Chapter IV

Discussion

4.1 Microsatellite isolation

There is variation in the relative frequency of various microsatellite motifs in different species. To determine which motifs are relatively common in *P. monodon* genome, the abundance of different tri-and tetranucleotide repeats in the *P. monodon* genome were examined. The library A was constructed first and screened with 5 different microsatellite probes, $(GAA)_n$, $(GATA)_6$, $(GGAT)_6$, $(GGAA)_5$ and $(CACC)_5$. Results revealed that $(GAA)_n$ and $(GATA)_n$ repeats were particularly prevalent in the genome of *P. monodon* whereas $(GGAT)_n$ and $(GGAA)_n$ were rare and $(CACC)_n$ was not existent.

To isolate more microsatellite containing clones, libraries B, C and D were then constructed and screened with combination of abundant target repeat types, $(GAA)_8$+$(GATA)_6$ at low stringent conditions. Four additional repeat types, $(CAT)_n$, $(ATG)_n$, $(TCAG)_n$ and $(CATA)_n$ were also abundantly examined in *P. monodon* genome. Due to the combination of $(CAT)_8$, $(ATG)_8$, $(TCAG)_5$ and $(CATA)_6$ probes for library screening, a particular repeat type of positive clones could not be identified precisely. By sequencing positive clones, it was found that the percentage of sequenced clones that contained $(CAT)_n$ or $(ATG)_n$ were rare while none of sequenced clones contained $(TCAG)_n$ or $(CATA)_n$.

The number of positive clones depends on a relative proportion of target microsatellite repeats in the genome being studies (Khasa et al., 2000). Tiptawannukul (1996) have found 8.3% positive clones by hybridizing *P. monodon* partial genomic library with $(GT)_{15}$. As in other organisms, dinucleotide repeats are much more abundant than tri- or tetranucleotide repeats in *P. monodon* genome. In this study, library screening with $(GAA)_n/(GATA)_n$ repeats yielded positive clones between 0.24 to 1.32%.

From the sequencing data of all libraries, the library A yielded the highest percentage of clones containing microsatellites of the repeat type that was used to screen the libraries. Panaud, Chen and McCouch (1995) reported the optimal washing temperature for oligonucleotide hybridization. After washing the plaque lifts with 2X SSC, 0.1% SDS at Tm-8.2°C, more than 85% of sequenced clones contained the corresponding microsatellite sequences to the probe. In this study, high stringent conditions of hybridization and washing at Tm-5°C showed 76% of sequenced clones containing the corresponding microsatellite sequences to the probe. Low stringent conditions of hybridization and washing at Tm-10°C revealed that about 40% of sequenced clones from both libraries B and C contained the corresponding microsatellite sequences to the probe while that of library D contain only 9% of the corresponding microsatellites. Thus accurately identified microsatellite clones with the screening conditions in this study can be obtained using high stringent conditions.

Although libraries B, C and D were screened with the same conditions and probe, differences of proportion of positive clones among these libraries were found. The actual proportion of microsatellite DNA clones depended on the restriction

enzymes chosen and the sizes of the fragment from the particular genome. This was due to a nonrandom distribution of microsatellites (Smith and Devey, 1994). Khasa et al. (2000) isolated microsatellites with non-enriched library in alpine larch, *Larix lyallii* Parl. Only one microsatellite clone was obtained from *Sau*3 AI colonies screened with [$(CA)_{15}$+$(GA)_{15}$] probes. Apparently this was due to the fact that most *Sau*3AI fragments carrying $(AC)_n$ sites exist in the 1-6 kb range which were not available in 300-700 bp size fragments used for cloning. Furthermore, little methodologic error, especially temperature for library screening occurred in that study. Panaud et al. (1995) reported that the slope of the washdown curve around the Tm-7.6°C is the greatest of the 13 oligonucleotide used;$(GA)_{15}$, $(GT)_{15}$, $(CGG)_{11}$, $(ATC)_{11}$, $(TTG)_{11}$, $(TGG)_{11}$, $(CAG)_{11}$, $(TCT)_{11}$, $(ATT)_{11}$, $(CTTT)_8$, $(GATA)_8$, $(GAGG)_8$, $(CTAG)_8$, so small differences in the washing temperature are likely to cause large artifact. The small error of hybridizing and washing temperatures could cause large errors on microsatellite isolation.

