to my co-advisor, Prof. Dr. Timothy William Flegel for his guidance and finalcial support throughout this thesis, and also Dr. Siriporn Pongsomboon for her great helps, guidances and suggestions in my thesis.

My gratitude is also extended to Associate Professor Dr. Aran Incharoensakdi, Associate Professor Dr. Siriporn Sittipraneed, and Dr. Teerapong Buaboocha for serving as thesis committee, for their available comments and also useful suggestions. I am very grateful to Dr. Sirawut Klinbunga and Dr. Suwit Wuthisuthimethavee for their sugesstion on genetic linkage map

Thanks are also expressed to all my friends of the Biochemistry Department especially in lab 728 for their helps in the laboratory and friendships that help me enjoy and happy through out my study.

Finally, I would like to express my deepest gratitude to my parents and my lovely sister for their love, care, understanding and encouragement extended throughout my study.

I wish to acknowledge to contributions of the Local Graduate Scholarships (LGS), the National Science and Technology Development Agency, NSTDA for my financial support.

Contents

Page

Page

Page

List of Tables

List of Figures

Page

List of Abbreviations

°C	Degree celsius
µl	microlitre
µM	micromolar
bp	Base pair
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dTTP	deoxythymidine triphosphate
EDTA	ethylene diamine tetraacetic acid (disodium salt)
PIC	polymorphic information content
h_{exp}	expected heterozygosity
h_{obs}	observed heterozygosity
mg	milligram
min	minute
ml	millilitre
mM	millimolar
mmol	millimole
mtDNA	mitochondrial DNA
cM	centiMorgan
θ	recombination fraction
ng	nanogram
OD	optical density
PCR	polymerase chain reaction
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulfate

sec	second
T_m	melting temperature
TE	tris EDTA
Tris	tris (hydroxy methyl)aminomethane
U	unit
V	volt
w/v	weight/volume

CHAPTER I

INTRODUCTION

1.1 General Introduction

Shrimps are the most valuable fisheries, particularly in terms of value of foreign exchange earning. This is because of a strong market demand, with the highest prices coming in the international markets. The United State of America and Japan are the two major consumer markets that have caused the rapid growth of the shrimp industry.

Thailand is the world's leader in shrimp exporter. In the year 2000, shrimp-exporting values reached the highest for last ten-year period with a total export of 249,633 metric tons worth 107,890 million baht (Table1.1). The majority of the harvest shrimps was the black tiger shrimp from culture. Since 2001, Thailand shrimp exporting value was decrease due to several problems including the spread of diseases across vast geographic locations and over exploitation of female broodstock in *P. monodon* culture. Nowadays, the shrimp species mainly cultured in Thailand has switched to the white shrimp, *Litopenaeus vannamei* (Table 1.2) (source: http://www.shrimpcenter.com). Because *L. vannamei* feeds on organisms, which grow naturally in the pond, it is cheaper to feed than *P. monodon*. White shrimp breed in captivity better than *P. monodon*, and has uniform growth rate. Hatchery survivals are high, from 50 to 60% comparing to 20–30% survival rate of *P. monodon*. Throughout Latin America, hatcheries maintain captive stocks of *L. vannamei* broodstock, some of them are pathogen-free, or

pathogen-resistant and some are in captivity for almost 30 years (Source: http://www.shrimpnews.com/Species.html). It is predicted that the white shrimp could become the dominant species in Thailand within the next couple of years. However, it is a non-native species, thus, the broodstock of white shrimp have to be imported from aboard mainly from Hawaii. The market sizes of white shrimp are smaller than those of black tiger shrimp, leading to lower price.

Table 1.1 Cultured shrimp production in Thailand between 1994 – 2004 (Source: Thai Customs Department cited in Shrimp Culture Newsletter)

Year	Exporting value (million baht)	Quantity of export (metric tons)
1994	49,846	190,014
1995	51,270	178,272
1996	43,976	163,545
1997	47,584	138,505
1998	58,343	156,176
1999	87,579	240,529
2000	107,890	249,633
2001	98,680	255,568
2002	73,947	212,091
2003	71,847	234,277
2004	67,289	240,841

Table 1.2 The shrimp production in Thailand between 2002-2004 and the prediction in 2005-2006 (Source: http://www.shrimpcenter.com)

Year	Black tiger shrimp (metric tons)	White shrimp (metric tons)	Percentage of black tiger shrimp to total shrimp (%)
2002	230,000	20,000	92.0
2003	120,000	200,000	37.5
2004	80,000	245,000	24.6
2005	80,000	320,000	20.0
2006	120,000	360,000	25.0

To maintain the production of *P. monodon*, the domestication and genetic improvement of black tiger shrimp are urgently needed for sustainable shrimp culture. Domestication will provide captive broodstock while genetic selection could select traits of economically importance such as disease resistant fast-growing *P. monodon*.

1.2 Taxonomy of *P. monodon*

The taxonomic definition of the black tiger shrimp is as follows (Bailey-Brook and Moss, 1992):

Phylum Arthopoda

Subphylum Crustacea

Class Malacostraca

Subclass Eumalacostraca

Superorder Eucarida

Order Decapoda

Suborder Natantia

Infraorder Penaeid

Superfamily Penaeoidae

Family Penaeidae Rafinesque, 1815

Genus Penaeus Fabricius, 1798

Subgenus *Penaeus*

Species *monodon*

Scientific name: *Penaeus monodon* Fabricius, 1798.

English common name: black tiger shrimp or giant tiger shrimp

It has also four synonyms:

Penaeus carinatus Dana, 1852

P. caeruleus Stebbings, 1905

P. monodon var. *manilenis* Villaluz and Arriola, 1938

P. bubulus Kubo, 1949

1.3 Morphology

Externally the shrimp can be divided basically into the thorax and abdomen (Figure 1.1). The thorax (or head) is covered by a single, immobile carapace, which protects internal organs and support muscle origin. The rostrum, extending beyond the tip of the attenular peduncle, is sigmoidal in shape, and has 6-8 dorsal and 2-4 ventral

teeth, mostly 7 and 3, respectively. The carapace is carinate with the adrostral carina, almost reaching to or not as far as the epigastric or first tooth. The gastro-orbital carina occupies the posterior one-third to one half distance between the post-orbital marine of the carapace and the hepatic spine. The hepatic carina is prominent and slightly curved, extending behind the antennal spine. The antennular flagellum is sub-equal or slightly longer the peduncle. The walking legs or pereiopods are the thoracic appendages. Exopods are present on the 1st and 4th but absent on the 5th pereiopods. Gills are formed from sac-like out growths of the base of the walking legs and sit in branchial chambers on either side of the thorax. The carapace extends laterally to cover the gills completely. The abdomen has the obvious segmentation of invertebrates. Pair of swimming legs or pleopods arises appendages of each of the six abdominal segments. A tail fan comprise of a telson, which bears the anus, and two uropods attaches to the last (6th) abdominal segment. The telson has deep median groove, without dorso-lateral spines. A rapid ventral flexion of the abdomen with the tail fan produces the quick backward dart characteristic of shrimp (Anderson, 1993).

Shrimps grow by periodically releasing their cuticle secreted by the epidermal cell layer, consisting of chitin and proteins. Molting starts when the epidermis detaches from the cuticle layer and begins to secrete a new cuticle. The new cuticle is soft and is stretched to accommodate the increased size of the shrimp immediately after molting. In *P. monodon*, mating occurs just after the female molts. At this time the male can insert the spermatopore through the soft cuticle of the thelycum (an organ for internal storage

of spermatophores). Shrimps spawn directly into the sea water, and the eggs are fertilized by the stored spermatozoa at the moment of spawning (Anderson, 1993).

A live black tiger shrimp has the following characteristic coloration: carapace and abdomen are transversely banded with red and white, the antennae are greyish brown, and the pereiopods and pleopods are brown with crimson fringing setae. In shallow brackish water or when cultured in ponds, the color changes to dark and, often, to blackish brown (Motoh, 1981: cited in Solis, 1988).

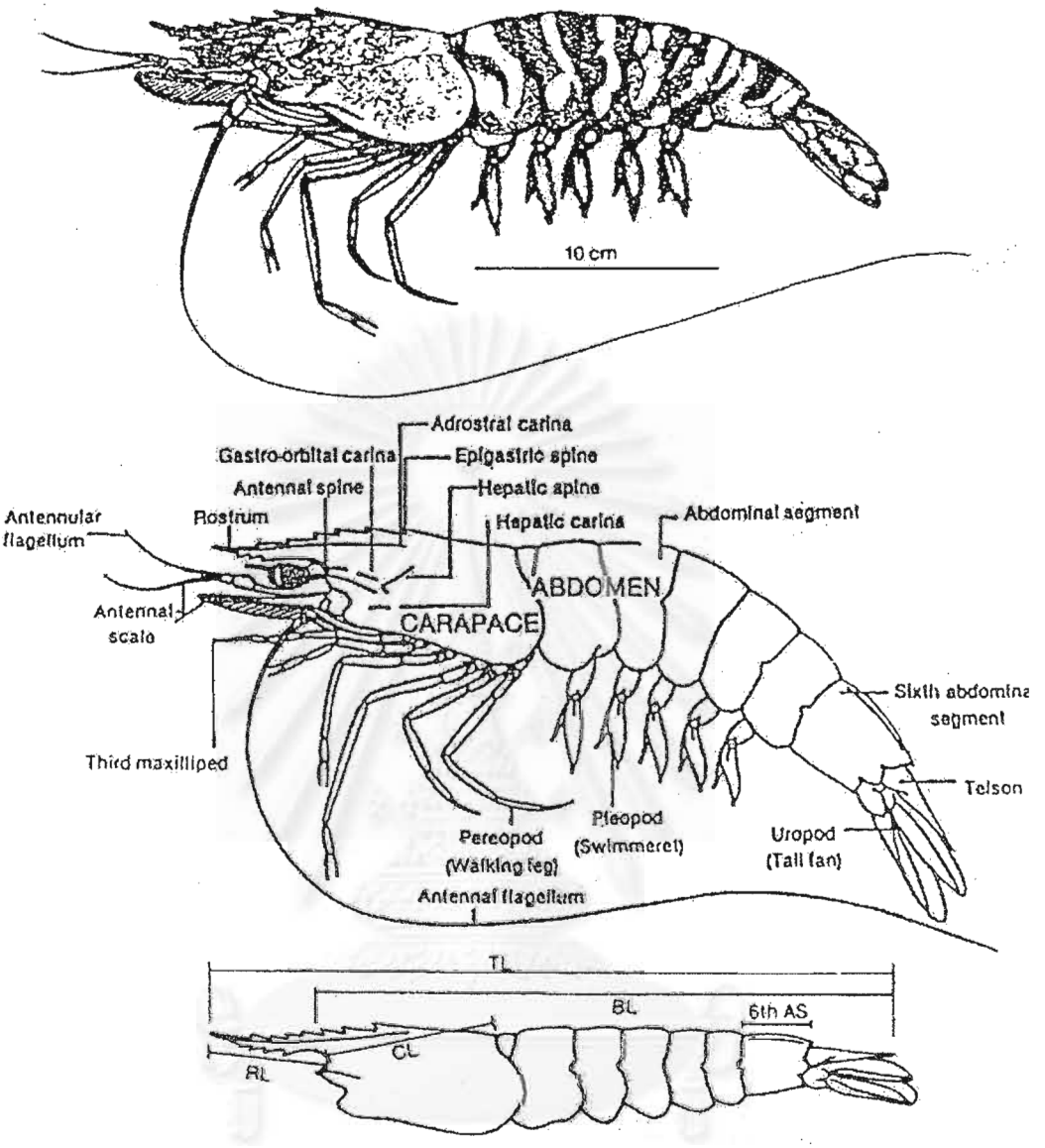


Figure 1.1 Upper: Adult female of *P. monodon*, Middle: Lateral view of *P. monodon* showing important parts, Lower: Methods of measurement of *P. monodon* (RL, rostrum length; CL, carapace length; TL, total length; BL, body length, 6^{th} AS, length of the 6^{th} abdominal segment). Motoh (1985)

1.4 Life cycle

The penaeid life cycle includes several distinct stages found in a variety of habitats. Juveniles often prefer brackish waters of estuaries and coastal wetlands, while adults are usually found offshore at higher salinities and greater depths. Larval stages inhabit surface water offshore, with an on-shore migration as they develop.

Development of penaeid shrimps is complex. It begins with a larva hatching from the fertilized egg to the first stage, nauplius, which occurs about 12 hours later (Figure 1.2). The larval stages consist of three to six nauplii, three protozoea and two or three mysis substages depending on shrimp species. This larval development period varies with temperature and feeding levels but is usually 10-14 days. Mysis II larvae molts to become post larvae (PL) with have all appendages and organs seen in adults. Larvae exhibit planktonic behavior with antennal propulsion for swimming in nauplii, antennal and thoracic propulsion in mysis, and abdominal propulsion in megalopa. Nauplii utilize yolk granules within their body while the feeding starts in protozoea and mysis. At the mysis stage, larvae have five pairs of functioning pereiopods. The carapace now covers all of the thoracic segments. The mysis swims like adults. After this stage, larvae metamorphose to the post-larvae with a full complement of functioning appendages. The post-larvae continue to be molting as they grow. They migrate shoreward and settle in nursery areas closed to shore or estuaries, before develop to juvenile and sub-adults, which more tolerate to variety of environmental factors. Sub-adults migrate back to the

sea where they finally mature and have the first corpulation and spawn. The life span of penaeid shrimp is approximately 2 years (Anderson, 1993; Solis, 1988).

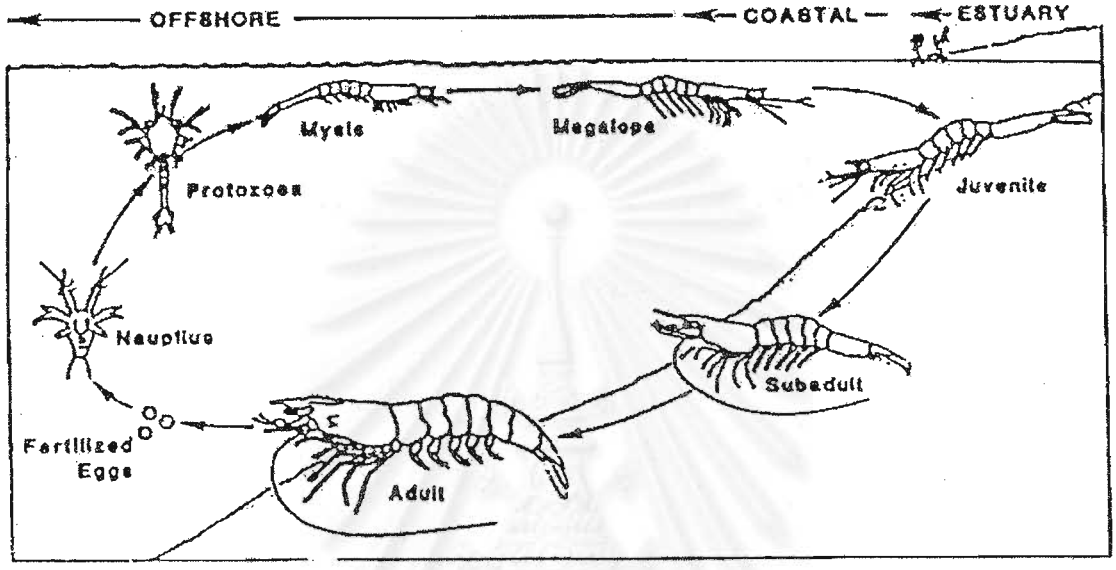


Figure 1.2 The life cycle of the black tiger shrimp, *P. monodon*, with stages in different habitats (Baily-Brook and Moss, 1992)

1.5 Distribution

The tiger shrimp, *P. monodon* Fabricius, is one of the largest penaeid shrimps. It is widely distributed throughout the greater part of the Indo-West Pacific region; South Africa, Tanzania, Kenya, Somalia, Madagascar, Saudi Arabia, Oman, Pakistan, India, Bangladesh, Sri Lanka, Indonesia, Thailand, Malaysia, Singapore, Philippines, Hong Kong, Taiwan, Korea, Japan, Australia and Papua New Guinea. Generally, *P. monodon* is distributed from 30^{O}E. to 155^{O}E. longitude and from 35^{O}N. to 35^{O}S. latitude (Motoh, 1985). The fry, juvenile and adolescent stages of *P. monodon* inhabit inshore areas and

mangrove estuaries. On the other hand, the adults basically inhabit offshore areas up to 160 m. in depth (Figure 1.3).

The reason for the widespread culture of this species is that it has several advantages over other penaeid shrimps. First of all, it has a higher growth rate in comparison to *P. japonicus, P. semisulcatus, P. peniculatlls, P. stylelotris* and *Metapenaeus* spp.. The second feature of this species is that it is a euryhaline animal. Its growth is not retarded even when reared in salinities near to freshwater. Furthermore, *P. monodon* has a wide temperature tolerance. It can tolerate temperatures as low as 12^{o}C or as high as 36^{o}C for a short period of time. The general tolerance of this species is reflected in the large geographic range over which it is cultured. The third feature in favour of this species is that it is omnivorous rather than carnivorous, therefore it needs a comparatively low protein feed. This is reflected in low production costs. Fourth, simple culture facilities such as clay bottom ponds are more than adequate for grow-out and generally give good results for intensive culture. The last important feature is that *P. monodon* has a high survival rate in culture. These characteristics make *P. monodon* an ideal culture species.

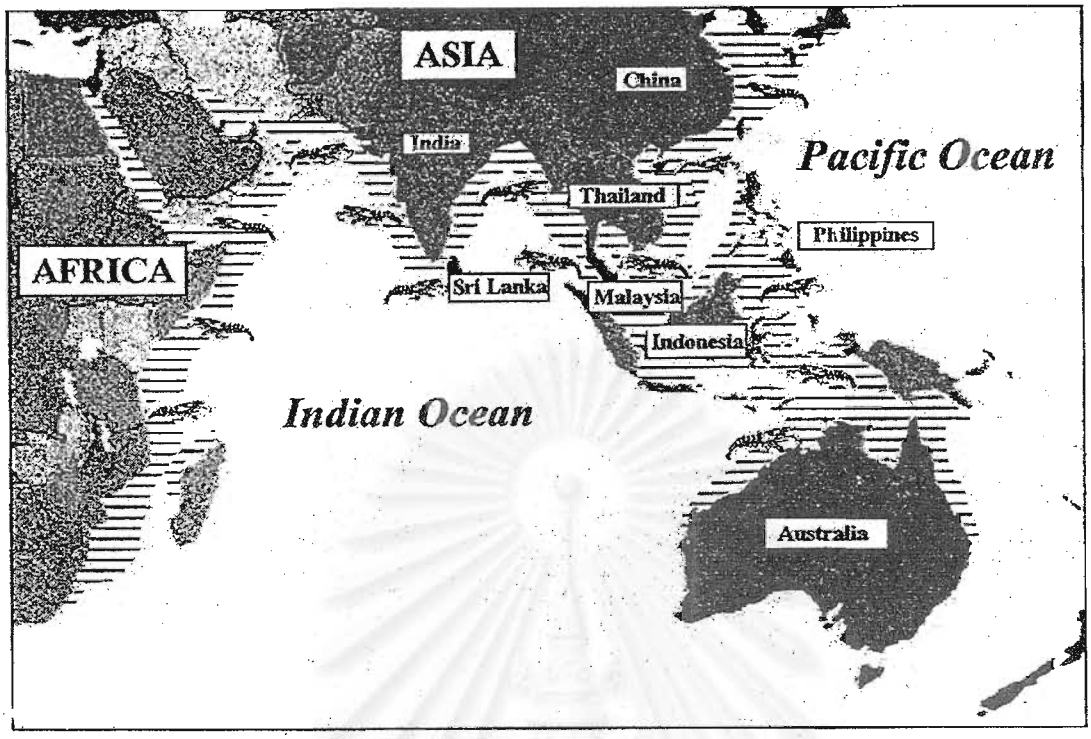


Figure 1.3 Geographic distributions of *P. monodon* in the Indo-West Pacific regions (Grey *et al.*, 1983)

1.6 Exploitation

The *P. monodon* farming in Thailand is one of the important industrial aquaculture. This causes the increasing use of wild female broodstock, because the farming of this species relies entirely on wild females for the seed production. This causes a lack of wild female broodstock. Currently, the culture of this species decline due to the spread of disease while the cultures of *P. vannamei* increase instead. *P. vannamei* has been domesticated and achieved genetics selection for fast growing and disease-resistant to get desirable traits. Domestication allows convenient disease prevention and control. The genetic selection may also be achieved through this process. As a result, domesticated *P. monodon* should be substituted for the use of wild

females as the only source of postlarvae production in the shrimp industry. These would decrease the risk of a loss of the production.

1.7 Genetic markers

The development of DNA-based genetic markers has had a revolutionary impact on aquaculture genetics. With DNA markers, it is theoretically possible to observe and exploit genetic variation in the entire genome. The application of DNA markers has allowed rapid progress in aquaculture investigations of genetic variability and inbreeding, parentage assignments, species and strain identification, and the construction of high-resolution genetic linkage maps for aquaculture species. Well-designed studies using these genetic markers will undoubtedly accelerate identification of genes involved in quantitative trait loci (QTL) for marker-assisted selection (Liu, 2004).

Several marker types are highly popular in aquaculture genetics. In the past, allozyme and mtDNA markers were commonly used in aquaculture genetics research. More recent marker types that are finding service in this field include restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellite, single nucleotide polymorphism (SNP), and expressed sequence tag (EST) markers. These markers exhibit variation in the degree of polymorphisms, thus, they can be used in different applications (Table 1.1)

Table 1.3 Summarizes the basic properties of different marker types

Marker type	Acronym or alias	Requires prior molecular information?	Mode of inheritance	Type	Locus under investigation	Likely allele numbers	Polymorphism or power	Major applications
Allozyme		Yes	Mendelian, codominant	Type I	Single	2–6	Low	Linkage mapping, population studies
Mitochondrial DNA	mtDNA	No[a]	Maternal inheritance	—		Multiple haplotypes		Maternal lineage
Restriction fragment length polymorphism	RFLP	Yes	Mendelian, codominant	Type I or type II	Single	2	Low	Linkage mapping
Random amplified polymorphic DNA	RAPD, AP-PCR	No	Mendelian, dominant	Type II	Multiple	2	Intermediate	Fingerprinting for population studies, hybrid identification
Amplified fragment length polymorphism	AFLP	No	Mendelian, dominant	Type II	Multiple	2	High	Linkage mapping, population studies
Microsatellites	SSR	Yes	Mendelian, codominant	Mostly Type II	Single	Multiple	High	Linkage mapping, population studies, paternity analysis
Expressed sequence tags	EST	Yes	Mendelian, codominant	Type I	Single	2	Low	Linkage mapping, physical mapping, comparative mapping
Single nucleotide polymorphism	SNP	Yes	Mendelian, codominant	Type I or type II	Single	2, but up to 4	High	Linkage mapping, population studies?
Insertions/deletions	Indels	Yes	Mendelian, codominant	Type I or type II	Single	2	Low	Linkage mapping

[a] Conserved PCR primers can be adopted from sequence information from a related species.

1.7.1 Type I versus type II markers and polymorphic information content (PIC)

Molecular markers are classified into two categories: type I are markers associated with genes of known function, while type II markers are associated with anonymous genomic segments (O'Brien, 1991; Table 1.3). Under this classification, most RFLP markers are type I markers because they were identified during analysis of known genes. Likewise, allozyme markers are type I markers because the protein they encode has known function. RAPD markers are type II markers because RAPD bands are amplified from anonymous genomic regions via the polymerase chain reaction (PCR). AFLP markers are type II because they are also amplified from anonymous genomic regions. Microsatellite markers are type II markers unless they are associated with genes of known function. EST markers are type I markers because they represent transcripts of genes. SNP markers are mostly type II markers unless they are developed from expressed sequences. Insertion and deletion (indel) are becoming more widely used as markers since they often are discovered during genomic or transcriptomic sequencing projects; they can be either type I or type II markers depending on whether they are located in genes.

The significance of type I markers was not fully appreciated in the early stages of aquaculture genetics, though it is becoming clear that these markers are extremely important. In addition to their functions as markers in population studies, type I markers are becoming very important in studies of genetic linkage and QTL mapping. Type I markers have utility in studies of comparative genomics, genome evolution, candidate

gene identification, and enhanced communication among laboratories. Due to evolutionary constraints on the genome, many genes and their organization are conserved among species. Comparative genomics deals with the similarity and differences found among genomes. Type I markers serve as a bridge for comparison and transfer of genomic information from a map-rich species into a relatively map-poor species. Such interspecific comparisons can also be made based on type II markers, but the extent to which the comparison can be made is limited to closely related taxa. The requirement for such comparisons lies in sequence conservation. For the most frequently used microsatellite markers, such comparative studies depend on conservation of the flanking sequences used for the design of PCR primers. In contrast, sequence conservation within genes are high, allowing type I markers to serve as anchor points for genomic segments to be compared among species. For instance, if 15 genes are located between type I markers A and B in shrimp, it is likely that the majority of the 15 genes also reside between markers A and B in crayfish, even though the exact number of genes, gene order, and orientation are not necessarily identical (Liu, 2004).

In general, type II markers such as RAPDs, microsatellites, and AFLPs are considered to be non-coding and therefore selectively neutral. Such markers have found widespread use in population genetic studies whose characterizations of genetic diversity and divergence within and among populations are based on assumptions of Hardy–Weinberg equilibrium and selective neutrality of the markers employed (Brown and Epifanio, 2003). Type II markers also have proven useful in aquaculture genetics for

species, strain and hybrid identification, in breeding studies, and more recently as markers linked to QTL.

The usefulness of molecular markers can be measured based on their polymorphic information content (PIC, Botstein *et al.*, 1980). PIC refers to the value of a marker for detecting polymorphism in a population. PIC depends on the number of detectable alleles and the distribution of their frequencies.

1.7.2 Allozyme markers

Allozymes are allelic variants of proteins produced by a single gene locus, and are of interest as markers because polymorphism exists and because they represent protein products of genes and are thus type I markers. Since the 1960s, starch gel electrophoresis of allozymes has been the most commonly employed molecular method in fishery genetics (Ryman and Utter, 1987; Hillis *et al.*, 1996). Still in widespread use, allozymes were among the earliest markers used in aquaculture genetics (May *et al.*, 1980; Seeb and Seeb, 1986; Johnson *et al.*, 1987; Liu *et al.*, 1992, 1996; Morizot *et al.*, 1994).

Amino acid differences in the polypeptide chains of the different allelic forms of an enzyme reflect changes in the underlying DNA sequence. Depending on the nature of the amino acid changes, the resulting protein products may migrate at different rates (due to charge and size differences) when run through a starch gel subjected to an electrical field. Differences in the presence/absence and relative frequencies of alleles are used to quantify genetic variation and distinguish among genetic units at the levels of

populations, species, and higher taxonomic designations. Allozymes found use in aquaculture for tracking inbreeding, stock identification, and parentage analysis. In a few cases, correlation existed between certain allozyme markers and performance traits (Hallerman *et al.*, 1986; McGoldrick and Hedgecock, 1997). Their use in linkage mapping has been demonstrated in studies of salmonids (Pasdar *et al.*, 1984; May and Johnson, 1993) and poeciliids (Morizot *et al.*, 1991), but the limited number of available allozyme loci precludes their use in large-scale genome mapping. Disadvantages associated with allozymes include heterozygote deficiencies due to null (enzymatically inactive) alleles and the amount and quality of tissue samples required. In addition, some changes in DNA sequence are masked at the protein level, reducing the level of detectable variation. Some changes in nucleotide sequence do not change the encoded polypeptide (silent substitutions), and some polypeptide changes do not alter the mobility of the protein in an electrophoretic gel (synonymous substitutions). Although 75 isozyme systems representing several hundred genetic loci are available (Murphy *et al.*, 1996), the relatively modest number of loci usually employed and the low number of alleles (usually two or three) exhibited by most loci tend to keep the PIC of these markers fairly modest. Low levels of genetic variation revealed in many allozyme studies of marine fish populations (e.g. Siddell *et al.*, 1980; Mork *et al.*, 1985; Crawford *et al.*, 1989) prompted a continued search for markers with greater genetic resolution. In spite of their strength as codominant type I markers, ease of use, and low cost, their use in aquaculture genetics has become limited.

1.7.3 Mitochondrial DNA markers

Studies of vertebrate species generally have shown that sequence divergence accumulates more rapidly in mitochondrial than in nuclear DNA (Brown, 1985). This has been attributed to a faster mutation rate in mtDNA that may result from a lack of repair mechanisms during replication (Wilson *et al.*, 1985) and smaller effective population size due to the strictly maternal inheritance of the haploid mitochondrial genome (Birky *et al.*, 1989). Almost the entire mtDNA molecule is transcribed except for the approximately 1 kb control region (D-loop), where replication and transcription of the molecule is initiated. In general, non-coding segments like the D-loop exhibit elevated levels of variation relative to coding sequences such as the cytochrome b gene (Brown *et al.*, 1993), presumably due to reduced functional constraints and relaxed selection pressure.

Analyses of mtDNA markers have been used extensively to investigate stock structure in a variety of fishes including eels (Avise *et al.*, 1986), bluefish (Graves *et al.*, 1992), red drum (Gold *et al.*, 1993), snappers (Chow *et al.*, 1993), and sharks (Heist and Gold, 1999). Mitochondrial markers are quite popular among aquaculture geneticists, in part due to their use in identification of broodstocks (e.g. Benzie *et al.*, 2002). In the early days of molecular analysis, the high levels of mtDNA polymorphism relative to allozymes were exploited for population differentiation. Technically, mtDNA markers are RFLP markers except that the target molecule is mtDNA rather than nuclear genomic DNA (Cronin *et al.*, 1993). Although mtDNA loci can exhibit large numbers of alleles per loci, the limited number of markers available on the mtDNA molecule positions its PIC values

higher than those for allozymes but lower than highly variable nuclear markers such as RAPDs, microsatellites, AFLPs, and SNPs.

Due to its non-Mendelian mode of inheritance, the mtDNA molecule must be considered a single locus in genetic investigations (Avise, 1994). In addition, because mtDNA is maternally inherited, the phylogenies and population structures derived from mtDNA data may not reflect those of the nuclear genome due to gender-biased migration (Birky *et al.*, 1989) or introgression (Chow and Kishino, 1995). In addition, mtDNA markers are subject to the same problems that exist for other DNA-based markers, such as back mutation (sites that have already undergone substitution are returned to their original state), parallel substitution (mutations occur at the same site in independent line ages), and rate heterogeneity or mutational hot spots (large differences in the rate at which some sites undergo mutation when compared to other sites in the same region).

1.7.4 Restriction fragment length polymorphism (RFLP)

RFLP markers (Botstein *et al.*, 1980) were regarded as the first shot in the genome revolution (Dodgson *et al.*, 1997), marking the start of an entirely different era in the biological sciences. The procedures and principles of RFLP markers are summarized in figure 1.4. Restriction endonucleases are bacterial enzymes that recognize specific 4, 5, 6, or 8 base pair (bp) nucleotide sequences and cut DNA wherever these sequences are encountered, so that changes in the DNA sequence due to insertion or deletion, base substitutions, or rearrangements involving the restriction sites can result in the gain,

loss, or relocation of a restriction site. Digestion of DNA with restriction enzymes results in fragments whose number and size can vary among individuals, populations, and species. Traditionally, fragments were separated using Southern blot analysis (Southern, 1976), in which genomic DNA is digested, subjected to electrophoresis through an agarose gel, transferred to a membrane, and visualized by hybridization to specific probes.

Most recent analyses replace the tedious Southern blot method with techniques based on the polymerase chain reaction (PCR). If flanking sequences are known for a locus, the segment containing the RFLP region is amplified via PCR. If the length polymorphism is caused by a relatively large (> approx. 100 bp depending on the size of the undigested PCR product) indel, gel electrophoresis of the PCR products should reveal the size difference. However, if the length polymorphism is caused by base substitution at a restriction site, PCR products must be digested with a restriction enzyme to reveal the RFLP. In addition, PCR products can be digested with restriction enzymes and visualized by simple staining with ethidium bromide due to the increased amount of DNA produced by the PCR method.

The potential power of RFLP markers in revealing genetic variation is relatively low compared to more recently developed markers. Indels and rearrangements of regions containing restriction sites are perhaps widespread in the genomes of most species, but the chances of such an event happening within any given locus under study should be rare. Similarly, in a given genome of 109 base pairs, approximately 250,000

restriction sites should exist for any restriction enzyme with a 6-bp recognition sequence (that accounts for $1.5x10^6$ bp or 0.15% of the entire genome). Base substitutions within these restriction sites must be widespread as well, but again, the chances that such base substitutions would occur within the locus under study would be relatively small.

The major strength of RFLP markers is that they are codominant markers, i.e., both alleles in an individual are observed in the analysis. Because the size difference is often large, scoring is relatively easy. The major disadvantage of RFLP is the relatively low level of polymorphism. In addition, either sequence information (for PCR analysis) or probes (for Southern blot analysis) are required, making it difficult and time-consuming to develop markers in species lacking known molecular information.

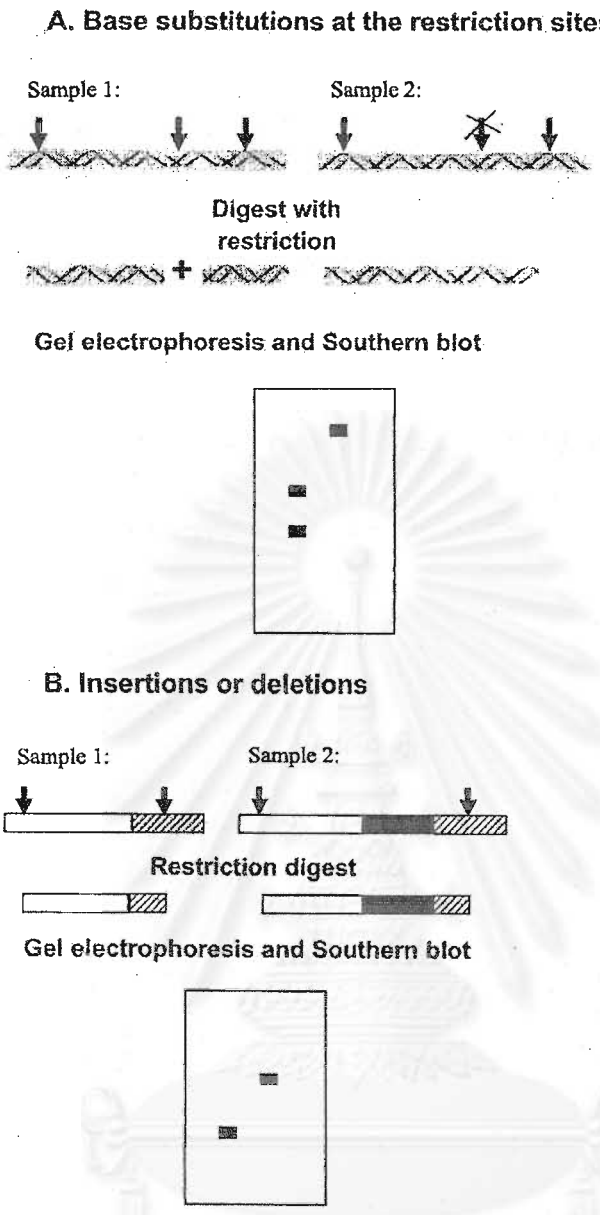


Figure 1.4 Molecular basis of restriction fragment length polymorphism (RFLP). Genomic DNA is digested with a particular restriction enzyme (arrows). (A) Base substitution within the restriction site can knock out the site (conversely, new sites can arise by base substitutions as well). Loss of a restriction within the locus of interest leads to loss of one fragment and increase of fragment size that can be resolved by gel electrophoresis. (B) Insertion of a piece (black bar) between two restriction sites within the locus leads to an increase in the fragment size that can be resolved by gel electrophoresis (reversely a deletion should reduce the fragment size). In both cases, classical means of detection relies on Southern blot using specific probes.

1.7.5 Random amplified polymorphic DNA (RAPD)

RAPD procedures were first developed in 1990 (Welsh and McClelland, 1990; Williams *et al.*, 1990) using PCR to randomly amplify anonymous segments of nuclear DNA with an identical pair of primers 8–10 bp in length (Figure 1.5). When the two primers bind to sites close enough (often less than 2000 base pairs) on opposite strands of DNA (indicated by arrowed segments with base pairing), a PCR product results. Random primers of 10 bases by chance have about 1000 perfect binding sites on each strand of genomic DNA for a genome size of $1x10^9$ base pairs. It should benoted that non-perfect binding of primers to genomic DNA templates (e.g., 9 out of 10 bases pair with genomic DNA, as shown with the primer on the right) may also lead to production of PCR products if the 3' end of the primer has strong base-pairing. Because the primers are short and relatively low annealing temperatures (often 36–40°C) are used, the likelihood of amplifying multiple products is great, with each product (presumably) representing a different locus. Because most of the nuclear genome in vertebrates is non-coding, it is presumed that most of the amplified loci will be selectively neutral. Genetic variation and divergence within and between the taxa of interest are assessed by the presence or absence of each product, which is dictated by changes in the DNA sequence at each locus. RAPD polymorphisms can occur due to base substitutions at the primer binding sites or to indels in the regions between the sites. The potential power is relatively high for detection of polymorphism; typically, 5–20 bands can be produced using a given primer pair, and multiple sets of random primers can be used to scan the

entire genome for differential RAPD bands. Because each band is considered a bi-allelic locus (presence or absence of an amplified product), PIC values for RAPDs fall below those for microsatellites and SNPs, and RAPDs may not be as informative as AFLPs because fewer loci are generated simultaneously.

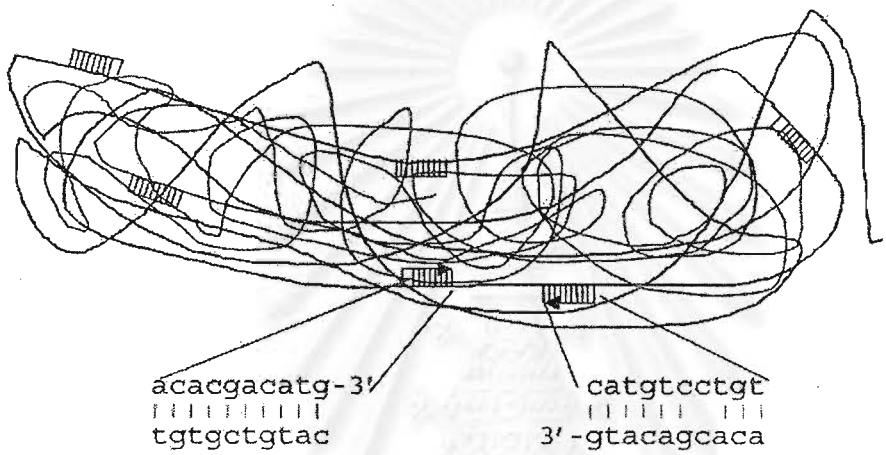


Figure 1.5 Schematic presentation of the RAPD procedure. Genomic DNA (indicated by long strings of lines) is used for PCR using two arbitrary short primers of identical sequences (indicated by short segments annealing to their complementary sites in the genome either perfectly or non-perfectly) under low annealing temperatures.

RAPD markers are inherited as Mendelian markers in a dominant fashion and scored as present/absent. A RAPD band is produced by homozygotes as well as heterozygotes, and though band intensity may differ, variations in PCR efficiency makes scoring of band intensities difficult. As a result, distinguishing homozygous dominant from heterozygous individuals is not generally possible. In addition, it is difficult to

determine whether bands represent different loci or alternative alleles of a single locus, so that the number of loci under study can be erroneously assessed. This is especially true if the RAPD is caused by deletion or insertion within the locus rather than at the primer binding sites. In breeding studies, the number of RAPD bands seen in the F1 generation should equal the sum of the bands seen in the parents, assuming parental homozygosity at each locus; polymorphic RAPD then segregates in a 3:1 ratio in F2 populations

RAPDs have all the advantages of a PCR-based marker, with the added benefit that primers are commercially available and do not require prior knowledge of the target DNA sequence or gene organization. Multi locus amplifications can be separated electrophoretically on agarose gels and stained with ethidium bromide, although higher resolution of bands has been achieved with discontinuous polyacrylamide gel electrophoresis (dPAGE) and silver staining (Dinesh *et al.*, 1995), a somewhat costlier and more labor-intensive method. Other advantages of RAPDs are the ease with which a large number of loci and individuals can be screened. RAPD makers have been used for species identification in fishes (Partis and Wells, 1996) and mollusks (Klinbunga *et al.*, 2000; Crossland *et al.*, 1993), analysis of population structure in black tiger shrimp (Tassanakajon *et al.*, 1998) and marine algae (Van Oppen *et al.*, 1996), analysis of genetic impact of environmental stressors (Bagley *et al.*, 2001), and analysis of genetic diversity (Wolfus *et al.*, 1997;Hirschfeld *et al.*, 1999; Yue *et al.*, 2002).

1.7.6 Amplified Fragment Length Polymorphisms (AFLPs)

The AFLP technique is also a PCR based technique, however unlike RAPDs, specific restriction fragments are targeted for amplification. Genomic DNA is digested with two different restriction enzymes, normally a hexacutter and a tetracutter (recognizing a four base pair sequence). Specific adapter molecules are then ligated to the ends of the restriction fragments. The oligonucleotide primers for the PCR reactions correspond to these adapter molecules. A greater specificity is achieved by adding an extra nucleotide to the PCR primer corresponding to the internal nucleotide following the restriction site. A preliminary amplification reaction is carried out using this +1 primer allowing the amplification of only a subset of the population of digested molecules. This initial amplification is allowed by a second amplification using an oligonucleotide primer with three extra bases, therefore only a fraction of the originally amplified fragments is subsequently amplified i.e. those processing precisely the combination of internal bases determined by the primers. Different combinations of three bases sequence at each restriction site lead to analysis of different fraction of the genome. In order to detect the amplified fragments, one of the +3 primers has radioactive or fluorescent label attached, the amplification products are then run out on a polyacrylamide sequencing gel and visualized by autoradiography or by scanning in a special apparatus designed to detect fluorescent compounds. The general idea behind the AFLP technique is to obtain information simultaneously for many restriction fragment polymorphisms but in an easily manageable fashion, therefore although initial digestion, ligation and amplification

reactions involve the whole genome, various steps within the protocol lead us to analyze only a small proportion of genome (Lin and Kuo, 1995; Vos *et al.*, 1995).

AFLPs have been shown to detect high levels of polymorphism in many different organisms. The polymorphism in this case has the same basis as for RFLPs, namely mutations in or around the restriction sites of the enzymes used in the initial digest. In spite of the complexity of the protocol for carrying out AFLP analysis and the preparation of sequencing gels, many reports of mapping and tagging using the AFLP technique have been published recently. The AFLP technique, like RAPDs, could be easily automated for high sample throughput (Cervera *et al.*, 1996; Hill *et al.*, 1996).

1.7.7 Microsatellites

Microsatellites consist of multiple copies of tandemly arranged simple sequence repeats (SSRs) that range in size from 1 to 6 base pairs (e.g., ACA or GATA; Tautz, 1989; Litt and Luty, 1989). Abundant in all species studied to date, microsatellites have been estimated to occur as often as once every 10 kb in fishes (Wright, 1993). Microsatellites tend to be evenly distributed in the genome on all chromosomes and all regions of the chromosome. They have been found inside gene coding regions (e.g. Liu *et al.*, 2001), introns, and in the non-gene sequences. The best known examples of microsatellites within coding regions are those causing genetic diseases in humans, such as the CAG repeats that encode polyglutamine tract, resulting in mental retardation. Most microsatellite loci are relatively small, ranging from a few to a few hundred repeats. The relatively small size of microsatellite loci is important for PCR-facilitated genotyping.

Generally speaking, microsatellites containing a larger number of repeats are more polymorphic, though polymorphism has been observed in microsatellites with as few as five repeats (Karsi *et al.*, 2002).

