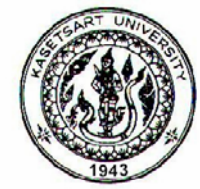


THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Agricultural Biotechnology)
DEGREE

Agricultural Biotechnology
FIELD

Interdisciplinary Graduate Program
PROGRAM

TITLE: Characterization of EST-linked Microsatellites and Centromere Mapping in Günther's Walking Catfish *Clarias macrocephalus*

NAME: Ms. Chantapim Sukkorntong

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR
(Assistant Professor Supawadee Poompuang, Ph.D.)

THESIS CO-ADVISOR
(Professor Uthairat Na-Nakorn, Ph.D.)

THESIS CO-ADVISOR
(Mr. Prapansak Srisapoome, Ph.D.)

GRADUATE COMMITTEE CHAIRMAN
(Associate Professor Pongthep Akratanakul, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN
(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

CHARACTERIZATION OF EST-LINKED MICROSATELLITES AND CENTROMERE MAPPING IN GÜNTHER'S WALKING CATFISH *CLARIAS MACROCEPHALUS*

CHANTAPIM SUKKORNTONG

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Agricultural Biotechnology)
Graduate School, Kasetsart University
2009

Chantapim Sukkorntong 2009: Characterization of EST-linked Microsatellites and Centromere Mapping in Günther's Walking Catfish *Clarias macrocephalus*. Master of Science (Agricultural Biotechnology), Major Field: Agricultural Biotechnology, Interdisciplinary Graduate Program. Thesis Advisor: Assistant Professor Supawadee Poompuang, Ph.D. 93 pages.

Expressed sequence tag (EST) databases potentially are a valuable source for developing gene-associated microsatellite markers. EST sequences of fish tissues are found to contain microsatellite repeats, particularly in the untranslated regions (UTRs). A total of 2,029 *Clarias macrocephalus* EST sequences were screened for di-, tri-, tetra-, and pentanucleotide repeat, 113 of which contained microsatellite repeats. Forty-three loci (37.7%) contained dinucleotide repeats, 44 loci (38.5%) contained trinucleotide repeats, 22 loci (19.3%) contained tetranucleotide repeat motifs and four loci (3.5%) contained for pentanucleotide repeats. EST sequences of known genes indicated that 32 microsatellite loci were found in the 3′ UTR region, seven loci in the 5′ UTR region and 13 loci in the open reading frame (ORF) of known genes, e.g., vitellogenin, myosin light chain, troponin, and parvalbumin. Primers were designed and synthesized for 41 loci. Fourteen loci were polymorphic with the number of alleles ranging from 2-15 alleles per locus and the observed and expected heterozygosities ranging from 0.47 to 1.0 and from 0.427 to 0.8819 per locus respectively. Eleven microsatellites contained dinucleotide core sequences, two loci contained trinucleotide repeat motifs and one locus contained a pentanucleotide repeat motif. Cross-species amplifications of ten primer pairs were observed in African catfish *Clarias gariepinus*, five in striped catfish *Pangasius hypophthalmus* and black ear catfish *P. larnaudii* and four in Mekong giant catfish *Pangasianodon gigas*.

Eleven EST-linked microsatellites and 33 microsatellites derived from genomic DNA were mapped in relation to their centromeres in two gynogenetic diploid families of walking catfish. Twenty-six loci showed high microsatellite-centromere recombination with a frequency greater than 0.67 and three loci displayed recombination frequencies greater than 0.9. The recombination frequency data suggested that microsatellites were randomly distributed within genome of walking catfish and supported the use of these markers for constructing linkage maps. Gene-centromere distances calculated under the assumption of complete interference, ranged from 2.15 cM to 46.8 cM with average distance of 31 cM. Mapping information obtained in this study proves useful for improving the initial linkage map of walking catfish.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deep appreciation to my advisor, Assistant Professor Dr. Supawadee Poompuang for her teaching, helpful guidance and encouragement which enable me to carry out my work successfully. I am also very grateful to Professor Dr. Uthairat Na-Nakorn and Dr. Prapansak Srisapoome, my co-advisor, for helpful discussion for the completion of my thesis.

I would like to thank Ms.Srijanya Sukmanomon, Dr. Dutrudi Panprommin, and all members in a laboratory of Aquaculture Genetics at Department of Aquaculture, Faculty of Fisheries, for sharing useful ideas and helping my experiments. My sincere thanks are extended to all members in a laboratory of Fish Molecular Biology (A311), A312 and my lovely friends (Jay, Au, Pong, King, Tai, Tik, Teuy, Keang and others) for their kindness and friendship.

This project was partially supported by the Center of Excellence on Agricultural Biotechnology, Postgraduate Education and Research Development Office, Commission on Higher Education, Ministry of Education, and the Graduate School, Kasetsart University.

Finally, I wish to express my deepest gratitude to my parents and all members in my family included Bow and Nam for their love, understanding, support and encouragement, which enable me to accomplish my study.