Since less than 1.3% of clones with small inserts in conventional libraries were positive for the most common tri- and tetranucleotide repeats, $(GATA)_n$ and $(GAA)_n$. Enrichment protocols had also been developed to increase the proportion of clones in a given library containing the microsatellite motif of interest. Increasing in the proportion of recombinant clones containing microsatellite reduced the time and expense of screening large numbers of colonies (Gibbs et al., 1997). It is especially valuable for enrichment of lower copy number repeats such as tri- and tetranucleotide (Kijas et al., 1994). In this study, two methods were used to enrich microsatellites, selection of recombinant microsatellite clones from the genomic library and selection of DNA fragments containing microsatellites from genomic DNA.

The first enriched library comprised a significant number of redundant clones (86%). The problem with the enrichment protocols was that they required amplification of DNA molecules being screened. This causes many copies of each clone to be presented and become a problem when the same clones were isolated and sequenced multiple times. Therefore, a success of this enrichment requires that the pool of recombinant clones input from a conventional library into the enrichment process contains many different microsatellite loci. The proportion of redundant clones will then be minimized. Paetkau (1998) isolated microsatellites from enriched libraries based on microsatellite selection of recombinant clones from the M13 genomic library. The redundant clones did not cause a problem for $(AC)_n$ while that of $(ATGG)_n$ and $(ATAG)_n$ had this problem. The proportion of redundant clones should in general reflect the number of microsatellites in fraction of the genome, but many also vary between libraries. (Gibbs et al., 1997).

This enrichment protocol did not increase the numbers of clones in a library containing microsatellite motifs of interest when compared to conventional libraries. The difference of both enriched and conventional libraries is the method to isolate microsatellite-containing clones. The conventional method isolated microsatellite-containing clones from hybridization with oligonucleotide microsatellite probe while the enrichment protocol used the principle of streptavidin-coated magnetic bead capture of biotin-labeled probe hybridized to recombinant plasmid DNA from small insert genomic library. The advantage of this enrichment protocol permits isolation microsatellite-containing clones from a large number of clones of the initial library construction in less time than is required for the conventional protocol which involve many rounds of filter hybridization.

Microsatellite enrichment based on selection of recombinant DNA from genomic library yielded 34 unique clones. This enriched library was limited by the presence of microsatellite-containing clones in the initial library. An alternative enrichment method involves selection of microsatellite-containing genomic DNA fragments before cloning. This alternative method is based on a precloning enrichment of microsatellite sequences using synthetic oligonucleotide probes hybridized to complementary microsatellite core sequences in digested genomic DNA and binding to streptavidin-coated magnetic beads. Two libraries enriched for $(GAA)_n$ and $(GATA)_n$ were prepared. The redundant clones did not cause a problem in this enrichment protocol. It provided clones with different insert sizes. This was due to the fact that only a population of sequences enriched for microsatellites were cloned.

As determined by sequencing, the highest level of enrichment (97.1%) was achieved for $(GATA)_n$ selection of DNA fragments from the genomic DNA. Lower efficiency (64.71%) was obtained for $(GATA)_n$ selection of recombinant clones from the genomic library and 41.2% for $(GAA)_n$ selection of DNA fragments from the genomic DNA. For both $(GATA)_n$ enrichment, more than 64% of clones sequenced were contained microsatellites. This corresponds to at least 53-fold enrichment when compared to 1.21% of microsatellite-containing clones identified by screening a non-enriched genomic library. The enriched library had $(GAA)_n$ clone at frequency of 41% while only 0.24% was observed from a representative non-enriched library. Enrichment thus provided over 171-fold increase in the frequency of $(GAA)_n$ clones. Previous publication indicated the achieved enrichment rate about 20-70% depending on types of microsatellite sequences in the genome being studied (Kijas et al., 1994; Fischer and Bachmann, 1998; Cordeiro et al., 1999; Kolliker et al., 2001).