Microsatellite polymorphism is based on size differences due to varying numbers of repeat units contained by alleles at a given locus. Microsatellite mutation rates have been reported as high as 10^{-2} per generation (Weber and Wong, 1993; Crawford and Cuthbertson, 1996), and are believed to be caused by polymerase slippage during DNA replication, resulting in differences in the number of repeat units (Levinson and Gutman, 1987; Tautz, 1989). Direct studies of human families have shown that new microsatellite mutations usually differed from the parental allele by only one or two repeats (Weber and Wong, 1993), favoring a stepwise mutation model (Estoup and Cornuet, 1999). Regardless of specific mechanisms, changes in numbers of repeat units can result in a large number of alleles at each microsatellite locus in a population.

Microsatellites are inherited in a Mendelian fashion as codominant markers (Fig. 1.6). In this figure shows the two pairs of mating used to produce the F1. In the first pair, the female has genotype AB at the locus; the male has genotype CD; and their F1 (one female individual) has genotype of AD. In the second pair, the female has genotype BC at the locus; the male has genotype CC; and their F1 (one male individual) has genotype BC. The F2 individuals produced from the two F1 individuals (AD and BC) have all four possibilities of genotypes under independent segregation: AB, AC, BD, and CD. Note

that both alleles are present for co-dominant markers. If only one band is observed, it is homozygous at the locus.

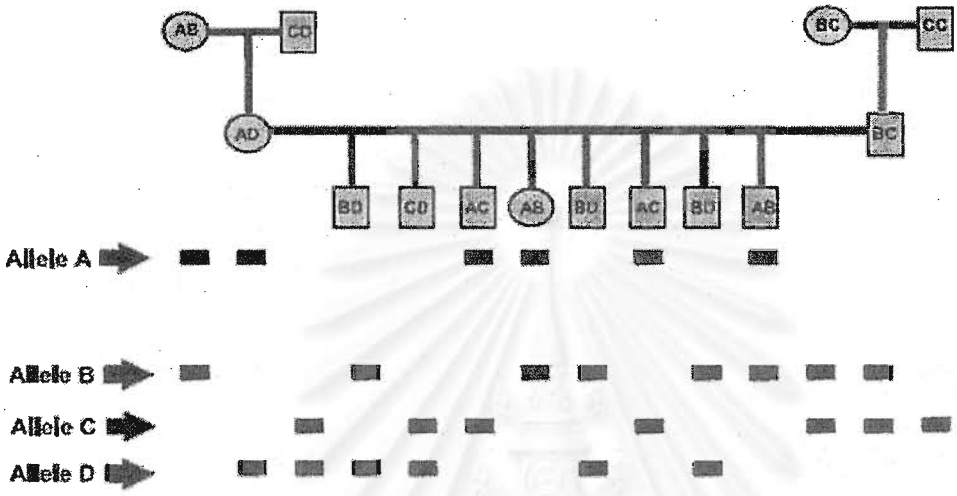


Figure 1.6 Schematic presentation of co-dominant marker inheritance. (Details were described in text)

Other strengths of microsatellite markers in addition to their abundance, include even genomic distribution, small locus size, and high polymorphism. However, use of microsatellite markers involves a large amount of up-front investment and effort. Each microsatellite locus has to be identified and its flanking region sequenced for the design of PCR primers. For most efficient marker development, microsatellite-enriched genomic DNA libraries are made (Ostrander *et al.*, 1992; Kijas *et al.*, 1994). Due to polymerase slippage during replication, small size differences between alleles of a given microsatellite locus (as little as 2 bp in a locus comprised of dinucleotide repeats) are possible. Because of this, PCR-amplified microsatellite DNA was traditionally labeled

radioactively, separated on a sequencing gel, and then exposed on X-ray film overnight (Sambrook *et al.*, 1989). Significant increases in the number of samples that can be typed in a day have been achieved by using automated fluorescent sequencers coupled with computer imaging systems (O'Reilly and Wright, 1995).

The large number of alleles per locus results in the highest PIC values of any DNA markers. Microsatellites recently have become an extremely popular marker type in a wide variety of genetic investigations, as evidenced by the recent debut of the journal Molecular Ecology Notes, dedicated almost entirely to publishing primer and allele frequency data for newly characterized microsatellite loci in a wide range of species. Over the past decade, microsatellite markers have been used extensively in fisheries research including studies of genome mapping, parentage, kinships, and stock structure (O'Connell and Wright, 1997); a cursory online literature search produced over 500 entries since 1998 involving the utilization of microsatellites in such studies.

1.8 Genetic markers in penaeid shrimps

Dating back to the mid-1960s, electrophoretic techniques for protein separation followed by histochemical localization of specific enzymes allowed geneticists, for the first time, to survey natural populations for the amount of variation in a random sample of primary gene products. As investigators turned to biochemical techniques. The interest in externally visible polymorphisms declined. The enzyme polymorphisms posses significant advantages, including codominant expression (Hedgecock et al., 1982).

The identification of the genetic diversity in penaeid shrimp is through the examination of allozyme variability indicates relatively few allozyme polymorphisms. Low levels of genetic variation and little geographic differentiation within wild penaeid shrimp species have been reported (Garcia *et al.*, 1994). Using three penaeid shrimps from the Gulf of Mexico, *Penaeus aztecus*, *P. setiferus* and *P. duorarum*, Lester (1979), could not demonstrate the significant differences among locality differentiation of allozyme frequencies. However, Benzie *et al.* (1992) found significant allozyme frequency differences among Australian populations of *P. monodon*. However, mtDNA analysis indicated higher levels of variation among the Australian populations of *P. monodon*, suggesting that DNA analysis would provide a better source of markers for penaeid shrimps.

A randomly amplified polymorphic DNA (RAPD) analysis has been demonstrated as useful markers in penaeid shrimp breeding programs (Garcia and Benzie, 1995). The percentage of RAPD markers generated that were polymorphic (6-7%) was similar to that observed in other organisms. A RAPD marker specific to the population originated from Ecuador was reported by Garcia *et al.* 1994. Tassanakajon *et al.* (1997) reported that the primer UBC 428 generated a RAPD marker that was found only in *P. monodon* originating from Satun-Trang, located in the Andaman Sea. A RAPD marker generated by UBC 268 was also found specific to the population of *P. monodon* from the Gulf of Thailand (Klinbunga *et al.*, 2001). These results suggested the potential use of these markers as population specific markers. The RAPD analysis of wild populations of Thai *P.*

monodon suggested high level of polymorphism and the existence of population differentiation (Tassanakajon *et al.*, 1998).

Mitochondrial DNA restriction fragment length polymorphism (mtDNA-RFLP) was successfully used for determination of genetic diversity in 3 wild populations of *P. monodon* (Klinbunga *et al.*, 1999). Recently, PCR-RFLPs of 2 mtDNA genes (16S rDNA and COI-COII) of *P. monodon* samples from the 2 coastal areas of Thailand showed high average nucleotide diversity within samples of Thai *P. monodon* compared to several marine organisms (Klinbunga *et al.*, 2001).

For AFLP markers, Pérez *et al*, (2004) construct of sex-specific linkage maps for the white shrimp *Litopenaeus vannamei* . They used 103 primer combinations to obtain 741 segregating bands. From them, the female map covered 2,771 cM and was 24% longer than the male map (2,116 cM long)

The use of microsatellite markers in penaeid shrimps was first demonstrated by Wolfus *et al.* (1997). A microsatellite locus was employed to determine the allelic inheritance within 14 families of *P. vannamei* stocks. Meehan *et al.* (2003) developed 136 primers that flanked microsatellites in *P. vannamei*. They found that 68% were polymorphic in cultured shrimp, with PIC values ranging from 0.195 to 0.873. Of these markers, they are being used along with other markers to construct a linkage map for *P. vannamei*. For the black tiger shrimp, Tassanakajon *et al.* (1998) and Pongsomboon (2002) isolated and characterized 21 microsatellite loci from *P. monodon* genomic libraries. Microsatellite markers were also used to illustrate geographic population

differentiation of *P. monodon* in Thailand and Australia (Supangul *et al.*, 2000, Brooker et al., 2000). Wilson et al. (www.aims.gov.au, 2004) constructed an initial genetic linkage map of *P. monodon* from 3 reference families using a several markers. This linkage map has 36 linkage groups for male map and 29 linkage groups for female map covering a total genetic distance of 878 and 630 cM for male and female maps, respectively. Recently, Wuthisuthimethavee *et al.* (2003) developed 102 microsatellite markers and 27 microsatellite markers were successfully placed into the genetic linkage map forming 9 linkage groups (Wuthisuthimethavee *et al.*, 2005). At present, *P. monodon* genetic linkage maps have not yet cover a whole genome. A more defined *P. monodon* map require sufficient markers including microsatellites to create larger number of linkage groups covering a long genetic distance.

1.9 Bioinformatics mining of type I microsatellites

In Expressed Sequence Tag (ESTs) project of several plants and animals, a bioinformatics was used for searching type I microsatellites located in the transcripts. In 2000, Scott *et al.* analyzed microsatellites derived from ESTs of grape *Vitis vinifera*. They searched a total of 5,000 ESTs in database and found that 10 of 16 developed microsatellite loci were polymorphic. The polymorphic microsatellite loci derived from the 3' untranslated region (3' UTR) showed higher polymorphic than 5' UTR. Thiel *et al.* (2003) exploited EST databases for the development and characterization of gene-derived microsatellite markers in barley *Hordeum vulgare* L. They screened 311 primer

pairs flanking microsatellite loci. Only 76 EST-derived microsatellite markers were integrated into a barley genetic consensus map. 3' UTR yielded a higher portion of polymorphic microsatellite (64%) than 5' UTR did. Recently, a genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton was constructed (Han *et al.*, 2004). Nine hundred and thirty three ESTs were found to contain microsatellites; 544 (58.4%) primer pairs were developed. However, only 99 (18.2%) were polymorphic and segregated in their inter specific backcross mapping population. Of these loci, hexa- and tri-nucleotide repeats were the most frequency, representing 40.1% and 30.0%, respectively. For aquaculture species, Yue *et al.*, (2004) developed 36 type I microsatellites from ESTs database of common carp (*Cyprinus carpio* L.) and found that they can successfully cross-species amplified in silver crucian carp (*Carassius auratusgibelio*). Thirty-four microsatellite loci were polymorphic in common carp with average allele number of 7.3 per locus (range from 2 to 15 alleles per locus). Cross-species amplification showed that 15 from 34 loci were polymorphic in silver crucian carp. Serapion *et al.* (2004), searched type I microsatellites from a total of 43,033 ESTs database of channel catfish (*Ictalurus punctatus*). The 4,855 ESTs containing microsatellites were identified. Dinucleotides CA/TG and GA/TC repeats were the most abundant microsatellites. The clustering result showed that 1,312 of these ESTs fell into 569 contigs, and the remaining 3,534 ESTs were singletons.

Up to date, the large-scale EST project of *P. monodon* is conducting at Shrimp Molecular Biology and Genomics Laboratory, Chulalongkorn University. At the end of

2004, the *P. monodon* EST database of this project contained about 10,100 EST sequences. In this thesis, type I microsatellite markers were developed by conducting a computer searching on the shrimp EST database in order to reveal microsatellite sequences and primer pairs were designed at each microsatellite locus. These type I microsatellite markers were used to further extend the previous *P. monodon* genetic linkage map consisting almost of AFLP markers. A define genetic linkage map covering the whole genome distance will allow identification of quantitative trait loci (QTL) that link to phenotypic traits of interest.

1.10 Objective of the thesis

The aim of this thesis is to obtain sufficient useful type I microsatellite markers in *P. monodon* for application in shrimp breeding and genetic mapping. We developed newly 50 type I microsatellite markers from shrimp ESTs database. These EST-derived microsatellite markers were tested to determine the level of polymorphism and microsatellite allelic patterns. The microsatellite markers were also applied for genetic linkage mapping.

CHAPTER II

MATERIALS AND METHODS

2.1 Equipments

- Autoclave LS- 2D (Rexall industries Co. Ltd., Taiwan)
- Automatic micropipettes P10, P20, PI00, P200 and P1000 (Gilson Medical Electronic S.A., France)
- -20°C Freezer (Sharp, Thailand)
- -80°C Freezer (Thermoforma, USA)
- DNA sequencer (MegaBACE 1000 DNA Analysis system, Amersham Bioscience)
- Light box (Shandon, Japan)
- Microcentrifuge tube 0.5, 1.5 ml (Bio-RAD Laboratories, USA)
- Multichannel micropitte 20-300 µl (Eppendorf, Germany)
- Multichannel micropitte 0.5 - 20 µl (Eppendorf, Germany)
- Orbital shaker (Stuart Scientific, England)
- PCR Therma! cyclers (Eppendorf, Germany)
- PCR Workstation Model#P-036 (Scientific Co., USA)
- Pipette tips 10, 200, 1000 µl (Bio-RAD Laboratories, USA)
- Power supplies: PowerPac3000 Power Supply (Bio-Rad Laboratories, USA)
- Refrigerated microcentrifuge Kubota 1300 (Kubota, Japan)
- Spectrophotometer DU 650 (Beckman, USA)

- Strip tube 0.2 ml (Axygen, USA)
- Vertical sequencing gel electrophoresis apparatus (Hoefer, USA)

2.2 Chemicals

- Absolute ethanol (Merck, Germany)
- Acrylamide (Merck, Germany)
- Acetic acid gracial (Merck, Germany)
- Ammonium persulfate (Promega, USA)
- Boric acid (Merck, Germany)
- Bromophenolblue (Merck, Germany)
- DyNAzyme II DNA polymerase (Finnzymes, Finland)
- Ethylene diamine tetraacetic acid, disodium salt dihydrate (Fluka, Switzerland)
- 100 mM dATP, dCTP, dGTP, dTTP, (Promega Corporation Medison, USA)
- Formaldehyde solution, 37 to 41 % HCOOH (BDH Chemical Ltd., England)
- Formamide (Gibco BRL Technologies Co., USA)
- *N,N*-methylenebisacrylamide (Amersham, England)
- Oligonucleotide primers (Proligo, Japan)
- Silver nitrate (BDH Chemical Ltd., England)
- Sodium carbonate (Ajax Chemicals, Australia)
- Sodium hydroxide (Merck, Germany)
- Sodium thiosulfate (UNIVAR, Australia)

- *N,N,N',N'*-tetramethylethylenediamine (Gibco BRL Technologies Inc., USA)
- Urea (Ajax Chemicals, Australia)
- Xylene cyanol (Sigma, USA)

2.3 Downloading Shrimp ESTs from the EST database

All black tiger shrimp ESTs data as of 1 January 2005 were downloaded from the shrimp EST database (http://pmonodon.biotec.or.th). All sequences were displayed in a FASTA format and further used for searching microsatellite loci. After downloading, the sequences were saved as text and stored in a local computer.

2.4 Searching and identification of EST containing microsatellites

After the ESTs data were downloaded into the local computer, all possible combination of dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide were searched by using a software, modified Sputnik II (http://wheat.pw.usda.gov/ITMI/EST-SSR/LaRata/). Single nucleotide repeats were not included because they are not very useful as polymorphic markers (Serapion *et al.*, 2004). Sputnik was run on DOS operation. The command was used as sputnik.exe [input file] > [output file]. Strings of oligo sequences were used to search microsatellites: 6 repeats were used for dinucleotide; 4 repeats were used for trinucleotide; 3 repeats were used for tetranucleotide and pentanucleotide according to Stalling *et al.* (1991).

Searching results from modify Sputnik II may contain redundancy clones. To obtain the different EST clones, EST containing microsatellites were masked by using Repeatmasker (Smit A., unpublish) and then clustered and assembled by using TIGR Gene Indices Clustering Tools (TGICL) (Pertea *et al.*, 2003) with the CAP3 (Huang and Medah, 1999).

2.5 Determination of gene identities of EST containing microsatellites

BLASTX program was used to compare a gene identity of ESTs containing microsatellite against the non-redundant database. When accumulated probability of sequence similarity was less than 10^{-4}, the tentative identities were established.

2.6 Shrimp samples

The black tiger shrimp (*P. monodon*) broodstock was wild-caught alive from Trang (N=48) in February 1998. Pleopods were excised from freshly killed shrimp individuals and immediately placed on dry ice. Specimens were stored at -80°C until required. Genomic DNA of these specimens was prepared by Supungul (1998) and used to determine polymorphism of microsatellite loci developed in this study.

For genetic linkage mapping analysis, the reference pedigree family was produced at the Australian Institute of Marine Science (AIMS), Townsville Facility, between 1995-1997. The first generation was wild broodstock (G0) captured from

inshore waters off the North Queensland coast in 1995. The offspring were raised over 17 months to produce the first generation of domesticated broodstock (G1). The G1 were spawned in March 1997 and their offspring, G2, were raised in nursery tanks to 3 months. Due to a disease outbreak in the facilities, all animals were harvested at that point. Offspring from several families were snap-frozen in liquid nitrogen for future DNA preparation. Pleopods were also preserved from the parental generations (Wilson *et al.*, 2002).

2.7 DNA extraction

2.7.1 The black tiger shrimp broodstock (Supungul, 1998)

Genomic DNA was extracted from a pleopod of each shrimp using a phenol-chloroform method modified from that of Davis *et al.* (1986). As soon as possible after removing from a -80°C freezer, a pleopod was transferred into a 1.5 ml microcentrifuge tube containing 400 µl of pre-chilled extraction buffer (100 mM Tris-HCl pH 9.0, 100 mM NaCl, 200 mM sucrose, 50 mM EDTA, pH 8.0) and briefly homogenized with a pre-chilled glass homogenizer. A 40% SDS solution was added to a final concentration 1.0% (W/V). The resulting mixture was then incubated at 65°C for 1 hour following by an addition of 10 µl of proteinase-K solution (30 mg/ml) and 5 µl of a RNase solution (10 mg/ml). The mixture was further incubated at the same temperature for 3 hour. To remove proteins, 91 µl of 5 M potassium acetate, pH 6.5 was added, thoroughly mixed and incubated at 4°C for 10 min prior to centrifugation at 10,000 revolution per minute

(rpm) for 10 min at 4°C. The supernatant was decanted to a sterile microcentrifuge. An equal volume of buffer-equilibrated phenol-chloroform-isoamyl alcohol (25:24:1 v/v) was added and gently mixed. The mixture was then centrifuged at 12,000 rpm for 10 min. The upper aqueous phase was carefully transferred to a new microcentrifuge tube. One-tenth volume of 3 M sodium acetate, pH 5.5 was added. DNA was precipitated by an addition of 2 volume of ice-cold absolute ethanol and kept at -20°C overnight. The precipitated DNA pellet was recovered using a cut tip and briefly wash twice with 70% ethanol. The pellet was air-dried and resuspended in 300 µl of TE buffer (10 mM Tris-HCl, pH 7.4 and 1 mM EDTA). The DNA solution was incubated at 37°C for 1-2 hours for complete redissolved and kept at 4°C until further use.

2.7.2 The reference family (Wilson et al., 2002)

DNA was prepared from parents and 41 offspring at the Australian Institute of Marine Science (AIMS) as follows. Frozen pleopod or tail muscle samples were weighed and placed in a glass homogenizer with 1 ml homogenisation buffer (100 mM Tris-HCl pH 9.0, 100 mM EDTA, 1% SDS, 100 mM NaCl) per 100 mg tissue. After grinding on ice, the homogenate was placed at 70°C for 60 min prior to addition of 250 µl of 5 M potassium acetate, pH 6.5 per 1 ml grinding buffer used, mixed thoroughly, incubated on ice for 30 min to precipitate proteins and centrifuged at 25,000xg for 15 min at 4°C. The supernatant was decanted. A 0.8 volume of isopropanol was added. The mixture was left at room temperature for 15 min to precipitate DNA. Genomic DNA was either spooled out or precipitated by centrifugation at 25,000xg for 15 min at 20°C. The precipitate was

washed with 70% ethanol, air dried and resuspended overnight at room temperature in 1 ml of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 μg/ml RNase. The genomic DNA was then further purified on Qiagen Genomic-tips according to manufacturer's instructions (Qiagen) and resuspended in 10 mM Tris-HCl pH 7.5, 1 mM EDTA. The reference DNA samples were sent to the laboratory at the Department of Biochemistry, Faculty of Science, Chulalongkorn University in the dried form. The reference DNA samples were redissolved with distilled water and kept at 4°C until further used.

2.8 Spectrophotometric measuring of DNA concentration

DNA concentration is estimated by spectrophotometry at the OD_{260}. An absorbance or optical density (O.D.) of 1.0 corresponds to 50 μg/ml of double-stranded DNA. DNA concentration of each specimen is estimated in μg/ml by OD_{260} X dilution factor x 50. An estimation of the purity of a sample can be obtained by calculating the ratio of the O.D. at 260 and 280 nm. For a pure preparation of DNA $OD_{260/280}$ should be 1.8. Ratios significantly less than this value indicated contamination of sample with protein or phenol (Kirby, 1992).

2.9 Design and synthesis of PCR primer pairs

PCR primers were designed from non-repetitive flanking regions of the microsatellite loci using OLIGO 4.0 (National BioScience). The criteria for primer designing were; 1) primers placed as closed to the repeat array as possible, 2) primers

were 19-25 base in length with melting temperature of 55-70°C, Tm = 2 (A+T) + 4 (G+C), 3) base distribution of the primers were random 4) the difference of the melting temperature of any primer pair was not exceed 4°C 5) the PCR product was 100 – 300 base pair in length. The primers were synthesized from commercial company (Proligo, Japan).

2.10 Microsatellite analysis

In preliminary analysis of the variation of microsatellite, primer pairs were diluted to appropriate concentration and tested for their ability to support amplification from 5 individuals genomic DNA by the polymerase chain reaction (PCR). PCR products were separated by denaturing polyacrylamide gel electrophoresis and detected by silver staining. The microsatellite loci showing at least 4 alleles from preliminary analysis were chosen for further characterization of those polymorphic loci and construction of a genetic linkage map. PCR amplification of the microsatellite loci was performed as described for the preliminary analysis except that the forward primer was 5'-labeled with fluorescent dye. Size-separation of PCR products was carried out by using an automated DNA sequencer (MegaBACE 1000 DNA Analysis system, Amersham Biosciences). The data was analyzed by the use of Genetic Profiler supplied by Amersham, Bioscience.

2.10.1 PCR amplification of microsatellites

PCR amplification was conducted in 10 μl reaction volume; the master mix contained 1X PCR buffer (Finnzymes) with 1.5 mM $MgCl_2$, 4 pmole of each primer, 2 nmole of dNTP, 50 ng of genomic DNA as template and 0.4 U of DyNAzyme DNA polymerase (Finnzymes). The PCR program was 5 min at 94°C followed by 35 cycles of 94°C for 30 s, annealing temperature (depending on each pair of primer) for 30 s and 72°C for 30 s.

2.10.2 Silver staining detection of microsatellite amplification

After the amplification was completed, each reaction was added with 4 μl of the stopping/loading dye (10 mM NaOH, 99% formamide, 0.1% bromophenol blue and 0.1% xylene cyanol). The PCR products were denatured at 95°C for 5 min, loaded onto 6% denaturing polyacrylamide gel (57 g acrylamide, 3 g *N,N'*methylenebisacrylamide, 7.66 M urea per 1 litre in IX TBE) and electrophoresed with IX TBE buffer at 50 W for 2-4 h, depending on PCR product size. A silver staining protocol used to stain PCR products in polyacrylamide gel was that described by Soot-anan (1999). After electrophoresis, gel was submerged for 30 min in 10% acetic acid. The gel was then briefly rinsed 3 times with ultrapure water for 2 min each and incubated in a staining solution (0.1% silver nitrate containing 1.5 ml/liter of 37% formaldehyde) for 30 min. The excess silver ions were eliminated by brief rinse with the ultrapure water. Gel was developed in a cold (8-10°C) developing solution (3% sodium carbonate containing 1.5 ml/liter of 37% formaldehyde and 200 μl/liter of 10 mg/ml sodium thiosulfate) and agitated until the first

bands were visible. The developing solution was poured off and the fresh developing solution was added. The gel was continued development until optimal image intensity was obtained. The developing reaction was stopped by using 10% acetic acid with agitating for 5 min. The gel was rinsed 2 times with the ultrapure water for 2 min each, the gel was dried by air dry.

2.10.3 PCR product separation and detection in an automated DNA sequencer

The PCR products were 10-fold diluted with ultrapure water. Each loading sample was comprised of 2 μl of the diluted PCR product, 0.125 μl of ET400-R size standard marker, and 2.875 μl of loading solution (70% formamide, 1mM EDTA) (Amersham, Biosciences). The loading sample were denatured at 94°C for 4 min and immediately placed on ice. Denaturing samples were loaded into an automated DNA sequencer and electrophoresed with 1X LPA buffer (Amersham, Bioscience) at 10 kV for 75 min. The allele scoring of PCR products was performed by the use of the computer program, Genetic Profiler (Amersham, Biosciences). The estimation of allele size was conducted by comparing with a DNA standard. The ET400-R size standard consisting of 20 double-stranded DNA fragments in which one strand is end labeled with an energy transfer (ET) dye. The following double-strand fragments, with sizes given in base pairs, comprise the standard: 60, 90, 100, 120, 150, 160, 170, 190, 200, 220, 250, 270, 290, 300, 310, 330, 350, 360, 380, 400.

2.11 Polymorphism analysis of microsatellite loci

Microsatellite loci were characterized by genotyping about 35-48 unrelated wild-caught *P. monodon* broodstocks from the Trang located in the Andaman sea. A genotype of each *P. monodon* individual was scored from an electrophoretically observed patterns of each locus. Allele sizes were scored in terms of numbers of base pair which were determined by comparison with the ET400-R size standard. When the stutter bands were present, the highest peak chromatogram was considered to be the actual allele (Stoner, Quattre and Weissman, 1997). The statistic used to calculate polymorphism is as follows:

2.11.1 Allele number and frequency

Allele number was determined by direct count of a particular allele. For diploid taxa, the frequency of a particular allele can be calculated as

$$P = \frac{2N_{AA} + N_{Aa}}{2N}$$

Where P is the frequency of the A allele, N is the total number of individuals and N_{AA} and N_{Aa} are the number of homo- and heterozygotes for such a locus (Nei, 1978).

2.11.2 Observed heterozygosity

The observed heterozygosity (h_{obs}) was determined as number of observed heterozygote individuals divided by the number of all individuals (Weimnay *et al.*, 2000).

2.11.3 Expected heterozygosity

When the population is in Hardy-Weinberg equilibrium, the expected heterozygosity (h_{exp}) was estimated using the equation;

$$h_{exp} = 1 - \sum p_i^2$$

Here P_i is the frequency of i th allele (Nei, 1978).

2.11.4 Polymorphic information content

Polymorphism information content (PIC) of each microsatellite locus was calculated as described by Botstein *et al.* 1980.

$$PIC = 1 - \left(\sum_{i=1}^{n} p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2 p_i^2 p_j^2$$

Where P_i is the frequency of the i-th allele, P_j is the frequency of the j-th alleles and n is the number of different alleles in the sample.

2.12 Application in shrimp genome mapping

Parents and 41 offspring of the reference family were genotyped using 50 microsatellite loci. The amplification, separation and visualization was observed by using an automated DNA sequencer Megabace (Amersham, Bioscience) as described in section 2.10. The genotyping data derived from the amplification of 50 microsatellite loci was sent to molecular Animal Genetic Centre, University of Queensland, Australia for linkage mapping analyses. The two-way pseudo-testcross strategy with the software program MAPMAKER/exp (F2 backcross model) was applied in the analyses. A LOD

score of 3.5 and maximum θ=0.30 were set as linkage threshold for grouping markers. Groups were then analysed using multipoint mapping functions to define the most likely map orders. Map distance in centiMorgans was calculated using Haldane's mapping function (Wilson et al., 2001). The 50 microsatellite markers were placed on the preliminary genetic map of *P. monodon* along with AFLP markers and microsatellite markers.

CHAPTER III

RESULTS

3.1 Searching for ESTs containing microsatellites

A total of 10,100 EST sequences in *P. monodon* EST database were searched for microsatellites by using modified Sputnik II software. A complete search revealed that 1,381 EST (13.7%) sequences contained microsatellite sequences. These EST containing microsatellites were clustered for determination of redundancy clones by using TGICL clustering tools. Of the 1,381 ESTs containing microsatellites, 513 clones fell into 129 contigs. The remaining 868 ESTs were singletons (see Table 3.1). In summary, 997 different EST sequences containing microsatellites were identified.

Table 3.1 Summary of analysis of *P. monodon* ESTs containing microsatellites

Total number of EST analyzed	10,100
Number of ESTs containing microsatellite	1,381
Number of ESTs included in contigs	513
Number of contigs	129
Number of singletons	868
Total number of unique genes containing microsatellites	997

3.2 Characteristics of microsatellite sequences of *P. monodon*

3.2.1 microsatellite classes

Microsatellites were classified into 3 categories, perfect, imperfect and compound repeats (Weber, 1990). Perfect repeats are uninterrupted stretches of repeat units, while imperfect repeats have one to three intervening bases with repeat sequence on either side. Compound repeats consist of several different repeat types and are separated by less than three bases. By classification with these definitions, a total of 2,165 loci were identified from 1,381 EST sequences that distributed in shrimp EST database. The classification result showed that perfect microsatellites were found at the highest frequency with 1,670 (78.1%) loci following by imperfect (362 loci, 16.7%), and compound microsatellites (362 loci, 16.7%) (Figure 3.1).

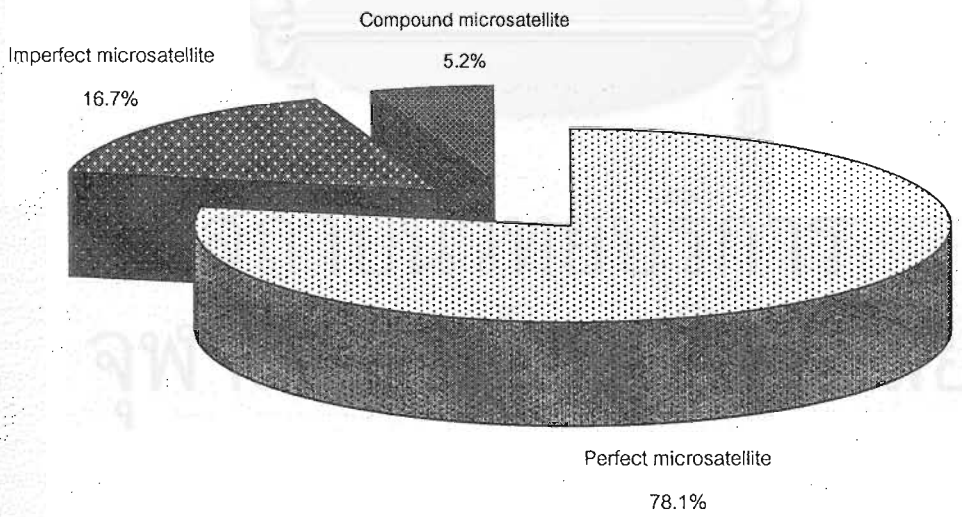


Figure 3.1 Distribution of microsatellite classes in *P. monodon*: perfect, imperfect, and compound microsatellites

3.2.2 Distribution of microsatellite types

From a total of 2,165 loci, trinucleotide repeats were the most abundant within shrimp ESTs, accounting for 934 repeats (40.6%). The second most abundant microsatellite type was dinucleotide repeats which represented 782 repeats (34.0%). Tetranucleotide and pentanucleotide repeats were found at lower frequencies with 433 repeats (18.8%) and 152 repeats (6.6%), respectively (Table 3.2). Several microsatellite loci contained more than one type of repeats, which called compound microsatellites. Compound microsatellites may be compose of dinucleotide close to trinucleotide, reflecting that 1 locus has 2 repeat types. In this study, 113 loci contained more than one type of repeats, thus, the sum of microsatellite types found in this study were 2,301 repeat types, not being 2,165 loci.

Table 3.2 Distribution of microsatellite repeats type in *P. monodon* ESTs

Repeat type	No. of microsatellite found in ESTs	Percentage of total microsatellite found in ESTs
Dinucleotide repeat	782	34.0
Trinucleotide repeat	934	40.6
Tetranucleotide repeat	433	18.8
Pentanucleotide repeat	152	6.6
Total of microsatellite repeats	2,301	100

Note: Of 2,165 microsatellite loci, 113 loci were compound microsatellites, which composed more than one type of microsatellites at each locus.

Of the dinucleotide repeats, AT/TA was the most abundant type, accounting for 40.4% of all dinucleotide repeats found in the shrimp ESTs. AG/CT was the second most abundant dinucleotide repeat type, accounting for 31.2% of all dinucleotide repeats. The AC/GT repeat was found at lower frequency, at 28.0%, while the CG/GC repeat was rare (0.4 %) (Figure 3.2).

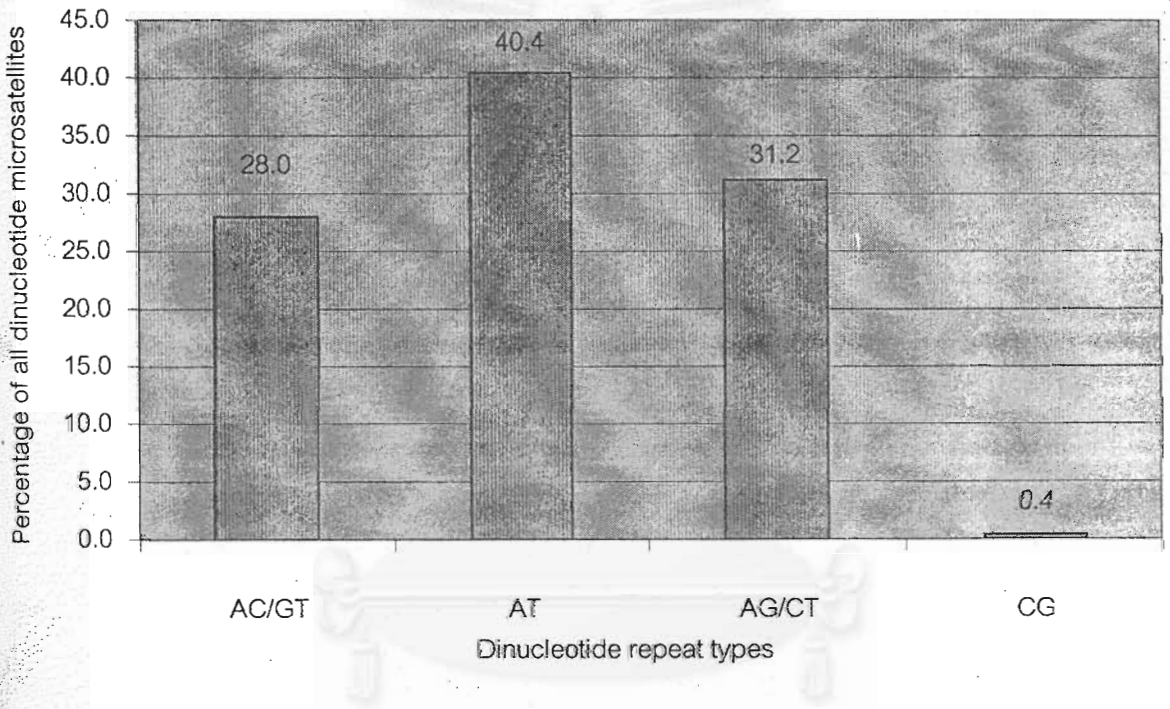


Figure 3.2 Distribution of dinucleotide microsatellites in *P. monodon* ESTs

From 20 possible types of trinucleotide repeat surveyed, the most abundant was AAT, with 39.4% of all trinucleotide microsatellites (Figure 3.3). ATC and AAG repeat were found at 12.6% and 11.6%, respectively. The other 6 types of trinucleotide repeat, AAC, ACC, ACG, ACT, AGC, and AGG repeats were found at less than 10%. AGT and

CCG repeats were rare, which was less than 1%. Another 9 possible types of trinucleotide repeats were not found.

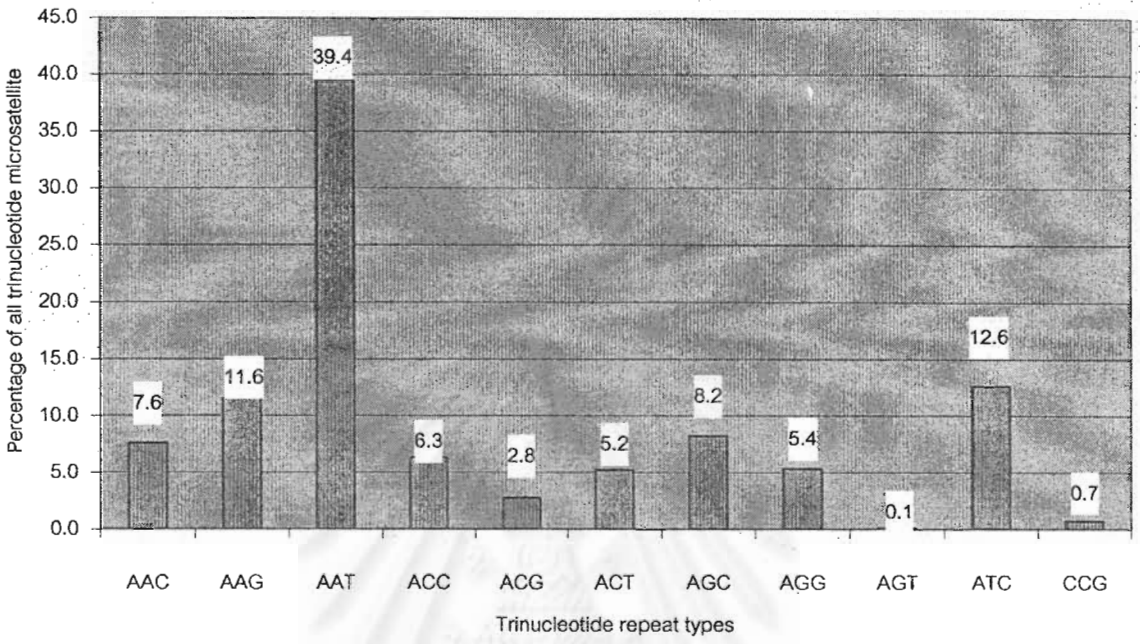


Figure 3.3 Distribution of trinucleotide microsatellites in *P. monodon* ESTs

A total of 28 types of tetranucleotide repeat were found in this study (from 79 possible types). The two most abundant of tetranucleotide repeats were AAAG and AAAT, accounting for 20.6% and 20.1%, respectiveiy. The intermediate abundant tetranucleotide repeats were ACAT (8.8%), AGAT (7.9%), AAAC (7.2%), and ACAG (6.5%). The low abundant tetranucleotide repeats were AATG (3.5%), ACTC (3.5%), AAGG (2.5%), AAGT (2.5%), AATC (2.5%), and AGGG (2.3%). The remaining 16 types of tetranucleotide repeats, when combined, represented only 12.2% of all tetranucleotide repeats (Figure 3.4).

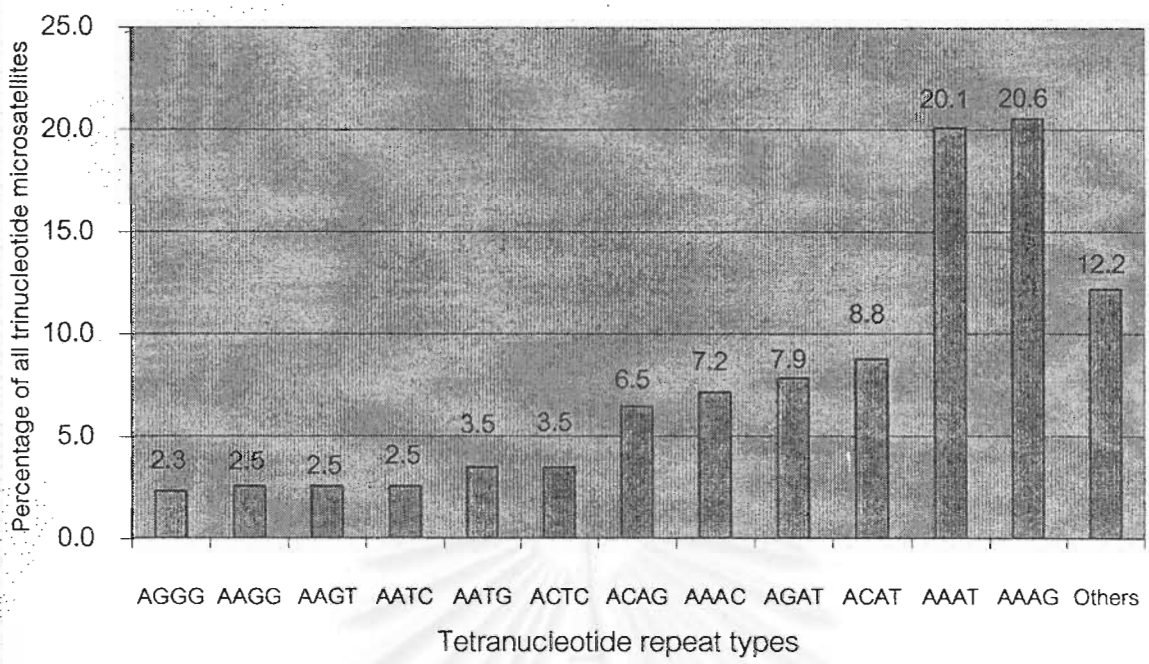


Figure 3.4 Distribution of tetranucleotide microsatellites in *P. monodon* ESTs

For the pentanucleotide repeats, only 34 types were found in this study. The AAAAG (25.7%) repeats were the most common, following by AAAAT (16.4%), AAAAC (9.9%), and AATAT (5.9%). The pentanucleotide repeats found at less than 5 % were AACCT (3.9%), AGAGG (3.9%), AAGAG (3.3%), ACTAT (3.3%), AAAGG (2.0%), AACAC (2.0%), AACAT (2.0%), AGGAT (2.0%). The remaining 22 types of pentanucleotide repeats, when combined, accounted for 19.7% of all pentanucleotide repeats (Figure 3.5).

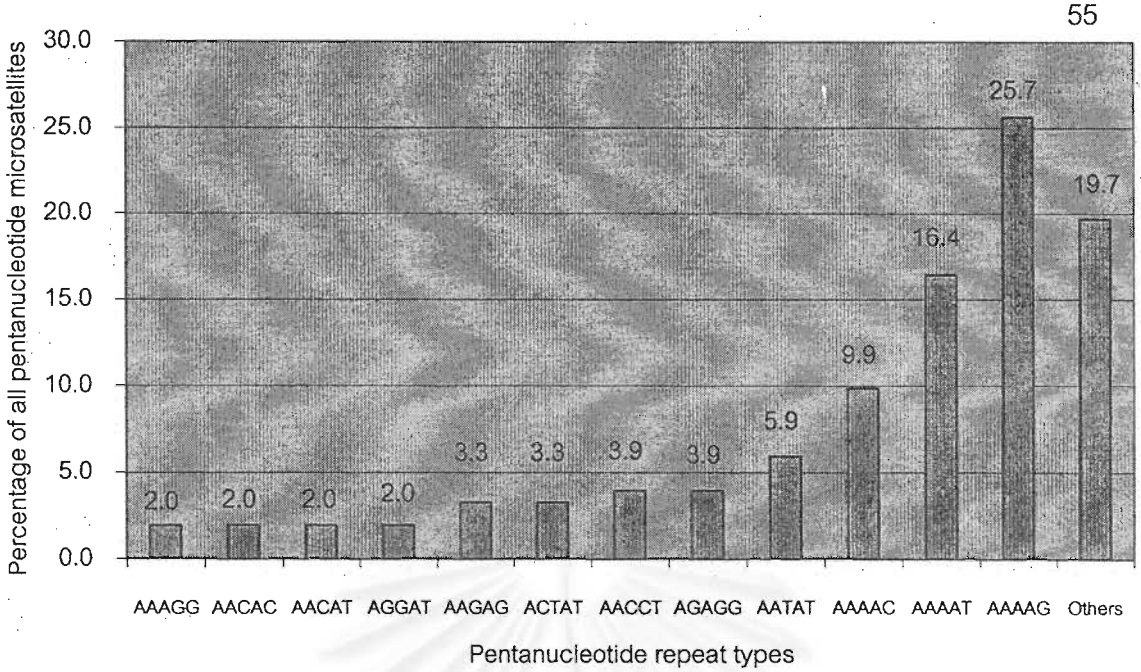


Figure 3.5 Distribution of pentanucleotide microsatellites in *P. monodon* ESTs

3.2.3 BLASTX results of EST containing microsatellites

The high proportion of ESTs containing microsatellite within the black tiger shrimp EST database offered a unique opportunity for the developments of type I microsatellite markers. As the microsatellite-associated genes are identified and polymorphism is validated, the microsatellite are converted into type I polymorphic markers. The EST containing microsatellites were blasted against GenBank to compare gene identities by using BLASTX program. After 997 EST sequences were blasted, the blasted results were classified into three groups, known gene product, hypothetical protein, and unknown gene product. Known gene product was homoloque to protein with known function. Hypothetical protein was homologous to genes of unknown function. Unknown gene product was a novel sequence which yielded no match or had poor matches (E-value >

10^{-4}) against the NCBI non-redundant database. Base on these criteria, the results of the 997 unique black tiger shrimp EST microsatellite-containing genes revealed that 116 EST sequences (11.6%) were known gene products; 372 ESTs (37.3%) were hypothetical proteins; 509 ESTs (51.1%) were unknown gene (Figure 3.6). The summary of known gene and hypothetical protein identities with EST containing microsatellite was shown in appendix A.

