CHAPTER IV

DISCUSSION

Since genetic markers are valuable tools for aquaculture improvement and stock definition and giant tiger prawn (*P. monodon*) is the most commercially important prawn species of Thailand fishery industry, this study aimed to potentially isolate and characterize the microsatellite loci in *P. monodon* genome that could be used as the powerful genetic markers. Our work began with the isolation of giant tiger prawn genomic DNA in order to construct a genomic library using a plasmid vector pUC18. The screening of the library for microsatellites containing clones was performed by colony hybridization with synthetic oligonucleotide $(GT)_{15}$ probe. Subsequently, nucleotide sequences of positive clones were determined by double strand alkaline sequencing. PCR primers were designed from the flanking regions of microsatellites. Allele length and polymorphism for each microsatellite locus were assessed for the development of genetic markers for use in DNA typing of *P. monodon*.

Many procedures have been reported for the extraction of chromosomal DNA from eucaryote. For example, a salting out procedure described by Millar et al. (1988) used high salt concentration to precipitate high molecular weight DNA. That procedure was initially used for the extraction of chromosomal DNA from the pleopods of *P. monodon*. The extracted DNA was contaminated with high amount of protein, thus made the DNA not suitable for cloning (Data not shown). The Proteinase K - phenol - chloroform extraction procedure described by Davis et al. (1986) had been successfully used to extract the honey bee, *Apis cerena*, chromosomal DNA (Uthaisang, 1994). The procedure was also successfully adapted for extraction of the DNA from *P. monodon* in this study. Good quality DNA obtained was suitable for molecular procedures for the development of microsatellite DNA markers. However, this procedure was time-consuming with long proteinase digestion and the organic solvent used was hazardous. The processing of the procedure would be a serious impediment for the studies involving population surveys or pedigree analyses by PCR which require the processing of hundreds of samples.

In the past, one approach to characterize microsatellites was to cloning and screening a large insert genomic library (Litt and Luty, 1989). The drawbacks of this approach were the requirement of many blot hybridizations and the difficulty of sequencing the relatively large subclones (Ostrander et al.,

1992). To characterize microsatellites in *P. monodon* genome, DNA fragments between 300 and 700 bp in length were chosen for cloning using a plasmid vector. DNA fragments of these lengths were short enough to be easily sequenced, yet relatively large enough to provide adequate sequence information up- and down-stream from an average-sized microsatellite region. However, the disadvantage of the small insert genomic was the frequency of $(GT)_n$ repeats in this library type was low (about 1 per more than 100 clones) (Ostrander et al., 1992). Consequently, large numbers of clones must be screened to obtain a significant pool of markers.

In the screening of microsatellites by colony hybridization, Denhardt's reagent was used to block the nonspecific attachment of the probe to the surface of the nylon membrane. Frequently, this agent was used in combination with detergent such as SDS and denatured and fragmented salmon sperm or yeast DNA (Sambrook et al., 1989). Because microsatellites were also found in those DNA as reported by Brooker et al. (1996) and Brown et al. (1995), high background of hybridization might occur. In this study, suppression of background hybridization was achieved by prehybridization the filters with a blocking reagent consisting of 2x Denhardt's reagent, 0.5% (W/V) SDS and 100 μg/ml yeast tRNA in place of denatured DNA.

Microsatellites have been investigated in several species of animal, but mainly in mammal. Very little information has been published about the sequence characteristics of invertebrates. Most of the studies in mammal have demonstrated that the dinucleotide repeat $(GT)_n$ predominated with the next abundant type being $(CT)_n$, such as in the study of rat and human genomes by Beckmann and Weber (1992) and the study of mammalian genomes by Moore et al. (1991) (cited in Ma et al., 1996). O'Reilly and Wright (1995) also concluded that $(GT)_n$ repeats were the most abundant repeats in the genomes of teleost fish with the second most abundant type being $(CT)_n$. For these reasons, screening was conducted with a synthetic $(GT)_{15}$ oligonucleotide probe. However, the results from *P. monodon* genomic library screening with synthesized oligonucleotide $(GT)_{15}$ in this study suggested that the second most dinucleotide repeat type was probably $(AT)_n$ microsatellites since more $(AT)_n$ microsatellites were picked up than $(CT)_n$ microsatellites. This was similar to the results of *P. monodon* genomic library screening with a $(CT)_{15}$ probe by Chaunchom Muanprasitporn (personnal communication, 1996) whose DNA sequencing results showed that $(CT)_n$ microsatellites in *P. monodon* appeared to be rare. In contrast to the microsatellites in other Arthopods, Estoup et al. (1993) reported that honey bee and bumble bee had $(CT)_n$ microsatellites 2.3 and 12.5 times, respectively, more abundant than $(GT)_n$.