4.2 Characteristic of microsatellite loci

In this study, several repeat types were found in *P. monodon* genome. Dinucleotide $(GA)_n$ repeats were found in high proportion. The presence of $(GA)_n$ repeats indicated a high occurrence of this kind of repeat within the genome of *P. monodon*. However, data in this study contradict with those previously reported where $(GT)_n$ were found to be the most abundant microsatellites among dinucleotide repeats. Tassanakajon et al. (1998) isolated $(TG)_n$ and $(GA)_n$ from partial genomic libraries of *P. monodon*. Results indicated that $(TG)_n$ repeats were more abundant than $(GA)_n$ repeats. Similarly, Brooker et al. (2000) isolated $(TG)_n$ and $(GA)_n$ from the partial genomic library of *P. monodon* and found that $(TG)_n$ were the most abundant type, followed by $(TA)_n$ and $(GA)_n$, respectively. Xu et al. (1999) directly sequenced clones from the conventional library of *P. monodon* and found that $(TG)_n$ repeats were the most abundant followed by $(TA)_n$ and $(GA)_n$ repeats. In this study, about 87% of sequenced clones were screened for $(GAA)_n$ and $(GATA)_n$ repeats. The abundance of $(GA)_n$ over $(TG)_n$ may be due to the fact that $(GA)_n$ repeats often link with $(GAA)_n$ and $(GATA)_n$ repeats. Among clones containing microsatellite repeats which did not correspond to the probes used for library screening, $(GA)_n$-containing clones were found the most frequent in the library screened with $(GAA)_n$, $(GAA)_8$ and $(GATA)_6$. $(GA)_n$ repeats were not predominant in the library screened with the remaining 7 oligonucleotide probes. Another reason for the high frequency of $(GA)_n$ repeats may have resulted from cross-hybridization between $(GAA)_n$/$(GATA)_n$ probes and $(GA)_n$ repeats. So the abundance of $(GA)_n$ over $(TG)_n$ and $(TA)_n$ repeats in this study may have resulted from the biases. From the sequencing data, $(TAA)_n$ repeats were found

the most frequency for trinucleotide repeats among non-corresponding microsatellite to the probes.

For tri- and tetranucleotide repeats, $(GAA)_n$ and $(GATA)_n$ were the most abundant. From the sequencing data, $(GAA)_n$ and $(GATA)_n$ were 3.6-7.6 times more abundant than $(ATG)_n$ and $(CAT)_n$. This number is overestimated because larger numbers of clones were screened for $(GAA)_n$ and $(GATA)_n$ repeats when compared to $(ATG)_n$ and $(CAT)_n$ repeats.

Genomic library screening with $(TCAG)_5$, $(CATA)_6$ and $(CACC)_5$ did not show any positive clone of these microsatellites. Upon sequencing of 152 clones, these repeat types were not found. In this study, all type of dinucleotide repeats were found except $(CG)_n$ repeats. Similarly, Xu et al. (1999) did not find $(CG)_n$ repeats in the 83 sequenced clones derived from a *P. monodon* genomic library without probe screening.

The average repeat length of di-, tri- and tetranucleotide microsatellites were 24.5, 15.8, 7.9, respectively in conventional libraries while that of respective microsatellites were 28.0, 19.1 and 20.8 in enriched libraries. Notably, average repeat length of tetranucleotide repeats found in the enriched libraries was 2.6 times greater than that of conventional libraries. This was resulted from the very large tetranucleotide repeats isolated from the $(GATA)_n$ library selected DNA fragments from the genomic DNA. Disregarding this, the average repeat length of tetranucleotide repeats was the shortest similar to that of the conventional libraries.

The most common repeat length of $(TG)_n$ and $(CT)_n$ in *P. monodon* reported by Tassanakajon et al. (1998) were 30-35 and 12-17 repeats, respectively. The maximum array size found in *P. monodon* microsatellite reported by Tassanakajon et al. (1998) was 371 bp. In this study the maximum array size was 264 bp. The average repeat length of white pine for $(TG)_n$ was 13.6 repeats (Echt et al., 1996) whereas that for dinucleotide repeats of polychaete was 16.6 for imperfect repeats and 12.0 for perfect repeats (Weinmayr, Vautrin and Solignac, 2000). The average repeat length of white clover for dinucleotide repeats was 12.3 and that of trinucleotide repeats was 29 repeats (Kolliker et al., 2001). Gibbs et al. (1997) reported that the average repeat length of chicken was 16.2 for tri- and tetranucleotide repeats and 20.3 for dinucleotide repeats. Therefore, dinucleotide repeats found in *P. monodon* were larger than those reported in other species. Among all libraries in this study, there were differences in the average repeat length of microsatellite sequences. The use of different screening conditions and differences in the construction of the libraries have affected the length of obtained microsatellite loci (Temnykh et al., 2000).