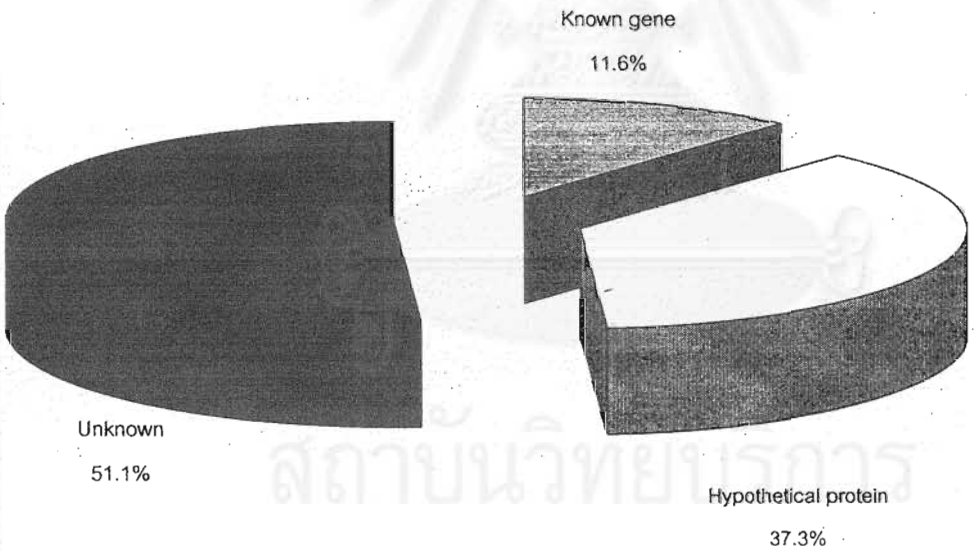


Figure 3.6 BLASTX results of 997 different EST sequences. Significant match was infered with the E-value cut off of 10^{-4}

The cDNA sequences could be divided into three parts, namely, the 5'-untranslated region (5'-UTR), the coding sequence (CDS), and the 3'-untranslated region (3'-UTR). The 5'-UTR is a region upstream of the translation start site, which does not code for protein product. The CDS is a region from the translation start site to the stop site, which code for a protein product. The 3'-UTR is a region downstream of the stop site of translation, which does not code for protein product. To determine whether the microsatellites were within the genes or in the untranslated regions, the EST clones were aligned against the protein database using blastx program. In this study, 721 microsatellite loci from cDNAs of known genes and hypothetical protein genes were identified. The blastx results indicated that the 573 microsatellite loci were located within the genes. However, the other 148 loci could not be located. Of the 573 microsatellite loci, 8 loci (1.4%) were located in the 5'-UTR; 536 loci (93.5%) were located in the coding sequence (CDS) and 29 loci (5.1%) were located in the 3'-UTR (Figure 3.7). An example of sequence alignment of microsatellite repeats that located within the gene coding sequence are shown in Figure 3.8. In this example, if the microsatellite repeats in query sequence are located in 2 regions, bases 122 to 143 and bases 450 to 470. The microsatellite repeats locating at bases 122 to 143 correspond to the coding sequence (CDS) of subject protein. We can define that this microsatellite array is located in the coding sequence. For the second array, bases 450 to 470, although we never known that this array indeed located within the gene, but according to the sequence alignment, the second array was predicted to be in the CDS and was defined by a parenthesis with

"Pd". The predicted locations of microsatellites were excluded from the summary of the existing location in genes (Figure 3.7).

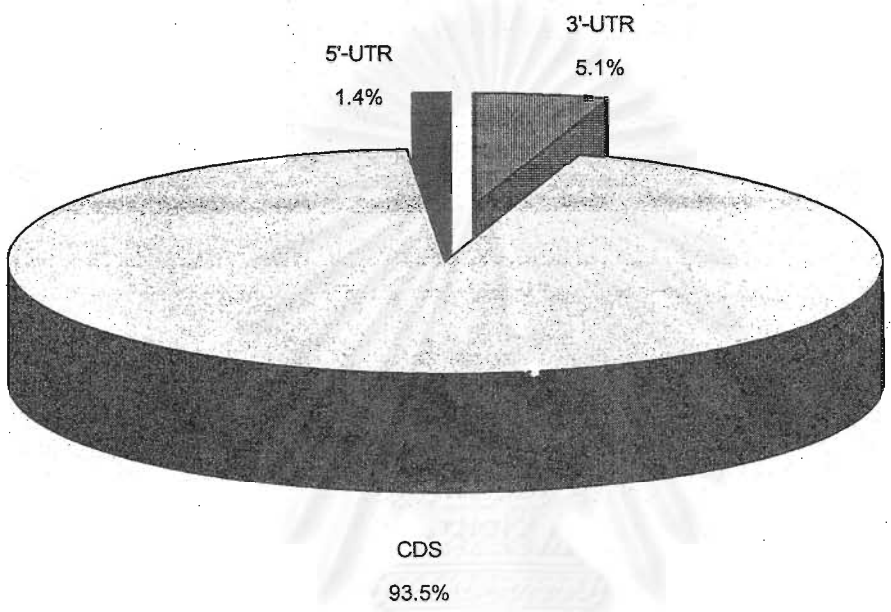


Figure 3.7 Microsatellites existed in region within gene. The positions of microsatellite repeats were located by sequence alignment with homologous gene by BlastX program.

>ref|XP_393238.1| similar to ENSANGP00000014057 [Apis mellifera]
Length = 205

Score = 116 bits (290), Expect = 1e-37
Identities = 54/107 (51%), Positives = 80/107 (74%), Gaps = 1/107 (0%), Frame = +2

```
Query: 83   MAEEENVNIEPYEEEEEEEPTNYKPPPEKSLEEMINQDQDDESLRKYKETLLGGSAGGAV 262
            M+E     +++  EEE E E +NYKPPPEK++E+++   D++DESLRKYKETLLG +  G +
Sbjct: 1    MSELNTDHVDSVEEELEVE-SNYKPPPEKTIEQILETDKEDESLRKYKETLLGEAKSGGI 59

Query: 263  IVDPSNPKRVLVRKLVLVAEDQPEMVLDLTTALDTLKDKPFTIKEGI 403
            +VDP++P++V+V+KL L    D+P+M LDLT  L  LK + F IKEG+
Sbjct: 60   VVDPNDPRKVIVKKLALCVADRPDMELDLTGDLSQLKKQTFVIKEGV 106
```

Figure 3.8 Blast result of query EST to subject protein in NCBI database. A box represented a microsatellite array that homologue to subject protein.

3.3 Polymorphism of developed microsatellites

3.3.1 Primer design

Out of 2,165 loci of ESTs containing microsatellites, 154 (7.11%) primer pairs could be designed. The remaining sequences could not designed primer pairs due to several reasons, such as, too short DNA sequence flanking the microsatellites and inappropriate flanking sequences for primer designs.

Of the 154 primer pairs, 126 (81.82%) pairs were successfully amplified in 5 unrelated individual shrimps with different condition of annealing temperature. Annealing temperature of polymorphic loci were shown in Table 3.3.

3.3.2 Detection of polymorphism

For preliminary analysis of polymorphism, flanking primers were used for amplification of 5 unrelated individual shrimps by Polymerase Chain Reaction (PCR). The PCR products were separated by denaturing polyacrylamide gel electrophoresis and visualized by silver staining. PCR products displayed either one or two major fragments, consistent with single locus detection of either homozygotes or heterozygotes, respectively (Figure 3.9). Genotype patterns of BTPm11 and BTPm3 show 7 alleles and 6 alleles, respectively. Both loci were chosen to further characterize and genotype with reference family for genetic linkage mapping. From preliminary analysis of polymorphism, 56 (36.36%) loci showed at least 4 alleles. The markers that showed at least 4 alleles were chosen to characterize and genotype reference family for linkage mapping according to section 2.11 and 2.12, respectively.

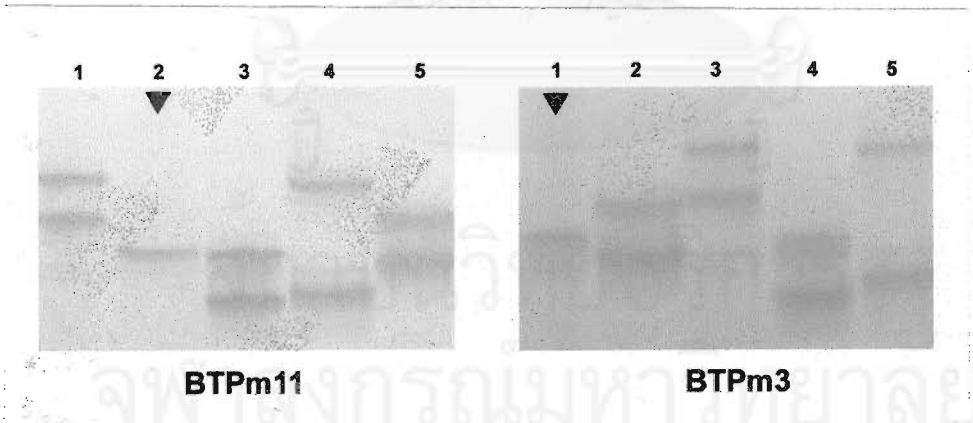


Figure 3.9 Genotype patterns of 5 unrelated individual shrimps (lane 1-5) at 2 microsatellite loci, BTPm11 and BTPm3. The PCR products were run on 6 % denaturing polyacrylamide gel and visualized by silver staining. Arrowheads indicate homozygote genotypes. Unmark lanes were heterozygotes.

3.3.3 Characteristic of microsatellite

DNA samples isolated from *P. monodon* originating from Trang province by Supangul (1998) were used to determine polymorphism of polymorphic loci from preliminary screening. PCR amplification of the microsatellites was performed as described for the preliminary analysis except that the forward primer was 5'-labeled with fluorescent dye. Each locus was amplified with 35-48 wild-caught *P. monodon* genomic DNA from Trang province. Allele scoring of the PCR products were carried out by using an automated DNA sequencer (MegaBACE 1000 DNA Analysis system, Amersham Biosciences). The data was analyzed by using computer program, Megabace™ Genetic Profiler version 2.0 (Amersham, Biosciences). The size of alleles was estimated by comparing with the ET-400R size standard.

Fifty-six polymorphic loci selected from preliminary analysis of polymorphism were tested for allelic segregation using the reference family from the Australian Institute of Marine Science (AIMS) under the Shrimp International Genetic linkage Map Project (http://www.aims.gov.au/shrimpmap). Three of them showed null alleles which may resulted from mutation at the priming site(s) (Figure 3.10). The other 3 loci showed parent's alleles incompatible with offspring's alleles (Figure 3.11). Those 6 loci were not further characterized. The remaining of 50 polymorphic loci (32.47% of all primer designed) were characterized by genotyping with 35-48 samples originating from Trang. Examples of the chromatogram of BTPm38 and BTPm40 loci were shown in Figure 3.12.

Genotype patterns and allele sizes of the remaining microsatellite loci were shown in appendix B and C, respectively. The polymorphic information related to 50 microsatellite loci developed in this study is summarized in Table 3.4. These include number of samples, number of alleles, size range of allele, heterozygosity and polymorphic information content (PIC).

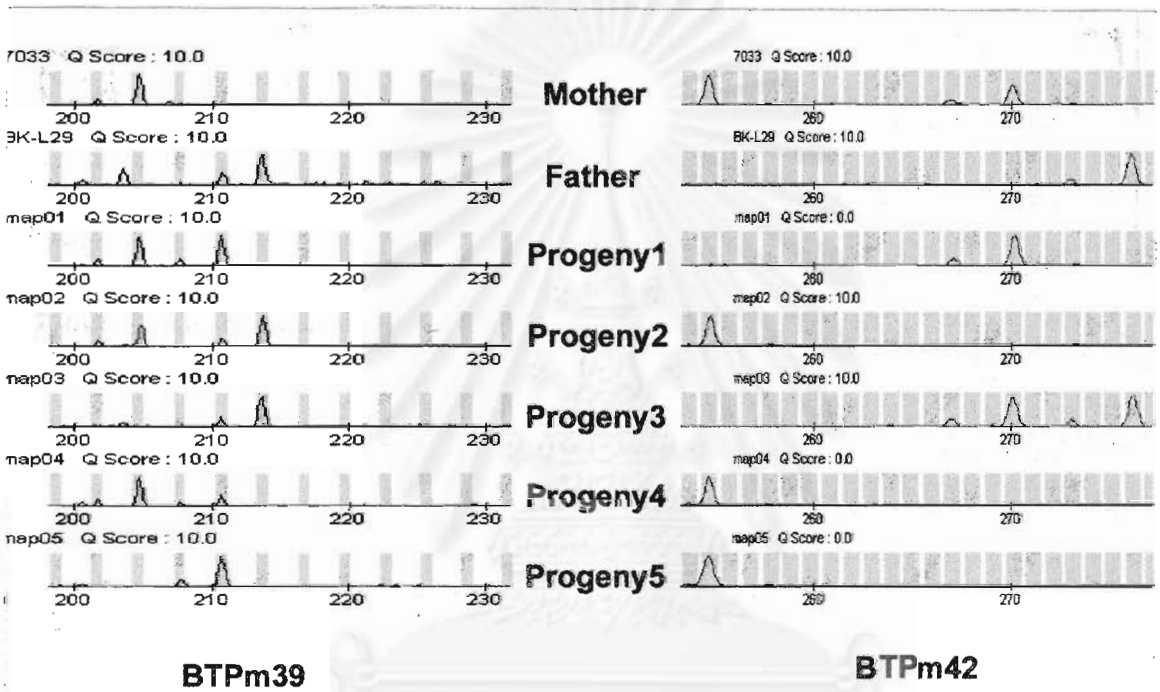


Figure 3.10 Examples of genotype patterns of null allele at 2 markers, BTPm39 and BTPm42. Genotyping was performed using MegaBACE 100 DNA Analysis System. Left panel showing allelic segregation at the BTPm39 in which a null allele occurred in the mother. Right panel showing allelic segregation at the BTPm42 in which a null allele occurred in the father.

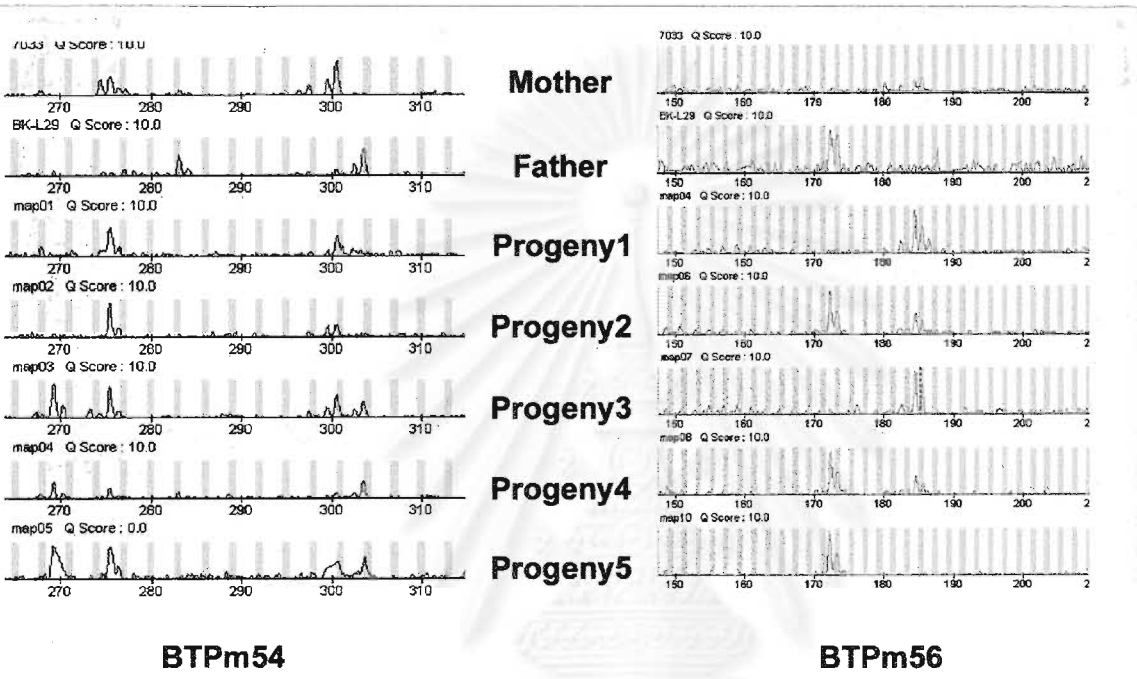


Figure 3.11 Examples of genotype patterns of incompatible allele at 2 markers, BTPm54 and BTPm56. Left panel is a BTPm54. Right panel is a BTPm56. The genotypes of progenies were not inherited in a Mendelian fashion.

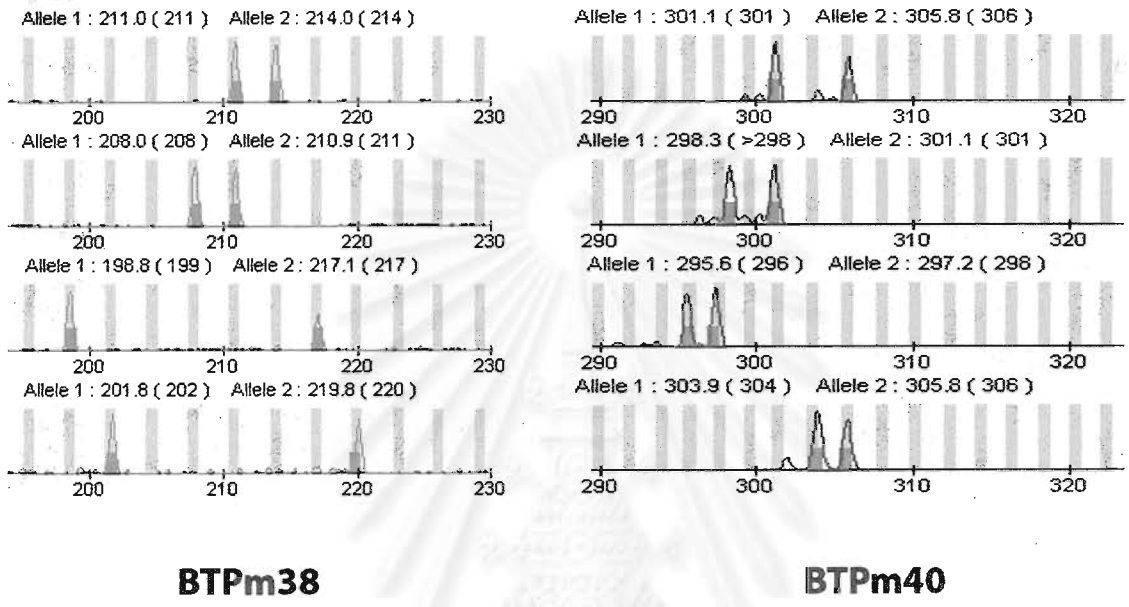


Figure 3.12 Example of chromatograms of 4 unrelated individual *P. monodon* characterized at 2 markers, BTPm38 and BTPm40. Allelic scoring was performed automatically using MegaBACE 100 DNA Analysis System. Polymorphism was tested against 35-48 unrelated individual shrimps.

Table 3.3 Repeat sequences, expected PCR products, and annealing temperature of 50 polymorphic microsatellite markers derived from *P. monodon* ESTs

Marker name	EST name	Repeats sequences	Expected PCR product (bp)	Annealing temperature (°C)
BTPm1	MBRU18	$(AT)_7$	278	54
BTPm2	LPN-177	$(AT)_6$	140	46
BTPm3	HPA-177	$(GT)_8$	253	50
BTPm4	ES03-071	$(CT)_{18}$	231	46
BTPm5	HPO-039	$(AG)_7$	285	50
BTPm6	LPN-149	$(AC)_{19}$ imp*	169	50
BTPm7	OV-292	$(AAT)_{16}$	293	48
BTPm8	ES02-217	$(AAT)_{46}$	242	50
BTPm9	HCH-344	$(AG)_{15}$ imp*	198	50
BTPm10	136	$(AAT)_{27}$ imp*	298	48
BTPm11	HPA-223	$(AT)_{10}$	253	50
BTPm12	LPN-221	$(ATT)_{27}$ imp*	295	50
BTPm13	ES01-099	$(AT)_{30}$	149	50
BTPm14	ES01-297	$(AT)_7$	150	48
BTPm15	ES03-088	$(ATCT)_{13}AC(ATCT)_{13}$	147	50
BTPm16	ES03-287	$(AAT)_{16}$ imp*	233	50
BTPm17	HCH-1036	$(AG)_4AT(AG)_5$	242	52
BTPm18	HPA-336	$(AT)_5AA(AT)_3$	155	48
BTPm19	HPO-198	$(AG)_{46}$	159	50
BTPm20	HPO-464	$(AC)_{11}(AT)_7$	221	50
BTPm21	HPO-592	$(AG)_{21}$ imp*	266	50
BTPm22	HPO-885	$(AG)_9$	224	48
BTPm23	OV-209	$(ATT)_{13}$ imp*	108	50
BTPm24	OV-985	$(AG)_{11}$	116	50
BTPm25	LPN-154	$(AAG)_8$	276	50

Table 3.3 (continued)

BTPm26	HCH-452	$(AT)_{12}CT(AT)_2$	143	48
BTPm27	HPA-684	$(AAT)_7$	145	48
BTPm28	LPY-182	$(AC)_{16}(AGAT)_{12}(AG)_{24}$	249	48
BTPm29	HCH-826	$(AG)_9$	162	50
BTPm30	HPAN-218	$(AC)_{42}$ imp*	273	50
BTPm31	HPAN-580	$(AAT)_{29}$	195	50
BTPm32	HPO-089	$(ATT)_6$	206	50
BTPm33	LPNN-485	$(AC)_7$-$(AC)_{14}$-$(AC)_9$-$(AC)_7$	193	48
BTPm34	LPNN-902	$(AC)_{27}$	124	50
BTPm35	LPNN-945	$(AT)_9$	134	50
BTPm37	LPV-502	$(GT)_{12}$	230	48
BTPm38	LPY-652	$(AC)_{17}$	217	50
BTPm40	ES03-133	$(AG)_{29}$	294	50
BTPm41	HCH-639	$(CT)_{11}$	311	50
BTPm43	HPA-632	$(ACT)_4(ATT)_{27}G(ATT)_6$	296	50
BTPm44	HPO-031	$(AT)_8$	156	50
BTPm45	HPO-772	$(ATT)_{12}$	212	50
BTPm46	OV-860	$(AC)_{16}(AGAT)_{12}(AG)_{22}$	153	50
BTPm47	ES01-301	$(GT)_{43}$ imp*	244	48
BTPm48	ES02-283	$(AT)_6$	130	50
BTPm49	ES02-402	$(AG)_{11}$	134	50
BTPm50	ES03-231	$(ATT)_{10}$ imp*	167	50
BTPm51	MBRU6	$(AC)_8$	272	46
BTPm52	OV-036	$(AG)_{19}$	158	50
BTPm53	OV-1034	$(AG)_8$	184	46

* imperfect microsatellites

Table 3.4 Polymorphism of 56 type I microsatellites markers of *P. monodon*

Marker name	EST name	No of samples	No. of alleles	Size range of alleles (bp)	h_{obs} *	h_{exp} **	PIC ***
BTPm01	MBRU18	48	26	244-324	0.875	0.924	0.908
BTPm02	LPN-177	46	11	127-141	0.87	0.788	0.753
BTPm03	HPA-177	45	25	234-270	0.867	0.915	0.899
BTPm04	ES03-071	35	7	226-238	0.486	0.766	0.723
BTPm05	HPO-039	47	22	277-330	0.872	0.931	0.916
BTPm06	LPN-149	42	7	155-176	0.738	0.797	0.757
BTPm07	OV-292	45	18	281-336	0.911	0.91	0.892
BTPm08	ES02-217	45	16	199-257	0.933	0.893	0.874
BTPm09	HCH-344	41	11	260-293	0.878	0.823	0.789
BTPm10	136	38	11	293-325	0.737	0.859	0.829
BTPm11	HPA-223	38	11	293-325	0.333	0.824	0.791
BTPm12	LPN-221	45	7	294-312	0.844	0.748	0.694
BTPm13	ES01-099	41	4	126-134	0.634	0.591	0.496
BTPm14	ES01-297	46	4	142-150	0.739	0.671	0.593
BTPm15	ES03-088	46	9	115-175	0.848	0.767	0.725
BTPm16	ES03-287	42	12	226-252	0.952	0.823	0.792
BTPm17	HCH-1036	41	18	215-252	0.902	0.914	0.895
BTPm18	HPA-336	48	11	126-165	0.792	0.706	0.666
BTPm19	HPO-198	40	6	140-159	0.525	0.742	0.694
BTPm20	HPO-464	39	10	167-196	0.667	0.849	0.818
BTPm21	HPO-592	48	27	230-321	0.854	0.926	0.911
BTPm22	HPO-885	45	14	216-251	0.644	0.677	0.638
BTPm23	OV-209	43	9	96-121	0.674	0.713	0.671
BTPm24	OV-985	45	12	110-168	0.511	0.717	0.664
BTPm25	LPN-154	35	9	267-285	0.857	0.798	0.760
BTPm26	HCH-452	42	3	137-141	0.143	0.257	0.237
BTPm27	HPA-684	46	5	143-153	0.587	0.548	0.437
BTPm28	LPY-182	46	25	211-271	0.87	0.948	0.934
BTPm29	HCH-826	36	10	152-166	0.583	0.687	0.651

Table 3.4 (continued)

BTPm30	HPAN-218	47	19	248-303	0.787	0.897	0.880
BTPm31	HPAN-580	46	14	171-210	0.935	0.882	0.861
BTPm32	HPO-089	47	15	238-270	0.957	0.893	0.873
BTPm33	LPNN-485	47	6	184-196	0.574	0.632	0.557
BTPm34	LPNN-902	42	4	119-125	0.5	0.58	0.519
BTPm35	LPNN-945	41	6	124-134	0.756	0.735	0.690
BTPm36[A]	LPV-267	-	-	-	-	-	-
BTPm37	LPV-502	37	9	217-268	0.135	0.807	0.768
BTPm38	LPY-652	47	12	199-235	0.872	0.884	0.863
BTPm39[A]	ES02-195	-	-	-	-	-	-
BTPm40	ES03-133	44	11	290-320	0.773	0.814	0.779
BTPm41	HCH-639	45	27	261-317	0.933	0.954	0.940
BTPm42[A]	HPA-049	-	-	-	-	-	-
BTPm43	HPA-632	46	19	279-339	0.891	0.894	0.876
BTPm44	HPO-031	46	16	142-176	0.696	0.785	0.762
BTPm45	HPO-772	45	30	179-267	0.889	0.949	0.935
BTPm46	OV-860	44	12	151-177	0.75	0.76	0.714
BTPm47	ES01-301	47	26	189-300	0.915	0.96	0.948
BTPm48	ES02-283	36	7	135-154	0.556	0.513	0.434
BTPm49	ES02-402	45	7	126-154	0.422	0.418	0.396
BTPm50	ES03-231	48	5	166-174	0.333	0.468	0.428
BTPm51	MBRU6	44	10	275-306	0.409	0.805	0.772
BTPm52	OV-036	36	7	153-178	0.361	0.514	0.480
BTPm53	OV-1034	36	7	135-154	0.306	0.28	0.267
BTPm54[B]	ES01-052	-	-	-	-	-	-
BTPm55[B]	OV-645	-	-	-	-	-	-
BTPm56[B]	LPN-024	-	-	-	-	-	-

*Observed heterozygosity **Expected heterozygosity ***Polymorphic Information Content

[A]marker showed null alleles

[B]marker showed parent's alleles incompatible with offspring's alleles

The number of alleles per locus was 3-30 alleles with an average of 12.6. The average observed heterozygosity of 50 polymorphic loci was 0.698 with a maximum of 0.957 for the BTPm32 and a minimum of 0.135 for the BTPm37.The average expected heterozygosity was 0.759 with a maximum of 0.960 for the BTPm47 and a minimum of 0.257 for the BTPm26. For the PIC, the average PIC of 50 polymorphic loci is 0.723 with a maximum of 0.948 for the BTPm47 and a minimum of 0.237 for the BTPm26. In this study, most loci with high allele number also showed high values of heterozygosity and PIC. It should be noted that for the BTPm37 marker, the observed heterozygosity was much lower than the expected heterozygosity and PIC.

After 50 polymorphic type I microsatellite markers were blasted against non-redundant NCBI database. The results showed that 36% and 54% of microsatellite repeats were located in coding sequence (CDS) and were unknown, respectively. Another 10% were predicted to be located within the genes, which indicated in parenthesis by Pd. Table 3.5 showed transcribed microsatellite location within gene, microsatellite repeat types, and polymorphism information.

Table 3.5 Transcribe 50 polymorphic type I microsatellites and their associated genes.

Marker name	Repeat type	h_{obs}*	h_{exp}**	PIC***	Gene identity	Repeat location
BTPm1	$(AT)_7$	0.875	0.924	0.908	unknown	unknown
BTPm2	$(AT)_6$	0.870	0.788	0.753	unknown	unknown
BTPm3	$(GT)_8$	0.867	0.915	0.899	hypothetical protein	CDS
BTPm4	$(CT)_{18}$	0.486	0.766	0.723	unknown	unknown
BTPm5	$(AG)_7$	0.872	0.931	0.916	unknown	unknown
BTPm6	$(AC)_{19}$ imp*	0.738	0.797	0.757	unknown	unknown
BTPm7	$(AAT)_{16}$	0.911	0.910	0.892	hypothetical protein	CDS
BTPm8	$(AAT)_{46}$	0.933	0.893	0.874	serine/ threonine protein kinase	CDS
BTPm9	$(AG)_{15}$ imp*	0.878	0.823	0.789	unknown	unknown
BTPm10	$(AAT)_{27}$ imp*	0.737	0.859	0.829	unknown	unknown
BTPm11	$(AT)_{10}$	0.333	0.824	0.791	actin-alpha	CDS
BTPm12	$(ATT)_{27}$ imp*	0.844	0.748	0.694	unknown	unknown
BTPm13	$(AT)_{30}$	0.634	0.591	0.496	ENSANGP00000024044	CDS
BTPm14	$(AT)_7$	0.739	0.671	0.593	hypothetical protein	CDS
BTPm15	$(ATCT)_{13}AC(ATCT)_{13}$	0.848	0.767	0.725	agCP4447	CDS (Pd)
BTPm16	$(AAT)_{16}$ imp*	0.952	0.823	0.792	unknown	unknown
BTPm17	$(AG)_4AT(AG)_5$	0.902	0.914	0.895	unknown	unknown
BTPm18	$(AT)_5AA(AT)_3$	0.792	0.706	0.666	unknown	unknown
BTPm19	$(AG)_{46}$	0.525	0.742	0.694	mariner transposase	CDS
BTPm20	$(AC)_{11}(AT)_7$	0.667	0.849	0.818	unknown	unknown
BTPm21	$(AG)_{21}$ imp*	0.854	0.926	0.911	hypothetical protein	CDS
BTPm22	$(AG)_9$	0.644	0.677	0.638	unknown	unknown
BTPm23	$(ATT)_{13}$ imp*	0.674	0.713	0.671	butyrate response factor 1	5'-UTR (Pd)
BTPm24	$(AG)_{11}$	0.511	0.717	0.664	unknown	unknown
BTPm25	$(AAG)_8$	0.857	0.798	0.760	unknown	unknown
BTPm26	$(AT)_{12}CT(AT)_2$	0.143	0.257	0.237	unknown	unknown

Table 3.5 (continued)

BTPm27	$(AAT)_7$	0.587	0.548	0.437	unknown	unknown
BTPm28	$(AC)_{16}(AGAT)_{12}(AG)_{24}$	0.870	0.948	0.934	unnamed protein product	CDS
BTPm29	$(AG)_9$	0.583	0.687	0.651	conserved hypothetical protein	CDS (Pd)
BTPm30	$(AC)_{42}$ imp*	0.787	0.897	0.880	unknown	unknown
BTPm31	(AAT)29	0.935	0.882	0.861	hypothetical protein	CDS
BTPm32	(ATT)6	0.957	0.893	0.873	unnamed protein product	CDS
BTPm33	$(AC)_7$-$(AC)_{14}$-$(AC)_9$-$(AC)_7$	0.574	0.632	0.557	hypothetical protein	CDS
BTPm34	$(AC)_{27}$	0.500	0.580	0.519	unknown	unknown
BTPm35	$(AT)_9$	0.756	0.735	0.690	unknown	unknown
BTPm37	$(GT)_{12}$	0.135	0.807	0.768	unknown	unknown
BTPm38	$(AC)_{17}$	0.872	0.884	0.863	hypothetical protein	CDS
BTPm40	$(AG)_{29}$	0.773	0.814	0.779	hypothetical protein	CDS (Pd)
BTPm41	$(CT)_{11}$	0.933	0.954	0.940	hypothetical protein	CDS
BTPm43	$(ACT)_4(ATT)_{27}G(ATT)_6$	0.891	0.894	0.876	hypothetical protein	CDS
BTPm44	$(AT)_8$	0.696	0.785	0.762	ENSANGP00000024044	CDS
BTPm45	$(ATT)_{12}$	0.889	0.949	0.935	hypothetical protein	CDS
BTPm46	$(AC)_{16}(AGAT)_{12}(AG)_{22}$	0.750	0.760	0.714	t-complex polypeptide 1	3'-UTR (Pd)
BTPm47	$(GT)_{43}$ imp*	0.915	0.960	0.948	unnamed protein product	CDS
BTPm48	$(AT)_6$	0.556	0.513	0.434	unknown	unknown
BTPm49	$(AG)_{11}$	0.422	0.418	0.396	unknown	unknown
BTPm50	$(ATT)_{10}$ imp*	0.333	0.468	0.428	unknown	unknown
BTPm51	$(AC)_8$	0.409	0.805	0.772	unknown	unknown
BTPm52	$(AG)_{19}$	0.361	0.514	0.480	unknown	unknown
BTPm53	$(AG)_8$	0.306	0.280	0.267	unknown	unknown

*Observed heterozygosity **Expected heterozygosity ***Polymorphic Information Content

+Coding sequence #not definable: expected location

3.4 Application of type I microsatellites in shrimp genome mapping

The aim of genetic mapping is to locate genetic markers at a particular position on the chromosome. Two loci (or genetic markers) are thus said to be link if the parental allele combinations are preserved more often than would be expected by random segregation. The degree of linkage, or genetic distance, between two loci is a function of the frequency of recombination. The measurement of map distance is expressed in Morgan (or cM for centiMorgan) and is defined as the expected number of crossover between two loci on chromosome (Boyd, 1998).

The international collaboration between difference parties on genetic mapping of *P. monodon* was initiated in 1997 (Wilson *et al.*, 2002). Construction of the genetic map of *P. monodon* was carried out using a reference family produced at the Australian Institute of Marine Science (AIMS). *P. monodon* family consisted of two parents and 41 offsprings were genotyped with 56 microsatellite loci. The result of genotype patterns showed that 44 loci could be used for the genetic analysis. The remaining 12 markers were discarded due to several reasons, such as null alleles (3 loci), insertion/deletion (3 loci), and the same genotype pattern of parents at each locus (6 loci). An example of the same genotype patterns of parents at a locus was showed in Figure 3.13. The amplification, separation, and allele scoring were carried out as described in Section 2.10. Data on allelic inheritance of 44 microsatellite loci for the *P. monodon* family were illustrated in Appendix D.

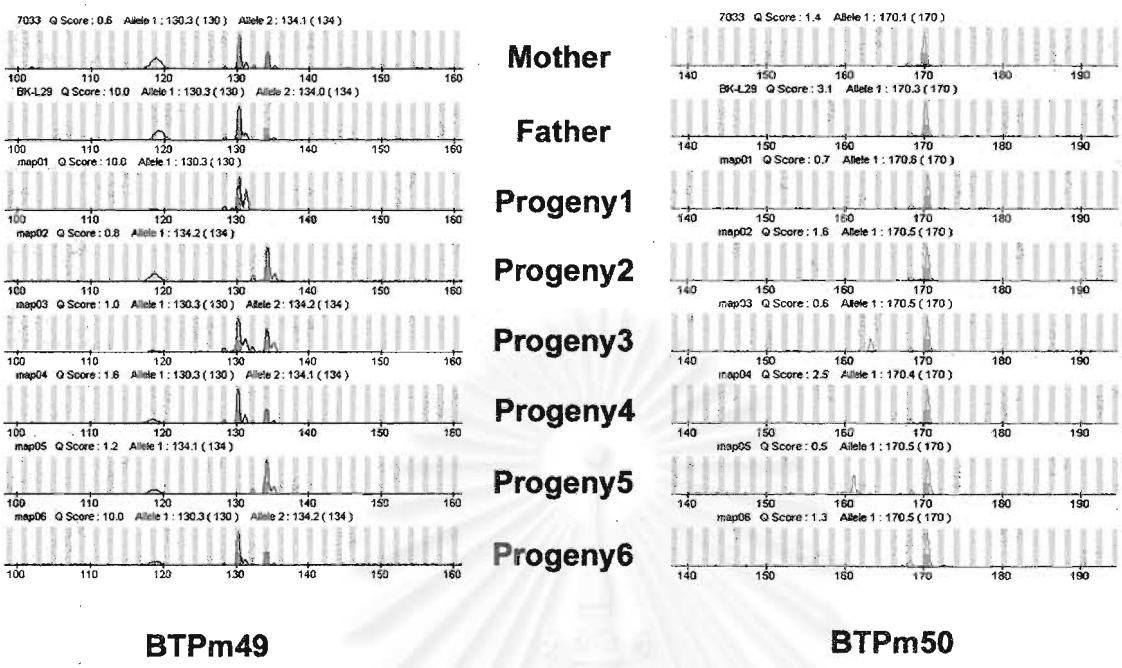


Figure 3.13 Examples genotype pattern of the references family showing identical genotypes of the Father and the Mother

Genotype data was sent to Molecular Animal Genetics Center, University of Queensland, Australia for linkage mapping analysis. All 44 loci were analyzed along with the previous markers, AFLP, SNP, etc., of both female and male maps. Thirty-six loci were placed on male and/or female linkage maps as shown in Figures 3.14 and 3.15, respectively. Eight loci could not placed in any of the existing linkage groups. The allele scoring of microsatellite was performed as described previously for AFLP (Wilson *et al.*, 2002). Overall, the male map comprised 157 markers and formed 47 linkage groups with a total genome length of 1,101.0 cM. Out of 44 microsatellite loci developed in this study, 26 markers were placed on the male map. For the female map, 111 markers were placed

on the map and formed 36 linkage groups with a total genome length of 891.4 cM. Twenty-five microsatellite loci in this study, accounting for 18 markers were placed on the female map. Table 3.6 summarized the current status of *P. monodon* genetic linkage maps and compared with those from the previous studies. Information of *P. monodon* map is available on *P. monodon* shrimp map web site (http://www.aims.gov.au/shrimpmap).

Table 3.6 Summary of *P. monodon* genetic linkage male and female maps

	Male map		Female map	
	Previous*	Current	Previous*	Current
Number of markers on the map	112	157	77	111
Number of linkage groups	36	47	29	36
Total length of linkage groups (cM)	878	1,101	630	891.4
Average space between 2 markers (cM)	7.8	7	8.2	8
Average length of linkage group (cM)	24.4	23.4	21.7	24.8
Average no. of markers per linkage group	3.1	3.3	2.7	3.1

*Status release on March 2004

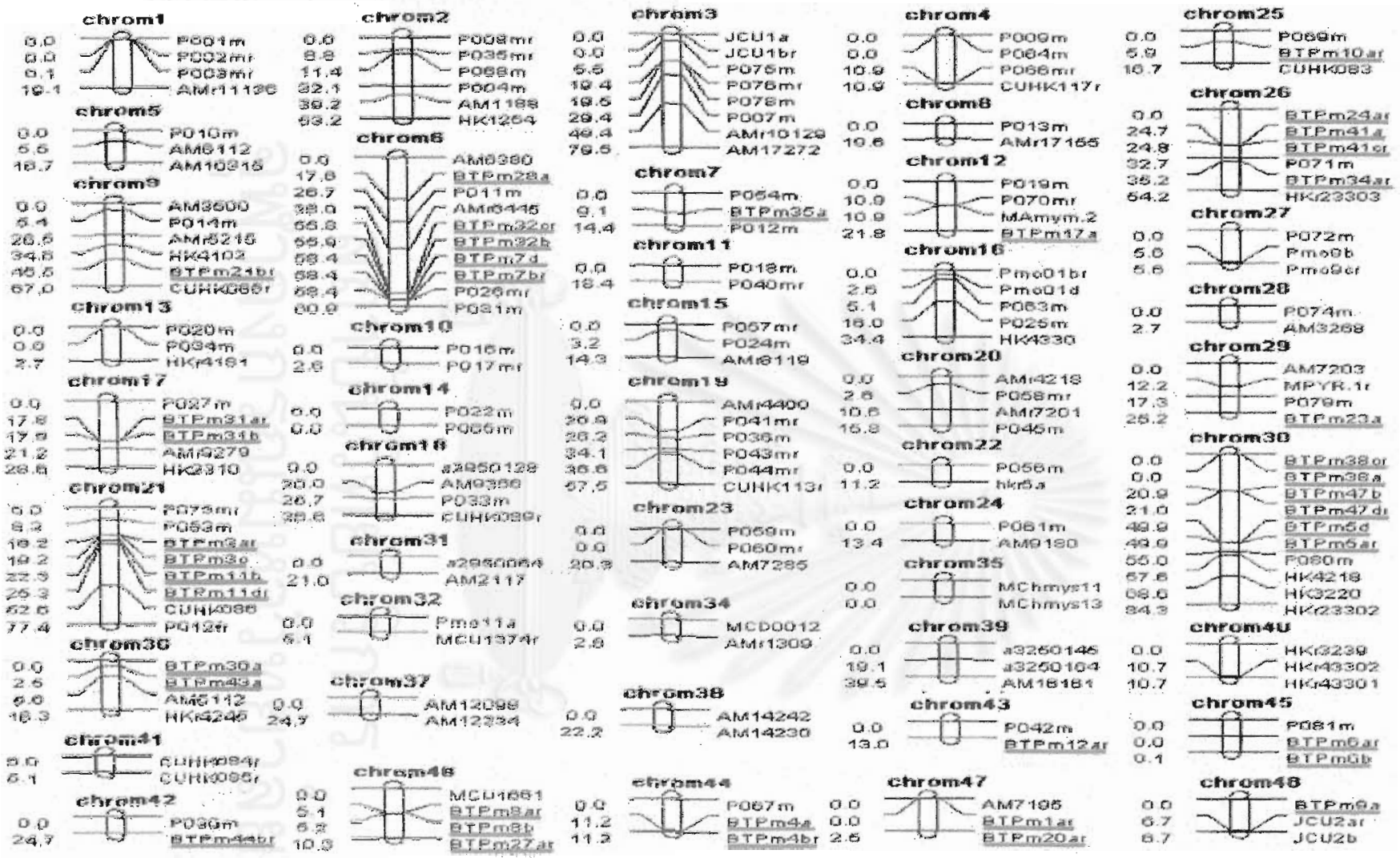


Figure 3.14 A genetic linkage map of male *P. monodon*. Red lines indicated type I microsatellite markers developed in this study.