Chantapim Sukkorntong

April, 2009

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp	=	Base pairs
cM	=	Centi Morgan
cm	=	Centimeter
dNTPs	=	Deoxydinucletide triphosphate
df	=	Degrees of freedom
EDTA	=	Ethylenediamine tetraacetic acid
HCl	=	Hydrochloric acid
M	=	Molar
µg	=	Microgram (s) (10^{-6})
µl	=	Microliter (s)
µM	=	Micromolar
mg	=	Milligram (s)
ml	=	Milliliter (s)
mM	=	Millimolar
$MgCl_2$	=	Magnesium chloride
ng	=	Nanogram (s) (10^{-9})
NaCl	=	Sodium chloride
PCR	=	Polymerase Chain Reaction
pmol	=	Picomole (s) (10^{-12})
pH	=	Logarithm of reciprocal of hydrogen (H) ion
Rep. Type	=	Repeat type
Seq ID	=	Sequence Identification
SSRs	=	Simple Sequence Repeats
SSRs_Seq	=	Simple Sequence Repeats Sequence
Taq	=	*Thermus aquaticus*
TBE	=	Tris-borate EDTA buffer
TE	=	Tris-EDTA buffer
TNES	=	Tris-NaCl-EDTA-sodium dodesylsulfate
Tris	=	Tris (hydroxylmethyl) methylamine

CHARACTERIZATION OF EST-LINKED MICROSATELLITES AND CENTROMERE MAPPING IN GÜNTHER'S WALKING CATFISH *CLARIAS MACROCEPHALUS*

INTRODUCTION

Genetic mapping is an integral part of genome research which the main purpose is to learn about the genetic structure of the chromosome. A genetic map of an organism is a model showing the linear arrangement of a group of genes or genetic markers along the chromosome. Knowledge of gene linkage has led to the idea of making a genetic map based on recombination of homologous chromosomes during meiosis. Microsatellites have become the most widely used DNA markers for the construction of genetic maps in various species of animal and plant. Genetic map is essential for quantitative trait loci (QTLs) mapping and the application of marker-assisted selection.

Gene-centromere mapping is the technique used to determine the position of a centromere on the genetic map. In species with a small number of available genetic markers, it is difficult to construct a genetic map, but the distances of these markers from their centromere can be estimated. The method of gene-centromere mapping is derived from tetrad analysis in simple eukaryotic organism such as fungus, *Neurospora*. In fish, gene-centromere mapping can be practiced using gynogenesis techniques to produce gynogenetic diploid families. Gene-centromere mapping provides a powerful approach for construction of female linkage maps.

Walking catfish (*Clarias macrocephalus*) is one of the important food fish species in Thailand. Walking catfish has superior flesh quality and commands high retail prices. Although culture of walking catfish has been well established in Thailand for over 30 years, the potential of the industry is limited by its slow growth rate and disease susceptibility. From the aquaculture point of view, molecular approach could be applied for genetic improvement of walking catfish. The application of this method, however, requires information of its genome including a

genetic map and gene sequences. The first generation of genetic map based on AFLP markers has been generated for walking catfish, but the marker density was low on the map. The map consisted of 31 linkage groups while walking catfish has 27 chromosomes. The number of linkage groups is greater than chromosomes indicating that there are at lease four gaps in the map. To fill these gaps in the linkage map, a large numbers of markers are needed so that the number of linkage groups would equal the number of chromosomes.

Although microsatellite loci have been isolated for walking catfish, the number is too small for improving the resolution of the existing map. A search of the GenBank database revealed that, only 67 microsatellite sequences were published for *C. macrocephalus*. This lack of marker data has been a major constraint to effectively utilize genetics information for research on walking catfish genome.

Recently, gene sequence information obtained from expressed sequence tags (ESTs) of liver and muscle tissues has been published for walking catfish. EST databases potentially are a valuable source for developing type I microsatellite markers. EST sequences of fish tissues are found to contain microsatellite repeats, particularly in the untranslated regions (UTRs). Because mutation occurs more frequent in non-coding sequences of the genome than those in coding sequences, it is more likely that microsatellites found within UTRs of ESTs will be polymorphic. Most studies suggested that isolation of EST-linked microsatellite loci can provide an effective way to develop polymorphic microsatellite markers.

In this research, I characterized new microsatellite loci by searching the walking catfish EST database. Primers were developed and tested for their informativeness and application in population and mapping studies. Developed primers were also tested for cross-species amplification in other clariid and pangasiid catfishes. These EST-linked microsatellites and other microsatellites derived from genomic DNA library were mapped to their centromeres in gynogenetic diploids.

OBJECTIVES

1. To search for microsatellites from the EST sequences of muscle and liver tissues of walking catfish and develop primers for PCR amplification.

2. To characterize the EST-linked microsatellite loci and test for cross-species amplification.

3. To estimate microsatellite-centromere distances based on half-tetrad analysis in walking catfish diploid gynogenetic families.