Classification of microsatellite repeats based on Weber's criteria revealed that perfect microsatellites were the predominant categories. For all conventional libraries, 46% of the repeat sequences were perfect, 31% were imperfect and 23% were compound. For all enrichment libraries, 69% of the repeat sequences were perfect, 20% were imperfect and 11% were compound. These results were similar to a previous report in *P. monodon* (Xu et al., 1999) where perfect microsatellite was found to be the most abundant in *P. monodon*. Perfect microsatellites were also the most abundant class in *P. vannamei* (Garcia and Alcivar Warren, 1996), fish and other vertebrate species (Cooijmans et al., 1997). Compound repeats associated with either

mono-, di-, tri-, tetra-, penta- or hexanucleotide repeats and those associated with dinucleotide repeats were the most frequent. Both conventional and enriched libraries $(GA)_n$ and $(AT)_n$ were the most frequent repeat associated with compound microsatellites. Tassanakajon et al. (1998) reported that $(GT)_n$ repeats were frequently associated with $(AT)_n$ repeats in compound microsatellites of *P. monodon*. In polychaete, the most abundantly associated microsatellite were also dinucleotide repeats, which associated with $(AT)_n$ (Weinmayr, Vautrin and Solignac, 2000). In wheat $(GT)_n$ microsatellites were most frequently associated with $(GA)_n$ microsatellites, $(TCT)_n$ microsatellites with $(TGG)_n$ microsatellites (Ma, Roder and Sorrells, 1996).

4.3 Efficiency of marker development

Of 225 microsatellite loci, 27 loci were suitable for primer designs. These 27 usable loci were derived from 25 clones. Therefore, this means that, 16.44% of the sequenced clones was promising to be useful genetic markers. Usable microsatellites in *P. monodon* were very difficult to obtain owing to the large and complex repeat arrays. Many clones contained only one side of the unique flanking sequences make them unavailable for primer design. Many microsatellites contained short repeat units but those short 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.

In total, 27 primer pairs were tested, twenty-six primer pairs (96%) provided amplified fragments of the expected size in *P. monodon* based on sequence data.

Moore et al (1999) reported that isolation of microsatellite markers in *P. japonicus* was very inefficient due mainly to the highly repetitive sequences in the genome. Most genomic clones contained very long repeats and unique sequence on either side of repeat could not be obtained. Out of 226 sequenced clones, only 13 primer pairs were designed accounting for 5.75% of the sequenced clones that give rise to useful genetic markers. Brooker et al. (2000) reported that out of 28 sequenced clones of *P. monodon*, only 3 microsatellite loci were usable markers with polymorphic products. This means that 10.71% of sequenced clones give useful genetic markers.

When compared the efficiency of microsatellite marker development of *P. monodon* with other species such as *Pectnaria koreni, Dicentrarchus labrax*, the proportion of usable clones varies enormously between species. Approximately 30-55% of sequenced clones of those species were successful for primer designs. (Garcia de leon et al., 1995; Byran et al., 1997; Pestsova, Ganal and Roder, 2000; Weinmayr, Vautrinamd Solignac, 2000).

The development of microsatellite markers from enrichment by selection of DNA fragments from the genomic DNA was not successful due to complexity of the protocol used for enrichment. The largest proportion of usable clones per sequenced clones was obtained from the $(GAA)_n$ library selected DNA fragments from the genomic DNA (4 in 34) but only 1 primer pair supported amplification of the *P. monodon* genomic DNA. The proportion of usable clones per sequenced clones of the $(GATA)_n$ library selected DNA fragments from the genomic DNA was 4 in 68 but there was no primer pair supported the amplification success of *P. monodon* genomic DNA. The failure in amplification of microsatellite markers developed from

enrichment approach based on selection from genomic DNA fragments occurred from unauthentic insert DNAs. These DNA fragments were created during amplification of inserts with PCR before ligated into a vector. With the compatible ends, microsatellite-containing fragments were jointed and amplified creating large inserts of unauthentic DNA sequences. As a consequence, the average repeat length of tetranucleotide repeats in this enriched library was much larger than those found in the conventional libraries. The average repeat length of tetranucleotide repeats from the latter was between 6-22 repeats while that of $(GATA)_n$ selected DNA fragments from the genomic DNA was 31 repeats. Moreover sequences of either adapters occurred within the insert sequences. To decrease unauthentic DNA sequences, the amount of ssDNA for a PCR reaction of the library of $(GAA)_n$ selection was decreased 5 times when compared to the $(GATA)_n$ library selection. All sequenced clones from $(GAA)_n$ selection contained no adapter sequences within the insert. The majority of repeat length of sequenced clones from $(GAA)_n$ selection was short. However, only 1 in 5 designed primers supported the amplification success of genomic DNA in expected size range. The failure to amplify *P. monodon* genomic DNA still occurred.