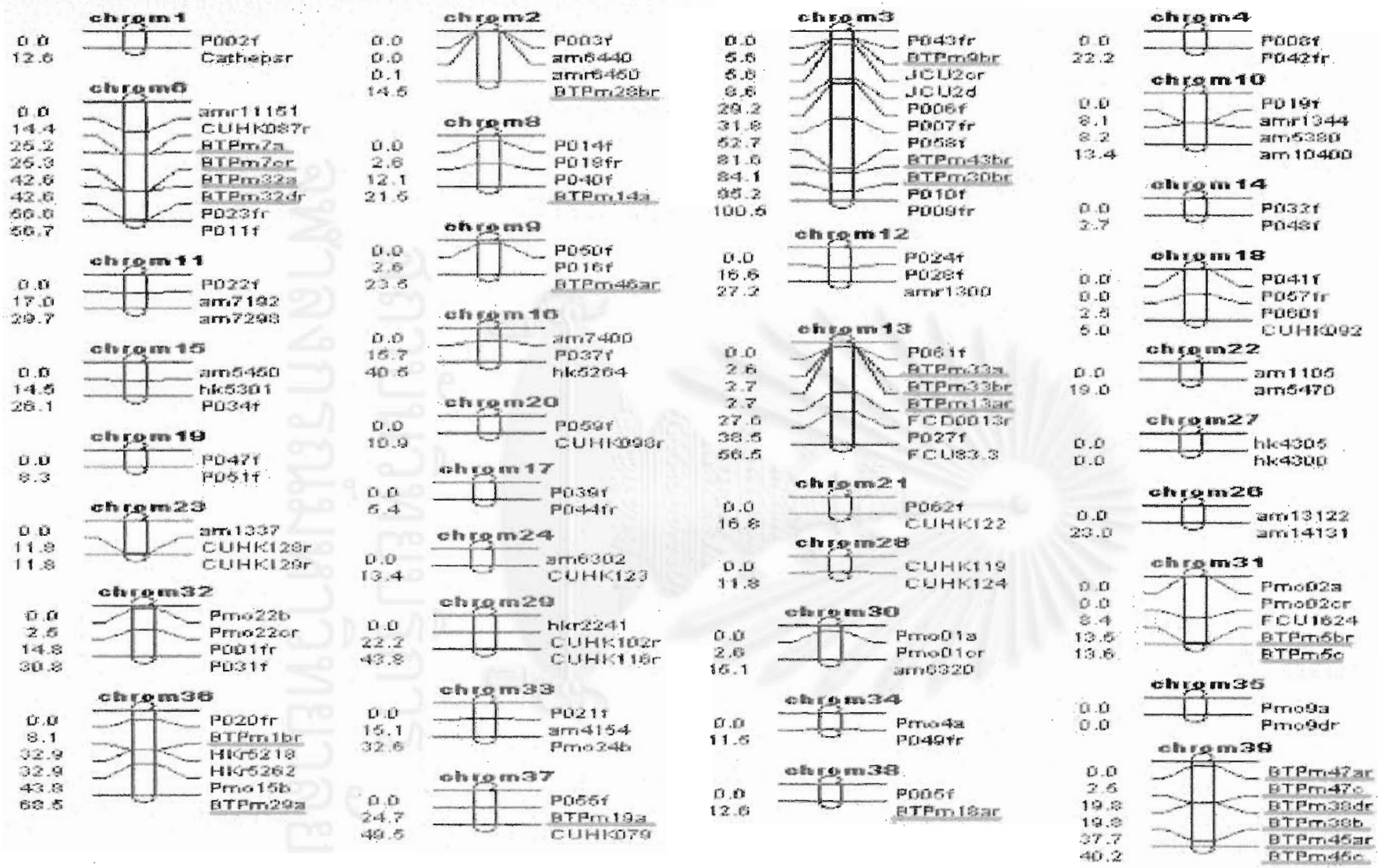


Figure 3.15 A genetic linkage map of female *P. monodon*. Red lines indicated type I microsatellite markers developed in this study.

CHAPTER IV

DISCUSSIONS

4.1 Microsatellite searching from *P. monodon* EST database

Data mining have used as a tool for development of type I microsatellite markers in several plant and animal species, such as arabidopsis (Morgante *et al.*, 2002), cotton (Saha *et al.*, 2003; Han *et al.*, 2004), festuca (Saha *et al.*, 2004), grapes (Scott *et al.*, 2000), medicago (Eujayl *et al.*, 2004), soybean (Gao *et al.*, 2003), atlantic cod (Delghandi *et al.*, 2003) and pacific oyster (Sekino *et al.*, 2003). Most of *P. monodon* microsatellite markers published were type II microsatellite markers. Generally, these type II markers were developed by sequencing clones containing microsatellites and microsatellite primers were designed from DNA flanking microsatellite sequences. Up to date, there are not any published type I microsatellite markers of *P. monodon*. This thesis is the first study that use bioinformatics analysis to identify microsatellites containing EST clones from the EST database to develop type I microsatellite markers of *P. monodon*.

For data mining of microsatellite sequences, the ESTs sequences from *P. monodon* EST database were downloaded to a local computer and after that a computer program was used to identify ESTs containing microsatellites. There are several computer programs for identification of microsatellite sequences (Table 4.1). Among several programs available in the public domain, the MicroSAtellite (MISA) search module has some features that are useful for EST quality control and for designing the primer pairs to

flank that microsatellites in a batch file. MISA has been used in several studies, in different laboratories (Thiel *et al.*, 2003; Kota *et al.*, 2001; Vershney *et al.*, 2002; Yu *et al.*, 2004; Khiestkina *et al.*, 2004). In this study we used a modified Sputnik II program to search microsatellites within EST sequences. Sputnik can show start site, stop site, repeat type, and length of microsatellites. For the *P. monodon* EST project, each clone was sequenced once, so the EST database that downloaded into a local computer was draft sequences. MISA is not suitable to search for microsatellites containing EST clones in this study because of draft sequences of the EST data. DNA sequences of primer pairs that are designed from draft EST sequences by MISA may not be true bases until editing of bases around microsatellites is performed. In this study, a modified Sputnik II program was used to identify microsatellite sequences and after that DNA sequences of flanking region at each microsatellite locus was corrected. Then primer design was performed.

Table 4.1 Computer programs for identification of microsatellite sequences

Script or Program	References
MIcroSAtellite (MISA)	http://pgrc.ipk-gatersleben.de/misa/; Thiel *et al.*, 2003
SSRFinder	Gao *et al.*, 2003
BuildSSR	Rungis *et al.*, 2004
SSR Identification Tool (SSRIT)	Kantety *et al.*, 2002
Tandem Repeat Finder (TRF)	Benson *et al.*, 1999
Tandem Repeat Occurrence Locator (TROLL)	Castelo *et al.*, 2002
CUGIssr	http://www.genome.clemson.edu/projects/ssr
Sputnik	C. Abajian; http://abajian.net/sputnik/index.html
Modified Sputnik	Morgante *et al.*, 2002
Modified Sputnik II	http://wheat.pw.usda.gov/ITMI/EST-SSR/LaRata/
SSRSEARCH	ftp://ftp.gramene.org/pub/gramene/software/scripts/ssr.pl

4.2 Characteristic of microsatellite loci

In this study, several repeat types were found in *P. monodon* EST database. Trinucleotide AAT repeats were found in the highest proportion (39.4% of trinucleotide repeat, 16.0% of EST database). The presence of AAT repeats indicated a high expression of this kind of repeat within the genome of *P. monodon*. The second most common repeat was dinucleotide AT repeats (40.4% of dinucleotide repeat, 13.7% of EST database). Several studies isolated microsatellites from genomic libraries of *P. monodon*. For dinucleotide repeat, data in this study contrast with those previously reports in which genomic GT repeats were found to be the most abundant microsatellites

among dinucleotide repeats. Xu *et al.* (1999) directly sequenced clones from partial genomic library of *P. monodon* and found that CT repeats were the most abundant (41.4%) followed by GT (35.6%) and AT repeats (19.5%). Tassanakajon *et al.* (1998) and Brooker *et al.* (2000) isolated GT and AG repeats from partial genomic libraries of *P. monodon*. Results indicated that GT repeats were more abundant than AG repeats. Brooker *et al.*, (2000) also reported that the second most abundant type of dinucleotide was AT and followed by AG repeats. Pongsomboon *et al.*, (2002) isolated tri- and tetranucleotide from partial genomic libraries of *P. monodon* and found that AG repeats were the most abundant. Wuthisuthimethavee *et al.* (2003) isolated di-, tri-, and tretranucleotide microsatellites by probe hybridization to *P. monodon* genomic library. They found that AT repeats were the most common type, followed by ATC repeats. However, all reports except the first one, reported by Xu *et al.*, (1999) isolated microsatellites by hybridizing specific repeat types of microsatellite probes with shrimp DNA. This can lead to a bias against the selected specific repeat types. The information of those reports and this thesis can not be compared.

Although, the most abundant of microsatellites from genomic libraries were dinucleotide repeats, this study found that the most abundant microsatellites from EST database were trinucleotide repeats. This may be due to the fact that trinucleotide repeats expressed more than dinucleotide repeats. However this study found that the second most abundant microsatellites in EST database were dinucleotide repeats.

There are not any reports on the isolation of CG repeats in *P. monodon.* The occurrence of CG repeats in *P. monodon* EST database of this study was rare (0.4% of dinucleotide repeat, 0.1% of EST database). Serapion *et al.*, (2004) used a bioinformatics approach to survey type I microsatellites in channel catfish EST database and found that CG repeats was rare, representing only 0.2%.

Tetra- and pentanucleotide repeats were found at low frequencies, compared with di- and trinucleotide repeats. These were concordant with the previous reports. By screening genomic *P. monodon* libraries with oligonucleotide microsatellite probes, the frequencies of positive clones for tetra- and pentanucleotide probes were much lower than dinucleotide. The positive clones of $(GT)_n$ and $(CT)_n$ probes were 85.8% and 14.2%, respectively (Tassanakajon *et al.*, 1998). Pongsomboon *et al.*, (2002) reported that the positive clones of $(AGAT)_n$ and $(ATGG)_n$ probes were 1.21 and 0.04 %, respectively. In this study, AAAG and AAAAG repeat were found at the highest proportion of tetra- and pentanucleotide repeats, respectively. The second highest proportion of tetranucleotide repeat was AAAT repeats. However, the frequency of AAAT repeats was nearly equal to AAAG repeats. The remaining of tetranucleotides were found at much lower frequencies. For pentanucleotide, although several repeat types were also identified but most were found at less than 6%.

Classification of microsatellite repeats based on Weber's criteria revealed that perfect microsatellites were the predominant categories. These results were similar to a previous report in *P. monodon* (Pongsomboon *et al.*, 2002) where perfect microsatellite

was found to be the most abundant. Perfect microsatellites were also the most abundant class in *P. vannamei* (Meehan *et al.* 2003). Unlike Wuthisuthimethavee *et al.*, (2003)'s report, compound microsatellites were found at greater frequency than perfect microsatellites. Compound microsatellites are the microsatellites at any locus that contain at least 2 types of microsatellite sequences e.g., $(GT)_n(GA)_n$, $(GATA)_n(ATT)_n$. Among compound microsatellites of this study, the association with dinucleotide repeats were the most frequent.

The BlastX results of EST containing microsatellites showed that approximately a half of EST containing microsatellites, were known genes and hypothetical proteins whereas another half were unknown. Of known genes and hypothetical proteins, microsatellites existed in coding region (CDS) more than 3'- and 5'-UTR. The existence of microsatellites in CDS region inferred that the expressed proteins contain a string of amino acid, which depend on microsatellite sequences. Most of microsatellites containing EST clones could be used to define the location of microsatellite repeats in genes, except some of them because they did not located within the region homologous to the known protein sequences.

4.3 Efficiency of marker development

Of 2,165 microsatellite loci found in the black tiger shrimp EST database, 154 loci were suitable for primer designs. Usable microsatellites in *P. monodon* were very difficult to obtain owing to the large and complex repeat arrays. Many clones contained short

flanking region that resulted from long repeat arrays of microsatellite sequences and/or the close range of microsatellites to the cloning site of vectors. These led many clones to contain only one side of the unique flanking sequences, which make them unavailable for primer design. Many microsatellites contained short repeat units but those loci were separated among each locus by a short tract of non-tandem repeat array or degenerate microsatellite-like motif, so the flanking sequences contained unsatisfied stringency requirements for primer designs.

When 154 primer pairs were tested, one hundred and twenty-six primer pairs (81.8%) were successfully amplified in 5 unrelated individual shrimps. Most of them provided amplification fragments of the expected size in *P. monodon* genomic DNA based on sequence data of EST clones. Some of them provided amplification fragments of the unexpected size in *P. monodon* genomic DNA or amplification completely failed. Previous studies reported that the successful amplification rates of microsatellite primers derived from both genomic DNA and EST sequences of various species were about 60–90% (Thiel *et al.*, 2003; Kota *et al.*, 2001; Yu *et al.*, 2004;Saha *et al.*, 2004; Cordeiro *et al.*, 2001; Gupta *et al.*, 2003). The successful amplification rates of EST-microsatellites depend on various factors. Possible explanations include: (i) one or both primers of the EST-SSR extend across a splice site; (ii) the presence of large introns in genomic DNA sequence; (iii) the use of questionable sequence information for primer development; and (iv) primers were derived from chimeric cDNA clones. Thus, the quality of the microsatellites-EST sequence for designing the primer pairs is important. In a cereal

ESTs, up to 9% of cereal ESTs sequences were low quality (Sreenivasulu *et al.*, 2002) and primer pairs should be designed carefully (Thiel *et al.*, 2003).

Furthermore, compared with genomic microsatellites, amplicon size more frequently deviated from expectation (Thiel *et al.*, 2003; Kota *et al.*, 2001; Yu *et al.*, 2004; Cordeiro *et al.*, 2001; Nicot *et al.*, 2004). This is probably results from the presence of introns and insertions-deletions (in-dels) in the corresponding genomic sequences, as was substantiated by sequence analysis (Saha *et al.*, 2004). This study probably found in-dels in the 3 markers, BTPm54, 55, and 56 which showed incompatible allele sizes when used in genotyping a *P. monodon* family,

In this study, null alleles (alleles that do not give a polymerase chain reaction product) of any microsatellite locus were observed by segregation analysis of the reference family. The results showed that 3 of polymorphic markers yielded null alleles, BTPm 36, 39, and 42. Null alleles of EST-microsatellite markers were observed in several species such as kiwifruit (Fraser *et al.*, 2004), rice (Cho *et al.*, 2000), spruce (Rungis *et al.*, 2004) and wheat (Gupta *et al.*, 2003; Eujayl *et al.*, 2001). In wheat, occurrence of null alleles is common and has been reported earlier using genomic microsatellites (Gupta and Varshney., 2000). Occurrence of null alleles can be explained by: (i) the deletion of microsatellite at a specified locus (Callen *et al.*, 1993); (ii) mutations (in-dels or substitutions) in the primer binding site (Lehman *et al.*, 1996). Occurrence of null alleles complicates the interpretation of segregation data because heterozygotes cannot be

identified. Null alleles result in deviation from the expected Mendelian segregation ratios (Fraser *et al.*, 2004).

4.4 Polymorphism analysis of microsatellite loci

Fifty microsatellite loci were characterized for genetic information content by genotyping *P. monodon* broodstock from Trang province (the Andaman sea of Thailand). Among these, the number of alleles at each of 50 polymorphic loci ranged from 4 to 30 with an average of 12.6. Expected heterozygosity ranged from 0.257 to 0.960 with an average of 0.759. In this thesis, we also used the data on microsatellite loci and their corresponding alleles to calculate the PIC in order to examine the extent of information on diversity that these markers can provide. The PIC value of this study ranges from 0.237 to 0.948 with an average of 0.723. Under PIC value, (1) PIC> 0.5 means locus with highly informative (2) 0.5>PIC>0.25 means locus with reasonably informative (3) PIC<0.25 mean locus with slightly informative. Loci with many alleles and a PIC near 1 are the most desirable for highly polymorphic markers (Botstein *et al.*, 1980).

In this study the degree of polymorphism of *P. monodon* type I microsatellite markers is comparable to that of type II markers. Wuthisuthimethavee *et al.* (2003) reported that the average PIC value of type II microsatellite markers of 30 loci was 0.835 with the highest at 0.926 and the lowest at 0.428. Thiel *et al.*, (2003) developed Type I microsatellite markers from barley (Hordeum vulgare L.) EST database. They found that these markers also showed high levels of polymorphisms. The average PIC value of 76

loci was 0.45 with the highest at 0.78 and the lowest at 0.28. However, the locus BTPm37 of this study showed low level of observed heterozygosity (0.135) and high PIC value (0.768) when 37 individuals *P. monodon* from Trang were investigated. Low level of observed heterozygosity found at the BTPm37 locus may possibly be resulted from null alleles. The null alleles of BTPm37 could not be detected by genotyping with the reference family used for genetic linkage mapping. The null alleles may be shown when BTPm 37 is used to genotype other families. "Null" alleles of the heterozygous status can cause an apparent excess of homozygosity. Moreover, null alleles may exist because of high polymorphisms affecting the efficiency of PCR priming or extension. The presence of null alleles at some loci has been documented for several species where pedigree analysis was feasible (e.g. Pemberton *et al.*, 1995; Paetkau and Strobeck, 1995).

The level of polymorphism of microsatellite loci in this study, as estimated by the number of alleles and PIC values, almost showed a positive correlation with the average number of microsatellite repeats. Thiel *et al.*, (2003) also found that the polymorphic of (Hordeum vulgare L.) type I markers depends on the number of repeats.

It is interesting to note that microsatellites, which locate in CDS, showed the highest polymorphism comparing to those located at the 5'- and 3'-UTR. These results correspond to those previously reported. Cordeiro *et al.*, (2001) reported that sugarcane (Saccharum spp.) type I microsatellite markers, which produced polymorphism, mainly located in the coding sequence region with only one locus located within the 5'-UTR. In

contrast to barley type I microsatellite markers (Thiel *et al.*, 2003), 3'-UTR yielded a higher proportion of polymorphic microsatellites than 5'-UTR did.

4.5 Application in shrimp genome mapping

Microsatellite markers, developed from genomic libraries, can belong to either the transcribed region or the non transcribed region of the genome, and rarely is there information available regarding their functions. By contrast, EST-derived microsatellite markers often have known or putative functions and are gene targeted markers with the potential of representing functional markers in those cases where polymorphisms in the repeat motifs affect the function of the gene in which they reside (Anderson and Lubberstedt., 2003). Putative functions for a significant proportion of EST-microsatellite markers have been reported (Thiel *et al.*, 2003; Yu *et al.*, 2002; Han *et al.*, 2004; Gao *et al.*, 2004). EST-microsatellite markers are one class of markers that can contribute to direct allele selection, if they are shown to be completely associated or even responsible for a targeted trait (Sorrells and Willson., 1997). For example, recently, a Dof homolog (DAG1 gene that showed a strong effect on seed germination in *Arabidopsis* (Papi *et al.*, 2000)) has been mapped on chromosome 1B of wheat by using wheat EST-microsatellite primers (Gao *et al.*, 2004). Similarly, Yu *et al.* (2004a) identified two EST-microsatellite markers linked to the photoperiod response gene (*ppd*) in wheat. Finally, mapping candidate genes can facilitate genome alignment across distantly related species (Yu *et al.*, 2004b).

In this study, 36 type I microsatellite markers integrated into *P. monodon* male and/or female maps. Construction of genetic linkage map is based on the determination of distance between 2 markers. The LOD score method was used in this study. For construction of genetic map, LOD score of 3.5 and maximum θ = 0.30 were set as linkage threshold for grouping markers. LOD score were considered the estimate of the linkage distance. The following is the formula for the LOD score:

$$\text{LOD score} = \log \left[\frac{\text{probabilit y of birth sequence with a given linkage value}}{\text{probabilty of birth sequence with no linkage}} \right]$$

Maximum distance parameter (θ) is the longest possible distance between 2 markers. Under these parameter, if 2 markers separated over 30 cM or LOD score lower than 3.5, these two markers were separated into 2 groups.

For male map, twenty-six type I microsatellites developed in this study were placed onto the male map, distributing into 18 linkage groups. The total length of the present male map was increase from 878 cM to 1,101 cM. The present map extended 223 cM in length from the previous. Chrom 42-48, were newly linkage groups. Type I microsatellites developed in this study were also placed in these 7 groups. For Chrom 30, type I microsatellites developed in this study made this group turn to be the longest group with 84.3 cM long. The length of this group was increased by 49.9 cM long when three markers of this study were added. The length of previous Chrom 9 was 34.6 cM

long. CUHK065r locus was out of the previous Chrom 9 with 32.4 cM apart when maximum θ was set at 0.30. BTPm21br acted as a bridge between CUHK065r and Chrom 9. This led CUHK065r to place onto Chrom 9. Like another group, Chrom 21, type I microsatellite markers of this study also acted as a bridge to form the markers of the previous map, which were separated over 30 cM to form into the same group.

For female map, eighteen type I microsatellite markers developed in this study were placed onto the female map. These 18 markers distributed into 11 groups. A total length of the present female map was increase from 630 cM to 891.4 cM, which was extended 261.4 cM in length. We found that Chrom 37-39, 3 groups, were newly groups, which included the markers of this study. Especially, Chrom 39 was only formed by type I microsatellite of this study. In Chrom 6, BTPm7 and BTPm32 acted as a bridge to evolve P011f into this group, and extended the length of Chrom6 from 14.4 cM to 56.7 cM. Like in Chrom3, type I microsatellite acted as a bridge in this group, and also turned this group to be the longest group of female map.

The genome size of shrimps, *P. aztecus, P. duorarum, P. setiferus,* and *L. vannamei* is approximately 70 % of the human genome ($2x10^9$ bp.), esimated to be about 2,000 cM (Chow *et al.*, 1991; Wilson *et al.*, 2002). In this study, the total length of both male and female maps covered about half of *P. monodon* genome size with average spacing between 2 markers of 7.0 and 8.0 cM for male and female maps, respectively. Additional markers are required to condense the existing map into 44 linkage groups corresponding to the number of haploid chromosome in *P. monodon* (Xiang *et al.*, 1993).

For male map, the number of linkage groups from this study was 47. The linkage groups were over the number of haploid chromosome, which were 44 linkage groups. This may be due to the lacking of bridge markers to combine the extra markers. For female map, the number of linkage groups from this study was 36. The lacking of the remaining groups may be due to the unavailable markers to cover the shrimp genome.

In conclusion, out of 50 polymorphic microsatellite markers developed in this study, 36 were mapped into previous *P. monodon* international linkage map. This means that 72% of polymorphic markers could be placed into the linkage map. BTPm1, 7, 9, 30, 32, 38, 43 and 47 can be placed on both male and female maps. Compared to other studies, Thiel *et al.* (2003) found that 38.0% of type I microsatellite markers derived from barley EST database were integrated into a barley genetic consensus map. These markers distributed into 7 genetic linkage groups. Han *et al.*, (2004) developed type I microsatellite markers of *Gossypium arboreum*, and they found that 99 primer pairs produced 118 microsatellite loci. Of these 118 loci, 111 loci were integrated into the backbone map. Han's report showed that 72 loci were allocated to the At subgenome and 37 loci were allocated to Dt subgenome of allotetraploid cotton, leaving two loci unallocated. However, genetic maps with saturation of useful markers is quite easy to find markers linked commercially important trait (Temnykh, 2000).

For the shrimp genome mapping project, the addition of new markers is still needed to populate several regions previous lacking in molecular markers. It is then intended to increase the density of this framework map by using more Type I (coding)

markers. In this study, we identified microsatellites from ESTs which are microsatellite markers in transcribed sequences (Type I marker). They can be associated with a gene of interest function. The polymorphic type I markers are more useful as anchorage points for comparative gene mapping than type II markers and can be directly used in selective breeding programs (Liu *et al.*, 1999).

CHAPTER V

CONCLUSIONS

1. Among 10,100 EST database of *P. monodon*, one thousand three hundreds and eighty-one microsatellite loci were identified

2. Trinucleotide AAT repeat and dinucleotide AT repeat were the most abundant repeat type in database.

3. The predominant category of microsatellite loci was perfect repeats whereas imperfect and compound microsatellites were found at a much lesser extent. Homology searches by BLASTX program revealed that the EST containing microsatellite clone represent 11.6% known gene products, 37.3% hypothetical protein, and 51.1% unknown gene products. Coding sequence–ESTs yield a higher portion of microsatellite location (93.5%) than 3'-UTR and 5'-UTR did.

4. All of 154 primer pairs, fifty polymorphic markers were developed. In general, microsatellite exhibited high levels of polymorphism. The average number of alleles per locus, PIC value, observed, and expected heterozygosities were 12.6, 0.723, 0.698, and 0.759, respectively.

5. A total of 36 polymorphic markers were integrated into male and/or female international genetic shrimp map along with AFLPs, microsatellite, and another markers.

REFERENCES

Anderson, I. 1993. *The veterinary approach to marine prawns.* In Brown (ed.), Aquaculture for veterinations: Fish husbandry and medicine. Amsterdam: Oxford Pergamon Press. pp. 271-290.

Anderson, J.R. and Lubberstedt, T. 2003. Functional markers in plants. *Trends. Plant. Sci.* 8: 554–560

Avise, J.C. 1994. *Molecular Markers, Natural History, and Evolution.* Chapman and Hall, New York, NY. 511 pp.

Avise, J.C., Helfman, G.S., Saunders, N.C., and Hales, L.S. 1986. Mitochondrial DNA differentiation in North Atlantic eels: population genetic consequences of an unusual life history pattern. *Proc. Natl. Acad. Sci. U. S. A.* 83: 4350–4354.

Bagley, M.J., Anderson, S.L., and May, B. 2001. Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. *Ecotoxicology* 10: 239–244.

Baily-Brook, J.H. and Moss, S.M. 1992. *Penaeid taxonomy, biology and zoo geography.* m A.W. Fast and L.J. Lester (eds), Marine shrimp culture: Principles and practices; Amsterdam: Elsrvier Science Publishers.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27: 573–580

Benzie, J.A., Ballment, E., Forbes, A.T., Demetriades, N.T., Sugama, K., and Moria, S. 2002. Mitochondrial DNA variation in Indo-Pacific populations of the giant tiger prawn, *Penaeus monodon. Mol. Ecol.* 11: 2553–2569.

Benzie, J.A.H., Frusher, S. and Ballment, E. 1992. Geographical variation in allozyme frequencies of populations of Penaeus monodon (Crustacea: Decapoda) in Australia. *Aus. Mar. Freshwat. Res.* 43: 715-725.

Birky, C.W., Fuerst, P., and Maruyama, T. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effect of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121: 613–627.

Botstein, D., White, R.L., Skolnick, M., and Davis; R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314–331.

Boyd, Y. 1998. Genetic mapping of the mouse genome. *Methods Enzymol.*14: 1120-1134.

Brooker, A. L., Benzie, J. A. H., Blair, D. and Versini, J. J. 2000. Population structure of the giant tiger prawn *Penaeus monodon* in Australian water, determined using microsatellite markers. *Mar. Biol.* 136: 149-157.

Brown, B., and Epifanio, J. 2003. Nuclear DNA. In: Hallermann, E.M. (Ed.), *Population Genetics: Principles and Applications for Fisheries Scientists*. American Fisheries Society, Bethesda, MD. 458 pp.

Brown, J.R., Bechenbach, A.T., and Smith, M.J. 1993. Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). *Mol. Biol. Evol.* 10: 326–341.

Brown, W.M. 1985. *The mitochondrial genome of animals.* In: MacIntyre, R.J. (Ed.), Molecular Evolutionary Genetics. Plenum, New York, NY, pp. 95–130.

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, RI., Mulley, IC.and Sutherland, G.R 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *Am. J. Hum. Genet.* 52: 922-927.

Carvalho, G.R and Pitcher, T. I 1995. *Molecular genetics in fisheries.* London: T.J.Press padstow. Ltd.

Castelo, A.T., Martins, W., and Gao, G.R. 2002. TROLL–Tandem Repeat Occurrence Locator. *Bioinformatics* 18: 634–636

Cho, Y. G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S. R, Park, W.D., Ayres, N. and Cartinhour, S. 2000. Diversity of microsatellites derived from genome libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100: 713-722.

Chow, S., and Kishino, H. 1995. Phylogenetic relationships between tuna species of the genus Thunnus (*Scombriidae: Teleosrei*): inconsistent implications from morphology, nuclear and mitochondrial genomes. *J. Mol. Evol.* 41: 741–748.

Chow, S., Clarke, M.E., and Walsh, P.J. 1993. PCR-RFLP analysis on thirteen western Atlantic snappers (subfamily Lutjaninae): a simple method for species and stock identification. *Fish. Bull.* 91: 619–627.

Chow, S., Dougherty, W.J., and Sanfer, P.A. 1990. Meitotic chromosome complements and nuclear DNA contents of four species of shrimps of the genus *Penaeus*. *J. Crustacean Biol.* 10: 29-36

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M., and Henry, R.J. 2001. Microsatellite markers from sugarcane Saccharum spp. ESTs cross transferable to erianthus and sorghum. *Plant Sci.* 160: 1115–1123

Crawford, A.M., and Cuthbertson, R.P. 1996. Mutations in sheep microsatellites. *Genome Res.* 6: 876–879.

Crawford, M.K., Grimes, C.B., and Buroker, N.E. 1989. Stock identification of weakfish, *Cynoscion regalis*, in the middle Atlantic region. *Fish. Bull.* 87: 205–211.

Cronin, M.A., Spearman, W.J., Wilmot, R.L., Patton, J.C., and Bickham, J.W. 1993. Mitochondrial DNA variation in chinook (*Oncorhyncus tshawytscha*) and chum salmon (*O. keta*) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. Can. *J. Fish. Aquat. Sci.* 50: 708–715.

Davis, L.G., Dibner, M.D. and Battey, J.F. 1986. *Preparation of genomic DNA*. In Basic method in molecular biology. Elsevier Science. New York. pp. 42-43.

Delghandi, M., Mortensen, A., and Westgaard, J.I. 2003. Simultaneous analysis of six microsatellite markers in Atlantic cod (*Gadus morhua*): a novel multiplex assay system for use in selective breeding studies. *Mar Biotechnol.* 5:141–148.

Dinesh, K.R., Chan, W.K., Lim, T.M., and Phang, V.P.E. 1995. RAPD markers in fishes: an evaluation of resolution and reproducibility. *Asia-Pac. J. Mol. Biol. Biotechnol.* 3: 112–118.

Dodgson, J.B., Cheng, H.H., and Okimoto, R. 1997. DNA marker technology: a revolution in animal genetics. *Poult. Sci.* 76: 1108–1114.

Estoup, A., and Cornuet, J. 1999. *Microsatellite evolution: inferences from population data.* In: Goldstein, D.B., Schlotterer, C. (Ed.), Microsatellites: Evolution and Applications. Oxford Univ. Press, New York, pp. 49–65.

Eujayl, I. Sledge, M.K., Wang. L., May, G.D., Chekhovskiy, K., Zwonitzer, J.C., and Mian, M.A. 2004. *Medicago truncatula* EST-SSRs reveal cross species genetic markers for Medicago spp. *Theor. Appl. Genet.* 108: 414–422

Fraser, L.G. Harvey, C.F., Crowhurst, R.N., De Silva, H.N. 2004. EST-derived microsatellites from Actinidia species and their potential for mapping. *Theor. Appl. Genet.* 108: 1010–1016

Gao, L.F., Jing, R.L., Huo, N.X., Li, Y., Li, X.P., Zhou, R.H., Chang, X.P., Tang, J.F., Ma, Z.Y., and Jia, J.Z. 2004. One hundred and one new microsatellite loci derived from ESTs EST-SSRs. in bread wheat. *Theor. Appl. Genet.*108: 1392–1400

Gao, L.F., Lifeng Gao1 , Tang, J., Jing, R., Huo, N., Li, H., and Jia, J. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.*12: 245–261

Gold, J.R., Richardson, L.R., Furman, C., and King, T.L. 1993. Mitochondrial DNA differentiation and population structure in red drum (*Sciaenops ocellatus*) from the Gulf of Mexico and Atlantic Ocean. *Mar. Biol.* 116: 175–185.

Gracia de Leon, F. J., Dallas, F., Chatain, B., Canonne, M., Versini, L., and Bonhomme, F. 1995. Development and use of microsatellite markers in seabass, *Dicentrarchus labrax* (linnaeus, 1758) (*Perciformes errandidae*). *Mol. Mar. Biol. Biotechnol.* 4: 62-68.

Graves, J.E., McDowell, J.R., and Jones, M.L. 1992. A genetic analysis of weakfish *Cynoscion regalis* stock structure along the mid-Atlantic coast. *Fish. Bull.* 90: 469–475.

Grey, D.L., Dall, W. and Baker, A. 1983. *A guide to the Australian Penaeid prawns*. Darwin: Northern Territory Government Printing Office

Gupta, P.K. and Varshney, R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163–185

Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H.S. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics* 270: 315–323

Hallerman, E.M., Dunham, R., and Smitherman, R.O. 1986. Selection or drift-isozyme allele frequency changes among channel catfish selected for rapid growth. *Trans. Am. Fish. Soc.* 115: 60–68.

Han, Z.G., Guo, W.Z., Song, X.L., and Zhang, T.Z. 2004. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. *Mol. Gen. Genom.* 272: 308–327

Hedgecock, D., Tracey, M.L. and Nelson, K. 1982. *Genetics*. In: Abele, L.G. (ed). Embryology, morphology and genetics. New York: Academic Press. pp. 283-355

Heist, E.J., and Gold, J.R. 1999. Microsatellite DNA variation in sandbar sharks (*Carcharhinus plumbeus*) from the Gulf of Mexico and mid-Atlantic bight. *Copeia* 1: 182–186.

Hillis, D.M., Moritz, C., and Mable, B.K. 1996. Molecular Systematics, 2nd ed. Sinauer Associates, Sunderland, MA. 655 pp.

Hirschfeld, D., Dhar, A.K., Rask, K., and Alcivar-Warren, A. 1999. Genetic diversity in the eastern oyster (*Crassostrea virginica*) from Massachusetts using RAPD technique. *J. Shellfish Res.* 18: 121–125.

Huang, X., and Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877

Johnson, K.R., Wright, J.E, and May, B. 1987. Linkage relationships reflecting ancestral tetraploidy in salmonid fish. *Genetics*. 116: 579–591.

Kantety, R.V., La Rota, M., Matthews, D.E., and Sorrells, M.E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48: 501–510.

Karsi, A., Patterson, A., Feng, J., and Liu, Z.J. 2002. Translational machinery of channel catfish: I. A transcriptomic approach to the analysis of 32 40S ribosomal protein genes and their expression. *Gene* 291: 177–186.

Khlestkina, E.K., Than, M.H.M., Pestsova, E.G., Roder, M.S., Malyshev, S.V., KORZUN, V., and Borner, A. 2004. Mapping of 99 new microsatellite-derived loci in rye *Secale cereale* L. including 39 expressed sequence tags.*Theor. Appl. Genet.* 109: 725–732

Kijas, J.M.H., Fowler, J.C.S., Garbett, C.A., and Thomas, M.R. 1994. Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *Bio-Techniques* 16: 656–662.

Kirby, L.T. 1992. *DNA Fingerprinting*: An introduction New York: W.H. Freemanand Company.

Klinbunga, S., Ampayup, P., Tassanakajon, A., Jarayabhand, P., and Yoosukh, W. 2000. Development of species specific markers of the tropical oyster (*Crassostrea belcheri*) in Thailand. *Mar. Biotechnol.* 2: 476–484.

Klinbunga, S., Penman, D.J., McAndrew, RJ. and Tassanakajon, A. 1999. Mitochondrial DNA diversity in three populations of the giant tiger shrimp *Penaeus monodon*. *Mar. Biotechnol.* 1: 113-121.

Klinbunga, S., Siludjai, D., Wudthijinda, W., Tassanakajon, A., Jarayabhand, P. and Menasveta, P. 2001. Genetic heterogeneity of the giant tiger shrimp (*Penaeus monodon*) in Thailand revealed by RAPD and mitochondrial RFLP analyses. *Mar. Biotechnol.* 3: 428-438.

Kota, R., Varshney, R.K., Thiel, T., Dehmer, K.J., and Graner, A. 2001. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *J. Hered.* 135: 145–151

Lehman, T. Hawley, W.A., and Collins, F.H. 1996. An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics* 144: 1155–1163

Lester, L.J. 1978. Population genetic of penaeid shrimp from the Gulf of Mexico. *J. Hered.* 70: 175-180.

Levinson, G., and Gutman, G.A. 1987. High frequency of short frame shifts in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* 15: 5323–5338.

Litt, M., and Luty, J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44: 397–401.

Liu, Q., Goudie, C.A., Simco, B.A., Davis, K.B., and Morizot, D.C. 1992. Gene-centromere mapping of six enzyme loci in gynogenetic channel catfish. *J. Heredity*. 83: 245–248.

Liu, Z.J., Li, P., Kocabas, A., Ju, Z., Karsi, A., Cao, D., and Patterson, A. 2001c. Microsatellite-containing genes from the channel catfish brain: evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene RAD23B. *Biochem. Biophys. Res. Commun.* 289: 317–324.

Liu, ZJ and Cordes, JF. 2004. DNA marker technologies and their application in aquaculture genetics. *Aquaculture*. 238: 1-37

May, B., and Johnson, K.R. 1993. Composite linkage map of salmonid fishes (Salvelinus, Salmo, and Oncorhynchus). In: O'Brien, S.J. (Ed.), Genetic Maps: Locus Maps of Complex Genomes. *Cold Spring Harbor*. 4: 309–317.

May, B., Stoneking, M., and Wright, J.E. 1980. Joint segregation of biochemical loci in Salmonidae: II. Linkage associations from a hybridized Salvelinus genome (S. namaycush X S. fontinalis). *Genetics*. 95: 707–726.

McGoldrick, D.J., and Hedgecock, D. 1997. Fixation, segregation and linkage of allozyme loci in inbred families of the Pacific oyster *Crassostrea gigas* (Thunberg): implications for the causes of inbreeding depression. *Genetics* 146: 321–334.

Meehan, D., Xu, Z., Zuniga, G., and Alcivar-Warren A. 2003. High Frequency and Large Number of Polymorphic Microsatellites in Cultured Shrimp, *Penaeus (Litopenaeus) vannamei* [Crustacea:Decapoda]. *Mar. Biotechnol.* 5: 311–330

Morgante, M., Hanafey. M., and Powell, W. 2002. Microsatellites are preferentially present with non-repetitive DNA in plant genomes. *Nat. Genet.* 30: 194–200

Morizot, D., Schmidt, M., and Carmichael, G. 1994. Joint segregation of allozymes in catfish genetic crosses: designation of *Ictalurus punctatus* linkage group I. *Trans. Am. Fish. Soc.* 123: 22–27.

Morizot, D.C., Slaugenhaupt, S.A., Kallman, K.D., and Chakravarti, A. 1991. Genetic linkage map of fishes of the genus Xiphophorus (Teleostei: Poeciliidae). *Genetics*. 127: 399–410.

Mork, J., Ryman, N., Sta°hl, G., Utter, F., and Sundness, G. 1985. Genetic variation in Atlantic cod (*Gadus morhua* L.) throughout its range. Can. *J. Fish. Aquat. Sci.* 42, 1580–1587.

Motoh, H. 1981. Studies on the fisneries bioligy of the giant tiger prawn, *panaeus monodon* in the Philippines. Technical paper no. 7. Aquaculture Department, Southeast Asia Fiaheries Development Center. Iloilo.

Motoh, H. 1985. Biology and ecology of *Penaeus monodon*, pp26-36. In: Proceeding of the first international conference on the culture of Penaeid prawns/shrimps (Taki, Y, Premavara, J.H. and Lobrera, J., eds.) Aquaculture Department, Southeast Asia Fiaheries Development Center. Iloilo.

Murphy, R.W., Sites Jr., J.W., Buth, D.G., and Haufler, C.H. 1996. *Proteins I: Isozyme electrophoresis*. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, 2nd ed. Sinauer Associates, Sunderland, MA, pp. 51–132.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 23: 341-369.

Nicot, N., Chiquet, V., Gandon, B., Amilhat, L., Legeai, F., Leroy, P., Bernard, M., and Sourdille, P. 2004. Study of simple sequence repeat SSR markers from wheat expressed sequence tags (ESTs). *Theor. Appl. Genet.* 109: 800–805

O'Brien, S.J. 1991. Molecular genome mapping: lessons and prospects. *Curr. Opin. Genet. Dev.* 1, 105–111.

O'Connell, M., and Wright, J.M. 1997. Microsatellite DNA in fishes. *Rev. Fish Biol. Fish.* 7: 331–363.

O'Reilly, P., and Wright, J.M. 1995. The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. *J. Fish Biol.* 47: 29–55.

Ostrander, E.A., Jong, P.M., Rine, J., and Duyk, G. 1992. Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proc. Natl. Acad. Sci. U. S. A.* 89: 3419–3423.

Paetkan, D. and Strobeck, C. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molec. Ecol.* 4: 519-520.

Papi, M., Sabatini, S., Bouchez, D., Camilleri, C., Costantino, P., and Vittorioso, P. 2000. Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. *Genes Dev.* 14: 28–33

Partis, L., and Wells, R.J. 1996. Identification of fish species using random amplified polymorphic DNA (RAPD). *Mol. Cell. Probes* 10: 435–441.

Pasdar, M., Philipp, D.P., and Whitt, G.S. 1984. Linkage relationships of nine enzyme loci in sunfishes (*Lepomis Centrarchidae*). *Genetics* 107: 435–446.

Pemberton, J.M., Slate, J., Bancroft, D.R., and Barrett, A. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.* 4: 249-252

Pérez, F., Erazoa, C., Zhinaulaa, M., Volckaertb, F., and Calderon J. 2004. A sex-specific linkage map of the white shrimp *Penaeus* (*Litopenaeus*) *vannamei* based on AFLP markers. *Aquaculture* 242: 105–118

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., and Quackenbush, J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 19: 651-652

Pongsomboon, S. 2002. Isolation and characterization of tri-and tetranucleotide microsatellite sequences for application to DNA typing in the black tiger shrimp *Penaeus monodon*. Ph.D. Thesis, Department of Biochemistry, Graduate school, Chulalongkorn University. 192 pp.

Pongsomboon, S., Whan, V., Moor, S. S. and Tassanakajon, A. 2000. Characterization of tri-and tetranucleotide micro satellites in the black tiger prawn, *Penaeus monodon*. *ScienceAsia*. 26: 1-8.

Rungis, D., Berube, Y., Zhuang, J., Ralph, S., Ritand, C.E., Ellis, B.E., Douglas, C., Bohlmann, J., and Ritland, K. 2004. Robust simple sequence repeat markers for spruce Picea spp.. from expressed sequence tags. *Theor. Appl. Genet.*109: 1283–1294

Ryman, N., and Utter, F. 1987. *Population Genetics and Fishery Management.* University of Washington Press, Seattle. 420 pp.

Saha, M.C., Mian, M.A., Eujayl, I., Zwonitzer, J.C., and Wang, L., May, G.D. 2004a. Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* 109: 783–791

Saha, M.C., Zwonitzer, J.C., Chekhovskiy, K., Hopkins, A.A., and Rouf Mian, M.A. 2004b. A high-density linkage map of tall fescue based on SSR and AFLP markers. *Theor. Appl. Genet.* (in press)

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* 100: 723–726

Seeb, J.E., and Seeb, L.W. 1986. Gene mapping of isozyme loci in chum salmon. *J. Heredity*. 77: 399–402.