LITERATURE REVIEW

Microsatellite DNA

Microsatellites have been the most widely used as marker of choice in various fields of genetic research. They become a powerful tool for generating genetic linkage maps, assessing genetic diversity of wild and cultured stocks of aquatic species as well as monitoring of genetic change in selection program. This wide range of applications is due to their abundance, neutrality, co-dominant expression, high levels of polymorphism, and PCR-based analysis. There are different names to refer to microsatellites such as simple sequence repeat (SSR), simple sequence length polymorphism (SSLP) and sequence-tagged microsatellite site (STMS). Microsatellites are non-coding repetitive DNA regions, comprising tandem repeated sequences of one to six nucleotides with the number of repeats ranges from 8 to 40 copies. Most of microsatellites (30-70%) found in genome of vertebrates are di-nucleotide repeats such as $(AC)_n$, $(AT)_n$, and $(CG)_n$ (Neff and Gross, 2001). Microsatellites with tri-, tetra-, and penta-nucleotide repeats are found at lower frequencies than the di-nucleotide repeats. Different number of repeat units results in microsatellite diversity among individuals in species or populations. The DNA replication slippage is thought to be the predominant mutation mechanism generating microsatellite variability. This mutation process occurs at the repetitive sequences when the new strand mis-pairs with the template strand, altering the repeat number of microsatellites. Microsatellites are distributed throughout the genome including gene-coding regions, introns and non-coding regions (Varshney, 2005).

Microsatellites are generally classified as type II markers because they are associated with anonymous genomic regions (O'Brien, 1991). The commonly used way to identify and characterize microsatellite loci is through the construction of small-insert enriched genomic libraries. Microsatellites are identified by screening genomic clones with oligo-nucleotide probes and their flanking regions are sequenced for the design of PCR primers. Conservation of microsatellite loci are often found among closely-related species. Cross-species amplifications of microsatellites have

been demonstrated in various species of teleost fish but the percentage of loci that amplify successfully may decrease with increasing genetic distance, for example, between *C. macrocephalus* and *C. batrachus* (Yue *et al*., 2003), and between clariids and pangasiids (Poompuang and Sukmanomon, 2003).

EST-linked Microsatellites

Most of microsatellites are type II markers for which no known function have been established. For microsatellites located in gene sequences, they are classified as type I markers (O'Brien 1991) which are useful for comparative gene mapping to study genome evolution (Liu *et al*.,1999; Siemon *et al.*, 2005; Yue, 2004). The expressed sequence tag (EST) approach provides researchers with a quick and effective way for identifying type I microsatellites. The expressed sequence tag is single-pass, partial sequences of randomly selected cDNA clones (Adams *et al*., 1991). These clones derived from mRNA in specific cells at specific development stage. Cloned cDNAs would be directionally sequenced from either the 5′ ends which contain more protein coding sequences or the 3′ ends, which often contain untranslated regions (UTRs). EST sequences of fish tissues are found to contain microsatellite repeats, particularly in the non-coding sequence. Because mutation occurs more frequent in non-coding sequences of the genome than those in coding sequences, it is more likely that microsatellites found within non-coding regions of ESTs will be polymorphic (Liu *et al*., 1999).

Expressed sequence tag (EST) databases potentially are a valuable source for developing gene-associated microsatellite markers. For example, approximately 1,753,760 EST sequences have been developed for zebrafish *Danio rerio*, 281,544 ESTs for rainbow trout *Onchorhynchus mykiss*, 617,585 sequences for Japanese medaka *Oryzias latipes*, 494,560 for Atlantic salmon *Salmo salar*, and EST 483,105 sequences for channel catfish *Ictalurus punctatus* (http://www.ncbi.nlm.nih.gov/). The EST-linked microsatellites can be searched for in these databases using computer programs. Locus-specific primers can be designed for PCR amplification. The abundance of type I microsatellites in fish genome has been reported from various

species. In channel catfish *Ictalurus punctatus*, analysis of 43,033 ESTs identified 4,855 (11.2%) EST containing microsatellites, of which 4,103 were unique sequences (Serapion *et al*., 2004). In Atlantic salmon *Salmo salar*, Vasemagi *et al*. (2005) identified 1,154 EST-associated microsatellites from 58,146 ESTs. A total of 75 loci were polymorphic and these microsatellite markers were tested for cross-species amplification in five other species from the family salmonidae brown trout, *Salmo trutta*; rainbow trout, *Oncorhynchus mykiss*; Arctic charr, *Salvelinus alpinus*; European grayling, *Thymallus thymallus*; and common whitefish, *Coregonus lavaretus*), resulting in the transferability of 133 gene-associated markers. In other study, Ng *et al*. (2005) identified 1,975 microsatellite sequences by screening the Genome Research on Atlantic Salmon Project (GRASP) EST database. Primers for seventy-six loci were developed. Cross-species amplification of these markers with other salmonids was also reported. Yue *et al*. (2004) identified 31 microsatellites from a testis cDNA library of common carp (*Cyprinus carpio*). These gene-associated microsatellites exhibited cross-species amplification in crucian carp (*Carassius auratus gibelio*). In turbot (*Scophthalmus maximus*), Chen *et al*. (2007) isolated twelve polymorphic microsatellite loci from 3,356 EST sequences from a spleen cDNA library. They also reported cross-species amplification of these marker loci in five other fish species including Japanese flounder (*Paralichthys olivaceus*), half-smooth tongue sole (*Cynoglossus semilaevis*), barfin flounder (*Verasper moseri*), sea perch (*Lateolabrax japonicus*) and red sea bream (*Chrysophrys major*).