To improve the enrichment protocol, cycles of PCR reaction for amplification of insert DNA should be decreased from 25 cycles for 1-5 cycles to avoid high concentration of microsatellite containing DNA fragments in a PCR reaction that may cause amplification of the unauthentic DNA template. In this study, no attempt was made to improve this enrichment protocol by decreasing PCR cycles. Generally, PCR cycles for amplification of insert DNA before ligation to a vector were 20-25 cycles. (Kijas et al., 1994; Burland, Barratt and Racey, 1998; Piertney et al., 1998).

The proportion of usable clones per sequenced recombinants obtained from the $(GATA)_n$ library selected recombinant clones from the genomic library was 3 in 34 sequence clones. Out of 3 primer pairs designed, 2 supported the amplification success of genomic *P. monodon* DNA. In summary, only 3 usable markers were developed from the enriched libraries.

4.4 Polymorphism analysis of microsatellite loci

Twenty-six microsatellite loci were characterized for genetic information content by genotyping *P. monodon* broodstock from Trad (the Gulf of Thailand). Among these, 5 were monomorphic among the tested individual (n=10-20). Microsatellite primers are locus-specific and are, therefore, generally consider to be single-locus primers (Prasad et al., 2000). Each of 21 polymorphic microsatellite was locus-specific.

The di-, tri- and tetranucleotide polymorphic markers that have been isolated included, perfect repeats (CUPmo 1, 9, 12, 13, 15, 16, 19, 20, 23); imperfect repeats (CUPmo 11, 14, 17, 21, 22, 24, 25); and compound repeats (CUPmo 2, 3, 4, 6, 7). Allele differences between consecutive alleles of microsatellite loci used in this study depended on repeat types. For the CUPmo 1, 11, 13, 15, 16, 17, 19, 21 and 24 (trinucleotide markers), consecutive alleles differed by 3 bases. For the CUPmo 9, 12, 14, 20, 22, 25 (dinucleotide markers) and the CUPmo 2, 3, 4, 6, 7 [$(GATA)_n$ linked to the different types of dinucleotide repeats], alleles were different by multiples of 2 bases. This suggests that polymorphism detected should occur within dinucleotide and (or) tetranucleotide repeats. For the CUPmo 23 (a tetranucleotide marker), consecutive alleles differed by 4 bases.

Analysis of polymorphic tri- or tetranucleotide microsatellites has the advantage of easier scoring of alleles because the alleles of tri- or tetranucleotide repeats showed lesser stutter bands than dinucleotide repeats. The appearance of shadow or stutter bands has been shown to be due to slipped-strand mispairing during PCR. Stutter bands tend to decrease with increasing unit length (Edwards et al., 1991). Among 21 polymorphic markers, 48, 28 and 24% were tri- or tetranucleotide, dinucleotide and compound of tetra- and dinucleotide microsatellites, respectively. All except 2 CUPmo3 and 25, polymorphic microsatellite markers generated clean and unambiguous allelic patterns. Different microsatellite allelic patterns of the CUPmo 3 and CUPmo 25 loci within heterozygotes resulted in difficulty of allele interpretation. Twenty-one polymorphic microsatellite markers exhibited high level of polymorphism.

The number of alleles at each of 21 polymorphic loci ranged from 4 to 33 with an average of 17.80. Sizes of allele length ranged from 83-450 bases. The average of observed and expected heterozygosities were 0.70 and 0.84, respectively. 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 in the present study ranges from 0.42 to 0.95 with an average of 0.82. Under PIC value, (1) PIC >.5 means locus with highly informative (2) 5>PIC>.25 means locus with reasonably informative (3) PIC<.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* microsatellite markers is consistent with that observed by Xu et al. (1999). They reported average PIC values at 7 loci of 0.80 among a set of 20 prawn individuals. The locus CUPmo 7 showed low level of heterozygosity of 0.21 when 43 individuals *P. monodon* from Trad were investigated. Low observed heterozygosity found at the CUPmo 7 locus may possibly be resulted from null alleles (Koorey et al., 1993). Cho et al. (2000) found uninformative of tri- and tetranucleotide loci for rice microsatellites. In this study, all of tri- and tetranucleotide loci showed high level of polymorphism.

The level of polymorphism, as estimated by the number of alleles, showed a positive correlation with the average number of microsatellite repeats. The PIC values showed a similar relationship with repeat length. It has been shown in early studies in human (Weber, 1990) and further confirmed in several different organisms, that variability of microsatellite markers correlates well with the length of tandem arrays. Although some authors have suggested that compound repeats may be less variable than single repeats (Rassmann et al., 1991). The 5 compound loci assayed here proved to be highly variable in *P. monodon*.