Sekino, M., Hamaguchi, M., Aranishi, F., and Okoshi, K. 2003. Development of novel microsatellite DNA markers from the Pacific oyster *Crassostrea gigas*. *Mar Biotechnol* 5: 227–233.

Serapion, J., Kucuktas, H., Feng, J., and Liu, Z. 2004. Bioinformatic mining of type I microsatellites from express sequence tags of chanelle catfish (*Ictalurus punctatus*). *Mar. Biotechnol.* 6: 364-377

Siddell, B.D., Otto, R.G., Powers, D.A., Karweit, M., and Smith, J. 1980. Apparent genetic homogeneity of spawning striped bass in the upper Chesapeake Bay. *Trans. Am. Fish. Soc.* 109, 99–107.

Solis, N.R. 1988. Biology and ecology. In Taki, Y. Premavara, J.H., and Lobrera, J.(eds.), *Biology and culture of Penaeus monodon*, Aquaculture Department, Southeast Asia Fisheries Development Certer.

Sootanan, P. 1999. Development of DNA typing in black tiger prawn *Penaeus monodon* Fabricius by microsatellite technique. M. Sc. Thesis, Department of Biochemistry, Graduate school, Chulalongkorn University. 116 pp.

Sorrells, M.E. and Wilson, W.A. 1997. Direct classification and selection of superior alleles for crop improvement. *Crop Sci.* 37: 691–697

Southern, E.M. 1976. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503–517.

Sreenivasulu, N., Kishor, P.V.K., Varshney, R.K., and Altschmied, L. 2002. Mining functional information from cereal genomes–the utility of expressed sequence tags. *Curr. Sci.* 83: 965–973

Stalling, R. L., Ford, A. F., Nelson, D., Tomey, D. C., Hildebrand, C. E. and Moyzis,R. K. 1991. Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics*. 10: 807-815.

Stoner, D.S., Quattro, J.M. and Weissman, 1. 1997. Highly polymorphic microsatellite loci in the colonial ascidian *Botryllus schlosseri. Mol. Mar. Biol. Biotechnol.* 6:163-171.

Supungul, P. 1998. Genetic variation and population structure of the black tiger prawn P. monodon in Thailand determined by microsatellite markers. M. Sc. Thesis, Department of Biochemistry, Graduate school, Chulalongkorn University. 68 pp.

Supungul, P., Sootanan, P., Klinbunga, S., Kamonrat, W., Jarayabhand, P. and Tassanakajon, A. 2000. Microsatellite polymorphism and the population structure of the black tiger shrimp (*Penaeus monodon*) in Thailand. *Mar.Biotechnol.* 2: 339-347.

Tassanakajon, A., Pongsomboon, S., Jarayabhand, P., Klinbunga, S., and Boonsaeng, V.V. 1998. Genetic structure in wild populations of black tiger shrimp (*Penaeus monodon*) using randomly amplified polymorphic DNA analysis. *J. Mar. Biotechnol.* 6: 249–254.

Tassanakajon, A., Pongsomboon, S., Rimphanitchayskit. V., Jarayabhand, P. and Boonsaeng, V. 1997. Random amplified polymorphic DNA (RAPD) markers for determination of genetic variation in wild populations of the black tiger prawn (*Penaeus monodon*) in Thailand. *Mol. Mar. Bioi. Biotechnol.* 6:110-115.

Tassanakajon, A., Tiptawonnukul, A., Supungul, P., Rimphanitchayakit, V., Cook, D.,Jarayabh, P., Klinbunga, S. and Boonsaeng, V. 1998. Isolation and characterization of microsatellite markers in the black tiger prawn *Penaeus monodon*. *Mol. Mar. Biol. Biotechnol.* 7: 55-61.

Tautz, D. 1989. Hyper variability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17: 6463–6471.

Thiel, T., Michalek, W., Varshney, R.K., and Graner, A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare L.*). *Theor Appl Genet* 106: 411–422

van Oppen, M.J.H., Klerk, H., Olsen, J.L., and Stam,W.T. 1996. Hidden diversity in marine algae: some examples of genetic variation below the species level. *J. Mar. Biol. Assoc. U. K.* 76: 239–242.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and Graner, A. 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* 7: 537–546.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeay, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407–4414.

Weber, J.L. 1990. Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. *Genomics* 7: 524-530.

Weber, J.L., and Wong, C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2: 1123–1128.

Weinmayr, G., Vautrin, D. and Solignac, M. 2000. Isolation and characterization ofhighly polymorphic microsatellites from the Polychaete Pectin aria koreni. *Mar. Biotechnol.*2: 92-99.

Welsh, J., and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18: 7213–7218.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18: 6531–6535.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., and Prager, E.M. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26: 375–400.

Wilson, K., Li, Y., Whan, V., Lehnert, S. Byme, K., Moor, S., Pongsomboon, S., Tassanakajon, A., Rosenberg, G., Ballment, E., Fayzi, Z., Swan, J., Kenway,M. and Benzie, J. 2002. Genetic mapping of the black tiger shrimp Penaeus

monodon with amplified fragment length polymorphism. *Aquaculture*. 204:297-309.

Wolfus, G.M., Garcia, D.K., and Alcivar-Warren, A. 1997. Application of the microsatellite techniques for analyzing genetic diversity in shrimp breeding programs. *Aquaculture* 152: 35–47.

Wright, J.M. 1993. DNA fingerprinting in fishes. In: Hochachka, P.W., Mommsen, T. (Eds.), Biochemistry and Molecular Biology of Fishes. Elsevier, Amsterdam, pp. 58–91.

Wuthisuthimethavee, S., Lumubolb P., Toojinda T., Tragoonrung, S., and Vanavichit, A. 2005. SSLP-based linkage map construction in black tiger shrimp (*Penaeus monodon* Fabricius). *ScienceAsia 31*: 87-93

Wuthisuthimethavee, S., Lumubolb P., Vanavichitc, A., and Tragoonrung, S. 2003. Development of microsatellite markers in black tiger shrimp (*Penaeus monodon* Fabricius). *Aquaculture* 224: 39–50

Xiang, J.H., Liu, R.Y., and Zhou, L.H. 1993. Chromosomes of marine shrimps with special reference to different techniques. Aquaculture 111: 321.

Yu, J.K., Dake, T.M., Singh, S., Benscher, D., Li, W., Gill, B., and Sorrells, M.E. 2004a. Development and mapping of EST-derived simple sequence repeat SSR. markers for hexaploid wheat. *Genome* 47: 805–818

Yu, J.K., La Rota, M., Kantety, R.V., and Sorrells, M.E. 2004b. EST-derived SSR markers for comparative mapping in wheat and rice. *Mol. Genet. Genomics* 271: 742–751

Yue, G., Li, Y., Chen, F., Cho, S., Lim, L.C., and Orban, L. 2002. Comparison of three DNA marker systems for assessing genetic diversity in Asian arowana (*Scleropages formosus*). *Electrophoresis* 23: 1025–1032.

Yuea, G.H., Hoa, M.Y., Orbana, L., and Komen J. 2004. Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* 234: 85–98

APPENDIX A

Transcribed microsatellites and their associated genes

Gene identity	Acc. number	Repeat unit	E-value	Location	Clone name
39 kDa antigen	AAB47253	(AC)13	2.00E-110	CDS	ES-N-S03-0379-W
39 kDa antigen	AAB47253	(AT)27	2.00E-11	CDS	ES-N-S03-0379-W
39 kDa antigen	AAB47253	(AT)24	6.00E-05	CDS	ES-N-S03-0413-W
39 kDa antigen	AAB47253	(AC)20	6.00E-05	CDS	ES-N-S03-0413-W
acetylcholinesterase	CAD32684	(AAAGG)3	1.00E-06	CDS	HPa-N-S01-0125-LF
acidic ribosomal protein P1	P02402	(AAGAG)3	2.00E-33	CDS	HC-V-S01-0238-LF
actin2	AAP36272	(AG)6	6.00E-19	3'-UTR	HC-N-S01-0165-LF
agCP4447	EAA10576	(AAG)14	1.00E-20	CDS(Pd)	ES-N-S03-0088-W
agCP4709	EAA15143	(ACAT)37	9.00E-17	CDS	ES-N-S02-0096-W
agCP4709	EAA15143	(AC)61	5.00E-23	CDS	ES-N-S03-0025-W
agCP4709	EAA15143	(AT)14	5.00E-23	3'-UTR	ES-N-S03-0025-W
agCP4709	EAA15143	(AC)68	1.00E-19	CDS	ES-N-S03-0073-W
agCP4709	EAA15143	(AC)58	3.00E-21	CDS	LP-Y-S01-0052-LF
agCP5044	EAA10663	(AT)6	2.00E-07	3'-UTR(Pd)	ES-N-S02-0472-W
agCP6523	EAA11105	(AAC)6	5.00E-07	CDS	ES-N-S02-0296-W
ankyrin	AAL92356	(AAT)6	2.00E-30	CDS	ES-N-S03-0002-W
ankyrin	AAL92356	(ATC)9	2.00E-30	CDS(Pd)	ES-N-S03-0002-W
ankyrin	AAL92356	(AAT)16	2.00E-30	CDS	ES-N-S03-0002-W
ankyrin	AAL92356	(AAT)38	2.00E-30	CDS	ES-N-S03-0002-W
ankyrin	AAL92356	(AAGT)3	4.00E-18	3'-UTR(Pd)	ES-N-S03-0068-W
ankyrin	AAL92356	(AAT)34	4.00E-18	CDS	ES-N-S03-0068-W
ATP-dependent RNA helicase	AAM43770	(AAT)35	2.00E-24	CDS	ES-N-S03-0085-W
ATP-dependent RNA helicase	AAM43770	(AAT)9	2.00E-24	CDS	ES-N-S03-0085-W
ATP-dependent RNA helicase	AAM43770	(AAT)9	8.00E-20	CDS	ES-N-S03-0158-W
ATP-dependent RNA helicase	AAM43770	(AAT)33	8.00E-20	CDS	ES-N-S03-0158-W
ATP-dependent RNA helicase	AAM43770	(AAT)24	1.00E-14	CDS	LP-N-S01-0346-LF
atrophin-1 related protein	AAC31120	(AG)21	2.00E-05	CDS	ES-N-S01-0093-W
Bmsqd-2	BAA07211	(AGC)4	6.00E-15	3'-UTR(Pd)	HC-N-S01-0182-LF
brain protein I3	NP_061242	(AG)12	1.00E-15	3'-UTR	HPa-N-N01-0233-LF
bromodomain adjacent to zinc finger domain	NP_038476	(AG)6	8.00E-19	CDS	OV-N-S01-0428-W
butyrate response factor 1	NP_004917	(AGG)10	3.00E-26	5'-UTR(Pd)	OV-N-S01-0209-W
cadherin domain(s)	NP_510685	(AG)37	5.00E-10	CDS	ES-N-S03-0141-W
cadherin domain(s)	NP_510685	(AAT)8	5.00E-10	CDS(Pd)	ES-N-S03-0141-W
calciumbinding protein	NP_071711	(AAAT)3	5.00E-05	3'-UTR(Pd)	HC-N-S01-0322-LF
CDH1-D	AAL31950	(ACGG)5	2.00E-23	CDS(Pd)	ES-N-S03-0508-W
CDNA FLJ12480 FIS	AAM43755	(AAT)30	2.00E-12	CDS	LP-V-S01-0267-LF
cell cycle control protein cwf2	EAA22781	(AC)6	6.00E-06	CDS	HC-N-S01-0179-LF
CG11931-PA	NP_608847	(AAT)22	8.00E-17	3'-UTR	ES-N-S03-0086-W
CG11931-PA	NP_608847	(ACT)9	8.00E-17	CDS	ES-N-S03-0086-W
CG16903-PA	NP_569980	(AAAAT)3	2.00E-06	CDS(Pd)	LP-V-S01-0338-LF
CG1697-PB	NP_525084	(AG)9	2.00E-10	CDS(Pd)	ES-N-S02-0306-W

(continued)

CG17271-PA	NP_650916	(AGC)8	5.00E-23	CDS	HC-H-S01-0009-LF
CG2839-PA	NP_608540	(AAAG)3	4.00E-21	CDS	LP-V-S01-0397-LF
CG2839-PA	NP_608540	(AAAG)3	4.00E-21	CDS	LP-V-S01-0397-LF
CG2839-PA	NP_608540	(AG)6	4.00E-21	CDS	LP-V-S01-0397-LF
CG2839-PA	NP_608540	(AAAG)3	4.00E-21	CDS	LP-V-S01-0397-LF
CG5792 protein	AAM33216	(AAT)35	3.00E-34	CDS	ES-N-S03-0039-W
CG5792 protein	AAM33216	(ATC)6	3.00E-34	CDS(Pd)	ES-N-S03-0039-W
CG5792 protein	AAM33216	(AAT)9	3.00E-34	CDS(Pd)	ES-N-S03-0039-W
CG5792 protein	AAM33216	(AAT)11	3.00E-34	CDS(Pd)	ES-N-S03-0039-W
CG7892-PE	NP_729316	(AG)9	5.00E-60	3'-UTR	LP-N-S01-0363-LF
chaperonin	EAA21693	(AGAGG)4	7.00E-08	3'-UTR(Pd)	LP-N-N01-1114-LF
chaperonin	EAA21693	(AAAT)3	2.00E-06	3'-UTR	LP-Y-S01-0012-LF
coiled-coil protein	AAO51721	(AAT)10	2.00E-08	CDS	HC-N-S01-0081-LF
conserved hypothetical protein	NP_779499	(AG)10	3.00E-05	CDS(Pd)	HC-H-S01-0826-LF
cyclic AMP-regulated protein	BAB85575	(AG)6	7.00E-10	5'-UTR(Pd)	HC-H-S01-0409-LF
cyclophilin-RNA interacting protein	CAC35733	(AGAT)7	1.00E-08	CDS	ES-N-S03-0015-W
cyclophilin-RNA interacting protein	CAC35733	(AT)10	1.00E-08	CDS	ES-N-S03-0015-W
cyclophilin-RNA interacting protein	CAC35733	(AG)37	1.00E-08	CDS(Pd)	ES-N-S03-0015-W
cyclophilin-RNAinteractingprotein	CAC35733	(AAAT)6	1.00E-18	CDS(Pd)	LP-Y-S01-0794-LF
cyclophilin-RNAinteractingprotein	CAC35733	(AG)6	1.00E-18	CDS	LP-Y-S01-0794-LF
cylicin	NP_776728	(ACAG)3	2.00E-10	CDS	HC-N-S01-0030-LF
cysteine rich protein	AAB05810	(AGC)18	8.00E-12	CDS	ES-N-S03-0570-W
cysteine rich protein	EAK95382	(AGC)11	3.00E-07	CDS	HPO-N-S01-0132-LF
cysteine rich protein	BAC34711	(AGC)9	9.00E-12	CDS	LP-N-N01-0021-LF
cytochrome c oxidase precursor	AAA34516	(AAAAG)4	1.00E-07	5'-UTR	ES-N-S03-0450-W
cytochrome P450-3	AAL38986	(AAG)7	4.00E-06	CDS	ES-N-S03-0112-W
cytochrome P450-3	AAL38986	(AG)24	4.00E-06	CDS	ES-N-S03-0112-W
cytochrome P450-3	AAL38986	(AAT)10	9.00E-09	CDS(Pd)	HC-N-S01-0288-LF
cytochrome P450-3	AAL38986	(AAAC)3	9.00E-08	CDS(Pd)	HC-N-S01-0288-LF
cytochrome P450-3	AAL38986	(AG)33	9.00E-10	CDS	HC-N-S01-0288-LF
DNA replication helicase dna2	AAL96717	(AAT)59	4.00E-35	CDS	HPO-N-S01-0036-LF
DNA replication helicase dna2	AAL96717	(AAT)10	4.00E-35	5'-UTR	HPO-N-S01-0036-LF
DNA-binding protein	NP_921613	(AG)17	1.00E-04	CDS	ES-N-S03-0493-W
DNA-binding protein	NP_921613	(AGAT)3	1.00E-04	CDS	ES-N-S03-0493-W
EAA55147	NP_608847	(AAAT)3	4.00E-09	CDS	OV-N-S01-0292-W
EAA55147	NP_608847	(AAT)18	4.00E-09	CDS	OV-N-S01-0292-W
ebiP200	EAA01989	(AAG)7	2.00E-12	CDS	ES-N-S03-0070-W
ebiP200	EAA01989	(AG)46	2.00E-12	CDS	ES-N-S03-0070-W
ebiP4168	EAA01624	(AT)34	3.00E-12	CDS	ES-N-S03-0074-W
ebiP4168	EAA01624	(ACT)35	3.00E-12	CDS	ES-N-S03-0074-W
ebiP4168	EAA01624	(AAGT)5	3.00E-12	CDS	ES-N-S03-0074-W
ebiP4168	EAA01624	(AT)31	1.00E-07	CDS	ES-N-S03-0106-W
ebiP4168	EAA01624	(ACTG)3	5.00E-09	CDS	LP-N-S01-0045-LF
ebiP4168	EAA01624	(AAGT)9	5.00E-09	CDS	LP-N-S01-0045-LF
ebiP4168	EAA01624	(ACT)21	5.00E-09	CDS	LP-N-S01-0045-LF
ebiP4168	EAA01624	(AT)25	5.00E-09	CDS	LP-N-S01-0045-LF
ebiP4168	EAA01624	(ACTC)11	5.00E-09	5'-UTR(Pd)	LP-N-S01-0045-LF

(continued)

ebiP9138	EAA00158	(AC)9	3.00E-42	3'-UTR	HPO-N-S01-0802-LF
ENSANGP00000000338	EAA02026	(AT)23	1.00E-17	CDS	ES-N-S01-0054-W
ENSANGP00000000338	EAA02026	(AT)30	1.00E-17	CDS	ES-N-S01-0054-W
ENSANGP00000000338	EAA02026	(AT)28	3.00E-09	CDS	ES-N-S02-0347-W
ENSANGP00000000338	EAA02026	(AT)6	3.00E-09	CDS	ES-N-S02-0347-W
ENSANGP00000000338	EAA02026	(ACAT)3	1.00E-13	CDS	LP-Y-S01-0054-LF
ENSANGP00000000338	EAA02026	(AT)32	1.00E-13	CDS	LP-Y-S01-0054-LF
ENSANGP00000000338	EAA02026	(AT)34	9.00E-16	CDS	LP-Y-S01-0107-LF
ENSANGP00000000338	EAA02026	(ACAT)3	9.00E-16	CDS	LP-Y-S01-0107-LF
ENSANGP00000004168	XP_322060	(AT)50	2.00E-13	CDS	LP-N-N01-0434-LF
ENSANGP00000004168	XP_322060	(AT)38	2.00E-10	CDS	LP-N-N01-1090-LF
ENSANGP00000008132	XP_311007	(AG)16	1.00E-10	CDS(Pd)	ES-N-S02-0153-W
ENSANGP00000012042	XP_322031	(AT)6	1.00E-30	3'-UTR(Pd)	LP-N-S01-0167-LF
ENSANGP00000014057	XP_393238	(AGG)7	1.00E-37	CDS	HPa-N-N01-1154-LF
ENSANGP00000017604	XP_311858	(AAT)16	5.00E-09	CDS(Pd)	HPO-N-S01-0225-LF
ENSANGP00000020684	XP_318175	(AAAT)5	7.00E-12	CDS(Pd)	HPa-N-N01-1164-LF
ENSANGP00000022728	XP_306301	(AT)6	6.00E-11	CDS	ES-N-S03-0386-W
ENSANGP00000022728	XP_306301	(AAAT)7	6.00E-11	CDS	ES-N-S03-0386-W
ENSANGP00000022728	XP_306301	(AT)22	6.00E-11	CDS	ES-N-S03-0386-W
ENSANGP00000022728	XP_306301	(AT)27	1.00E-20	CDS	HPO-N-S01-0624-LF
ENSANGP00000022728	XP_306301	(AT)33	1.00E-20	CDS	HPO-N-S01-0624-LF
ENSANGP00000022728	XP_306301	(ACT)13	1.00E-20	CDS	HPO-N-S01-0624-LF
ENSANGP00000022728	XP_306301	(ATC)15	1.00E-20	CDS	HPO-N-S01-0624-LF
ENSANGP00000022728	XP_306301	(AT)89	1.00E-23	CDS	LP-N-N01-1201-LF
ENSANGP00000022728	XP_306301	(AT)43	5.00E-13	CDS	LP-Y-S01-0291-LF
ENSANGP00000022785	XP_314418	(ACTAT)3	1.00E-24	CDS	ES-N-S03-0056-W
ENSANGP00000022785	XP_314418	(ACT)4	1.00E-24	CDS	ES-N-S03-0056-W
ENSANGP00000022785	XP_314418	(ACTAT)4	1.00E-24	CDS	ES-N-S03-0056-W
ENSANGP00000022785	XP_314418	(AT)66	1.00E-24	CDS	ES-N-S03-0056-W
ENSANGP00000022785	XP_314418	(AT)26	6.00E-10	CDS	ES-N-S03-0267-W
ENSANGP00000022785	XP_314418	(AT)41	2.00E-08	CDS	ES-N-S03-0366-W
ENSANGP00000022785	XP_314418	(AT)33	2.00E-09	CDS	HPa-N-N01-0309-LF
ENSANGP00000022995	XP_309825	(AT)6	9.00E-18	CDS	HC-H-S01-0055-LF
ENSANGP00000023184	XP_314872	(AAAC)3	1.00E-19	3'-UTR(Pd)	LP-N-N01-1243-LF
ENSANGP00000023184	XP_314872	(AAAC)3	2.00E-29	3'-UTR	LP-N-N01-1266-LF
ENSANGP00000023986	XP_316923	(AC)44	1.00E-13	CDS	HPa-N-N01-0663-LF
ENSANGP00000024044	XP_314419	(AT)30	7.00E-09	CDS	ES-N-S01-0099-W
ENSANGP00000024044	XP_314419	(AT)31	1.00E-07	CDS	ES-N-S02-0381-W
ENSANGP00000024044	XP_314419	(AT)33	5.00E-08	CDS	ES-N-S03-0110-W
ENSANGP00000024044	XP_314419	(AT)19	1.00E-04	CDS	ES-N-S03-0192-W
ENSANGP00000024044	XP_314419	(AT)30	2.00E-08	CDS	ES-N-S03-0367-W
ENSANGP00000024044	XP_314419	(AG)18	2.00E-08	CDS	ES-N-S03-0367-W
ENSANGP00000024044	XP_314419	(AT)43	3.00E-10	CDS	ES-N-S03-0423-W
ENSANGP00000024044	XP_314419	(AT)46	2.00E-17	CDS	ES-N-S03-0502-W
ENSANGP00000024044	XP_314419	(AT)32	2.00E-17	CDS	ES-N-S03-0502-W
ENSANGP00000024044	XP_314419	(AG)6	1.00E-04	5'-UTR(Pd)	ES-N-S03-0537-W
ENSANGP00000024044	XP_314419	(AT)30	1.00E-04	CDS	ES-N-S03-0537-W

(continued)

ENSANGP00000024044	XP_314419	(AT)6	2.00E-06	CDS	HPa-N-N01-1163-LF
ENSANGP00000024044	XP_314419	(ACTC)6	2.00E-05	CDS(Pd)	HPa-N-N01-1163-LF
ENSANGP00000024044	XP_314419	(AG)38	2.00E-07	CDS	HPa-N-N01-1163-LF
ENSANGP00000024044	XP_314419	(AC)12	2.00E-24	CDS	HPO-N-S01-0031-LF
ENSANGP00000024044	XP_314419	(ACTAT)3	2.00E-24	CDS	HPO-N-S01-0031-LF
ENSANGP00000024044	XP_314419	(AT)20	2.00E-24	CDS	HPO-N-S01-0031-LF
ENSANGP00000024044	XP_314419	(AT)52	2.00E-24	CDS	HPO-N-S01-0031-LF
ENSANGP00000024044	XP_314419	(AT)15	7.00E-05	CDS	LP-N-N01-0916-LF
ENSANGP00000024044	XP_314419	(AAT)6	7.00E-05	CDS	LP-N-N01-0916-LF
ENSANGP00000024247	XP_306744	(AT)40	1.00E-13	CDS	ES-N-S03-0424-W
ENSANGP00000024247	XP_306744	(AAGT)3	1.00E-13	5'-UTR(Pd)	ES-N-S03-0424-W
ENSANGP00000024247	XP_306744	(ACTAT)5	1.00E-13	CDS	ES-N-S03-0424-W
ENSANGP00000024316	XP_315306	(AGAT)23	8.00E-10	CDS	ES-N-S03-0270-W
ENSANGP00000024316	XP_315306	(AT)6	8.00E-10	CDS(Pd)	ES-N-S03-0270-W
ENSANGP00000024316	XP_315306	(AC)26	8.00E-10	CDS	ES-N-S03-0270-W
ENSANGP00000024316	XP_315306	(AC)36	5.00E-12	CDS	ES-N-S03-0388-W
ENSANGP00000024316	XP_315306	(AAAG)5	5.00E-12	CDS	ES-N-S03-0388-W
ENSANGP00000024412	XP_311205	(AG)31	3.00E-05	CDS	LP-N-N01-0786-LF
ENSANGP00000024412	AAL38986	(AG)41	4.00E-10	CDS	LP-V-S01-0227-LF
ENSANGP00000024462	XP_322061	(AAT)8	3.00E-10	CDS	HPO-N-S01-0526-LF
ENSANGP00000024462	XP_322061	(ATC)11	3.00E-10	CDS	HPO-N-S01-0526-LF
ENSANGP00000024462	XP_322061	(AT)39	3.00E-10	CDS	HPO-N-S01-0526-LF
ENSANGP00000024462	XP_322061	(AT)28	6.00E-11	CDS	HPO-N-S01-0621-LF
ENSANGP00000024462	XP_322061	(ACAT)8	6.00E-11	CDS	HPO-N-S01-0621-LF
ENSANGP00000025247	XP_306300	(AT)42	3.00E-10	CDS	LP-V-S01-0667-LF
ENSANGP00000025247	XP_306300	(AATG)3	3.00E-10	5'-UTR	LP-V-S01-0667-LF
ENSANGP00000025247	XP_306300	(ATC)4	3.00E-10	CDS	LP-V-S01-0667-LF
F-box family protein	NP_175273	(ATC)16	5.00E-09	CDS	HC-H-S01-0617-LF
fused~toxic gene	CAA67126	(AG)14	2.00E-18	5'-UTR(Pd)	OV-N-S01-0107-W
Gcap1 gene product	AAA68426	(AGAT)13	7.00E-06	CDS	ES-N-S03-0245-W
Gcap1 gene product	AAA68426	(AC)15	7.00E-06	CDS	ES-N-S03-0245-W
Gcap1 gene product	AAA68426	(AT)6	7.00E-06	CDS	ES-N-S03-0245-W
Gcap1geneproduct	AAA68426	(AC)13	6.00E-05	CDS	ES-N-S03-0495-W
Gcap1geneproduct	AAA68426	(AT)11	6.00E-05	CDS	ES-N-S03-0495-W
glutamate dehydrogenase	S42919	(AAG)7	4.00E-51	3'-UTR	LP-Y-S01-0787-LF
glycosyl transferase	AAM45349	(AAT)6	6.00E-08	5'-UTR	HPO-N-S01-0070-LF
G-protein-coupled receptor	AAO50842	(AT)6	5.00E-21	CDS	LP-N-N01-0002-LF
G-protein-coupled receptor	AAO50842	(AAT)18	5.00E-21	CDS	LP-N-N01-0002-LF
heat shock protein 10	AAP06016	(ACAT)3	4.00E-26	3'-UTR	HC-N-S01-0496-LF
heat shock protein 25	NP_509009	(AT)6	7.00E-30	3'-UTR(Pd)	ES-N-S02-0060-W
heat shock protein 25	AAB08736	(AT)6	2.00E-21	3'-UTR	HPO-N-S01-0423-LF
hepatocarcinogenesis-related transcription factor	JC4857	(AGG)6	8.00E-15	CDS(Pd)	OV-N-S01-0434-W
hepatocarcinogenesis-related transcription factor	JC4857	(AG)6	8.00E-15	CDS(Pd)	OV-N-S01-0434-W
homeobox-containing protein	AAO52354	(AAT)27	2.00E-24	CDS	ES-N-S03-0540-W
human xnp gene related protein 1, isoform b	AAN71844	(AAG)7	2.00E-08	CDS	HC-H-S01-0614-LF
hypothetical 51.8 Kd protein	CAC59772	(ACAT)3	1.00E-07	CDS	ES-N-S03-0132-W
hypothetical 51.8 Kd protein	CAC59772	(AC)6	1.00E-07	CDS	ES-N-S03-0132-W

(continued)

hypothetical 51.8 Kd protein	CAC59772	(AT)10	1.00E-07	CDS	ES-N-S03-0132-W
hypothetical 51.8 Kd protein	CAC59772	(ACT)6	1.00E-07	CDS-3'-UTR	ES-N-S03-0132-W
hypothetical protein	XP_207741	(AGGG)3	6.00E-11	CDS	ES-N-S01-0014-W
hypothetical protein	AAL88717	(AG)20	1.00E-34	3'-UTR(Pd)	ES-N-S01-0052-W
hypothetical protein	AAL88717	(AAT)12	1.00E-34	CDS	ES-N-S01-0052-W
hypothetical protein	AAL88717	(AAT)41	1.00E-34	CDS	ES-N-S01-0052-W
hypothetical protein	AAL88717	(ATC)9	1.00E-34	CDS	ES-N-S01-0052-W
hypothetical protein	AAL88717	(AAT)12	1.00E-34	CDS	ES-N-S01-0052-W
hypothetical protein	EAA13112	(AT)51	1.00E-14	CDS	ES-N-S01-0072-W
Hypothetical protein	AAO51254	(AAT)17	9.00E-11	CDS(Pd)	ES-N-S01-0122-W
Hypothetical protein	AAO51254	(AAT)10	2.00E-22	CDS	ES-N-S01-0122-W
Hypothetical protein	AAO51254	(AAT)6	2.00E-22	CDS	ES-N-S01-0122-W
hypothetical protein	T01285	(AG)26	2.00E-12	CDS	ES-N-S01-0144-W
hypothetical protein	EAA13112	(AT)29	9.00E-11	CDS	ES-N-S02-0005-W
hypothetical protein	AAL92981	(AAT)24	7.00E-45	CDS(Pd)	ES-N-S02-0071-W
hypothetical protein	AAL92981	(AAT)34	7.00E-45	CDS(Pd)	ES-N-S02-0071-W
hypothetical protein	AAL92981	(AAT)21	7.00E-45	CDS	ES-N-S02-0071-W
hypothetical protein	XP_326282	(AAAG)9	2.00E-05	CDS	ES-N-S02-0088-W
hypothetical protein	NP_702139	(AAT)11	2.00E-30	CDS(Pd)	ES-N-S02-0102-W
hypothetical protein	NP_042981	(AC)47	7.00E-18	CDS	ES-N-S02-0137-W
hypothetical protein	AAM34352	(AAT)46	1.00E-25	CDS	ES-N-S02-0195-W
hypothetical protein	NP_702139	(AAT)6	3.00E-33	CDS	ES-N-S02-0203-W
hypothetical protein	NP_702139	(AAT)29	3.00E-33	CDS(Pd)	ES-N-S02-0203-W
hypothetical protein	T01285	(AG)50	4.00E-12	CDS	ES-N-S02-0224-W
hypothetical protein	T01285	(AAAGC)4	4.00E-12	CDS	ES-N-S02-0224-W
hypothetical protein	AAM45347	(AAT)17	1.00E-22	CDS	ES-N-S02-0236-W
hypothetical protein	AAM45347	(AAC)9	1.00E-22	CDS	ES-N-S02-0236-W
hypothetical protein	NP_701307	(AT)10	5.00E-06	CDS	ES-N-S02-0254-W
hypothetical protein	NP_704773	(ATC)9	5.00E-05	CDS(Pd)	ES-N-S02-0257-W
hypothetical protein	EAA13112	(AT)42	1.00E-11	CDS	ES-N-S02-0343-W
hypothetical protein	EAA13112	(AC)15	1.00E-11	CDS	ES-N-S02-0343-W
hypothetical protein	AAM33220	(AT)6	4.00E-07	3'-UTR(Pd)	ES-N-S02-0350-W
hypothetical protein	NP_705132	(ATC)13	2.00E-15	CDS	ES-N-S02-0375-W
hypothetical protein	NP_705132	(AAT)7	2.00E-15	CDS(Pd)	ES-N-S02-0375-W
hypothetical protein	NP_704588	(AAT)19	1.00E-17	CDS	ES-N-S03-0018-W
hypothetical protein	NP_704588	(AAT)10	1.00E-17	CDS(Pd)	ES-N-S03-0018-W
hypothetical protein	NP_704588	(AAT)19	5.00E-17	CDS	ES-N-S03-0019-W
hypothetical protein	NP_704588	(AAT)10	5.00E-17	CDS(Pd)	ES-N-S03-0019-W
hypothetical protein	AAL88717	(AAT)41	3.00E-26	CDS	ES-N-S03-0026-W
hypothetical protein	AAL88717	(AAT)9	3.00E-26	CDS	ES-N-S03-0026-W
hypothetical protein	AAM45260	(AAT)35	6.00E-26	CDS	ES-N-S03-0029-W
hypothetical protein	AAM45260	(AAT)9	6.00E-26	CDS	ES-N-S03-0029-W
hypothetical protein	AAM45260	(AAT)11	6.00E-26	5'-UTR	ES-N-S03-0029-W
hypothetical protein	NP_700505	(ATC)13	5.00E-17	CDS	ES-N-S03-0035-W
hypothetical protein	NP_700505	(AAT)9	5.00E-17	CDS(Pd)	ES-N-S03-0035-W
hypothetical protein	AAL92183	(AGGG)3	3.00E-17	CDS(Pd)	ES-N-S03-0040-W
hypothetical protein	AAM44363	(AT)11	6.00E-23	CDS(Pd)	ES-N-S03-0041-W

(continued)

hypothetical protein	AAM44363	(AAT)14	6.00E-23	CDS(Pd)	ES-N-S03-0041-W
hypothetical protein	AAM44363	(AT)14	6.00E-23	CDS(Pd)	ES-N-S03-0041-W
hypothetical protein	AAM44363	(ATC)18	6.00E-23	CDS(Pd)	ES-N-S03-0041-W
hypothetical protein	AAM44363	(ACT)7	6.00E-23	CDS(Pd)	ES-N-S03-0041-W
hypothetical protein	AAM44363	(ATC)72	6.00E-23	CDS	ES-N-S03-0041-W
hypothetical protein	AAM34346	(AGGG)3	2.00E-18	CDS	ES-N-S03-0054-W
hypothetical protein	NP_704154	(AAT)30	7.00E-15	CDS	ES-N-S03-0055-W
hypothetical protein	NP_586311	(AAAG)5	2.00E-09	CDS	ES-N-S03-0075-W
hypothetical protein	AAL88717	(AAT)48	2.00E-25	CDS	ES-N-S03-0082-W
hypothetical protein	AAM44363	(ATC)12	7.00E-25	CDS(Pd)	ES-N-S03-0089-W
hypothetical protein	NP_702139	(AAT)11	5.00E-17	CDS	ES-N-S03-0097-W
hypothetical protein	NP_702139	(AAG)6	5.00E-17	CDS	ES-N-S03-0097-W
hypothetical protein	NP_702139	(AAT)8	7.00E-39	CDS	ES-N-S03-0113-W
hypothetical protein	NP_702139	(AAT)19	7.00E-39	CDS	ES-N-S03-0113-W
hypothetical protein	NP_702139	(AAT)25	7.00E-39	CDS	ES-N-S03-0113-W
hypothetical protein	NP_702139	(AAT)6	7.00E-39	CDS	ES-N-S03-0113-W
hypothetical protein	NP_702139	(AAT)13	7.00E-39	CDS(Pd)	ES-N-S03-0113-W
hypothetical protein	EAA55147	(AAT)32	8.00E-18	CDS	ES-N-S03-0119-W
hypothetical protein	NP_042981	(AAAC)3	2.00E-14	CDS	ES-N-S03-0133-W
hypothetical protein	NP_042981	(AG)19	2.00E-14	CDS(Pd)	ES-N-S03-0133-W
hypothetical protein	NP_705132	(ATC)13	1.00E-13	CDS(Pd)	ES-N-S03-0157-W
hypothetical protein	NP_042981	(AC)36	2.00E-19	CDS	ES-N-S03-0189-W
hypothetical protein	NP_042981	(AC)20	2.00E-19	CDS	ES-N-S03-0189-W
hypothetical protein	NP_042981	(AAAG)3	2.00E-19	CDS(Pd)	ES-N-S03-0189-W
hypothetical protein	NP_704154	(AAT)20	1.00E-15	CDS	ES-N-S03-0190-W
hypothetical protein	NP_704154	(AAC)6	1.00E-15	CDS	ES-N-S03-0190-W
hypothetical protein	NP_704154	(AAT)22	1.00E-15	CDS(Pd)	ES-N-S03-0190-W
hypothetical protein	NP_042981	(AC)39	2.00E-18	CDS	ES-N-S03-0215-W
hypothetical protein	AAM44363	(ATC)13	1.00E-15	CDS	ES-N-S03-0217-W
hypothetical protein	NP_701277	(ATC)13	1.00E-12	CDS	ES-N-S03-0242-W
hypothetical protein	NP_701277	(AAT)9	1.00E-12	CDS(Pd)	ES-N-S03-0242-W
hypothetical protein	NP_042981	(AC)67	2.00E-26	CDS	ES-N-S03-0247-W
hypothetical protein	AAM44363	(ATC)22	3.00E-21	CDS	ES-N-S03-0263-W
hypothetical protein	AAM44363	(AAT)9	3.00E-21	CDS	ES-N-S03-0263-W
hypothetical protein	AAM44363	(AATAG)3	3.00E-21	CDS(Pd)	ES-N-S03-0263-W
hypothetical protein	NP_042981	(AC)16	3.00E-24	CDS	ES-N-S03-0266-W
hypothetical protein	NP_042981	(AC)25	3.00E-24	CDS	ES-N-S03-0266-W
hypothetical protein	NP_042981	(AC)6	3.00E-24	CDS	ES-N-S03-0266-W
hypothetical protein	AAO53175	(AAT)22	8.00E-05	CDS(Pd)	ES-N-S03-0274-W
hypothetical protein	AAO53175	(AAT)10	8.00E-05	CDS	ES-N-S03-0274-W
hypothetical protein	AAO53175	(AATG)3	8.00E-05	CDS(Pd)	ES-N-S03-0274-W
hypothetical protein	NP_700925	(ATC)13	2.00E-13	CDS	ES-N-S03-0301-W
hypothetical protein	NP_700925	(AAT)9	2.00E-13	CDS(Pd)	ES-N-S03-0301-W
hypothetical protein	NP_042981	(AC)19	5.00E-18	CDS	ES-N-S03-0316-W
hypothetical protein	NP_042981	(AC)20	5.00E-18	CDS	ES-N-S03-0316-W
hypothetical protein	NP_042981	(AC)6	5.00E-18	CDS	ES-N-S03-0316-W
hypothetical protein	AAM44363	(ATC)13	9.00E-14	CDS	ES-N-S03-0337-W

(continued)

hypothetical protein	NP_042981	(AC)55	5.00E-17	CDS	ES-N-S03-0357-W
hypothetical protein	NP_042981	(AC)29	1.00E-13	CDS	ES-N-S03-0376-W
hypothetical protein	NP_042981	(AAAC)3	1.00E-13	CDS	ES-N-S03-0376-W
hypothetical protein	AAL88717	(AAT)10	1.00E-33	CDS	ES-N-S03-0392-W
hypothetical protein	AAL88717	(AAT)47	1.00E-33	CDS	ES-N-S03-0392-W
hypothetical protein	AAL88717	(ACT)8	1.00E-33	CDS	ES-N-S03-0392-W
hypothetical protein	NP_704173	(AG)27	2.00E-05	CDS	ES-N-S03-0414-W
hypothetical protein	XP_317980	(AT)12	9.00E-12	CDS	ES-N-S03-0416-W
hypothetical protein	XP_317980	(AT)17	9.00E-12	CDS	ES-N-S03-0416-W
hypothetical protein	XP_317980	(AAAT)7	9.00E-12	CDS	ES-N-S03-0416-W
hypothetical protein	XP_317980	(AC)6	9.00E-12	CDS(Pd)	ES-N-S03-0416-W
hypothetical protein	XP_329242	(AG)24	7.00E-08	CDS	ES-N-S03-0436-W
hypothetical protein	CAE59383	(AG)29	3.00E-08	CDS	ES-N-S03-0442-W
hypothetical protein	NP_042981	(ACAT)12	4.00E-29	CDS	ES-N-S03-0460-W
hypothetical protein	NP_042981	(AC)43	4.00E-29	CDS	ES-N-S03-0460-W
hypothetical protein	NP_042981	(AC)12	4.00E-29	CDS	ES-N-S03-0460-W
hypothetical protein	AAO52193	(AAT)11	1.00E-28	CDS(Pd)	ES-N-S03-0469-W
hypothetical protein	AAO52193	(AAT)51	1.00E-28	CDS	ES-N-S03-0469-W
hypothetical protein	AAO52193	(AAT)6	1.00E-28	CDS(Pd)	ES-N-S03-0469-W
hypothetical protein	NP_704206	(AAT)18	5.00E-06	CDS	ES-N-S03-0480-W
hypothetical protein	CAE59383	(AG)22	6.00E-09	CDS	ES-N-S03-0515-W
hypothetical protein	NP_703806	(AAT)25	4.00E-05	CDS	ES-N-S03-0517-W
hypothetical protein	NP_703806	(AAAG)3	4.00E-05	CDS	ES-N-S03-0517-W
hypothetical protein	NP_704154	(AAT)60	7.00E-19	CDS	ES-N-S03-0519-W
hypothetical protein	NP_704154	(AAT)11	7.00E-19	CDS(Pd)	ES-N-S03-0519-W
hypothetical protein	AAM44363	(AAT)20	9.00E-12	CDS	ES-N-S03-0533-W
hypothetical protein	AAM44363	(AAT)10	9.00E-12	CDS	ES-N-S03-0533-W
hypothetical protein	AAO52193	(AAT)17	7.00E-14	CDS	ES-N-S03-0534-W
hypothetical protein	XP_317980	(AT)32	5.00E-05	CDS	ES-N-S03-0541-W
hypothetical protein	NP_702139	(AAT)16	5.00E-16	CDS(Pd)	ES-N-S03-0544-W
hypothetical protein	AAO51477	(AAT)12	3.00E-14	CDS(Pd)	ES-N-S03-0554-W
hypothetical protein	AAO51477	(AAT)6	3.00E-14	CDS(Pd)	ES-N-S03-0554-W
hypothetical protein	AAO51477	(ATC)6	3.00E-14	CDS	ES-N-S03-0554-W
hypothetical protein	BAA74876	(AG)35	2.00E-06	CDS	ES-N-S03-0565-W
hypothetical protein	NP_700669	(AT)11	2.00E-06	CDS	ES-N-S03-0567-W
hypothetical protein	NP_700669	(AAT)9	2.00E-06	CDS	ES-N-S03-0567-W
hypothetical protein	BAA97098	(AAG)24	8.00E-29	CDS	ES-N-S03-0571-W
hypothetical protein	BAA97098	(AAC)14	8.00E-29	CDS	ES-N-S03-0571-W
hypothetical protein	AAL88717	(AAT)48	2.00E-29	CDS	ES-N-S03-0572-W
hypothetical protein	AAL88717	(AAT)9	2.00E-29	CDS	ES-N-S03-0572-W
hypothetical protein	T31613	(AAAAT)5	7.00E-07	CDS	ES-N-S03-0578-W
hypothetical protein	BAA97098	(AAAAG)6	6.00E-18	CDS	ES-N-S03-0579-W
hypothetical protein	BAA97098	(AAAG)3	6.00E-18	CDS	ES-N-S03-0579-W
hypothetical protein	AAL93605	(AGC)8	4.00E-12	CDS	HC-H-S01-0153-LF
hypothetical protein	AAM09333	(AAT)8	4.00E-23	3'-UTR(Pd)	HC-H-S01-0209-LF
hypothetical protein	AAM09333	(AAT)11	4.00E-23	CDS	HC-H-S01-0209-LF
hypothetical protein	AAL88717	(AAT)16	6.00E-43	CDS(Pd)	HC-H-S01-0265-LF