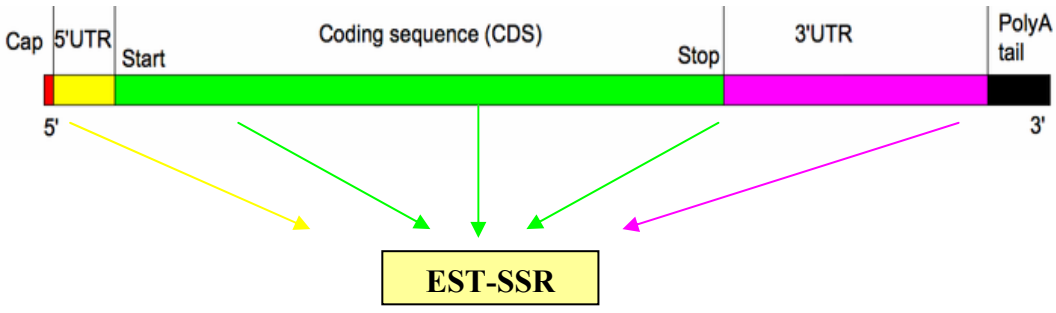


Figure 1 The locations of microsatellites within cDNA sequence (5′ and 3′ UTRs, and coding sequence (CDS) or open reading frame (ORF).

Centromeres and chromosomes

Centromeres are parts of eukaryotic chromosomes which play important role during mitosis and meiosis. They are attachment sites for spindle fibers that facilitate the separation and the movement of chromosomes to the opposite ends of the cell. The separation or disjunction of chromosomes is critical to the successful sexual reproduction of organisms. During the first and second meiotic divisions, improperly functioning centromeres result in non-disjunction for which gametes receiving the wrong number of chromosomes.

Centromeres appear as constricted region on each chromosome and they are most easily visualized during mitosis. Chromosomes can be classified by arm ratio and the position of the centromere as metacentric, submetacentric, acrocentric, and telocentric. For example, a chromosome with its centromere right in the middle and having two arms of same length is called metacentric. If the arms' lengths are unequal, the chromosome is submetacentric. If the centromere location is close to the end of the chromosome, it is acrocentric. One with its centromere located at the end of the chromosome is called telocentric.

The occurrence of crossovers during meiosis is highly dependent on the structure of a chromosome which can affect genetic map distances. For example, short arms and long arms display different levels of recombination (Danzmann and Gharbi, 2001). A method called gene-centromere mapping has been developed for estimating distances of genes from their centromere. It also allows researchers to determine the position of a centromere onto the genetic map and to assess chiasma interference along the chromosomes.

The principle of gene-centromere mapping is based on ordered tetrad analysis of widely used eukaryotic organisms such as yeast *Saccharomyces*, fungi *Neurospora*, and algae (Copenhaver, 2000). During sexual reproduction, all four haploid cells (tetrads) from a single meiotic division are contained in a single structure and remain separate from products from the other meiotic division. In the first meiotic division, the centromeres of homologous chromosomes separate from each other following

crossing-over between two strands of non-sister chromatids (half tetrad). The separation of sister chromatids during the second meiotic division results in four haploid cells (tetrads). The proportion of haploid genotypes is dependent on recombination between genes and their centromere. Analysis of ordered tetrads allows us to directly identify cross-over events and determine gene-centromere distances. In other organisms, gene-centromere mapping is based on half-tetrad analysis for which two meiotic products (or two from four strands) from a single meiosis are recovered together. In this case, analyzing one of the two products of the first meiotic division is analogous to ordered-tetrad analysis (Streisinger *et al*., 1986). Figure 2 illustrates the idea using half-tetrad analysis for mapping centromere. In the absence of crossing over, the two homologous chromosomes containing different alleles separate in the first division and all half-tetrad progeny are homozygous. When crossing over occurs in the diploid gamete between the locus and centromere, the sister chromatids are no longer identical. The two alleles do not segregate in the first meiotic division and all half-tetrad progeny are heterozygous.

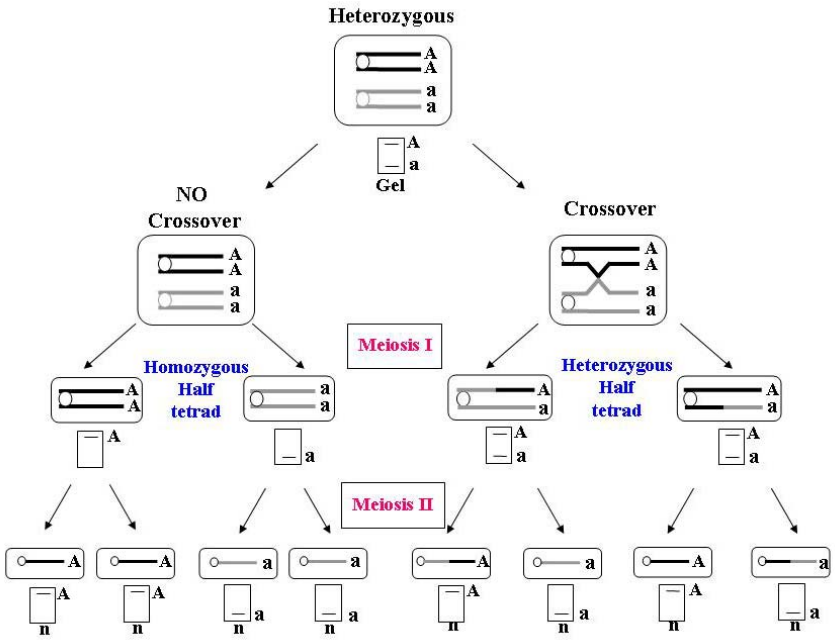


Figure 2 Half-Tetrad analysis

Source: modified from Kauffman *et al*. (1995)