The 21 polymorphic microsatellite loci in this study showed high level of polymorphism. Among highly polymorphic microsatellites, some microsatellites displayed extremely high level, not all do. These microsatellite loci exhibit variation in the level of polymorphism, thus, offer opportunity to select for microsatellite loci that exhibit the appropriate level of variation for a given research problem. The loci with extreme high level of polymorphism are the best for paternity and parentage that require Mendelian inheritance. Genetic mapping also requires loci with extreme high

level of polymorphism. Microsatellite loci that exhibit only 3 to 5 alleles may be preferred for population studies (Wright and Bentzen, 1994). Among 21 polymorphic microsatellite loci, 18 were usable loci. The remaining 3 loci were excluded from this study due to the ambiguous allelic patterns (CUPmo 3, CUPmo 25) and the very low level of observed heterozygosity (CUPmo 7).

Out of 21 polymorphic microsatellite loci, 9 were significantly deviated from Hardy-Weinberg expectations due to homozygote excess. Similar deviations from Hardy-Weinberg equilibrium at microsatellite loci have been reported for whiting (*Merlangius merlangus*, European sea bass (*Dicentrachus labrax*), Atlantic cod and rock fish (*Sebastes thompsoni*) (Bentzen et al., 1996; Garcia de Leon et al., 1997; Rico et al., 1997; Sekino et al., 2001). Supungul (2000) investigated genetic variation and differentiation of *Penaeus monodon* from 5 geographic locations of Thai waters using 5 microsatellite loci. In 19 of 25 possible tests, significant deviation from the Hardy-Weinberg expectations was observed following correction for multiple tests using the Bonferroni procedure ($P<0.001$). However, Mendelian segregation of these loci was conformed using pedigree sample ($P>0.05$).

Deviation from the expectations may be explained by several reasons. First, "null" (non-amplifying) alleles may be present. "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. Callen et al., 1993; Paetkau and Strobeck, 1995). A second artifact of microsatellite genotyping, particularly for analysis of

dinucleotide-repeat loci, is the appearance of "stutter" bands or ladders of bands typically smaller than the band of interest. Stutter bands complicate scoring and may lead to misclassification of heterozygotes for closely sized alleles as homozygotes with phenotypically similar band-patterns (O'Reilly and Wright 1995). Excess homozygosity may be observed if sampling ignores genetic structure within the sample sites. Variability in allele frequencies among year classes can also produce an apparent homozygosity excess (Bagley, Lindquist and Geller, 1999).

The presence of null alleles at the CUPmo 7 locus and difficult allele interpretation at the CUPmo 3 locus should have resulted in Hardy-Weinberg disequilibrium at these loci. The remaining 7 loci are difficult to conclude causes that the disequilibrium represented. Among these 7 loci, 3 loci show stuttered bands. Alleles of these loci were determined on the basis of the highest intensity, and therefore, accurate scoring and genotype assignation should not interfere to a large extent. In contrast, sampling errors of *P. monodon* from Trad ignoring subpopulations within a population may be the care. Specimens from this geographic sample may contain a mixture of adults from separate breeding stocks as a result of extensive transplantation of *P.monodon* larvae in Thailand annually. Generally, each wild *P. monodon* breeding female may spawn more than once a year and produce 248,000 to 811,000 eggs per spawning and this enormous number of eggs may be sufficient to replace part of the entire adult population (Supungul et al, 2000). Recently, Klinbunga et al. (2001) found biased female gene flow between Trad and Chumphon *P. monodon*. This may response for disequilibrium pronuced in *P. monodon* from Trad analysed by microsatellites in this study.

4.5 Multiplex analysis of microsatellite loci

Up to date, a few *P. monodon* microsatellites have been published (Tassanakajon et al., 1998; Xu et al., 1999). Most of penaeid microsatellite markers did not cross-species amplification (Moore et al., 1999). The limitation of available potential markers made it difficult for selections of compatible loci well-suited for multiplexing systems. As mention above, in this study, microsatellite loci for multiplex systems were chosen from 21 available markers, which developed in this study. The system for genotyping by these multiplex amplification sets were radioisotopic visualization of alleles length polymorphism in a single lane. Using a single labeled colour system, the range of allele lengths associated with each of these loci limits the number of loci that can be electrophoresed in a single lane (Wright and O'Reilly, 1995). The most critical step in the establishment of multiplex PCR is to choose the correct primer combinations and PCR conditions. In addition, varying salt concentration and annealing temperature has been found to be useful to increase specificity in a multiplex PCR (Schlotterer, 1998). In this study, up to 4 loci can be separated simultaneously. By selecting microsatellite loci with non-overlapping allele size, similar primer annealing temperature, and minimal interaction of the primer sets, 4 multiplex sets could be developed including 1 tetraplex, 2 triplex and 1 duplex, respectively. Twelve out of 21 available microsatellite markers were used to organize multiplex sets. The remaining microsatellite markers were not selected for multiplex analysis due to overlapping allele size and incompatible of microsatellite loci.