(continued)

hypothetical protein	AAL88717	(AAT)74	6.00E-43	CDS	HC-H-S01-0265-LF
hypothetical protein	AAM44363	(ATC)49	1.00E-28	CDS	HC-H-S01-0334-LF
hypothetical protein	T31613	(AG)6	6.00E-05	CDS	HC-H-S01-0543-LF
hypothetical protein	AAM44363	(ATC)60	1.00E-32	CDS	HC-H-S01-0628-LF
hypothetical protein	NP_179527	(AG)47	2.00E-06	CDS	HC-H-S01-0639-LF
hypothetical protein	EAK90721	(AAT)6	9.00E-34	CDS	HC-H-S01-0761-LF
hypothetical protein	T01285	(AG)46	3.00E-13	CDS	HC-N-S01-0059-LF
hypothetical protein	AAM44363	(ATC)74	2.00E-42	CDS	HC-N-S01-0132-LF
hypothetical protein	AAM44363	(ATC)4	2.00E-42	CDS	HC-N-S01-0132-LF
hypothetical protein	AAO51254	(AAT)12	1.00E-08	CDS	HC-N-S01-0278-LF
hypothetical protein	AAH55507	(AGG)7	4.00E-05	CDS(Pd)	HC-N-S01-0536-LF
hypothetical protein	AAH55507	(ATC)6	4.00E-05	CDS(Pd)	HC-N-S01-0536-LF
hypothetical protein	NP_701307	(AT)33	2.00E-08	CDS	HC-W-S01-0368-LF
hypothetical protein	NP_179527	(AG)28	9.00E-07	CDS	HC-W-S01-0680-LF
hypothetical protein	NP_179527	(AG)6	9.00E-07	CDS	HC-W-S01-0680-LF
hypothetical protein	BAA97098	(AAAG)3	7.00E-23	CDS	HC-W-S01-0719-LF
hypothetical protein	AAL88717	(AAC)6	9.00E-29	CDS	HPa-N-N01-0049-LF
hypothetical protein	AAL88717	(AAT)6	9.00E-29	CDS	HPa-N-N01-0049-LF
hypothetical protein	AAL88717	(AAT)9	9.00E-29	CDS	HPa-N-N01-0049-LF
hypothetical protein	AAL88717	(AAT)23	9.00E-29	CDS	HPa-N-N01-0049-LF
hypothetical protein	NP_042981	(ACAG)3	8.00E-22	CDS	HPa-N-N01-0313-LF
hypothetical protein	NP_042981	(AC)6	8.00E-22	CDS	HPa-N-N01-0313-LF
hypothetical protein	NP_042981	(AC)16	8.00E-22	CDS	HPa-N-N01-0313-LF
hypothetical protein	NP_042981	(ACAG)3	8.00E-22	CDS	HPa-N-N01-0313-LF
hypothetical protein	CAE59383	(AG)40	2.00E-11	CDS	HPa-N-N01-0573-LF
hypothetical protein	CAE59383	(ACCT)3	2.00E-11	CDS(Pd)	HPa-N-N01-0573-LF
hypothetical protein	CAE59383	(AG)10	2.00E-11	CDS	HPa-N-N01-0573-LF
hypothetical protein	EAA55147	(AAT)29	1.00E-14	CDS	HPa-N-N01-0580-LF
hypothetical protein	NP_703317	(AAT)22	2.00E-10	CDS	HPa-N-N01-0632-LF
hypothetical protein	BAC35524	(AC)6	6.00E-12	CDS	HPa-N-N01-0655-LF
hypothetical protein	BAC35524	(AGAT)3	6.00E-12	CDS	HPa-N-N01-0655-LF
hypothetical protein	EAA55147	(AAT)29	9.00E-14	CDS	HPa-N-N01-0656-LF
hypothetical protein	BAC35524	(AC)6	2.00E-12	CDS	HPa-N-N01-0696-LF
hypothetical protein	BAC35524	(AAAC)4	2.00E-12	CDS(Pd)	HPa-N-N01-0696-LF
hypothetical protein	BAC35524	(AGAT)3	2.00E-12	CDS	HPa-N-N01-0696-LF
hypothetical protein	AAL92597	(AG)40	1.00E-11	CDS	HPa-N-S01-0066-LF
hypothetical protein	AAL92597	(AAG)6	1.00E-11	CDS(Pd)	HPa-N-S01-0066-LF
hypothetical protein	XP_207700	(AG)40	7.00E-11	CDS	HPa-N-S01-0085-LF
hypothetical protein	XP_207700	(AAG)8	7.00E-11	CDS(Pd)	HPa-N-S01-0085-LF
hypothetical protein	XP_502130	(AAAT)3	3.00E-07	CDS	HPa-N-S01-0173-LF
hypothetical protein	AAL88717	(AAT)21	3.00E-21	CDS	HPa-N-S01-0177-LF
hypothetical protein	EAA00488	(AG)6	4.00E-20	3'-UTR	HPa-N-S01-0309-LF
hypothetical protein	BAA74876	(AG)40	2.00E-08	CDS	HPa-N-S01-0374-LF
hypothetical protein	NP_700900	(AG)10	2.00E-16	CDS	HPa-N-S01-0512-LF
hypothetical protein	CAF89508	(AC)22	1.00E-04	CDS	HPO-N-S01-0018-LF
hypothetical protein	EAL01937	(AAT)8	3.00E-07	CDS	HPO-N-S01-0201-LF
hypothetical protein	AAL92597	(AG)12	4.00E-11	CDS(Pd)	HPO-N-S01-0229-LF

(continued)

hypothetical protein	AAL92597	(AG)13	4.00E-11	CDS(Pd)	HPO-N-S01-0229-LF
hypothetical protein	AAL92597	(AG)6	4.00E-11	CDS(Pd)	HPO-N-S01-0229-LF
hypothetical protein	BAB31107	(AG)15	6.00E-05	CDS	HPO-N-S01-0241-LF
hypothetical protein	BAB31107	(AC)6	6.00E-05	CDS	HPO-N-S01-0241-LF
hypothetical protein	NP_042981	(AC)27	9.00E-11	CDS	HPO-N-S01-0336-LF
hypothetical protein	NP_042981	(AG)30	9.00E-11	CDS(Pd)	HPO-N-S01-0336-LF
hypothetical protein	EAA13112	(AT)32	4.00E-05	CDS	HPO-N-S01-0365-LF
hypothetical protein	XP_107847	(AGG)6	1.00E-08	CDS	HPO-N-S01-0384-LF
hypothetical protein	EAK90721	(AAT)54	1.00E-36	CDS	HPO-N-S01-0394-LF
hypothetical protein	BAB31201	(AC)13	5.00E-06	CDS	HPO-N-S01-0425-LF
hypothetical protein	AAO52564	(AAC)43	2.00E-15	CDS	HPO-N-S01-0471-LF
hypothetical protein	AAM33695	(AAT)6	3.00E-10	CDS	HPO-N-S01-0477-LF
hypothetical protein	AAM33695	(AC)18	3.00E-10	CDS(Pd)	HPO-N-S01-0477-LF
hypothetical protein	EAA55147	(AAT)33	8.00E-18	CDS	HPO-N-S01-0496-LF
hypothetical protein	XP_400202	(ATC)9	1.00E-07	CDS	HPO-N-S01-0498-LF
hypothetical protein	XP_502130	(ATC)7	5.00E-13	CDS	HPO-N-S01-0505-LF
hypothetical protein	XP_502130	(AAT)15	5.00E-13	CDS	HPO-N-S01-0505-LF
hypothetical protein	XP_502130	(AAT)7	5.00E-13	CDS	HPO-N-S01-0505-LF
hypothetical protein	BAA97098	(AAAG)4	1.00E-24	CDS	HPO-N-S01-0570-LF
hypothetical protein	BAA97098	(AAAG)8	1.00E-24	CDS	HPO-N-S01-0570-LF
hypothetical protein	BAA97098	(AAAG)3	1.00E-24	CDS	HPO-N-S01-0570-LF
hypothetical protein	BAA97098	(AG)13	1.00E-24	3'-UTR(Pd)	HPO-N-S01-0570-LF
hypothetical protein	EAK90160	(AG)65	2.00E-16	CDS	HPO-N-S01-0592-LF
hypothetical protein	BAA97098	(AAAG)3	4.00E-32	CDS	HPO-N-S01-0627-LF
hypothetical protein	BAA97098	(AAAG)6	4.00E-32	CDS	HPO-N-S01-0627-LF
hypothetical protein	BAA97098	(AAAG)3	4.00E-32	CDS	HPO-N-S01-0627-LF
hypothetical protein	AAM44363	(ATC)11	2.00E-16	CDS(Pd)	HPO-N-S01-0750-LF
hypothetical protein	CAE59383	(AG)27	1.00E-10	CDS	HPO-N-S01-0772-LF
hypothetical protein	CAE59383	(ACAG)6	1.00E-10	CDS	HPO-N-S01-0772-LF
hypothetical protein	EAA55147	(AAT)16	1.00E-05	CDS	HPO-N-S01-0783-LF
hypothetical protein	AAO52546	(AAT)53	9.00E-25	CDS	HPO-N-S01-0812-LF
hypothetical protein	BAA97098	(AAT)7	2.00E-07	CDS	HPO-N-S01-0826-LF
hypothetical protein	AAL87169	(AGG)6	4.00E-05	CDS	HPO-N-S01-0831-LF
hypothetical protein	AAL87169	(ATC)10	4.00E-05	CDS	HPO-N-S01-0831-LF
hypothetical protein	AAL87169	(AAGC)6	4.00E-05	CDS	HPO-N-S01-0831-LF
hypothetical protein	NP_179527	(AAAG)5	5.00E-05	CDS(Pd)	HPO-N-S01-0844-LF
hypothetical protein	NP_042981	(ACAG)3	9.00E-07	CDS	HPO-N-S01-0866-LF
hypothetical protein	AAL92213	(AAT)18	2.00E-13	CDS	HPO-N-S01-0939-LF
hypothetical protein	NP_042981	(AC)9	6.00E-22	CDS	LP-N-N01-0485-LF
hypothetical protein	NP_042981	(AC)14	6.00E-22	CDS	LP-N-N01-0485-LF
hypothetical protein	AAO50754	(AAT)33	5.00E-16	CDS	LP-N-N01-0553-LF
hypothetical protein	XP_317980	(AT)26	5.00E-07	CDS	LP-N-N01-0573-LF
hypothetical protein	NP_297344	(CG)11	2.00E-13	CDS(Pd)	LP-N-N01-1116-LF
hypothetical protein	NP_297344	(AC)15	2.00E-13	CDS	LP-N-N01-1116-LF
hypothetical protein	ZP_00293802	(AC)6	1.00E-13	3'-UTR(Pd)	LP-N-S01-0048-LF
hypothetical protein	ZP_00293802	(AG)6	1.00E-13	3'-UTR(Pd)	LP-N-S01-0048-LF
hypothetical protein	EAA13112	(AT)42	2.00E-11	CDS	LP-N-S01-0242-LF

(continued)

Hypothetical protein	NP_042981	(AC)39	1.00E-14	CDS	LP-N-S01-0326-LF
hypothetical protein	NP_042981	(AC)6	5.00E-23	CDS	LP-V-S01-0148-LF
hypothetical protein	AAS38609	(AG)38	3.00E-20	CDS	LP-V-S01-0459-LF
hypothetical protein	NP_701571	(AT)25	9.00E-05	CDS	LP-Y-S01-0027-LF
hypothetical protein	AAO50882	(AG)29	1.00E-05	CDS	LP-Y-S01-0445-LF
hypothetical protein	BAA74876	(AG)24	1.00E-04	CDS	LP-Y-S01-0529-LF
hypothetical protein	NP_705489	(AAT)15	5.00E-07	CDS	LP-Y-S01-0931-LF
hypothetical protein	XP_207715	(AAAAG)3	1.00E-09	CDS(Pd)	OV-N-S01-0023-W
hypothetical protein	XP_207715	(AAGAG)10	1.00E-09	CDS	OV-N-S01-0023-W
hypothetical protein	XP_148915	(ACAT)20	2.00E-07	CDS	OV-N-S01-0269-W
hypothetical protein	XP_148915	(AATG)3	2.00E-07	CDS	OV-N-S01-0269-W
hypothetical protein At2g19370	T01285	(AG)49	2.00E-12	CDS	ES-N-S02-0120-W
hypothetical protein CaO19.64	EAK90721	(AAT)9	9.00E-34	CDS	HC-H-S01-0761-LF
hypothetical protein CaO19.64	EAK90721	(AAT)11	9.00E-34	CDS	HC-H-S01-0761-LF
hypothetical protein CaO19.64	EAK90721	(AAAAG)3	9.00E-34	CDS	HC-H-S01-0761-LF
hypothetical protein CaO19.64	EAK90160	(AG)42	9.00E-11	CDS	HC-H-S01-0795-LF
hypothetical protein CaO19.64	EAK90721	(AAT)96	1.00E-51	CDS	HPa-N-N01-0482-LF
hypothetical protein CaO19.64	EAK90721	(AAT)24	5.00E-44	CDS(Pd)	LP-V-S01-0553-LF
hypothetical protein CaO19.64	EAK90721	(AAT)13	5.00E-44	CDS	LP-V-S01-0553-LF
Hypothetical protein CBG02739	CAE59383	(AG)6	3.00E-29	CDS	LP-Y-S01-0105-LF
Hypothetical protein CBG02739	CAE59383	(ACAG)7	3.00E-29	CDS	LP-Y-S01-0105-LF
Hypothetical protein CBG02739	CAE59383	(ACT)9	3.00E-29	CDS	LP-Y-S01-0105-LF
Hypothetical protein CBG02739	CAE59383	(AG)12	3.00E-29	CDS	LP-Y-S01-0105-LF
Hypothetical protein CBG02739	CAE59383	(AG)65	3.00E-29	CDS	LP-Y-S01-0105-LF
hypothetical protein CBG03194	CAE59749	(AG)38	6.00E-06	CDS	HPa-N-N01-0880-LF
hypothetical protein CBG09105	CAE64412	(AAG)7	1.00E-05	CDS	ES-N-S03-0430-W
hypothetical protein CBG14637	CAE68712	(AG)33	4.00E-09	CDS	ES-N-S03-0243-W
hypothetical protein CBG14637	CAE68712	(AG)34	1.00E-07	CDS	HC-N-S01-0357-LF
hypothetical protein Chut02000723	ZP_00311107	(AT)6	2.00E-21	3'-UTR	HC-H-S01-0715-LF
hypothetical protein F44G3.3	T22200	(AAAT)3	6.00E-10	CDS(Pd)	ES-N-S02-0460-W
hypothetical protein F44G3.3	T22200	(AAAT)19	6.00E-10	CDS(Pd)	ES-N-S02-0460-W
hypothetical protein MG06804.4	EAA55147	(AAT)16	1.00E-12	CDS(Pd)	ES-N-S02-0216-W
hypothetical protein MG06804.4	EAA55147	(ACT)9	8.00E-10	3'-UTR(Pd)	ES-N-S02-0443-W
hypothetical protein MG06804.4	EAA55147	(ATC)10	8.00E-10	3'-UTR(Pd)	ES-N-S02-0443-W
hypothetical protein MG06804.4	EAA55147	(ATC)14	8.00E-10	3'-UTR(Pd)	ES-N-S02-0443-W
hypothetical protein MG06804.4	EAA55147	(ACT)68	8.00E-10	CDS	ES-N-S02-0443-W
hypothetical protein MG06804.4	EAA55147	(AT)28	8.00E-10	CDS(Pd)	ES-N-S02-0443-W
hypothetical protein Tfus02001166	ZP_00293802	(AAG)27	4.00E-06	CDS	ES-N-S03-0536-W
hypothetical protein Tfus02001166	ZP_00293802	(AAAAG)3	2.00E-05	CDS	LP-Y-S01-0275-LF
hypothetical protein with Thr stretches	EAK88124	(ACT)9	3.00E-11	CDS	HPO-N-S01-0442-LF
KIAA0853 protein	NP_055885	(AG)31	4.00E-12	CDS	ES-N-S03-0535-W
KIAA0853 protein	NP_055885	(AG)18	4.00E-12	CDS	ES-N-S03-0535-W
krueppel-like protein	CAC42230	(AT)29	4.00E-05	CDS	HC-H-S01-0566-LF
Ku autoantigen	AAD49720	(ACC)4	4.00E-13	CDS	LP-N-N01-0976-LF
latency associated transcript (LAT)	NP_297344	(AC)6	5.00E-13	CDS	HPa-N-N01-0408-LF
LD30931p	AAR96176	(AAG)4	7.00E-66	CDS	635-1
liver regeneration-related protein	AAP92532	(AGG)7	3.00E-05	CDS	ES-N-S03-0218-W

(continued)

liver regeneration-related protein	AAP92535	(AG)11	3.00E-07	CDS	ES-N-S03-0239-W
liver regeneration-related protein	AAP92535	(AG)6	3.00E-07	CDS	ES-N-S03-0239-W
liver regeneration-related protein	AAP92532	(AGG)9	1.00E-10	CDS	HC-N-S01-0475-LF
LOC126147protein	AAH21210	(AG)9	1.00E-04	CDS(Pd)	ES-N-S03-0188-W
LRRGT00012	AAQ96225	(AAT)4	2.00E-09	CDS(Pd)	ES-N-S03-0237-W
LRRGT00165	AAS66256	(AG)27	1.00E-06	CDS	ES-N-S03-0427-W
LRRGT00165	AAS66256	(AT)17	7.00E-10	CDS(Pd)	HPO-N-S01-0794-LF
LRRGT00165	AAS66256	(AG)20	7.00E-10	CDS	HPO-N-S01-0794-LF
mariner transposase	AAC52011	(AC)6	2.00E-13	3'-UTR(Pd)	HPO-N-S01-0198-LF
mariner transposase	AAC52011	(AAAT)15	2.00E-13	CDS	HPO-N-S01-0198-LF
mariner transposase	AAL92597	(AT)20	4.00E-11	CDS(Pd)	HPO-N-S01-0229-LF
membrane protein	BAB78478	(ACAG)3	4.00E-08	CDS	HPO-N-S01-0643-LF
metal transporter	NP_249988	(ATC)27	9.00E-08	CDS	ES-N-S03-0206-W
mucin-like glycoprotein 900	EAK89192	(ACT)36	3.00E-23	CDS	ES-N-S03-0585-W
mucin-like glycoprotein 900	EAK89192	(ACT)6	3.00E-23	CDS	ES-N-S03-0585-W
NADH dehydrogenase subunit 4	NP_038296	(AAAG)3	2.00E-36	CDS	LP-N-N01-0693-LF
NADH dehydrogenase subunit 4	NP_038296	(AAAAT)3	2.00E-36	CDS	LP-N-N01-0693-LF
NADH dehydrogenase subunit 4L	NP_038297	(AAAG)3	5.00E-29	3'-UTR(Pd)	LP-N-N01-0591-LF
nuclear autoantigenic sperm protein; NASP	XP_216505	(AAAG)3	2.00E-11	CDS(Pd)	OV-N-S01-1127-W
nucleolar phosphoprotein	I51618	(AT)6	4.00E-14	5'-UTR	OV-N-S01-0141-W
OG1426	ZP_00055454	(AAAT)3	9.00E-05	CDS(Pd)	HPa-N-N01-0456-LF
ORF	BAA00447	(ACAG)3	1.00E-11	CDS(Pd)	ES-N-S03-0186-W
ORF	BAA00447	(AGGC)3	1.00E-11	CDS	ES-N-S03-0186-W
ORF	BAA00447	(AG)10	1.00E-11	CDS(Pd)	ES-N-S03-0186-W
ORF	XP_134237	(ACAG)3	4.00E-06	CDS	HPO-N-S01-0008-LF
ORF	XP_134237	(AGAT)6	4.00E-06	CDS	HPO-N-S01-0008-LF
ORF	XP_134237	(AG)19	4.00E-06	CDS	HPO-N-S01-0008-LF
ORF	BAA00447	(ACAG)3	3.00E-10	CDS	LP-Y-S01-0969-LF
ORF	BAA00447	(AGAT)3	3.00E-10	CDS	LP-Y-S01-0969-LF
ORF	BAA00447	(AAAG)4	3.00E-10	5'-UTR(Pd)	LP-Y-S01-0969-LF
ORF	BAA00447	(AAAT)6	3.00E-10	5'-UTR(Pd)	LP-Y-S01-0969-LF
ORF	BAA00447	(AATAT)3	3.00E-10	5'-UTR(Pd)	LP-Y-S01-0969-LF
ORF	BAA00447	(ACAG)3	3.00E-10	CDS	LP-Y-S01-0969-LF
peroxisomal targeting signal type 2 receptor	NP_17274	(ATC)7	7.00E-05	CDS	LP-N-S01-0384-LF
phosphatidy linositol 3-kinase 3	AAO52301	(AAT)21	3.00E-09	CDS	HPa-N-N01-0250-LF
PMAV	AAQ75589	(AT)31	7.00E-09	3'-UTR(Pd)	HPa-N-N01-1159-LF
polyprotein precursor	AAK84670	(AAAAT)4	9.00E-22	3'-UTR	LP-V-S01-0004-LF
polyprotein precursor	AAK84670	(AAAAT)4	1.00E-85	3'-UTR	LP-V-S01-0575-LF
pre-mRNA splicing SR protein	NP_492875	(AC)6	3.00E-13	CDS	OV-N-S01-0636-W
prespore-specific protein	AAD16881	(AAT)16	2.00E-15	CDS	ES-N-S02-0124-W
prespore-specific protein	AAD16881	(AAT)11	8.00E-16	CDS	ES-N-S03-0385-W
prespore-specific protein	AAD16881	(AAT)15	8.00E-16	CDS	ES-N-S03-0385-W
prespore-specific protein	AAD16881	(AAT)11	8.00E-16	CDS	ES-N-S03-0385-W
prosaposin	BAA95677	(AAAAG)3	2.00E-10	3'-UTR(Pd)	HC-H-S01-0460-LF
prosaposin	NP_776586	(AAAAG)3	1.00E-12	3'-UTR	LP-N-N01-0903-LF
prosaposin	BAA95677	(AAAAG)3	2.00E-14	3'-UTR	LP-N-S01-0124-LF
proteasome alpha type 6	AAH61438	(AAAG)3	6.00E-28	3'-UTR	HC-N-S01-0440-LF

(continued)

protein kinase	NP_472959	(AACG)14	2.00E-05	CDS	HC-H-S01-0656-LF
protein kinase	NP_472959	(AAG)14	2.00E-06	CDS(Pd)	HC-H-S01-0656-LF
protein kinase	CAD98686	(AC)17	1.00E-12	CDS	HPa-N-N01-0134-LF
protein kinase	CAD98686	(AG)51	1.00E-12	CDS	HPa-N-N01-0134-LF
protein translation factor	P42678	(AC)6	2.00E-50	3'-UTR	LP-N-N01-0181-LF
proteinase inhibitor	S45677	(AAAT)3	2.00E-08	3'-UTR(Pd)	HC-H-S01-0324-LF
putative nuclear protein	T01285	(AG)40	4.00E-09	CDS(Pd)	HPO-N-S01-0697-LF
putative protein	NP_510538	(AG)34	3.00E-09	CDS	HPO-N-S01-0667-LF
putative transport protein	NP_706726	(AC)6	1.00E-17	5'-UTR(Pd)	LP-Y-S01-0395-LF
putative transport protein	NP_706726	(AC)9	1.00E-17	5'-UTR(Pd)	LP-Y-S01-0395-LF
pva1	CAA63219	(AATG)4	2.00E-14	CDS	LP-Y-S01-0473-LF
reverse transcriptase	XP_358982	(AGG)9	4.00E-08	CDS	HC-N-S01-0503-LF
Rho	AAT46562	(AAG)4	1.00E-103	CDS	HC-H-S01-0968-LF
rhodopsin	P35356	(AAGGC)3	6.00E-11	CDS	LP-Y-S01-0001-LF
rhoGAP protein	NP_700545	(AAAT)5	7.00E-05	CDS	ES-N-S02-0222-W
rhoGAP protein	NP_700545	(AAT)11	7.00E-05	CDS	ES-N-S02-0222-W
rhoGAP protein	NP_700545	(AAAT)3	7.00E-05	CDS	ES-N-S02-0222-W
ribosomal protein S2	XP_204804	(AG)6	3.00E-11	CDS	HPO-N-S01-0657-LF
ribosomal protein S27	P55833	(AAC)5	2.00E-19	3'-UTR	HPa-N-N01-0709-LF
ribosomal protein S28	AAP21778	(AAAC)3	4.00E-20	3'-UTR	LP-N-N01-1036-LF
RIKEN cDNA 1110001M19	XP_110931	(ATC)8	3.00E-08	CDS(Pd)	ES-N-S02-0157-W
RIKEN cDNA 2010003O14	NP_079838	(AATG)3	6.00E-39	3'-UTR(Pd)	LP-N-S01-0425-LF
RIKEN cDNA 2310003C23	XP_130746	(AT)6	2.00E-11	3'-UTR	HC-H-S01-0175-LF
RIKEN cDNA 2310003C23	XP_130746	(AG)23	2.00E-11	3'-UTR	HC-H-S01-0175-LF
RIKEN cDNA 4933408N05	XP_314418	(AC)42	3.00E-30	CDS-3'-UTR	ES-N-S01-0030-W
RIKEN cDNA 4933408N05	XP_314418	(ATATC)3	3.00E-30	CDS	ES-N-S01-0030-W
RIKEN cDNA 4933408N05	XP_314418	(AT)60	3.00E-30	CDS	ES-N-S01-0030-W
RIKEN cDNA 4933408N05	XP_314418	(AT)9	3.00E-30	CDS	ES-N-S01-0030-W
RIKEN cDNA 4933408N05	XP_314418	(ACT)16	3.00E-30	CDS	ES-N-S01-0030-W
RIKEN cDNA 4933408N05	XP_314418	(AT)46	3.00E-30	CDS	ES-N-S01-0030-W
RIKEN cDNA 4933408N05	XP_150119	(AT)43	1.00E-11	CDS	ES-N-S02-0063-W
RIKEN cDNA 4933408N05	XP_150119	(AC)24	1.00E-11	CDS	ES-N-S02-0063-W
RIKEN cDNA 4933408N05	XP_150119	(AT)6	1.00E-11	CDS(Pd)	ES-N-S02-0063-W
RIKEN cDNA 4933408N05	XP_150119	(AT)27	8.00E-11	CDS	ES-N-S03-0094-W
RIKEN cDNA 4933408N05	XP_150119	(AC)10	8.00E-11	CDS	ES-N-S03-0094-W
RIKEN cDNA 4933408N05	XP_150119	(AC)18	4.00E-12	CDS	ES-N-S03-0128-W
RIKEN cDNA 4933408N05	XP_150119	(AT)22	4.00E-12	CDS	ES-N-S03-0128-W
RIKEN cDNA 4933408N05	XP_150119	(AAAT)8	4.00E-12	CDS	ES-N-S03-0128-W
RIKEN cDNA 4933408N05	XP_150119	(AAT)6	4.00E-12	5'-UTR	ES-N-S03-0128-W
RIKEN cDNA 4933408N05	XP_150119	(AT)6	3.00E-09	CDS	HPa-N-N01-0599-LF
RIKEN cDNA 4933408N05	XP_150119	(AT)21	3.00E-10	CDS	HPa-N-N01-0599-LF
RIKEN cDNA 5730583A19	XP_148950	(AG)39	7.00E-07	CDS	ES-N-S02-0260-W
RIKEN cDNA 5730583A19	XP_148950	(AC)12	7.00E-07	CDS	ES-N-S02-0260-W
RIKEN cDNA 9330155M09gene	NP_796228	(AAAT)5	2.00E-07	CDS	LP-Y-S01-0724-LF
RIKEN cDNA 9330155M09gene	NP_796228	(AGAT)3	2.00E-07	CDS	LP-Y-S01-0724-LF
RIKEN cDNA C030013D06	XP_147855	(ACAT)3	1.00E-08	CDS	ES-N-S02-0046-W
RIKEN cDNA C030013D06	XP_147855	(ATC)6	1.00E-08	CDS	ES-N-S02-0046-W

(continued)

RIKEN cDNA C030013D06	XP_147855	(AC)15	1.00E-08	CDS	ES-N-S02-0046-W
RIKEN cDNA C030013D06	XP_147855	(AT)17	1.00E-08	CDS	ES-N-S02-0046-W
RIKEN cDNA C030013D06	XP_147855	(AT)42	5.00E-11	CDS	ES-N-S02-0133-W
RIKEN cDNA C030013D06	XP_147855	(AT)6	5.00E-11	CDS	ES-N-S02-0133-W
RIKEN cDNA C030013D06	XP_147855	(ACGCG)3	6.00E-08	5'-UTR	ES-N-S03-0034-W
RIKEN cDNA C030013D06	XP_147855	(AACT)6	6.00E-08	CDS	ES-N-S03-0034-W
RIKEN cDNA C030013D06	XP_147855	(AT)29	6.00E-08	CDS	ES-N-S03-0034-W
RIKEN cDNA C030013D06	XP_147855	(AC)9	4.00E-21	CDS(Pd)	ES-N-S03-0136-W
RIKEN cDNA C030013D06	XP_147855	(ACAT)3	4.00E-21	CDS	ES-N-S03-0136-W
RIKEN cDNA C030013D06	XP_147855	(AT)32	9.00E-12	CDS	ES-N-S03-0333-W
RIKEN cDNA C030013D06	XP_147855	(AC)25	9.00E-12	CDS	ES-N-S03-0333-W
RIKEN cDNA C030013D06	XP_147855	(AT)15	6.00E-09	CDS	ES-N-S03-0462-W
RIKEN cDNA C030013D06	XP_147855	(AT)9	6.00E-09	CDS	ES-N-S03-0462-W
RIKEN cDNA C030013D06	XP_147855	(AT)52	5.00E-14	CDS	HC-H-S01-0589-LF
RIKEN cDNA C030013D06	XP_147855	(AT)6	9.00E-17	CDS	LP-N-N01-0350-LF
RIKEN cDNA C030013D06	XP_147855	(AT)27	9.00E-17	CDS	LP-N-N01-0350-LF
RIKEN cDNA C030013D06	XP_147855	(AATAT)3	9.00E-06	CDS	LP-V-S01-0516-LF
RIKEN cDNA C030013D06	XP_147855	(AT)22	2.00E-06	CDS	LP-V-S01-0532-LF
RIKEN cDNA C030013D06	XP_147855	(AT)39	2.00E-09	CDS	LP-Y-S01-0108-LF
RNA helicase	NP_700682	(AAT)13	2.00E-05	CDS	ES-N-S03-0176-W
RNA-binding protein 6	AAP42142	(AAAG)5	2.00E-08	5'-UTR(Pd)	HPa-N-S01-0466-LF
RNA-binding protein 6	AAP42142	(ATC)18	2.00E-08	CDS	HPa-N-S01-0466-LF
serine/threonine protein kinase	AAL92711	(AAT)23	1.00E-17	CDS	ES-N-S02-0217-W
serine/threonine protein kinase	AAL92711	(AAT)8	1.00E-17	CDS(Pd)	ES-N-S02-0217-W
serine/threonine protein kinase	AAL92711	(AAT)9	1.00E-17	CDS(Pd)	ES-N-S02-0217-W
similar to Ac1873	XP_342268	(AAAAT)3	9.00E-09	3'-UTR(Pd)	LP-Y-S01-0299-LF
similar to Dictyostelium discoideum	AAL92253	(AAT)9	1.00E-06	CDS	LP-N-N01-1030-LF
similar to Dictyostelium discoideum	AAL92253	(AAT)6	1.00E-06	CDS	LP-N-N01-1030-LF
similar to RIKEN cDNA 1810054O13	XP_393044	(AAG)6	4.00E-10	CDS	HPa-N-N01-0153-LF
similar to RIKEN cDNA 1810054O13	XP_393044	(AG)11	4.00E-10	CDS	HPa-N-N01-0153-LF
small nuclear ribonucleoprotein auxiliary factor	AAD28792	(AAAG)7	7.00E-22	CDS(Pd)	LP-V-S01-0564-LF
survival motor neuron protein	O02771	(AAG)8	4.00E-15	CDS	OV-N-S01-0645-W
techylectin-5B	BAA84189	(AAAAT)3	7.00E-18	3'-UTR	LP-V-S01-0587-LF
testis-determining factor	AAO50779	(ATC)11	2.00E-10	CDS	ES-N-S03-0290-W
testis-determining factor	AAO50779	(AAT)35	2.00E-10	CDS	ES-N-S03-0290-W
testis-determining factor	AAO50779	(AATT)7	2.00E-10	CDS(Pd)	ES-N-S03-0290-W
testis-determining factor	AAO50779	(AC)10	4.00E-06	CDS(Pd)	ES-N-S03-0477-W
TIP120 protein; KIAA0829 protein	NP_060918	(AG)6	7.00E-21	CDS(Pd)	OV-N-S01-0176-W
TIP120 protein; KIAA0829 protein	NP_060918	(AG)6	1.00E-12	3'-UTR(Pd)	OV-N-S01-0178-W
transcription factor BTF 3a	JC1235	(ATC)4	2.00E-52	CDS	LP-N-N01-0576-LF
tumor suppressor pHyde	AAM45136	(AC)10	7.00E-27	3'-UTR(Pd)	HPa-N-N01-0002-LF
tyrosine kinase	AAO50855	(ATC)25	6.00E-27	CDS(Pd)	ES-N-S03-0582-W
tyrosine kinase	AAO50855	(AAT)37	6.00E-27	CDS	ES-N-S03-0582-W
U1 small ribonucleoprotein 1SNRP	AAF19255	(AG)44	2.00E-18	CDS	HPa-N-S01-0632-LF
ubiquitin specific protease 9, X chromosome	NP_033507	(AGC)13	8.00E-71	CDS	OV-N-S01-0352-W
ubiquitin-conjugating enzyme	AAP36617	(ACC)6	7.00E-06	3'-UTR	LP-N-N01-0726-LF
ubiquitin-conjugating enzyme	AAP36617	(AGCC)3	7.00E-06	3'-UTR	LP-N-N01-0726-LF

(continued)

ubiquitin-conjugating enzyme E2	NP_076074	(AGCC)3	3.00E-16	3'-UTR(Pd)	HC-H-S01-0296-LF
ubiquitin-conjugating enzyme E2	NP_076074	(ACC)6	3.00E-16	3'-UTR(Pd)	HC-H-S01-0296-LF
unnamed protein product	CAG03719	(AT)9	4.00E-14	CDS	ES-N-S01-0042-W
unnamed protein product	CAG03719	(AC)42	4.00E-14	CDS	ES-N-S01-0042-W
unnamed protein product	CAG03719	(AT)41	4.00E-14	3'-UTR	ES-N-S01-0042-W
unnamed protein product	BAC86958	(AT)27	7.00E-13	CDS	ES-N-S02-0138-W
unnamed protein product	BAC86958	(AAAT)5	7.00E-13	CDS	ES-N-S02-0138-W
unnamed protein product	BAC86958	(AACCT)4	7.00E-13	CDS(Pd)	ES-N-S02-0138-W
unnamed protein product	CAF96706	(AC)37	3.00E-14	CDS	ES-N-S02-0141-W
unnamed protein product	CAF96706	(AGGC)14	3.00E-14	CDS	ES-N-S02-0141-W
unnamed protein product	CAG13317	(AGAT)14	1.00E-14	CDS	ES-N-S02-0221-W
unnamed protein product	CAG13317	(AG)31	1.00E-14	CDS	ES-N-S02-0221-W
unnamed protein product	CAG13317	(ACAG)3	1.00E-14	CDS	ES-N-S02-0221-W
unnamed protein product	CAG13317	(AGAT)7	1.00E-14	CDS	ES-N-S02-0221-W
unnamed protein product	BAC04594	(ACT)25	4.00E-07	3'-UTR(Pd)	ES-N-S02-0447-W
unnamed protein product	BAC04594	(AC)20	4.00E-07	CDS	ES-N-S02-0447-W
unnamed protein product	BAC04594	(AATT)3	4.00E-07	3'-UTR(Pd)	ES-N-S02-0447-W
unnamed protein product	BAC04594	(AT)15	4.00E-07	CDS	ES-N-S02-0447-W
unnamed protein product	BAC04594	(AAAT)11	4.00E-08	3'-UTR(Pd)	ES-N-S03-0163-W
unnamed protein product	BAC04594	(AC)27	4.00E-08	CDS	ES-N-S03-0163-W
unnamed protein product	BAC87611	(AC)38	5.00E-14	CDS	ES-N-S03-0201-W
unnamed protein product	BAC30584	(AC)6	2.00E-19	CDS	ES-N-S03-0296-W
unnamed protein product	BAC30584	(ACTC)9	2.00E-19	CDS	ES-N-S03-0296-W
unnamed protein product	BAC30584	(ATGC)3	2.00E-19	CDS	ES-N-S03-0296-W
unnamed protein product	BAC85241	(ACGG)7	1.00E-05	3'-UTR(Pd)	ES-N-S03-0345-W
unnamed protein product	BAC36719	(AC)11	1.00E-07	CDS	ES-N-S03-0451-W
unnamed protein product	BAC36719	(AC)11	1.00E-07	CDS	ES-N-S03-0451-W
unnamed protein product	BAC36719	(ACAT)3	1.00E-07	CDS	ES-N-S03-0451-W
unnamed protein product	CAF88611	(ACC)13	6.00E-15	CDS	ES-N-S03-0472-W
unnamed protein product	CAF88611	(ATC)16	6.00E-15	CDS	ES-N-S03-0472-W
unnamed protein product	CAF88611	(AT)10	6.00E-15	CDS	ES-N-S03-0472-W
unnamed protein product	CAF88611	(ACT)6	6.00E-15	CDS	ES-N-S03-0472-W
unnamed protein product	CAF88611	(AC)18	6.00E-15	CDS(Pd)	ES-N-S03-0472-W
unnamed protein product	CAG14665	(AC)36	3.00E-09	CDS	ES-N-S03-0497-W
unnamed protein product	BAC35524	(AC)50	4.00E-14	CDS	ES-N-S03-0556-W
unnamed protein product	BAC35524	(AAT)11	4.00E-14	CDS(Pd)	ES-N-S03-0556-W
unnamed protein product	BAC87052	(AAAG)6	4.00E-11	CDS	HC-H-S01-0266-LF
unnamed protein product	CAF99231	(ATC)6	5.00E-20	CDS	HC-H-S01-0499-LF
unnamed protein product	CAF98009	(AACT)6	9.00E-31	3'-UTR	HPa-N-N01-0430-LF
unnamed protein product	CAF98009	(AT)24	9.00E-31	3'-UTR	HPa-N-N01-0430-LF
unnamed protein product	CAG13317	(AGAT)3	1.00E-06	CDS	HPO-N-S01-0012-LF
unnamed protein product	CAG13317	(AGAT)3	4.00E-09	CDS	HPO-N-S01-0072-LF
unnamed protein product	CAG13317	(AGAT)8	4.00E-09	CDS	HPO-N-S01-0072-LF
unnamed protein product	AAH01584	(AG)19	7.00E-05	CDS	HPO-N-S01-0089-LF
unnamed protein product	EAA15143	(AC)61	3.00E-22	CDS	HPO-N-S01-0103-LF
unnamed protein product	CAG13317	(AGAT)26	1.00E-15	CDS	HPO-N-S01-0177-LF
unnamed protein product	CAG13317	(AT)9	1.00E-15	CDS	HPO-N-S01-0177-LF

(continued)

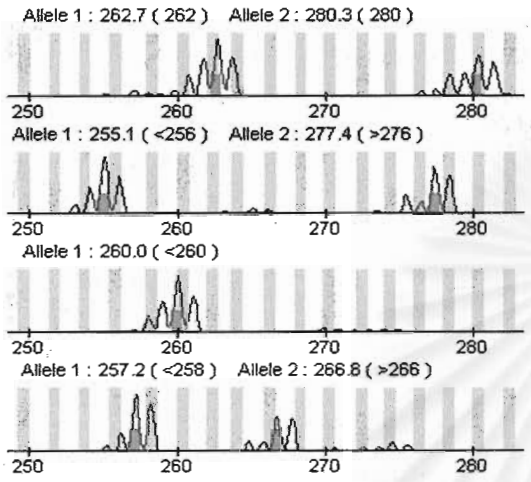
unnamed protein product	BAC30584	(AT)16	1.00E-08	CDS	HPO-N-S01-0194-LF
unnamed protein product	BAC30584	(ACAT)5	1.00E-08	CDS	HPO-N-S01-0194-LF
unnamed protein product	BAC35633	(AC)13	8.00E-05	CDS	HPO-N-S01-0511-LF
unnamed protein product	CAF88611	(ATC)24	2.00E-15	CDS	HPO-N-S01-0632-LF
unnamed protein product	BAC31392	(AG)22	4.00E-08	CDS	HPO-N-S01-0713-LF
unnamed protein product	BAC31392	(AC)13	4.00E-08	CDS	HPO-N-S01-0713-LF
unnamed protein product	CAG13317	(AGAT)23	2.00E-12	CDS	HPO-N-S01-0817-LF
unnamed protein product	CAG13317	(AAGAG)3	2.00E-12	CDS	HPO-N-S01-0817-LF
unnamed protein product	BAC35524	(AC)6	8.00E-10	CDS	HPO-N-S01-0818-LF
unnamed protein product	BAC35524	(AC)12	8.00E-10	CDS	HPO-N-S01-0818-LF
unnamed protein product	BAC35524	(ACACT)3	8.00E-10	CDS	HPO-N-S01-0818-LF
unnamed protein product	CAG03719	(AC)31	2.00E-09	CDS	HPO-N-S01-0837-LF
unnamed protein product	CAG13317	(AGAT)12	2.00E-08	CDS	HPO-N-S01-0879-LF
unnamed protein product	CAG13317	(AT)20	2.00E-08	CDS	HPO-N-S01-0879-LF
unnamed protein product	CAG13317	(AC)12	2.00E-08	CDS	HPO-N-S01-0879-LF
unnamed protein product	CAG06679	(AG)31	3.00E-09	CDS	HPO-N-S01-0892-LF
unnamed protein product	CAG13317	(AGAT)7	1.00E-13	CDS	HPO-N-S01-0926-LF
unnamed protein product	CAG13317	(AGAT)13	1.00E-13	CDS	HPO-N-S01-0926-LF
unnamed protein product	CAG13317	(AGAT)3	1.00E-13	CDS	HPO-N-S01-0926-LF
unnamed protein product	CAG13317	(ACAG)3	1.00E-13	CDS(Pd)	HPO-N-S01-0926-LF
unnamed protein product	CAF88611	(ATC)24	9.00E-20	CDS	HPO-N-S01-0991-LF
unnamed protein product	BAC04906	(AC)37	3.00E-13	CDS	LP-N-N01-0088-LF
unnamed protein product	BAC04906	(AC)14	3.00E-13	CDS	LP-N-N01-0088-LF
unnamed protein product	BAC85241	(AT)24	1.00E-10	CDS	LP-N-N01-0284-LF
unnamed protein product	BAC85241	(ACAT)12	1.00E-10	CDS	LP-N-N01-0284-LF
unnamed protein product	BAC31550	(AGAGG)7	4.00E-08	CDS	LP-V-S01-0295-LF
unnamed protein product	BAC31550	(ACGAG)4	4.00E-08	CDS	LP-V-S01-0295-LF
unnamed protein product	BAC31550	(AGAGG)3	4.00E-08	CDS	LP-V-S01-0295-LF
unnamed protein product	CAG07569	(AAGG)3	4.00E-10	CDS(Pd)	LP-Y-S01-0005-LF
unnamed protein product	CAG13317	(AGAT)12	8.00E-07	CDS	LP-Y-S01-0154-LF
unnamed protein product	CAG13317	(AC)16	8.00E-07	CDS	LP-Y-S01-0154-LF
unnamed protein product	CAG13317	(AG)24	8.00E-07	CDS	LP-Y-S01-0154-LF
unnamed protein product	CAF93717	(AC)47	7.00E-13	CDS	LP-Y-S01-0178-LF
unnamed protein product	CAG13317	(AC)16	6.00E-07	CDS	LP-Y-S01-0182-LF
unnamed protein product	CAG13317	(AG)24	6.00E-07	CDS	LP-Y-S01-0182-LF
unnamed protein product	CAG13317	(AGAT)12	6.00E-07	CDS	LP-Y-S01-0182-LF
unnamed protein product	BAC31047	(AC)24	5.00E-06	CDS	LP-Y-S01-0295-LF
unnamed protein product	BAC31047	(AC)24	6.00E-06	CDS	LP-Y-S01-0499-LF
unnamed protein product	CAG03719	(AC)18	8.00E-06	CDS	LP-Y-S01-0795-LF
unnamed protein product	CAG03719	(ACAT)3	8.00E-06	CDS	LP-Y-S01-0795-LF
unnamed protein product	BAC33523	(AG)6	3.00E-07	CDS	OV-N-S01-0027-W
unnamed protein product	BAC33523	(AAAG)8	3.00E-07	CDS	OV-N-S01-0027-W
unnamed protein product	BAC33523	(AG)12	3.00E-07	CDS	OV-N-S01-0027-W
unnamed protein product	XP_453836	(CCCG)3	4.00E-10	CDS(Pd)	HPa-N-S01-0335-LF
Vacuolar ATP synthase subunit G	Q25532	(AT)6	2.00E-27	3'-UTR(Pd)	LP-Y-S01-0492-LF
Wiscott-Aldrich syndrome protein	AAA85515	(ACCC)3	4.00E-08	CDS(Pd)	HC-W-S01-0314-LF
wsv209	NP_477731	(ACG)6	2.00E-35	CDS	ES-N-S03-0003-W