In fish and amphibians, two products of second meiotic division including the second polar bodies and the secondary oocytes can be recovered using gynogenesis techniques (Thorgaard *et al*., 1983). To produce diploid gynogenetic embryos, eggs are fertilized with UV-irradiated sperm and subjected to either, heat shock, cold shock, or pressure shock to suppress the release of second polar bodies. As a result, the fertilized egg contains an egg haploid nucleus and a second polar body haploid nucleus which fuse to form a diploid nucleus and a diploid zygote (Figure 3).

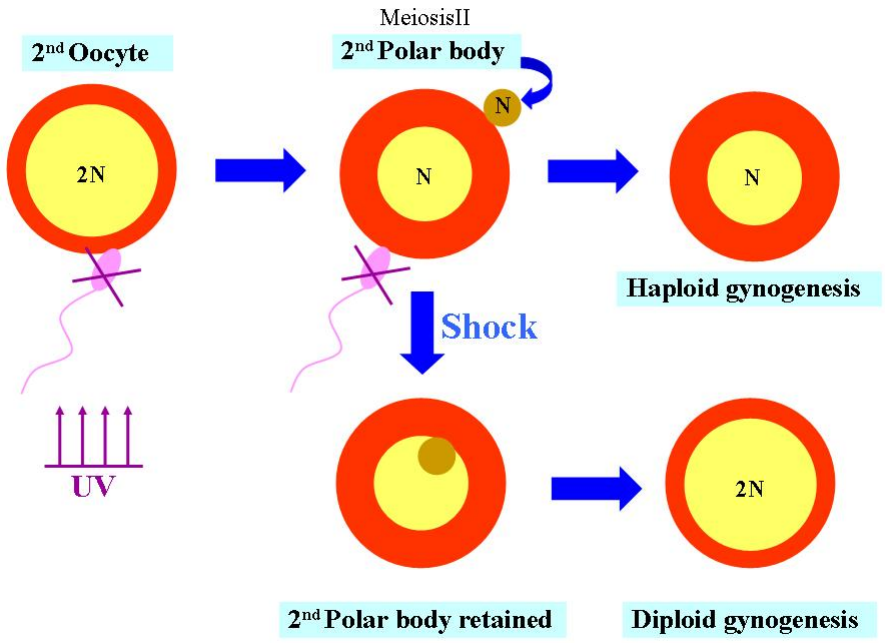


Figure 3 Induction of gynogenesis in fish.

Estimating gene-centromere distances

Most of statistical methods for centromere mapping are based on single locus gene-centromere linkage analysis. It is generally assumed complete chiasma interference or only one crossover takes place between non-sister chromatids (Thorgaard, 1983). Genes close to centromeres tend to segregate during the first meiotic division following crossover whereas the segregation of more distal genes occurs in meiosis II. The genotypes of gynogenetic progeny from a heterozygous female can be used to determine crossover event between the particular locus and its

centromere. In the absence of crossover, the genotype of the progeny will be homozygous. Alternatively, the genotype of the progeny will be heterozygous if there are crossovers. The gene-centromere distance can be estimated from the crossover frequency between the locus and the centromere by counting the number of heterozygous half-tetrads. The proportion of heterozygous progeny, therefore, represents the frequency of second division segregation (*y*) (Streisinger, 1986). For example, all progeny will be homozygous for a given genetic marker ($y = 0$) with no crossover between the locus and its centromere. All gynogenetic progeny will be heterozygous ($y = 100\%$), if a single crossover always takes place. The recombination fraction *y* of 2/3 suggests that the locus and centromere are unlinked. The presence of chiasma interference will result in *y* values greater than 2/3 (Kauffman, 1995).

The gene-centromere distance (x) in centiMorgans can be estimated using an appropriate mapping function. Three mapping functions have been described under different assumptions. The gene-centromere distance, $x = \frac{y}{2}$, if there is complete interference. For Kosambi's function, $x = \frac{-\left[\ln(1-y) - \ln(1-y) - \ln(1-y)\right] \times 100}{4}$ if 50% interference is assumed. For Haldane's equation which assumes no interference, $x = \frac{-\left[\ln(1-y)\right] \times 100}{2}$ (Li *et al.*, 2005).