Optimization of multiplex systems was performed, first using exactly the same PCR condition as with a single locus PCR. No multiplex set could be amplified under

these PCR conditions all 4 multiplex sets had to be further adjusted in PCR programs and components.

The multiplex PCR set A was a model for further optimization of multiplex systems in *P. monodon*. Changing the extension temperature had a major effect on the amount of the amplified products of larger loci. The CUPmo 15, the largest allelic size, gave weak products for extension at 72°C. At 65°C, it, however, gave strong products. Thus, lowering the extension temperature provided an advantage to increase amounts of amplified large alleles. To increase the amount of large products, the extension time was increased from 2 to 4 min. Increasing the extension time in the multiplex PCR consistently increased the amount of larger products. In multiplex PCR, as more loci are simultaneously amplified, the pool of enzyme and nucleotides becomes a limiting factor and complete polymerization of the products require longer extension time (Henegariu et al., 1997). Shorter annealing time, at 30 sec eliminated large nonspecific products, which occurred from interaction between loci to be co-amplified. When the CUPmo 11 locus was combined into the duplex PCR, KCl concentration needed to be increased. Many primer pairs producing shorter amplification products worked better at higher salt concentrations. An increase in salt concentration make longer DNA denature slower than shorter DNA, so shorter molecules will be preferentially amplified (Henegariu et al., 1997). Equimolar primer mixture usually did not provide optimal amplification at all loci, this test allowed the observation that too high or too low yields of some loci need to be optimized. Changing the proportion of various primers in the reaction had been done by increasing the quantity of primers for the loci providing the weak signal and decreasing the quantity of primers for the loci providing the strong signal. Multiplex

amplification resulted in different amounts of the amplification products. It is possible that differences in the ability of primers to find the target sequences and to properly anneal to them will eventually result in different efficiency of amplification where more efficient primer set prevails (Bercovich et al, 1999).

Non-specific products also occurred from coamplification of multiple loci in the same reaction. Some of these non-specific were fixed so they could complicate genotypic interpretation. And individual loci could be specifically amplification at different annealing temperature. Touchdown PCR was used instead of the regular PCR profiles. Touchdown PCR allow loci of varying optimal annealing temperature to be amplified simultaneously while suppressing the production of spurious artifact bands. This allowed only expected allele to be amplified (although some non-specific were still present in small quantities) (Rithidech, Dunn and Gordon, 1997; Fishback et al., 1999; Dean and Milligan, 1998). Some loci such as the CUPmo 15 and 16, non-specific products were shown in a single locus PCR but became invisible or less when the multiplex reaction and touchdown PCR were performed.

Amplification of the other 3 multiplex set, B, C and D were followed similar to the optimization procedure described for the multiplex PCR set A. All 4 multiplex sets were successfully amplified in the same PCR conditions, except that the amounts of primers and the annealing temperature depended on the nature of each primer set. It is possible that this standard protocol for amplification of multiplex sets obtained from this study may be applied to other sets of multiplex PCR. Only the ratio of primers and annealing temperature have to be optimized. This approach allows the ease to optimize of those new multiplex sets. These 4 multiplex PCR genotyping systems for

amplification of 12 microsatellite loci provide a rapid and powerful method for genotyping analysis in *P. monodon*. They generate consistent and reliable genotypic data. This approach facilitate the analysis of large samples of *P. monodon*.

4.6 Silver staining detection of microsatellite amplification

Typing of microsatellites typically requires separation of PCR products by denaturing polyacrylamide gel electrophoresis. The radioactive labeled method was commonly used for detection of microsatellite alleles by autoradiography after electrophoresis because of their sensitivity and efficiency. However, this method is expensive, hazardous and not appropriate in some laboratories. The non-isotopic methods, for example the silver staining detection, have been used as an alternative choice of the radioisotopic method.