(continued)

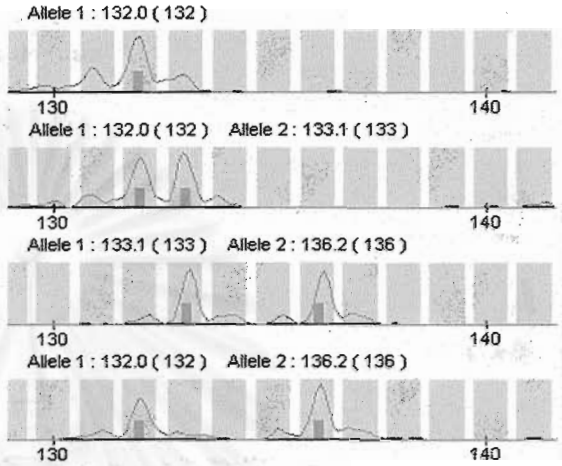
zinc finger protein	NP_187692	(ATC)27	3.00E-06	CDS	ES-N-S03-0319-W
zinc finger protein	NP_187692	(ATC)16	5.00E-10	CDS	LP-Y-S01-0357-LF
zinc finger protein	NP_175290	(AAGC)3	7.00E-17	CDS(Pd)	OV-N-S01-0320-W

APPENDIX B

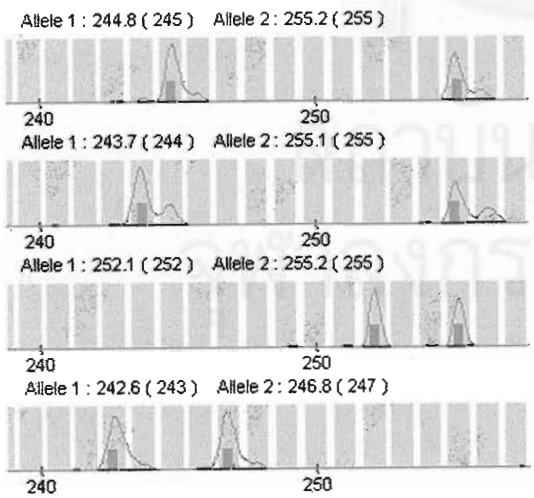
Genotype patterns obtain from wild broodstock *P. monodon* from Trang province with difference microsatellite loci



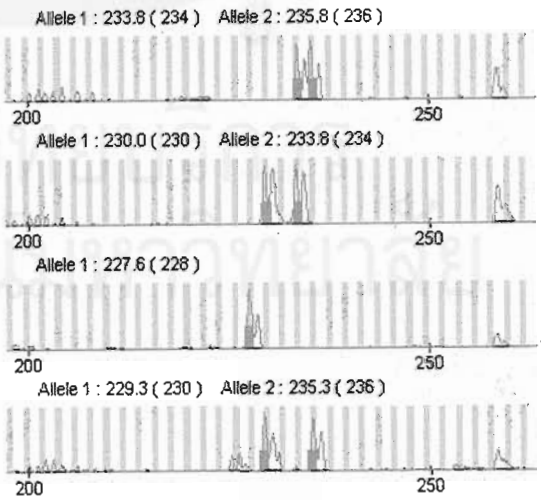
BTPm1



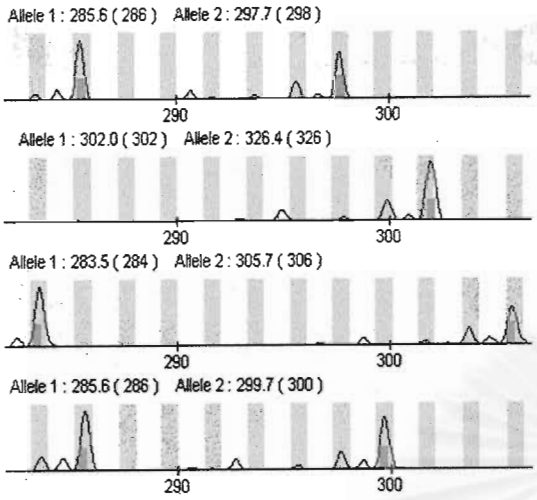
BTPm2



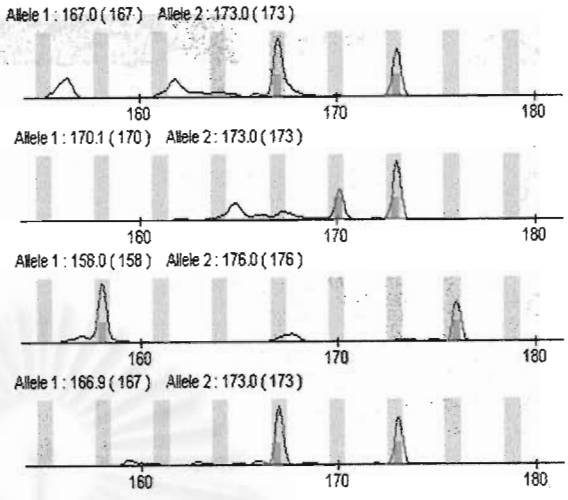
BTPm3



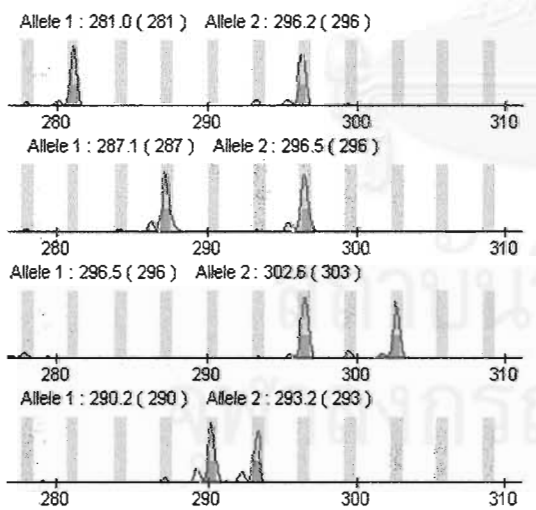
BTPm4



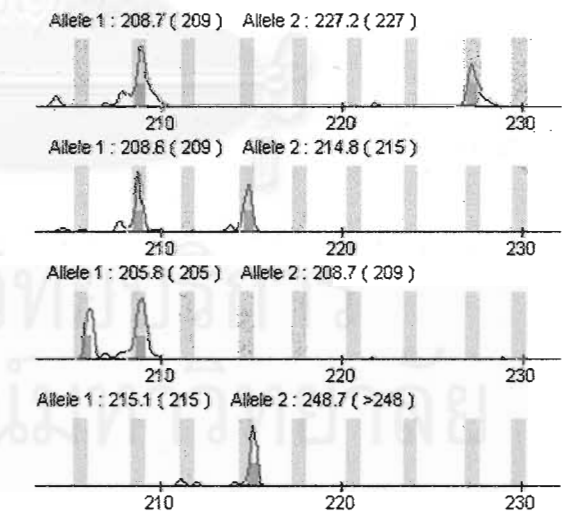
BTPm5



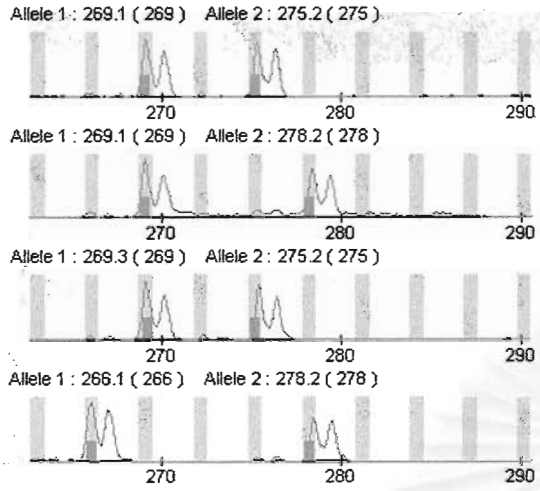
BTPm6



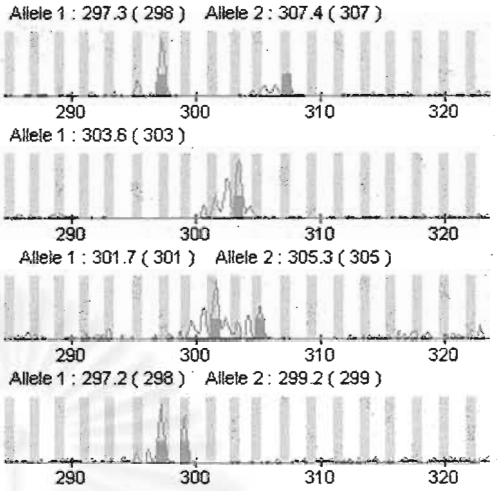
BTPm7



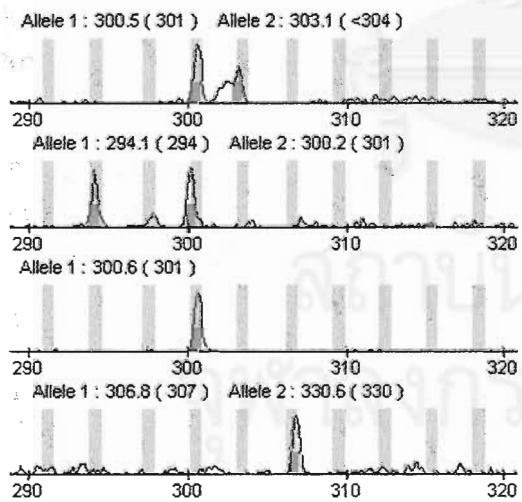
BTPm8



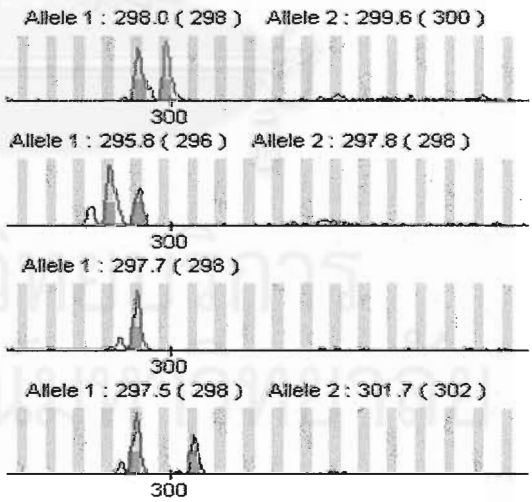
BTPm9



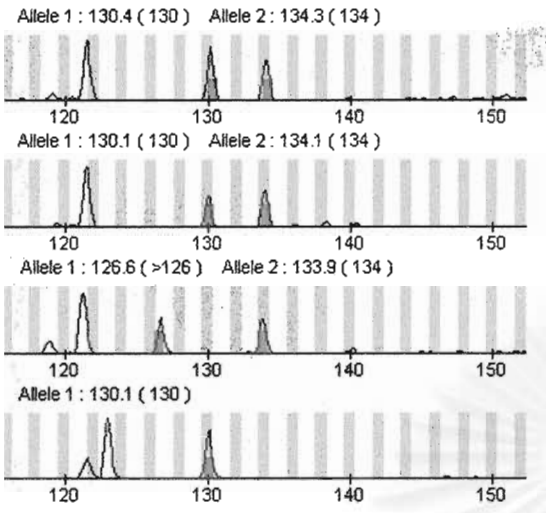
BTPm10



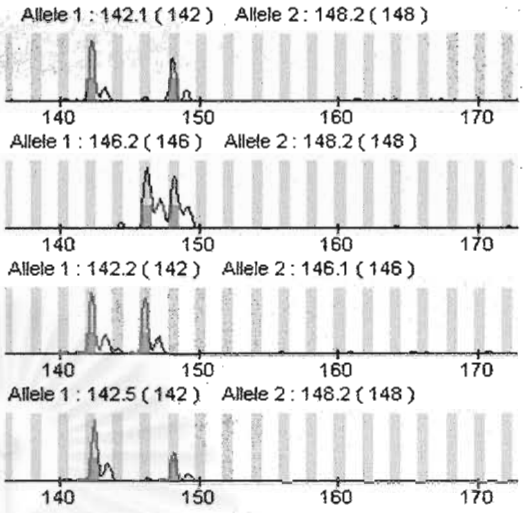
BTPm11



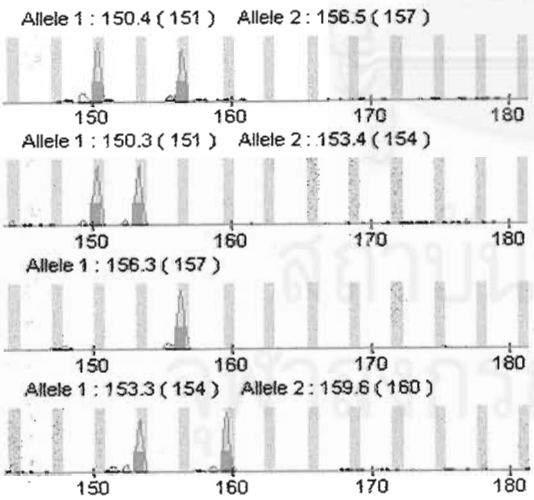
BTPm12



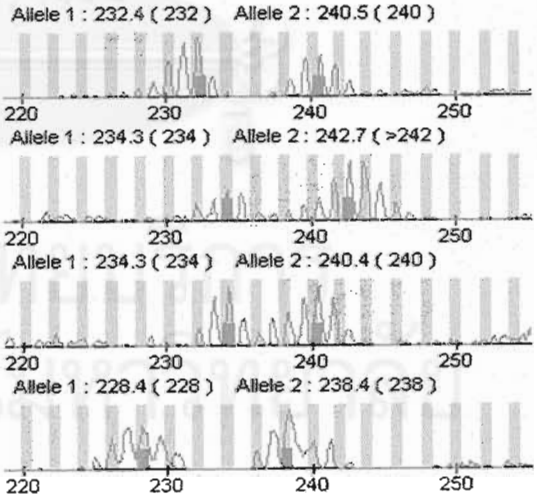
BTPm13



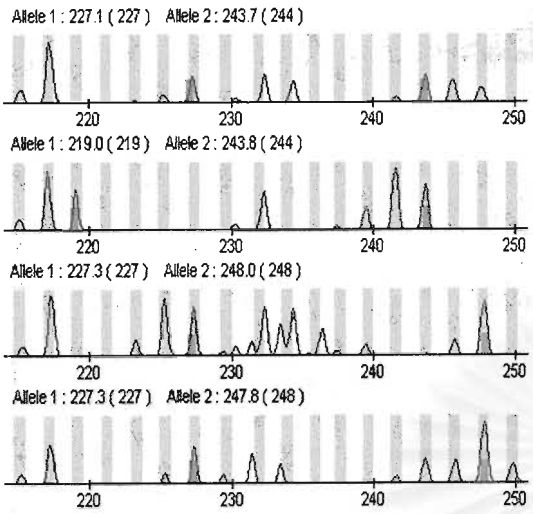
BTPm14



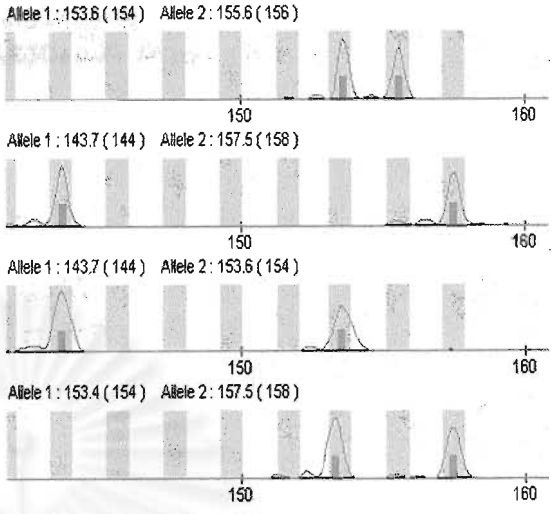
BTPm15



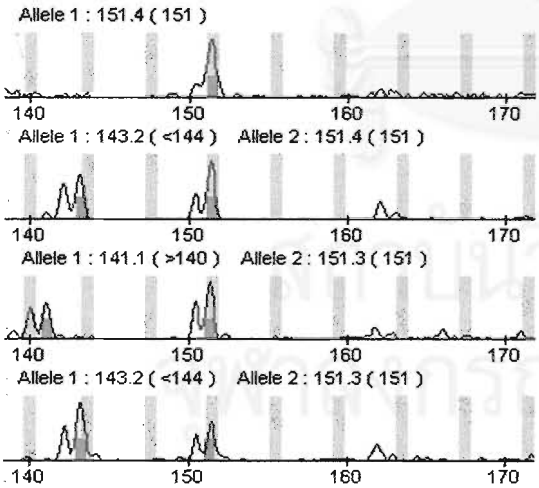
BTPm16



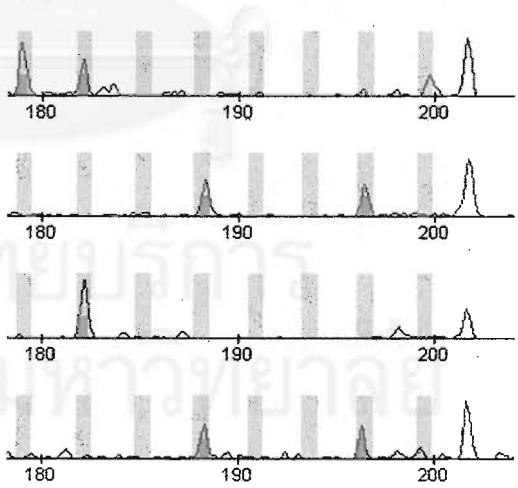
BTPm17



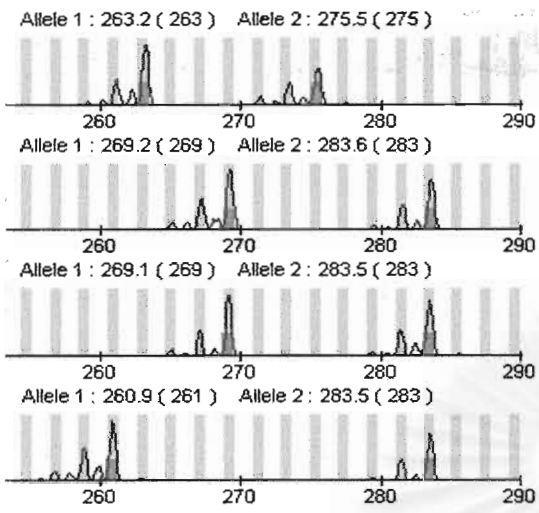
BTPm18



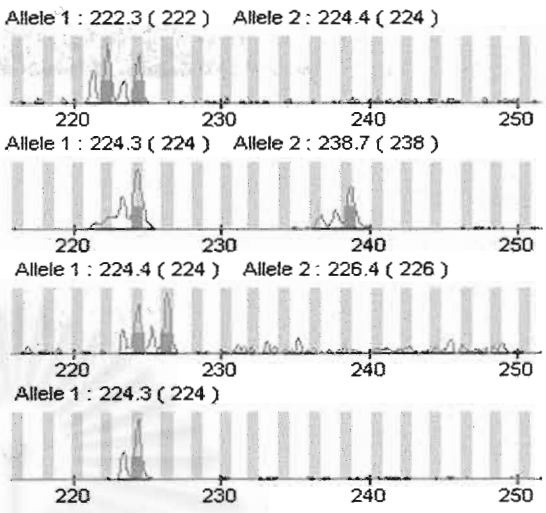
BTPm19



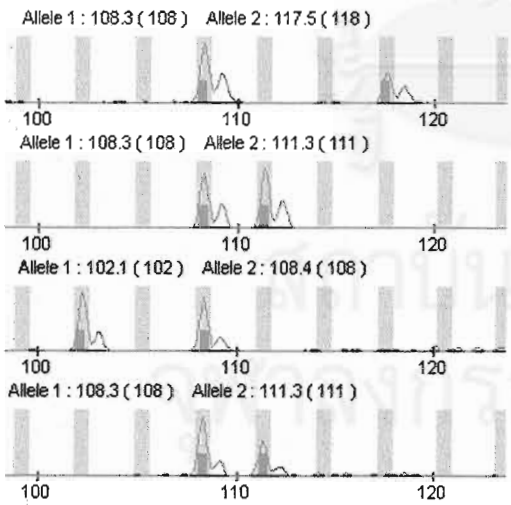
BTPm20



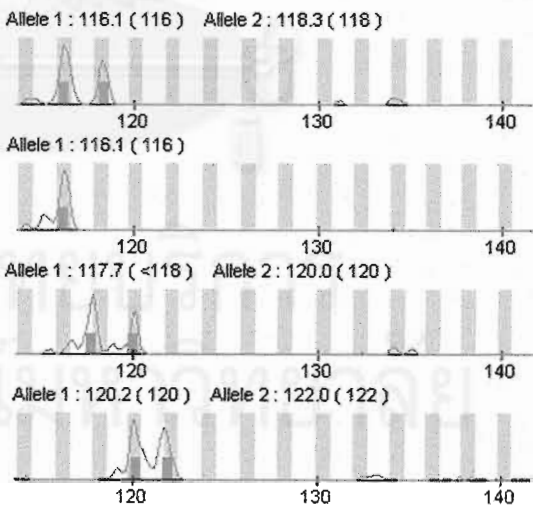
BTPm21



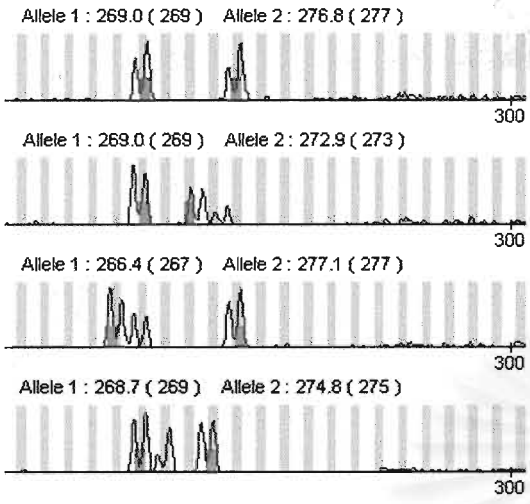
BTPm22



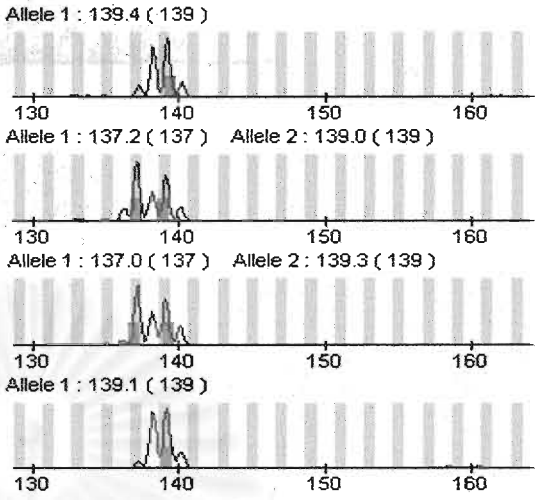
BTPm23



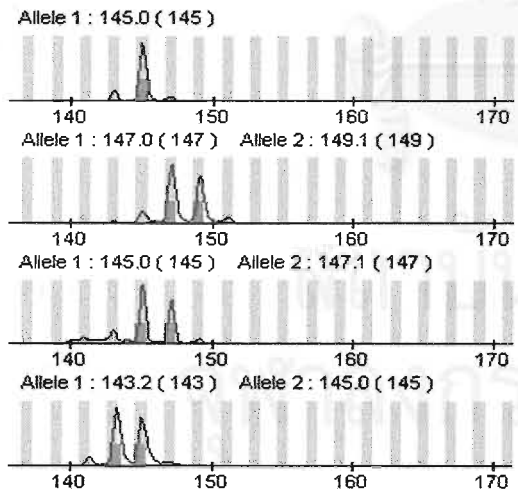
BTPm24



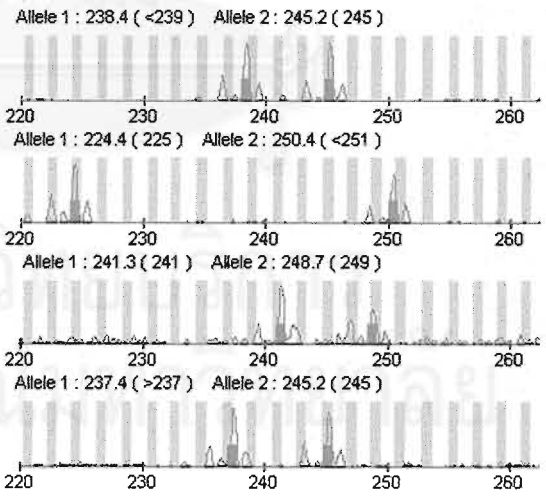
BTPm25



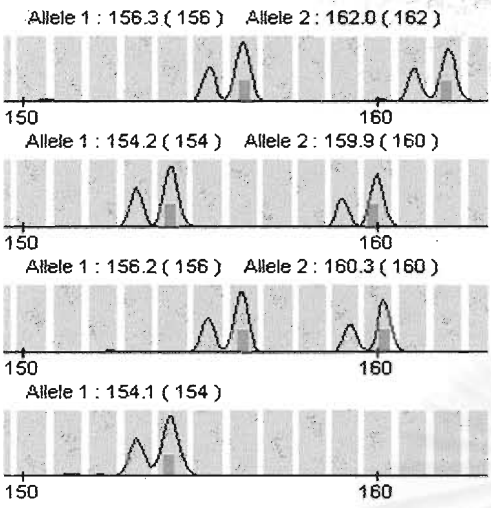
BTPm26



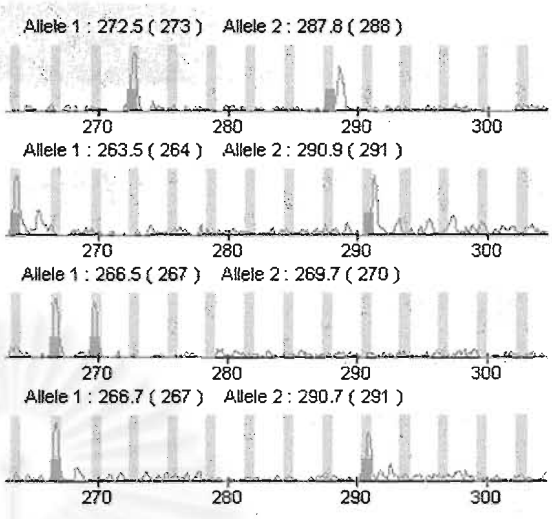
BTPm27



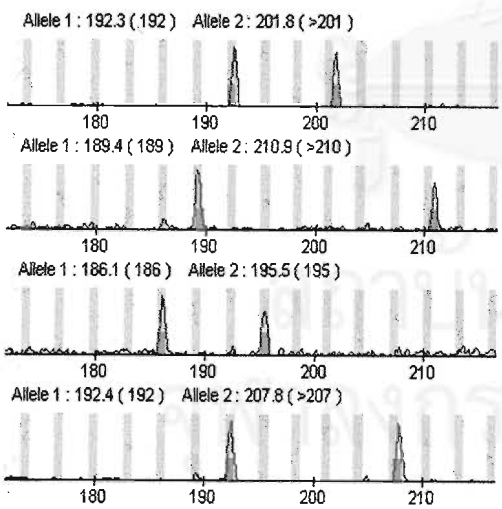
BTPm28



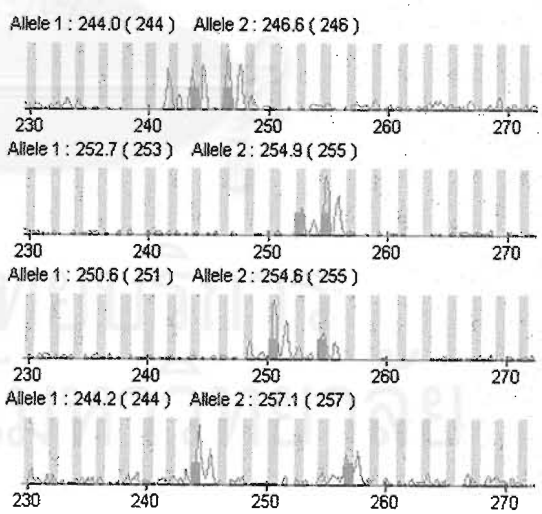
BTPm29



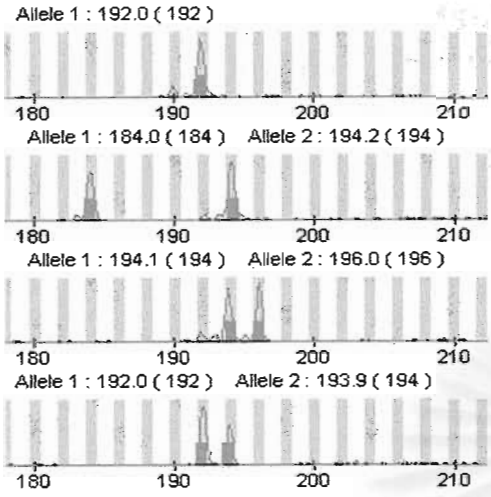
BTPm30



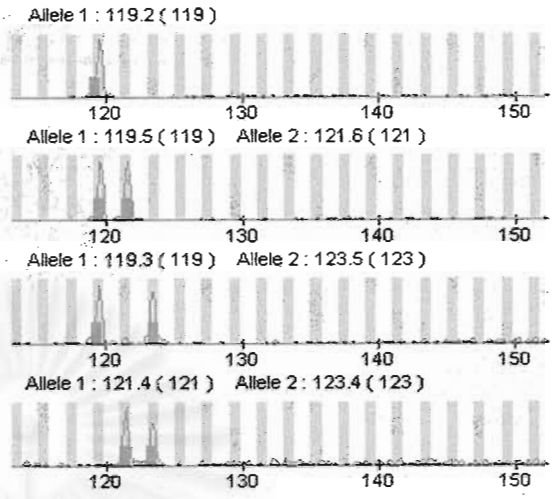
BTPm31



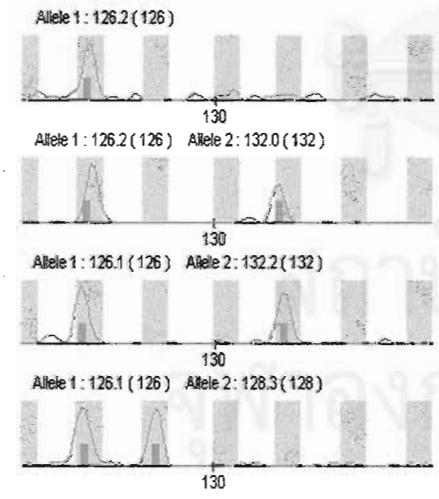
BTPm32



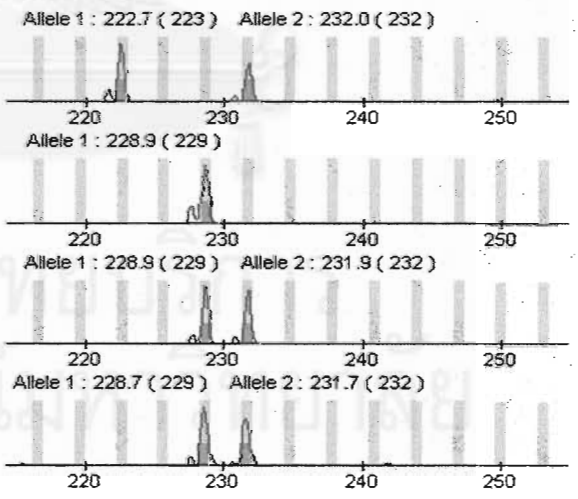
BTPm33



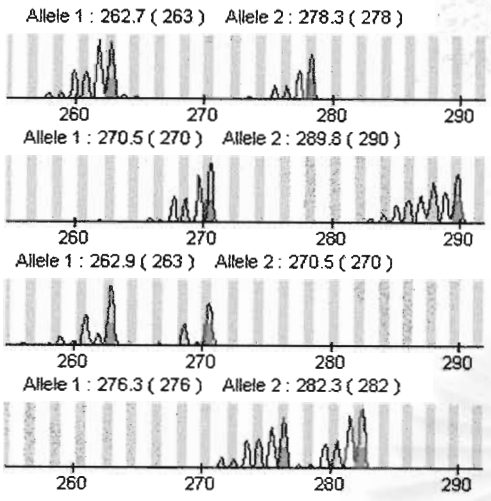
BTPm34



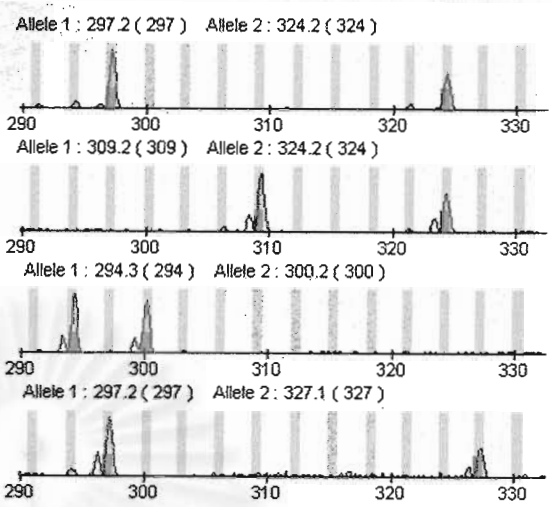
BTPm35



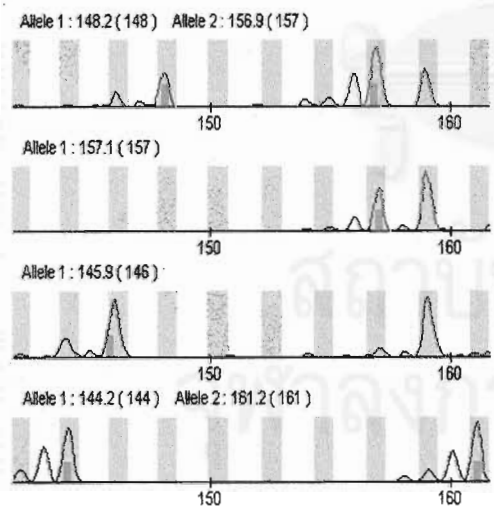
BTPm37



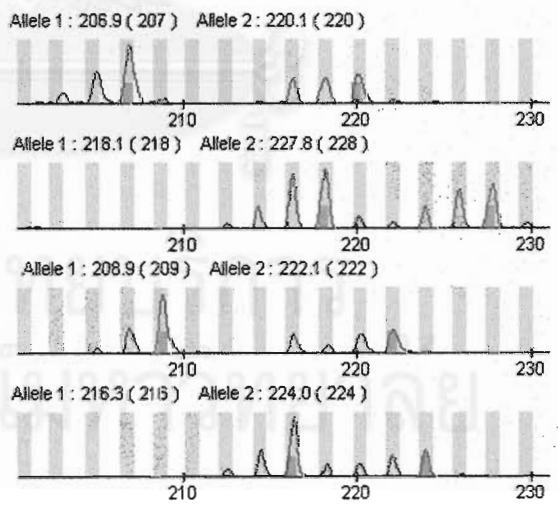
BTPm41



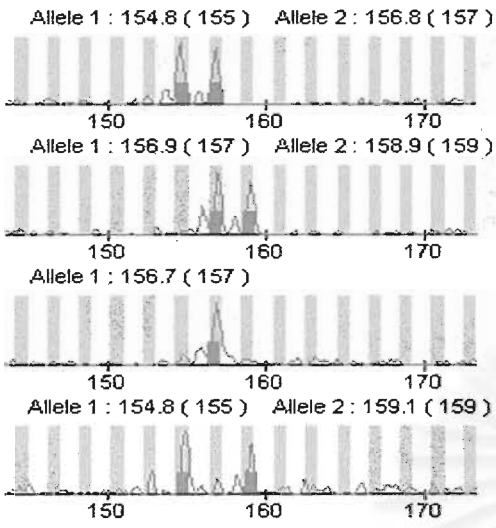
BTPm43



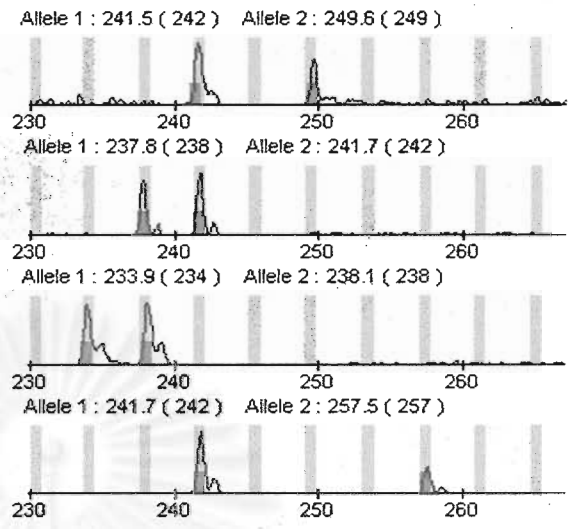
BTPm44



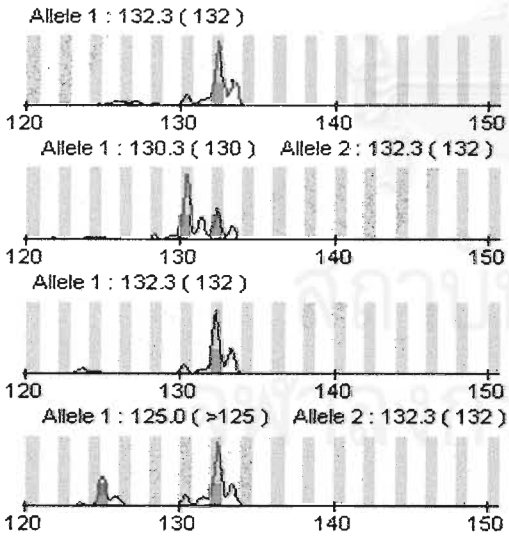
BTPm45



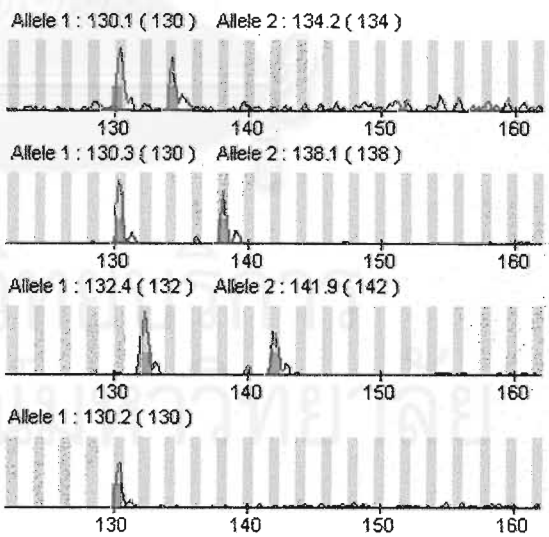
BTPm46



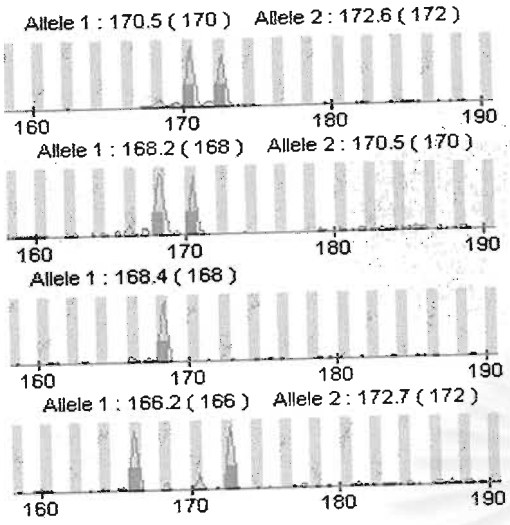
BTPm47



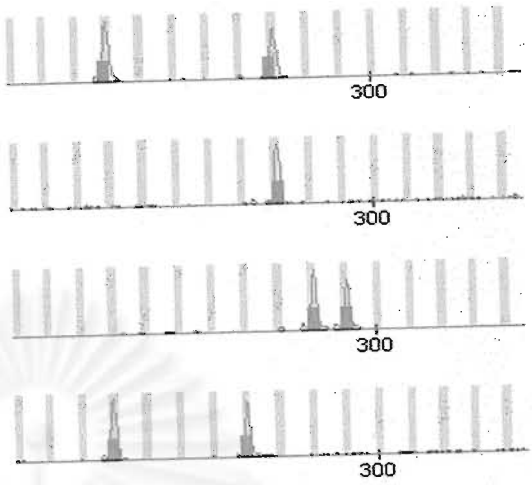
BTPm48



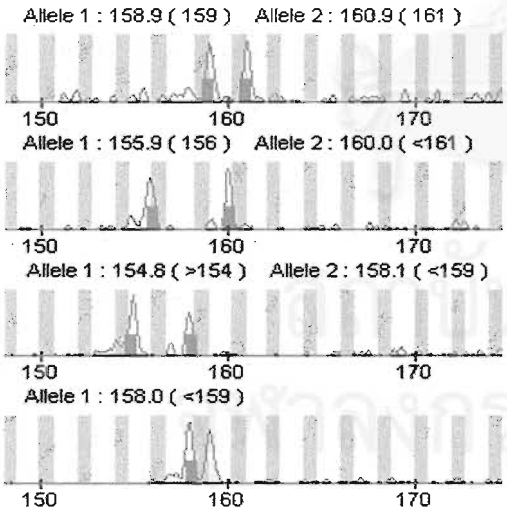
BTPm49



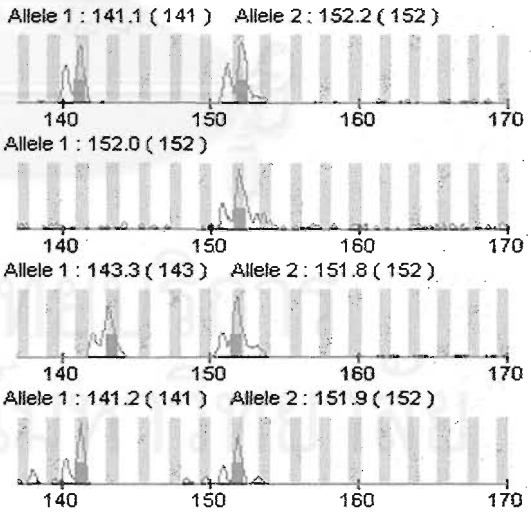
BTPm50



BTPm51



BTPm52



BTPm53

APPENDIX C

Genotypes of *P. monodon* individual from Trang province at 50 type I microsatellite markers