For multi-locus gene-centromere mapping, Zhao and Speed (1998) developed a method based on chi-square models for half-tetrad analysis. The models were tested using data from rainbow trout (Thorgaard *et al*., 1983) for ordering markers in relation to centromeres. The estimated map distances using Zhao and Speed's model were in agreement with that of Thorgaard *et al*. (1983). In mammalian species, Da *et al*. (1995) described a maximum likelihood approach for multi-locus gene-centromere mapping using first polar bodies and secondary oocytes. The methods were developed for all possible gene-centromere orders and take into account all crossover possibilities.

Gene-centromere mapping in fish

Fish provides an attractive system for gene-centromere mapping because chromosome manipulations have been successfully applied in many freshwater and marine species. Most gene-centromere mapping studies for fish employed gynogenetic diploids whereas the applications of triploids were also reported in a few species, e.g., medaka (Sato *et al*., 2001), and Japanese eel *Anguilla japonica* (Nomura *et al*., 2006).

Early work for centromere mapping in fish used allozymes, which have been the most widely employed protein markers for genetic research since 1970's. The applications of allozymes for mapping centromere, however, have some limitations due to the small number of available loci and their low levels of polymorphism. For example, only ten allozyme loci were mapped to their centromeres in gynogenetic rainbow trout *Oncorhynchus mykiss* (Thorgaard *et al*., 1983). The development of polymerase chain reaction (PCR) technology has allowed highly polymorphic DNA markers such as microsatellites, as well as RAPDs, AFLPs, to become powerful tools for gene-centromere mapping in addition to allozymes. In pink salmon *Oncorhynchus gorbuscha*, a total of 312 gene loci including nine allozyme loci, 34 microsatellite loci, 168 AFLPs and 101 anonymous DNA loci were mapped to their centromeres (Lindner *et al*., 2000). Johnson *et al*. (1996) have placed centromeres on the RAPD-based linkage map of zebrafish and consolidated the number of linkage groups to the number of chromosomes (25).

Gene-centromere mapping allows assessment of chiasma interference which determines the accuracy of gene-centromere distances. Chiasma interference refers to the event in which the occurrence of one crossover in a certain region prevents the occurrence of other crossovers in the adjacent region (Mohideen, 2000). If interference is complete, i.e., with only a single crossover event between non-sister chromatids, gene-centromere map distance will equal recombination frequency. In contrast, with the occurrence of multiple crossovers along the length of a chromosome, the recombination frequency is an underestimate of map distance.

Chiasma interference was found to be a common phenomenon for fish chromosomes as compared to other animal and plant species. Gene-centromere studies from various species of teleosts have reported high levels of chiasma interference with the *y* values (the proportion of heterozygous gynogenetic progeny) of marker loci greater than 0.667, i.e., rainbow trout (Thorgaard *et al*., 1983), ayu *Plecoglossus altivelis* (Taniguchi *et al*., 1987; Seki *et al*., 1989), channel catfish *Ictalurus punctatus* (Liu *et al*., 1992), tilapia *Oreochromis niloticus* (Hussain *et al*., 1994), zebrafish *Danio rerio* (Kauffman *et al*., 1995; Mohideen *et al*., 2000), pink salmon *Oncorynchus gorbuscha* (Lindner *et al*., 2000), Japanese eel *Anguilla japonica* (Nomura *et al*., 2006), yellow croaker *Pseudosciaena crocea* (Li *et al*., 2008), barfin flounder *Verasper moseri* (Lahrech *et al*., 2007) and turbot *Scophthalmus maximus* (Martinez *et al*., 2008).

Genome information of walking catfish

Despite being one of the best-studied fish in Thailand, research on walking catfish genome and other molecular investigations are still in early stage. The karyotype of walking catfish comprises 27 chromosome pairs with no morphologically distinct sex chromosomes (Donsakul and Magtoon, 1989), of which 15 are metacentric, six are submetacentric, two are subtelocentric, and four are acrocentric. The genome size in terms of recombination units (centiMorgan) of this fish is not known at present. Although a genetic map based on AFLP markers has been generated, the marker density was low on the map (Poompuang and Na-Nakorn, 2004). The map consisted of 134 loci placed into 31 linkage groups with the number of markers per linkage group varies from 3 to 14. The map spans an estimated 2,037.4 Kosambi cM, with an average spacing of 17.07 cM.

Because DNA markers are powerful tools for genome research and other genetic investigations, microsatellite loci have been isolated for walking catfish (Na-Nakorn *et al*., 1999; Sukmanomon *et al*., 2003; Yue *et al*., 2003), of which 67 sequences were reported on the GenBank database. The number of available microsatellite markers, however, is too small for further investigation of its genome. To increase information of walking catfish genes, Panprommin *et al*., (2007)

characterized 2,029 expressed sequence tags (ESTs) from muscle and liver cDNA libraries of adult female fish. These sequences have been deposited in the GenBank nucleotide database (dbEST) (http://www. ncbi.nlm.nih.gov/dbEST/index.html). A total of 1,334 EST clones showed significant sequence similarity to known genes in the databases, representing 303 genes from the liver library and 234 genes from the muscle library.