A rough estimate of the average distance between $(GT/AC)_n$ arrays occurred in the genome of giant tiger prawn was calculated by dividing the total length of screened DNA by the number of isolated microsatellites, with the assumption that $(GT/AC)_n$ sequences were distributed evenly throughout the prawn genome. The average distance between neighbouring $(GT)_n$ microsatellites in *P. monodon* genome was 92.8 kb. This figure was compared to those reported for various invertebrate species, fish and mammal as shown in Table 3.4. It appeared that the density of $(GT)_n$ microsatellites in *P. monodon* was much less abundant than in mammals (every 18-46 kb) and insect (honey bee, every 34 kb). However, $(GT)_n$ repeats in *P. monodon* were about as abundant as those of Atlantic salmon (every 90 kb) and European flat oyster (every 139 kb).

Seventy seven clones harboring $(GT)_n$ repeats embedded in non-repetitive segment were classified as perfect, imperfect and compound microsatellites according to Weber (1990). The predominant category of dinucleotide repeat $(GT)_n$ sequences in *P. monodon* genome was imperfect.

The most common size-class in all categories contained sequences of 30-35 repeats in length. By comparing to microsatellites from honey bee (7-9 repeats), human (12-15 repeats), pig (16-18 repeats), Atlantic cod (30-35 repeats) and rainbow trout (24-29 repeats), the proportion of the different

classes of repeat types in *P. monodon* was significantly different in length and composition from mammal species. They were approximately twice as big as the size of the most common size class reported in mammalian genome and four times of those in honey bee genome. They resembled more closely those of the fish species when the definition of an imperfect repeat was altered to incorporate four or five intervening bases rather than three as defined by Weber (1990).

Considering that slightly different procedures have been used to isolate microsatellites from different species, it is possible that conditions of screening may have been biased towards longer arrays. In this study, a $(GT)_{15}$ oligonucleotide probe and higher stringency of hybridization and washing temperatures of 65°C were used. Screening at a lower stringency may enhance detection of both long and short arrays as well as more degenerate dinucleotide arrays. However, this is unlikely to be the case, since low stringency screening efforts for mammals, as used by Weber (1990) and Wintero et al. (1992) had failed to reveal microsatelllites of over 30 repeats in length.

It remains to be seen, with subsequent isolation of microsatellites from other species, whether poikilotherms consistently have longer arrays than homiothermic species. The longer, degenerate microsatellite arrays found in teleost fishes may be caused by a predisposition to slippage owing to

fluctuations in temperature. The DNA polymerase may not function efficiently when subjected to changes in temperature of several degrees, such as occur in the Atlantic cod environment and that of the salmonid species (Brooker et al., 1994). This can also be the reason for the longer repeat arrays observed in giant tiger prawn.

An event such as described above suggested that the microsatellites in giant tiger prawn and those of other poikilotherms evolved independently on the same or similar evolution forces. It is, therefore, not surprising that they all have longer repeat arrays of microsatellites. While different forces acted on mammals resulting in shorter arrays of microsatellites.

Although 131 microsatellite loci isolated in this study contained partial or complete sequence data, approximately 52 percent of them were found to be imperfect. According to Weber (1990), imperfections within the repeat sequences tended to reduce the informativeness of the resulting polymorphisms relative to what would be expected for the total number of repeats. There were also other reasons for low number of primers in this study resulting from the unsuitability in primer design from flanking regions. Firstly, there were more than 80 percent of microsatellite loci had at least one cloning site located adjecent to microsatellite sequence. Secondly, clones contained two or more microsatellites had flanking regions too short to use for a primer design.