Allelic patterns of the multiplex PCR set B (CUPmo 4+13+19) from silver staining were similar to those of the isotopic detection. However, the latter produced an overall cleaner appearance than the former method. Two fragments were revealed for each allele at loci CUPmo 13 and 19 for silver staining while isotopic detection revealed only 1 band for each allele. This causes by differential electrophoretic mobility of two strands on a denaturing gel. The degree of differential migration of the opposite strands depends on a partial amplified sequences being separated. Thus, it varies with locus being amplified, with distance of fragments have been subjected to electrophoresis, with the concentration of gel matrix, to a certain degree and with the selection of primer for amplification. Strand-specific staining is not possible for silver staining so two fragments are seen for each allele. With isotopic detection, only one of

the two amplified product strands is labeled and displayed. This simplifies interpretation in that one fragment per allele is observed with isotopic rather than the two fragments per allele often seen with silver staining. Apparently, alleles of CUPmo 13 and 19 were easy to score with silver staining. However, it was difficult to assign alleles of CUPmo 4 when this staining method was used. Dinucleotide repeat loci generally reveal extensive repeat slippage. Detection of dinucleotide alleles with silver staining system produces several minor bands from stuttering of dinucleotide repeats and detection both strands of amplification products. For this reason, allele of CUPmo 4 compound of tetra- and dinucleotide repeats became more difficult to determine with the silver staining detection method. CUPmo 13 and 19 were trinucleotide loci which are less prone to stuttering. Only 2 bands for each allele were observed at loci CUPmo 13 and 19. In general, the silver staining detection method can be used for typing of microsatellites. Some but not all of microsatellite loci were difficult to determine alleles therefore results should be determined with cautions (Lins et al., 1996; Schlotterrer, 1998).

4.7 Application in shrimp genome mapping

Microsatellite loci with high polymorphism are desirable for genetic mapping. Loci with many alleles and a PIC near 1 are most desirable. Genotypes of either father or mother at the interested locus have to be heterozygous genotype. Among 10 microsatellite chosen for genetic mapping, number of allele varied from 8-29 and PIC were between 0.75-0.95. The number of markers for genetic mapping depends on the size of genome. In human, about 150 RFLP markers would be required (Botstein et al., 1980). Chow et al (1990) estimated the genome size of four penaeid species by

flow cytometry to be approximately 70% that of the human genome. Thus, high number of microsatellite loci need to coverage the genome of *P. monodon*. Moreover, intergeneric differences in prawn (25%) were more divergent than between some orders of mammals (23%). The extent of genome divergence is a likely cause of difficulties in applying microsatellites isolated from one penaeid species to other penaeid taxa.

Isolating a large number of usable microsatellites from penaeid species was extremely difficult. In addition, the large number of chromosomes and the relatively large size of the genome and the considerable effort will be required to develop the basic information for assisting prawn genetic improvement such as genome mapping with microsatellite markers.

Seven microsatellite loci were mapped into linkage groups with one or more markers. Microsatellite loci which segregated into the same linkage group, were closed one another together on the chromosome. The closer two loci are on the chromosome, the higher the probability that the two alleles on this chromosome will appear together upon segregation. CUPmo 1 and CUPmo 9 were mapped in more than 1 linkage group. The reason is differences of segregation between male and female *P. monodon*. This was corresponded to deviation from Hardy-Weinberg expectation at the CUPmo 1 locus. Although the CUPmo 9 locus was in agreement with the expectation. P-value of this locus on Trad *P. monodon* 0.0039 where the critical value to reject the null hypothesis was 0.0024. Increasing a larger number of specimens in the analysis would provide more accurate conclusion. If CUPmo 1 and CUPmo 9 loci are, however, located on the sex chromosome of *P. monodon*, homozygote excess is

obviously observed. Causing Hardy-Weinberg disequilibrium upon the number of homogametic and heterogametic individuals used in the experiments.

Moore et al. (1999) constructed AFLPs-based map of *P. japonicus*. There were 129 markers included in 44 linkage groups. The average length of linkage group was 29. For *P. monodom* map, there were 160 markers included in 60 linkage groups. The average length of linkage group was 25.9. In rice, full genome microsatellite-based maps are available. Microsatellite markers are evenly distributed throughout the genome and provided good distribution along 12 chromosomes with an average density of one microsatellite loci for every 6 cM. Genetic maps with saturation of useful markers is quite easy to find markers linked commercially important trait (Temnykh, 2000).

For the prawn genome mapping project, the addition of sufficient number of new markers is required to populate several regions previous lacking in molecular markers. It is then intended to increase the density of this framework map using Type I (coding) markers that can be used across families and can be directly used in selective breeding programs.