Chromosome manipulation techniques including gynogenesis and triploidy have been developed for walking catfish for the purpose of genetic improvement of this species (Na-Nakorn *et al.*, 1994). In addition, gynogenetic techniques provided a powerful tool for gene-centromere mapping in this species (Na-Nakorn *et al*., 2006). Results of gene-centromere mapping of five isozyme loci and four microsatellite loci in walking catfish gynogens were consistent with other studies in fish genomes where high recombination rates were observed.

MATERIALS AND METHODS

1. Searching Microsatellite-containing ESTs

A total of 2,029 *C. macrocephalus* EST sequences from a previous study (Panprommin *et al*., 2007) were screened for di-, tri-, tetra-, and pentanucleotide repeat using the Modified Sputnik II program (http://wheat .pw.usda.gov/ IDTMI/EST-SSR/LaRota/) (Varshney *et al*., 2005). The criteria used for searching microsatellite repeat were five repeats for di-, tri, and tetra-nucleotide and three for pentanucleotide. The program displayed the repeat type, the starting point, the ending point, sequence length (base pair), the repeat motif and the microsatellite sequence, respectively.

Because EST databases of walking catfish are highly redundant, microsatellite-containing ESTs were clustered using the Clustal W program (http://www.ebi.ac.uk/Tools/clustalw2/index.html) with 40-bp overlap and 95% identity criterion to identify homologous loci including contigs and sigletons. Contigs represent genes whose transcripts were sequenced more than once. Singletons represent genes whose transcripts were sequenced only once (Serapion *et al*., 2004).

2. Design of EST-linked Microsatellite Primers

Primers were designed for the flanking regions of each microsatellite locus by using commercially available software "OLIGO ^{TM} program" (National Bioscience, Inc.). One designed primer pair was for one locus. The high quality primers were selected from a number of designed primer pairs based on the following criteria: (1) the primers should have more than 50% for G-C content (2) if possible, G or C should locate at the end of the 3′OH region (3) forward primer does not match with reverse primer (4) sequence length of the primers should range from 18 to 25 base pairs and (5) forward and reverse primers should have nearly melting temperature (T_m) because T_m was needed to calculate appropriate annealing temperature for PCR reaction (Kamonrat, 1996).

3. Primer Analysis

3.1 DNA Extraction

DNA extraction was performed following the standard phenol-chloroform method (Taggart *et al*., 1992). DNA was extracted from fin clips of adult fish and stored at -20^{o}C before use. Quality and quantity of DNA were measured by the spectrophotometer. The quality of DNA was determined by comparing the proportion of A_{260}:$A_{280.}$ The values ranged from 1.65 to 1.85 were described as clean DNA. The value greater than 1.85 indicated RNA contamination. The value less than 1.65 indicated protein contamination (Kirby, 1992).

3.2 PCR Reaction

PCR amplification was performed in a 10 µl reaction volume as follows:

10 µg/µl of genomic DNA as a template	2.0 µl
5X PCR Buffer	2.0 µl
1.5 mM of MgCl2	0.6 µl
0.2 µM each of dNTPs	0.4 µl
0.5 µl M each of forward and reverse primers	1.0 µl
1U of Taq Polymerase	0.1 µl
dH_2O	3.9 µl

The PCR profile was as follows:

Step 1 pre-denaturation	95°C	5	min
Step 2 denaturation	95°C	30	sec
Step 3 annealing*	__°C	45	sec
Step 4 extension	72°C	1	min
Step 5 polishing	72°C	5	min
Step 6 cooling	4°C		

* Annealing temperature was locus-specific which can be calculated from the following equation:

$$T_A = [4(G+C)+2(A+T)]-5$$

The PCR profile was: pre-denaturation at 95°C for 5 min; then 35 cycles for denaturation, annealing, and extension; followed by 1 cycle for polishing in a PCT-100 Programmable Thermal Controller (MJ Research). Following amplification, reaction products were mixed with sequencing dye (0.1% bromophenol blue, 0.1% xylene cyanol, 2% of 0.5 M EDTA pH 8.0 and 98% formamide). The reaction mixtures were heated for 5 min at 95°C and eletrophoresed on 4.5% denaturing polyacrylamide gel in 1x TBE running buffer at 60 W for 2-3 h (depending on the size of PCR product) in sequencing gel apparatus (Bio-Rad). Gels were denatured at 100°C for 30 min before electrophoresis. Gels were visualized by silver staining. Allele sizes were determined according to an M13 sequence ladder.

3.3 Characterization of EST-linked microsatellite loci

Primers were tested on genomic DNA from thirty unrelated fish sampled from a hatchery population to assess the levels of heterozygosities, to test for deviation from Hardy-Weinberg expectation and to determine linkage disequilibrium between loci using the software POPGENE v.1.32 (Yeh *et al*.; 2000). The exact *P* values for tests of Hardy-Weinberg expectation were calculated by a Markov chain randomization method using ARLEQUIN version 3.11 (Excoffier *et al.*, 2005). A sequential Bonferroni correction method was used to adjust significance levels for multiple tests (Rice, 1989). Test for linkage disequilibrium was performed by comparing each microsatellite loci pair using Markov chain (Dememorization = 1000, batchs = 100 and Iteration per batch = 1000). The sequences of all microsatellites were compared for similarity with the sequences previously submitted to the GenBank database using BLAST programs (Altschul *et al*., 1990) at the National Center for Biotechnological Information (NCBI, http//www.ncbi.nlm.nih.gov).