Sample	Genotypes																
	BTPmo1	BTPmo2	BTPmo3	BTPmo4	BTPmo5	BTPmo6	BTPmo7	BTPmo8	BTPmo9	BTPmo10	BTPmo11	BTPmo12	BTPmo13	BTPmo14	BTPmo15	BTPmo16	BTPmo17
1	246/274	134/136	241/252	-	302/312	158/170	309/309	209/227	-	305/305	294/298	298/298	-	-	115/115	236/240	-
2	268/268	132/139	255/262	234/236	296/296	170/170	284/290	205/227	269/272	298/305	301/301	298/300	-	142/146	151/160	226/240	-
3	264/318	135/136	254/255	230/234	298/300	170/173	296/303	227/233	272/281	295/303	301/301	298/300	-	142/148	151/157	240/242	-
4	256/270	133/136	241/244	230/234	286/298	170/170	293/318	205/209	266/266	295/298	-	298/300	-	146/148	151/175	236/238	227/244
5	258/306	133/136	252/252	232/232	310/316	176/176	287/293	209/227	275/275	301/303	298/318	296/298	130/134	142/146	151/157	238/240	-
6	264/273	135/136	234/237	228/228	288/306	158/170	290/293	-	275/288	298/301	298/298	298/298	130/134	142/148	151/157	240/242	227/250
7	264/264	133/133	260/261	228/228	298/302	167/173	290/290	209/215	269/269	293/307	301/301	298/302	126/134	142/142	151/157	232/240	225/244
8	270/312	132/132	-	228/230	302/306	-	-	199/209	281/293	295/298	-	296/298	130/130	142/148	-	232/240	225/238
9	258/260	133/136	243/243	228/232	277/304	170/173	290/293	233/233	272/288	303/307	-	298/300	130/134	142/146	154/157	238/240	227/244
10	262/280	136/136	245/255	228/228	281/300	158/176	303/315	209/224	272/278	299/299	301/304	298/304	130/130	142/142	148/157	240/244	238/238
11	256/276	131/136	252/252	-	306/308	158/176	284/309	209/227	275/275	295/298	-	298/302	126/134	148/148	160/163	244/252	219/244
12	260/260	139/141	244/255	228/232	302/302	-	290/306	215/248	266/269	307/325	-	296/300	134/134	146/148	151/157	230/244	236/252
13	258/266	133/136	252/255	226/234	296/330	158/158	287/290	221/224	272/281	303/303	298/315	296/298	130/134	142/148	151/157	244/246	227/248
14	258/258	136/136	245/255	238/238	281/320	170/173	290/327	209/221	260/263	295/295	-	-	130/134	146/146	151/154	234/236	232/252
15	289/312	134/139	243/247	230/230	281/284	158/176	300/312	215/227	278/281	298/307	-	298/298	130/134	146/146	157/157	240/246	227/248
16	258/264	132/136	-	230/236	306/310	176/176	287/309	205/209	-	298/305	-	296/300	130/134	146/148	154/160	238/240	232/248
17	260/262	127/136	251/254	228/232	284/300	173/173	281/321	209/209	-	-	271/274	296/300	134/134	142/142	154/157	238/240	225/242
18	258/270	-	262/265	-	302/308	167/173	284/287	205/215	269/272	-	288/288	-	130/134	146/148	157/157	234/242	238/238
19	256/256	-	244/252	232/236	298/302	167/170	318/318	215/248	269/275	-	256/256	296/300	134/134	142/148	151/157	-	230/248
20	264/270	133/136	243/252	-	300/308	-	284/290	230/230	260/275	295/303	298/298	296/300	130/134	142/142	157/163	-	221/244
21	256/293	133/137	238/252	228/228	281/306	167/170	290/293	205/209	-	-	298/298	296/300	130/134	142/148	154/163	238/240	225/244
22	256/266	136/139	237/237	-	298/298	173/176	312/324	209/215	269/272	305/305	-	296/300	130/134	142/146	157/157	240/242	227/248
23	258/285	131/139	252/258	-	302/312	158/176	281/336	209/218	272/285	295/298	294/301	298/302	130/134	142/148	154/163	234/240	219/238
24	308/310	133/136	238/255	-	298/312	170/173	293/303	211/224	269/275	301/305	301/301	300/300	132/134	142/148	151/163	240/244	-
25	258/262	132/136	243/252	-	300/330	158/173	312/333	209/211	269/272	298/298	-	298/300	130/134	142/146	157/163	238/242	227/248
26	299/324	133/139	242/256	-	302/302	155/173	290/306	211/224	269/275	-	301/301	296/302	130/134	146/146	151/157	238/240	217/232
27	264/270	128/133	252/258	-	298/304	170/170	-	-	-	301/301	298/304	296/298	126/126	142/142	141/151	242/244	227/242
28	246/256	135/139	252/270	234/234	-	158/170	281/296	218/257	272/275	298/305	298/298	296/302	-	146/148	148/151	238/240	230/248
29	258/268	132/136	243/255	230/236	281/290	167/170	287/296	205/209	269/272	303/305	301/301	296/298	126/126	146/148	141/151	242/244	248/248
30	285/285	132/136	252/255	232/232	298/302	158/170	309/327	218/221	269/272	295/305	298/298	296/296	130/134	142/142	157/160	236/236	232/248
31	268/280	136/136	243/261	228/232	290/312	161/170	290/293	221/227	-	301/311	307/307	296/296	134/134	142/148	148/151	236/238	217/244
32	258/264	132/132	239/243	228/228	286/298	158/158	293/315	209/257	272/275	-	-	296/300	130/134	146/148	151/157	-	215/244
33	260/320	132/133	252/258	-	302/326	158/173	287/293	209/215	269/272	303/303	298/304	296/300	134/134	146/148	148/157	240/242	217/248
34	256/258	136/139	237/243	232/232	284/306	158/167	290/309	209/218	269/272	295/298	307/307	296/298	134/134	142/148	157/163	-	217/238
35	264/268	133/136	-	228/228	286/300	170/170	287/318	205/227	269/278	-	304/304	298/300	130/134	142/146	151/163	240/242	217/244
36	260/266	132/136	235/262	228/228	306/306	158/176	290/309	211/224	269/275	293/303	-	298/302	130/130	146/148	141/157	240/242	-
37	262/273	133/136	249/251	228/232	302/312	-	290/309	227/236	266/278	303/303	294/294	296/302	130/134	142/146	151/157	238/240	234/248
38	264/266	134/136	258/258	228/236	290/302	170/173	-	-	275/275	303/315	304/304	296/312	134/134	142/146	-	232/236	223/244
39	256/258	128/133	237/252	232/232	298/304	173/173	287/287	218/236	269/272	301/305	298/298	296/298	-	146/148	151/163	-	-
40	256/260	133/136	234/237	228/238	288/296	-	293/303	209/215	269/272	298/303	312/312	298/300	130/134	142/146	151/157	240/242	234/252
41	258/264	133/136	251/261	228/228	308/328	-	281/296	209/251	266/269	298/299	294/298	298/300	130/134	142/146	157/160	228/238	230/248
42	256/274	131/136	244/257	-	296/300	170/170	284/315	221/230	275/278	295/298	304/304	296/298	130/130	142/148	157/163	242/244	232/248
43	244/258	131/133	252/252	234/234	298/328	167/170	287/309	227/236	272/275	-	298/298	298/300	130/134	142/142	157/157	238/240	217/240
44	262/268	134/136	249/255	-	300/324	155/158	281/290	211/230	263/275	295/305	298/301	296/300	130/130	142/146	151/157	240/240	225/242
45	258/262	132/133	242/255	232/236	281/308	155/158	287/296	221/233	272/275	-	301/301	300/300	130/134	146/146	157/157	238/240	225/242
46	256/301	132/136	255/261	228/228	300/306	155/158	287/293	209/224	269/272	303/303	301/307	300/304	130/130	142/148	151/151	236/242	225/225
47	258/274	132/136	252/261	230/234	294/294	158/170	303/306	205/209	-	293/295	298/307	294/302	126/130	146/150	151/157	-	225/248
48	264/274	133/136	255/264	228/228	281/312	170/173	287/290	202/205	266/269	-	310/310	-	-	-	151/163	236/240	225/252

Sample	Genotypes															
	BTPmo18	BTPmo19	BTPmo20	BTPmo21	BTPmo22	BTPmo23	BTPmo24	BTPmo25	BTPmo26	BTPmo27	BTPmo28	BTPmo29	BTPmo30	BTPmo31	BTPmo32	BTPmo33
1	152/153	148/152	182/196	281/283	226/228	102/105	114/116	269/273	139/139	145/145	243/245	156/162	248/267	192/201	244/261	192/194
2	153/157	144/144	194/194	271/296	224/224	108/108	116/116	-	139/139	145/147	223/237	154/160	258/267	192/210	244/253	192/194
3	126/153	152/152	-	283/294	222/224	108/108	116/118	269/277	139/139	145/147	241/245	-	264/264	189/210	246/257	188/192
4	152/153	148/155	167/167	283/304	224/238	102/108	118/118	-	139/139	145/147	219/251	154/154	263/288	195/198	244/244	192/194
5	153/155	140/140	-	271/283	224/226	105/118	118/118	267/271	139/139	145/147	239/253	158/166	273/273	186/195	244/261	192/194
6	153/155	148/148	191/194	283/298	224/224	108/108	116/118	-	139/139	147/147	245/267	154/154	267/291	177/210	244/253	192/192
7	144/153	152/152	170/170	263/302	224/224	105/108	118/120	267/275	139/139	145/147	237/259	155/156	291/291	192/207	244/246	194/194
8	153/155	152/152	-	263/275	220/226	105/105	120/168	269/275	139/141	-	233/241	154/160	273/288	192/195	253/255	190/192
9	153/153	152/152	196/196	273/310	228/228	108/118	116/116	267/271	139/139	145/145	239/243	-	288/294	189/210	251/255	194/194
10	153/155	152/152	179/194	269/269	224/224	108/111	116/118	-	137/139	145/147	221/249	-	263/291	192/204	244/257	192/192
11	153/155	-	179/182	230/271	224/228	-	116/120	267/267	139/139	145/147	231/247	160/164	263/285	177/201	257/259	184/194
12	144/157	148/152	179/182	253/263	224/236	-	118/120	269/273	139/139	145/147	239/253	154/160	267/270	192/195	246/255	194/196
13	152/155	144/155	188/196	263/321	224/224	102/108	120/120	267/279	139/139	145/145	227/235	154/154	267/291	192/192	253/257	192/194
14	153/153	140/140	188/194	269/269	218/224	108/111	116/116	-	139/139	147/147	237/247	-	-	-	251/255	192/192
15	128/153	140/144	182/182	263/277	224/224	108/111	118/118	269/273	137/139	147/147	221/255	-	267/294	192/198	246/259	192/192
16	144/153	144/155	-	244/255	224/226	111/111	116/116	269/277	139/139	145/145	211/251	-	276/276	192/195	251/257	192/194
17	153/155	144/144	176/176	259/269	-	105/111	110/110	269/269	-	-	217/219	-	267/267	171/174	257/255	190/194
18	153/153	152/159	185/188	269/283	224/230	-	116/118	267/279	137/137	145/147	245/255	-	254/267	183/189	244/259	194/194
19	153/155	144/152	179/182	259/263	224/224	108/111	116/116	-	137/139	145/147	245/249	156/156	267/270	195/195	244/259	190/190
20	153/157	-	188/188	269/283	224/224	108/111	116/116	267/269	139/139	145/147	239/239	154/154	270/297	186/198	240/259	184/192
21	155/157	152/152	188/196	277/283	224/236	108/111	116/118	267/275	139/139	147/147	219/243	154/166	291/300	177/192	244/251	190/192
22	144/155	144/152	182/188	296/296	216/224	-	120/122	-	139/139	145/149	249/251	154/154	267/270	192/195	251/249	192/194
23	144/153	144/155	182/194	261/283	224/224	102/108	116/118	269/277	139/139	147/147	255/259	160/160	282/288	171/204	244/249	194/194
24	152/153	148/152	182/188	283/283	224/226	108/118	-	269/277	139/139	145/147	243/251	154/158	267/267	177/207	253/255	192/192
25	144/144	140/152	188/194	263/263	224/224	108/108	-	269/277	139/139	145/147	239/247	-	279/294	198/207	246/253	192/192
26	153/157	152/152	-	267/294	226/226	108/108	114/116	-	141/141	145/147	-	154/154	263/270	195/198	-	192/194
27	144/153	144/152	188/194	257/267	-	-	116/116	-	139/139	145/147	229/243	-	267/297	204/207	244/249	-
28	153/155	148/152	-	271/283	224/232	105/111	118/118	269/271	-	147/147	245/245	154/156	263/263	171/195	247/257	192/194
29	153/165	152/159	179/188	257/269	226/226	108/108	116/118	269/269	139/139	147/147	231/231	-	264/267	195/210	244/247	192/192
30	153/159	152/152	194/196	253/261	226/241	111/114	116/116	-	139/139	145/145	237/255	160/164	267/270	201/210	238/242	194/194
31	153/153	144/144	196/196	269/283	224/224	96/108	116/116	267/285	139/139	145/145	255/255	154/157	264/267	192/195	238/255	194/194
32	153/153	-	194/196	253/265	224/224	108/111	116/118	267/277	-	145/147	249/249	-	267/294	-	247/255	184/192
33	153/157	152/152	188/196	271/277	224/230	108/121	118/118	-	139/139	145/147	237/249	152/152	267/291	192/201	244/255	192/194
34	153/155	-	182/188	253/292	226/251	108/111	118/132	269/273	137/137	145/147	233/239	154/154	258/267	195/201	251/255	192/194
35	144/155	148/155	196/196	263/292	224/228	105/108	116/166	267/271	139/139	145/147	245/245	154/160	264/270	180/189	251/257	184/192
36	144/155	152/152	-	244/271	228/234	105/114	-	267/269	139/139	145/147	253/265	154/154	270/273	177/204	246/251	192/194
37	153/153	144/152	179/194	259/285	226/234	105/111	116/116	269/275	137/139	145/147	251/259	152/154	270/279	195/198	259/270	194/196
38	144/157	144/144	196/196	259/265	224/226	102/111	120/120	269/273	139/139	145/145	247/251	154/154	260/267	195/195	249/251	192/192
39	153/153	-	191/196	255/261	224/226	111/118	120/134	271/273	139/139	147/149	-	154/164	294/297	195/198	253/266	192/194
40	153/157	-	194/196	271/283	224/226	108/118	116/118	-	139/139	145/147	239/245	154/154	273/297	189/201	244/255	192/194
41	153/153	144/144	194/194	253/283	222/224	108/108	116/120	267/273	139/139	143/145	247/261	156/162	270/297	192/195	246/253	194/194
42	153/155	148/148	188/188	267/292	220/226	108/108	120/126	267/267	139/139	145/147	241/247	154/160	267/267	201/204	253/255	192/194
43	153/155	140/148	179/182	275/290	224/226	99/118	118/118	269/271	139/139	147/147	221/255	156/160	285/303	189/195	257/261	192/194
44	153/153	140/148	176/176	283/283	224/226	108/108	120/142	269/277	139/139	145/145	241/271	154/154	267/267	198/210	244/246	192/192
45	144/155	-	-	269/283	224/226	108/108	118/118	269/275	-	147/147	229/243	154/158	263/300	192/201	244/251	196/196
46	150/153	144/152	-	267/290	-	108/108	116/116	-	139/141	147/147	229/245	154/160	254/303	195/201	246/266	192/192
47	153/155	-	179/194	269/269	226/232	111/111	118/132	269/281	-	145/145	237/247	154/160	254/273	195/204	244/244	192/192
48	148/157	148/152	194/196	259/283	224/224	102/108	116/116	269/269	-	145/153	219/249	154/154	248/248	189/195	244/257	192/194

Sample	Genotypes																
	BTPmo34	BTPmo35	BTPmo37	BTPmo38	BTPmo40	BTPmo41	BTPmo43	BTPmo44	BTPmo45	BTPmo46	BTPmo47	BTPm48	BTPm49	BTPm50	BTPm51	BTPm52	BTPm53
1	121/125	126/128	226/232	223/226	301/301	267/267	279/297	157/166	197/228	153/157	242/257	-	126/130	170/170	-	155/159	152/152
2	119/121	130/134	220/220	211/214	301/301	278/284	288/297	155/161	214/224	159/165	234/269	132/132	130/132	168/170	275/291	159/159	-
3	121/123	-	-	208/214	-	267/267	294/294	155/157	195/216	157/159	223/223	130/132	130/134	168/170	291/300	-	152/152
4	121/121	124/126	229/229	208/211	304/306	263/278	294/318	157/157	216/220	157/159	273/277	130/132	130/138	168/170	297/297	159/159	152/152
5	-	126/130	229/229	211/214	304/306	294/301	303/324	150/153	207/220	159/175	242/249	132/132	132/142	168/168	297/297	-	152/152
6	-	126/126	251/251	208/211	300/306	282/311	297/321	148/157	201/236	157/159	269/292	130/132	130/130	170/170	281/281	157/159	152/152
7	121/121	130/134	223/223	199/217	301/301	284/307	297/321	157/157	218/228	155/157	261/273	132/132	130/130	170/170	291/291	159/161	152/154
8	121/121	126/132	223/223	202/220	301/306	303/315	303/318	157/157	179/220	155/157	261/277	125/132	130/130	170/170	294/297	161/161	152/152
9	121/121	126/126	-	223/226	298/301	284/290	318/324	155/164	209/222	155/155	238/242	132/132	130/130	166/166	291/291	159/159	141/152
10	119/119	126/126	229/229	211/214	296/298	301/303	294/321	155/168	197/222	-	234/238	130/132	130/130	168/170	294/294	153/159	148/152
11	121/121	126/126	251/251	208/235	304/306	284/295	294/315	150/161	214/214	157/167	246/253	132/132	130/142	170/170	291/294	-	152/152
12	121/121	126/128	226/226	208/211	294/306	309/317	297/300	155/157	216/224	157/159	257/261	132/132	130/132	170/170	294/294	157/178	135/152
13	123/125	126/132	-	211/214	298/304	270/290	297/321	161/164	207/243	157/157	238/261	132/134	130/130	170/170	294/294	-	135/152
14	121/121	126/128	232/232	208/211	-	294/303	303/333	155/168	216/216	155/163	261/269	132/132	130/130	170/170	291/291	159/159	-
15	119/121	128/128	251/251	199/217	296/304	303/305	297/324	146/157	214/214	157/161	223/246	132/132	130/130	170/172	291/291	159/159	152/152
16	119/123	126/130	232/232	202/220	304/306	267/292	306/339	161/168	-	157/159	214/246	130/130	130/130	168/170	275/288	159/159	141/152
17	119/119	-	229/229	214/217	-	-	-	-	203/203	-	230/230	130/132	130/134	168/168	291/291	-	-
18	-	-	232/232	-	300/300	280/303	285/297	157/164	211/218	157/177	238/253	132/132	130/130	166/172	275/294	-	-
19	-	-	217/217	229/229	301/301	299/303	297/300	157/157	-	155/159	230/300	130/130	130/130	170/170	288/288	-	-
20	121/121	-	263/263	214/217	290/304	263/303	300/327	157/164	195/245	157/171	230/277	130/132	130/132	170/170	-	-	-
21	121/121	126/130	-	214/226	301/301	261/307	321/330	148/174	216/220	155/159	242/281	130/130	-	170/170	285/294	-	152/152
22	119/119	126/132	268/268	202/208	301/306	307/307	297/300	157/164	220/230	159/161	230/238	132/132	126/130	168/170	294/294	159/159	-
23	119/123	126/130	232/232	211/220	301/306	263/270	312/318	157/168	214/238	151/161	234/285	132/132	130/130	170/170	294/300	159/163	-
24	121/121	128/130	-	211/214	-	-	297/324	146/146	199/267	157/159	223/249	130/132	130/134	170/170	288/294	159/159	-
25	119/121	126/130	-	199/217	306/306	-	309/324	142/142	201/239	155/157	206/206	130/132	-	168/168	297/297	159/159	-
26	121/121	130/134	223/223	220/229	301/306	276/282	294/300	164/168	211/259	155/159	202/281	132/132	-	170/170	281/294	155/159	152/152
27	-	128/134	226/226	211/220	301/304	295/303	297/327	148/157	214/218	157/159	218/285	128/132	130/130	170/170	288/294	-	152/152
28	121/121	128/132	223/223	211/214	290/304	276/305	294/294	157/157	203/211	159/159	206/234	-	130/130	170/172	275/306	-	152/152
29	121/123	126/130	-	211/211	304/304	261/282	294/297	157/164	226/226	-	285/285	132/132	130/134	170/172	278/291	159/161	143/152
30	121/123	126/126	229/229	214/214	304/308	292/303	297/300	155/155	209/249	151/155	253/296	130/132	130/130	170/170	294/294	157/161	141/152
31	119/121	128/130	-	199/226	306/306	267/294	294/297	157/157	226/245	157/159	269/277	130/130	130/132	166/168	294/294	155/159	-
32	123/125	126/132	-	205/217	300/301	270/303	297/324	-	216/224	157/157	265/285	130/132	132/132	168/170	288/291	159/159	-
33	121/121	128/130	229/229	211/211	301/304	295/303	297/321	157/157	216/234	155/155	265/273	130/132	130/138	170/170	294/294	161/161	152/152
34	121/121	126/128	223/223	214/217	300/304	282/286	288/297	157/157	195/216	155/157	206/214	128/132	130/130	170/172	291/291	159/159	152/152
35	121/123	126/126	-	208/214	294/304	282/313	294/300	157/164	205/216	157/159	253/257	126/132	130/130	170/170	297/297	159/159	152/152
36	121/121	128/130	223/229	211/211	300/306	294/307	300/303	164/172	238/243	157/157	-	132/132	130/134	170/170	297/297	157/159	141/152
37	119/121	-	223/232	217/217	298/306	292/311	300/309	157/166	195/216	157/159	249/277	132/132	130/154	170/170	288/288	159/161	152/152
38	121/123	128/132	-	205/217	301/306	261/284	291/297	157/161	201/228	155/159	210/223	-	130/132	170/170	-	-	141/152
39	121/123	126/126	229/229	202/211	301/312	280/305	324/327	157/170	216/241	161/161	242/257	130/132	130/130	170/170	291/294	161/161	152/152
40	119/121	126/126	229/229	217/226	298/304	265/309	303/327	157/157	220/241	155/157	223/265	126/132	130/130	170/170	291/300	159/159	152/152
41	119/119	126/130	229/229	199/217	294/306	305/307	300/327	148/157	216/218	155/155	234/277	130/132	130/132	172/172	285/300	159/159	152/152
42	119/121	126/132	229/232	220/229	301/312	297/313	297/297	157/161	234/249	157/157	269/273	130/132	130/130	170/170	294/294	157/159	152/152
43	119/123	128/132	229/229	226/229	301/306	292/307	315/333	157/161	218/234	157/163	246/269	130/132	130/130	170/170	294/294	159/159	152/152
44	119/121	126/132	223/223	205/217	301/301	303/305	297/297	157/168	218/224	157/157	206/269	132/132	130/130	168/170	294/294	159/159	152/152
45	121/123	126/132	223/223	220/235	301/306	299/305	294/330	157/157	214/216	155/157	189/206	130/132	130/130	170/170	278/278	159/159	152/152
46	121/121	126/128	232/232	211/217	304/320	265/299	333/333	157/176	216/243	157/159	218/230	130/132	130/130	170/170	288/306	159/159	145/152
47	-	124/124	232/232	214/226	294/301	272/284	285/303	144/161	-	155/155	246/289	130/132	126/130	168/170	294/294	159/159	152/152
48	121/121	-	229/232	205/214	296/301	267/309	-	153/153	213/245	-	249/265	130/132	130/130	170/174	-	155/155	152/152

APPENDIX D

Genotypes of *P. monodon* pedigree samples for 44 type I microsatellite markers

Sample	Genotypes																
	BTPmo1	BTPmo2	BTPmo3	BTPmo4	BTPmo5	BTPmo6	BTPmo7	BTPmo8	BTPmo9	BTPmo10	BTPmo11	BTPmo12	BTPmo13	BTPmo14	BTPmo15	BTPmo16	BTPmo17
K1	256/260	a/a	255/259	a/c	296/310	167/170	293/299	209/229	270/270	304/304	b/c	296/298	133/137	143/148	150/156	242/244	219/219
K2	260/260	a/b	252/255	b/c	280/300	170/176	293/299	209/229	270/276	304/304	-	296/296	133/137	143/148	150/156	242/242	219/244
K3	256/260	a/b	252/255	b/c	280/300	170/176	290/293	209/229	270/276	304/312	c/d	296/296	133/133	143/148	150/163	242/242	219/244
K4	256/266	a/b	252/264	a/c	280/296	170/176	293/299	209/229	270/276	304/312	a/d	296/296	133/133	143/143	150/156	242/242	219/244
K5	260/260	a/b	259/264	b/c	280/300	170/176	290/293	206/229	270/273	304/312	b/d	296/296	133/137	143/148	150/156	242/244	219/219
K6	260/266	a/a	252/264	a/c	300/310	167/170	290/293	206/229	270/270	304/304	a/d	296/298	133/137	143/143	150/156	242/242	219/244
K7	260/266	a/b	259/264	a/c	300/310	170/176	287/290	206/229	273/276	304/312	a/b	296/298	133/133	143/148	150/156	242/244	219/244
K8	260/260	a/a	255/259	b/c	280/300	167/170	290/293	209/229	270/273	304/312	b/c	296/296	133/137	143/148	150/156	242/244	219/219
K9	260/266	a/b	259/264	a/c	280/300	167/170	293/299	209/229	270/273	304/304	a/b	296/298	133/133	143/148	150/156	242/242	219/244
K10	260/266	a/a	259/264	b/c	280/296	170/176	287/299	209/229	270/273	304/312	a/b	296/298	133/133	143/143	150/156	242/242	219/244
K11	256/260	a/b	252/264	a/c	296/310	167/170	287/290	206/229	270/276	304/304	a/d	296/298	133/137	143/143	150/163	242/242	219/244
K12	256/266	a/b	252/255	a/c	280/300	170/176	293/299	209/229	270/273	304/312	c/d	296/296	133/137	143/148	150/156	242/242	219/244
K13	260/266	a/b	252/264	b/c	280/300	167/170	287/290	206/229	270/270	304/304	a/d	296/298	133/137	143/148	150/163	242/244	219/244
K14	260/266	a/a	255/259	a/c	296/310	170/176	290/293	206/229	270/273	304/312	b/c	296/298	133/137	143/148	150/156	242/244	219/219
K15	256/266	a/a	259/264	b/c	296/310	170/176	287/299	209/229	270/270	304/304	a/b	296/296	133/137	143/143	150/156	242/242	219/219
K16	260/266	a/a	252/264	a/c	280/296	170/176	293/299	206/229	270/273	304/304	a/d	296/298	133/137	143/143	150/163	242/244	219/219
K17	260/266	a/a	252/264	a/c	280/300	170/176	290/293	206/229	270/276	304/312	a/d	296/298	133/133	143/143	150/156	242/244	219/244
K18	256/266	a/b	252/264	a/c	280/296	170/176	290/293	209/229	270/270	304/312	b/c	296/298	133/133	143/143	150/156	242/242	219/244
K19	260/260	a/a	252/264	a/c	296/310	167/170	290/293	206/229	273/276	304/312	a/d	296/296	133/133	143/148	150/156	242/244	219/244
K20	256/260	a/b	252/264	a/c	280/300	170/176	287/299	206/229	270/276	304/312	a/d	296/298	133/133	143/148	150/156	242/244	219/244
K21	260/260	a/b	255/259	b/c	300/310	167/170	287/299	206/229	270/273	304/312	b/c	296/298	133/137	143/148	150/163	242/244	219/244
K22	256/266	a/b	255/259	b/c	280/300	170/176	287/290	209/229	270/273	304/304	b/c	296/298	133/133	143/148	150/156	242/242	219/219
K23	256/266	a/b	259/264	a/c	280/300	167/170	287/290	206/229	270/270	304/312	a/b	296/298	133/133	143/143	150/156	242/244	219/244
K24	256/266	a/b	252/264	a/c	280/300	167/170	290/293	209/229	270/273	304/304	-	296/298	133/133	143/148	150/163	242/242	219/244
K25	256/260	a/a	255/259	a/c	296/310	167/170	287/299	206/229	270/270	304/312	b/c	296/296	133/133	143/148	150/156	242/242	219/219
K27	256/260	a/a	259/264	a/c	296/310	167/170	290/293	209/229	270/273	304/312	a/b	296/298	133/137	143/143	150/163	242/244	219/219
K28	260/260	a/b	259/264	b/c	280/296	170/176	290/293	206/229	270/273	304/304	a/b	296/296	133/133	143/143	150/156	242/242	219/244
K29	260/260	a/a	252/264	b/c	300/310	167/170	287/290	209/229	270/276	304/312	a/d	296/298	133/133	143/148	150/156	242/242	219/219
K30	256/266	a/b	252/264	a/c	280/300	170/176	287/299	209/229	270/273	304/312	-	296/296	133/133	143/148	150/163	242/242	219/219
K31	256/266	a/b	252/264	a/c	280/300	170/176	287/299	206/229	270/270	304/312	a/d	296/296	133/137	143/143	150/156	242/244	219/244
K32	256/266	a/b	259/264	a/c	280/300	170/176	293/299	206/229	273/276	304/304	a/b	296/296	133/137	143/148	150/156	242/244	219/244
K33	256/266	a/b	259/264	a/c	296/310	170/176	290/293	209/229	270/273	304/312	a/b	296/298	133/133	143/143	150/163	242/244	219/244
K34	256/266	a/b	252/264	b/c	280/300	170/176	293/299	206/229	270/273	304/312	a/d	296/298	133/133	143/143	150/156	242/244	219/219
K35	260/260	a/b	259/264	a/c	280/300	170/176	290/293	206/229	273/276	304/304	a/b	296/298	133/137	143/143	150/163	242/244	219/219
K36	256/266	a/a	252/264	a/c	296/310	170/176	293/299	209/229	273/276	304/304	a/d	296/298	133/137	143/143	150/156	242/244	219/244
K37	260/266	a/b	259/264	a/c	300/310	170/176	293/299	206/229	270/273	304/304	a/b	296/298	133/137	143/143	150/156	242/244	219/219
K38	256/260	a/a	255/259	a/c	296/310	167/170	287/290	209/229	270/273	304/304	b/c	296/298	133/137	143/148	150/163	242/244	219/244
K39	256/260	a/a	255/259	a/c	296/310	170/176	287/299	206/229	270/270	304/304	b/c	296/298	133/133	143/148	150/156	242/244	219/244
K40	260/260	a/a	252/255	a/c	300/310	170/176	287/299	206/229	270/273	304/312	c/d	296/296	133/133	143/143	150/156	242/242	219/244
K41	256/260	a/a	259/264	a/c	296/310	170/176	287/290	209/229	270/273	304/304	a/b	296/296	133/137	143/148	150/156	242/244	219/219
K42	260/266	a/b	252/264	a/c	300/310	170/176	287/299	209/229	270/270	304/304	a/d	296/296	133/133	143/148	150/156	242/244	219/244
Mother	260/266	a/a	255/264	c/c	296/300	170/170	287/293	229/229	270/276	304/304	a/c	296/296	133/137	143/148	150/150	242/242	219/219
Father	256/260	a/b	252/259	a/b	280/310	167/176	290/299	206/209	270/273	304/312	b/d	296/298	133/133	143/143	156/163	242/244	219/244

Sample	Genotypes																
	BTPmo18	BTPmo19	BTPmo20	BTPmo21	BTPmo22	BTPmo23	BTPmo24	BTPmo25	BTPmo26	BTPmo27	BTPmo28	BTPmo29	BTPmo30	BTPmo31	BTPmo32	BTPmo33
K1	154/154	151/151	193/196	261/269	224/226	108/111	134/142	276/280	139/141	145/145	231/237	154/154	264/267	192/208	245/247	194/196
K2	154/154	151/151	196/196	253/261	224/226	111/111	134/142	276/280	139/141	145/145	237/246	154/160	267/279	192/208	247/260	194/196
K3	154/154	151/151	193/196	253/261	224/224	111/111	134/134	276/276	139/141	145/145	231/237	154/160	264/267	192/208	247/260	192/194
K4	154/154	151/155	193/196	253/261	224/224	108/111	134/134	276/280	139/141	145/145	231/246	154/154	267/279	192/211	247/260	192/194
K5	144/154	151/151	196/196	253/269	224/224	108/111	134/142	276/276	139/141	145/149	231/237	154/154	267/267	192/208	250/260	194/196
K6	144/154	151/155	196/196	261/269	224/224	108/111	134/142	276/280	139/141	145/145	246/246	154/160	264/279	192/208	250/260	194/196
K7	154/154	151/155	196/196	261/269	224/224	111/111	134/142	276/276	139/141	145/149	246/246	154/154	264/267	192/208	245/250	192/194
K8	144/154	151/151	196/196	253/261	224/224	108/111	134/142	276/276	139/141	145/145	237/246	154/160	264/267	192/211	250/260	194/196
K9	144/154	151/155	196/196	253/253	224/226	111/111	134/142	276/280	139/141	145/145	231/237	154/154	267/267	192/208	247/260	192/194
K10	154/154	151/155	196/196	261/269	224/224	111/111	134/142	276/280	139/141	145/145	231/246	154/154	267/267	192/211	245/247	192/194
K11	144/154	151/155	193/196	253/269	224/226	111/111	134/142	276/280	139/141	145/149	237/246	154/160	267/267	192/208	245/250	194/196
K12	144/154	151/155	193/196	253/269	224/226	108/111	134/134	276/276	139/141	145/145	231/237	154/154	264/267	192/208	245/247	194/196
K13	154/154	151/151	196/196	253/261	224/224	108/111	134/134	276/280	139/141	145/149	-	154/154	264/279	192/211	245/250	194/196
K14	154/154	151/151	193/196	253/269	224/224	111/111	134/142	276/280	139/141	145/149	231/246	154/154	267/267	192/211	250/260	194/196
K15	144/154	151/155	193/196	261/269	224/224	108/111	134/142	276/276	139/141	145/145	231/246	154/154	264/267	192/211	247/260	194/196
K16	144/154	151/155	196/196	261/269	224/226	108/111	134/142	276/276	139/141	145/149	231/246	154/160	264/279	192/211	247/260	194/196
K17	144/154	151/155	196/196	253/253	224/226	111/111	134/142	276/276	141/141	145/149	246/246	154/154	264/267	192/211	250/260	192/194
K18	154/154	151/155	193/196	253/269	224/226	111/111	134/142	276/276	139/141	145/145	231/246	154/160	267/267	192/211	250/260	192/194
K19	154/154	151/151	196/196	253/261	224/226	108/111	134/142	276/276	139/141	145/149	231/246	154/154	264/279	192/211	250/260	192/194
K20	144/154	151/155	193/196	253/253	224/224	108/111	134/142	276/276	139/141	145/149	231/246	154/160	264/267	192/208	245/247	192/194
K21	144/154	151/155	196/196	253/253	224/224	111/111	134/142	276/276	139/141	145/149	231/246	154/154	264/267	192/211	245/247	194/196
K22	154/154	151/155	193/196	253/269	224/226	111/111	134/142	276/276	139/141	145/145	231/237	154/154	267/279	192/208	245/250	192/194
K23	144/154	151/155	193/196	261/269	224/226	111/111	134/142	276/276	139/141	145/149	246/246	154/154	264/267	192/208	245/250	192/194
K24	154/154	151/151	193/196	253/253	224/224	111/111	134/142	276/280	141/141	145/145	237/246	154/160	267/279	192/211	250/260	192/194
K25	154/154	151/151	193/196	253/253	224/224	108/111	134/134	276/276	139/141	145/145	231/237	154/160	264/279	192/208	245/247	192/194
K27	144/154	151/155	193/196	253/253	224/224	111/111	134/142	276/276	139/141	145/145	231/237	154/160	264/267	192/211	250/260	194/196
K28	144/154	151/151	196/196	261/269	224/226	111/111	134/134	276/276	139/141	145/149	237/246	154/154	267/267	192/211	250/260	192/194
K29	144/154	151/155	196/196	253/253	224/226	111/111	134/142	276/280	139/141	145/145	237/246	154/154	267/279	192/211	245/250	192/194
K30	144/154	151/151	193/196	253/253	224/224	108/111	134/134	276/280	139/141	145/145	231/238	154/154	267/267	192/211	245/247	192/194
K31	144/154	151/155	193/196	253/261	224/224	108/111	134/142	276/276	139/141	145/149	231/246	154/154	264/267	192/208	245/247	194/196
K32	154/154	151/151	193/196	253/253	224/226	108/111	134/142	276/276	139/141	145/149	231/246	154/154	264/279	192/208	247/260	194/196
K33	154/154	151/151	193/196	253/269	224/224	108/111	134/134	276/276	139/141	145/145	231/246	154/154	264/267	192/211	245/250	192/194
K34	144/154	151/155	193/196	261/269	224/226	108/111	134/142	276/276	139/141	145/149	231/246	154/154	264/267	192/208	245/247	192/194
K35	154/154	151/155	196/196	253/269	224/226	108/111	134/142	276/276	139/141	145/149	237/246	154/154	267/267	192/211	250/260	194/196
K36	144/154	151/151	193/196	261/269	224/224	108/111	134/134	276/280	139/141	145/145	231/237	154/154	267/267	192/208	247/260	194/196
K37	154/154	151/151	196/196	253/253	224/226	111/111	134/142	276/280	139/141	145/149	231/237	154/154	267/279	192/208	247/260	194/196
K38	154/154	151/151	193/196	253/261	224/226	108/111	134/134	276/280	139/141	145/145	238/246	154/160	264/279	192/208	245/250	194/196
K39	144/154	151/155	193/196	253/261	224/226	111/111	134/142	276/276	139/141	145/149	231/246	154/154	264/279	192/208	247/260	192/194
K40	144/154	151/151	196/196	253/269	224/226	111/111	134/142	276/276	139/141	145/149	231/246	154/160	267/279	192/208	245/247	192/194
K41	154/154	151/151	193/196	253/261	224/224	108/111	134/142	276/280	139/141	145/145	237/246	154/160	267/267	192/211	245/250	194/196
K42	154/154	151/155	196/196	253/261	224/226	111/111	134/142	276/280	139/141	145/145	231/237	154/154	267/279	192/211	245/247	192/194
Mother	144/154	151/155	196/196	253/261	224/224	111/111	134/134	276/280	141/141	145/145	237/246	154/160	267/279	192/192	245/260	192/196
Father	154/154	151/151	193/196	253/269	224/226	108/111	134/142	276/276	139/139	145/149	231/246	154/154	264/267	208/211	247/250	194/194

Sample	Genotypes										
	BTPmo34	BTPmo35	BTPmo37	BTPmo38	BTPmo40	BTPmo41	BTPmo43	BTPmo44	BTPmo45	BTPmo46	BTPmo47
K1	121/123	131/131	230/253	211/214	301/301	288/299	294/297	156/157	243/251	157/159	258/266
K2	121/123	131/131	230/253	214/217	301/301	276/288	297/309	156/166	228/243	157/159	266/274
K3	121/121	126/130	230/253	211/214	301/301	255/299	294/297	156/166	243/251	157/157	258/266
K4	121/121	126/130	230/253	211/220	301/305	255/299	297/309	156/157	214/251	157/157	246/258
K5	121/123	126/131	230/253	214/217	301/305	288/299	297/297	157/166	228/243	157/159	266/274
K6	121/123	131/131	230/236	211/220	301/305	288/299	294/309	166/166	214/251	157/159	246/258
K7	121/121	131/131	230/253	211/214	301/305	255/276	294/297	156/157	243/251	157/159	258/266
K8	121/121	126/131	230/236	214/217	301/305	255/276	294/297	156/157	228/243	157/157	246/274
K9	121/123	131/131	230/253	214/217	301/301	288/299	297/297	157/166	228/243	157/157	266/274
K10	121/121	126/130	230/236	211/214	301/305	255/276	297/297	156/166	243/251	157/157	266/274
K11	121/121	131/131	230/236	211/214	301/305	276/288	297/297	156/166	243/251	157/157	258/266
K12	121/121	126/131	230/253	211/220	301/301	255/299	294/297	156/166	214/251	157/157	246/258
K13	121/121	126/130	230/236	211/220	301/301	255/299	294/309	166/166	214/251	157/159	246/258
K14	121/123	131/131	230/253	211/220	301/305	276/288	297/309	156/166	214/251	157/159	246/258
K15	121/123	126/130	230/236	211/214	301/305	276/288	294/297	156/166	228/251	157/157	258/266
K16	121/123	130/131	230/236	217/220	301/305	288/299	294/309	166/166	214/228	157/157	246/274
K17	121/123	126/131	230/253	214/217	301/301	255/299	297/297	157/166	214/251	157/159	266/274
K18	121/123	131/131	230/253	214/217	301/305	288/299	297/297	166/166	214/251	157/159	266/274
K19	121/123	126/131	230/253	211/220	301/301	288/299	294/309	156/157	214/251	157/157	246/258
K20	121/123	131/131	230/253	217/220	301/305	288/299	294/297	157/166	214/251	157/159	246/274
K21	121/123	131/131	230/236	211/220	301/301	276/288	294/297	156/157	214/228	157/159	246/258
K22	121/121	126/131	230/236	211/220	301/305	288/299	297/309	166/166	214/251	157/157	246/274
K23	121/121	131/131	230/253	211/220	301/301	255/299	294/297	156/166	228/243	157/161	246/258
K24	121/123	126/130	230/253	211/220	301/301	288/299	297/309	166/166	228/243	157/157	246/274
K25	121/121	126/130	230/253	211/214	301/305	255/299	294/309	156/156	243/251	157/159	258/266
K27	121/123	126/131	230/236	211/220	301/305	288/299	294/297	157/166	214/251	157/157	246/258
K28	121/121	130/131	230/253	211/214	301/301	255/299	297/297	156/157	214/251	157/157	246/258
K29	121/121	131/131	230/236	214/217	301/305	255/276	297/309	166/166	228/243	157/159	266/274
K30	121/121	130/131	230/253	217/220	301/301	255/299	297/297	156/166	214/228	157/157	246/274
K31	121/123	126/131	230/253	217/220	301/301	288/299	294/297	157/166	214/228	157/159	246/274
K32	121/123	126/130	230/236	217/220	301/305	276/288	294/309	157/166	228/243	157/157	274/274
K33	121/121	131/131	230/253	214/217	301/305	255/299	294/297	156/166	243/251	157/159	258/266
K34	121/123	131/131	230/253	217/220	301/301	288/299	294/297	166/166	214/228	157/159	266/274
K35	121/123	131/131	230/253	211/214	301/301	288/299	297/297	157/166	243/251	157/157	266/274
K36	121/121	126/130	230/236	214/217	301/301	255/299	297/297	157/166	228/243	157/157	266/274
K37	121/123	131/131	230/236	217/220	301/305	288/299	297/309	156/157	214/251	157/159	258/266
K38	121/121	130/131	230/236	217/220	301/301	255/299	294/309	157/166	214/228	157/159	246/258
K39	121/123	130/131	230/253	211/220	301/301	288/299	294/309	166/166	214/251	157/157	258/266
K40	121/121	130/131	230/253	211/214	301/301	255/299	297/309	156/166	-	157/159	258/266
K41	121/121	126/130	230/236	214/217	301/301	276/288	297/297	157/166	228/243	157/159	266/274
K42	121/121	131/131	230/253	211/220	301/305	255/276	297/309	157/166	214/251	157/157	258/266
Mother	121/121	130/131	230/230	211/217	301/305	276/299	297/309	156/166	214/243	157/159	246/266
Father	121/123	126/131	236/253	214/220	301/301	255/288	294/297	157/166	228/251	157/157	258/274

BIOGRAPHY

Mr. Chersak Maneeruttanarungroj was born on December 11, 1981. He graduated with the Bachelor of Science from the Department of Biochemistry at Chulalongkorn University in 2002. He has studied for the degree of Master of Science at the department of Biochemistry, Chulalongkron University since 2002.