Finally, flanking regions consisted mainly of the same bases as in the microsatellites. Therefore, PCR primers were only designed from the unique flanking sequences of 7 perfect microsatellites; 2 pairs for $(AT)_n$ and 5 pairs for $(GT)_n$ microsatellites. By testing 8 individual prawns from Kruntung (the Andaman Sea) and Angsila (the Gulf of Thailand), it was found that 3 out of 7 microsatellite primer pairs, namely, Pmo 18, Pmo 386 and Pmo 14, produced polymorphic fragments whose sizes were in the expected range while the others gave nonspecific amplifications.

The results of nonspecific amplifications, which were presented as multiple accessory bands, could still be detected at locus Pmo 1, 2, 62 and 380 after several optimization attempts (i.e. varying annealing temperatures or components of the PCR conditions). The occurrence of this problematic nonspecific bands could be explained as reasoned by O'Reilly and Wright (1995) that microsatellites frequently occur within or near large dispersed repetitive elements in humans (Beckmann and Weber, 1992) and Atlantic salmon (O'Reilly, unpublished data cited in O'Reilly and Wright, 1995). Therefore, if one or both of the PCR primers was located within the repetitive elements, either the pattern of nonspecific amplification or low yields of PCR products would be resulted.

In the present study, two polymorphic microsatellite loci, namely, Pmo18 and Pmo 386, exhibited unambiguously scorable alleles whose sequences differed only by even numbers of dinucleotide microsatellite sequences. This was undoubtedly from the reason that polymorphism of these loci was due to variation in the copy numbers of dinucleotide repeats. The appearance of shadow or stutter bands in those loci, which were visible as a ladder of bands, usually occurred in most dinucleotide microsatellite alleles, rather than a single discrete PCR product. Stutter bands were generally thought to be due to slippage-strand mispairing during PCR (Litt and Lutty, 1989; Tautz, 1989; Luty et al., 1990). The practical outcome of PCR stutter was that it caused problems scoring alleles. However, the alleles could be determined by analysing the number and relative intensity of bands in the stutter bands on the film (O'Reilly and Wright, 1995).

Although polymorphic products were obtained from Pmo 14 locus, the faint and ambiguous stutter ladders of PCR products resulting from the formation of primer-dimers artefacts caused the difficulty in allelic identification even with many optimization attempts and the unusual exposure time of one week. It is possible that this locus has a potential to be a genetic marker. However, with the possible high development cost and long time in optimizing Pmo 14 locus, the finding of new locus may be more worthwhile.

Locus Pmo 18 which contained 60 copies of dinucleotide repeats in the original clone, exhibited the highest number of alleles. This was concured with the study in human genome by Weber (1990) whose finding showed that marker informativeness clearly increased as the average number of repeats increased. It is, therefore, possible that this highly polymorphic microsatellite locus Pmo 18 can be used in a parentage analysis; i.e. the parents of superior progeny can be identified in mixed family rearing environments, thus enabling selective breeding to occur on commercial prawn farms.

From the results of Pmo 386 locus, 16 out of 19 individual templates were homozygous. The result can be explained by two possible reasons. First, the allozyme data in the population genetic study conducted by Sodsuk (1996) indicated that the low average heterozygosity of *P. monodon* was present in the samples collected from 11 different sites, from Kenya to Indochina to Philippines. However, with a few number of test samples used in this study (23 samples), it is possible that the results were biased. Secondly, as occurred in many studies of other species, there may be an occurrence of null alleles, the existing alleles in this microsatellite locus that are not observed using standard assays. Null alleles were caused by a mutation occurred in one of the priming sites (Koorey et al., 1993; Callen et al., 1993, cited in Pemberton et al., 1995). Unidentified polymorphism (base substitutions or deletions) at primer sites may

cause some microsatellite alleles to amplify poorly or not at all. These nonamplifying alleles in microsatellites have been identified in deer (Pemberton et al., 1995) with the detection through mismatchs between known mother-offspring pairs and by significant deviations from Hardy-Weinberg equillibrium. With a few samples in this study as mentioned above, null alleles could not be obviously identified. However, this should be a concern in future studies in population genetics.

In conclusion, the results suggested that the task of developing informative microsatellite markers for giant tiger prawn could be done with the available procedures. However, it may be more difficult than in other animals because of its microsatellite characteristics; imperfect and longer arrays. Two microsatellite markers were developed. The study showed that these markers were highly polymorphic and exhibited many alleles per locus which were suitable for future studies in population genetic, progeny identification and selective breeding program in this commercial prawn.