BLASTn program was used to search the nucleotide sequence homology. Significant homology would be considered only when *E*-values were less than 10^{-4}.

3.4 Cross-species Amplification

For cross-species amplification, fin clips were collected from five individuals of other four species of catfishes including the African catfish *Clarias gariepinus*, the striped catfish *Pangasius hypophthalmus*, the Black-ear catfish *P. larnaudii* and the Mekong giant catfish *Pangasianodon gigas*. DNA samples were isolated from the fin clips using a phenol-chloroform method as described in 3.1. PCR amplification was conducted in a 10 µl reaction volume, using the same master mix and the PCR profile for the amplification as described in 3.2.

3.5 Mendelian Inheritance and Segregation Analysis

Primers were tested on genomic DNA from twenty fish sampled from two control families (C-1 and C-2) of a hatchery population at the Department of Aquaculture, Faculty of Fisheries, Kasetsart University. Genotypes of progeny were tested for deviation from Mendelian expectations using chi-square test (χ^2) as follows:

$$\chi^2 = \Sigma \frac{(O - E - 0.5)^2}{E}$$

Where,

O is the observed value.

E is the expected value.

0.5 is the correction factor for adjusting χ^2 value when using the small numbers of sample.

4. Production of Gynogenetic Diploids

Ten female walking catfish and a male striped catfish were obtained from a domesticated stock kept at the Department of Aquaculture, Faculty of Fisheries, Kasetsart University. Gynogenesis was performed following the method described by Na-Nakorn *et al*. (1993). The females were hand-stripped 10 hours after injection with 25 μg/kg of LHRH analogue (Suprefect™) and 5 mg/kg of domperidone (Motilium™). Milt was collected from a male striped catfish four hours after injection with 10 μg/kg of LHRH analogue (Suprefect™) and 5 mg/kg of domperidone (Motilium™), and diluted to 1:1000 in Ringer's solution. Diluted milt was placed into Petri dishes and irradiated using a 30-W UV germicidal tube at a dose of 10.8 x 10^4ergs/mm^2. Irradiation was done at a distance of 30 cm. and lasted for 2 min. The irradiated sperm and the eggs were mixed and subjected to a cold shock at 7°C for 4 min. Fertilized eggs were incubated in flowing water at 28°C and fry were collected four days after hatching. Genomic DNA was extracted from fin clips of a female and 50 whole fry from each family using a standard phenol-chloroform method (Taggart *et al*., 1992) as described in 3.1.

5. Microsatellite-centromere Mapping

5.1 Genotyping

Two gynogenetic families consisting of a female and 50 offspring were genotyped at 55 microsatellite loci identified from two different approaches.

1) Fourteen EST-linked microsatellite primers designed from the EST sequences of *C. macrocephalus* deposited in the GenBank EST databases (Panprommin *et al*., 2007).

2) Forty-one microsatellite primers previously developed from genomic library of *C. macrocephalus* including 11 loci identified by Sukmanomon (*et al*. 2003)

and 30 loci (S. Poompuang unpublished data). The microsatellite sequences, the Accession number, and annealing temperatures were shown in Appendix Table 1.

5.2 The Proportion of Heterozygotes (*y*)

A ratio of 1:1 between two classes of homozygotes at each locus in two gynogenetic families was tested using a chi-square test The proportion of heterozygotes (*y*) was calculated using the equations based on Thorgaard *et al*., (1983) as follows:

$$y = 1 - \left[\frac{\text{No. homozygotes}}{\text{total}} \right].$$

For loci showing significant deviation from Mendelian expectation in the mapping families, *y* was estimated by using the equation: *y* = 1-(2 x number of most common homozygotes).

5.3 Estimating Gene-centromere Distances (*x*)

Gene-centromere recombination frequencies were estimated for loci at which females were heterozygous. Genotypes of two females and their offspring were scored and analyzed by counting the numbers of homozygous and heterozygous offspring.

Gene-centromere distance was estimated by pooling data from the two families using the following equation: $x = \left(\frac{y}{2} \right) \times 100$ where *y* was the number of heterozygotes, assuming complete interference (Thorgaard *et al*., 1983; Allendorf *et al*., 1986).

5.4 Linkage analysis

The “JoinMap” program (Stam,1993) was used for testing linkages among loci and grouping of related loci on the linkage groups. Initial grouping of markers was carried out with a minimum LOD score of 3.0 and a maximum distance of 50.0 cM. Because of interference observed in the genome of walking catfish, the map distances were estimated using the Kosambi function. Distance data from JoinMap were used to draw the map with “Mapchart” program (Voorrips, 2002).

RESULTS

1. Searching Microsatellites

A total of 2,029 EST sequences of liver and muscle tissues from walking catfish were screened for microsatellite sequences. Of 136 new microsatellites identified, 60 were from liver tissue and 76 were from muscle tissue. Cluster analysis resulted in 8 contigs and 87 singletons, of which 113 unique microsatellite sequences (61 loci of unknown genes and 52 loci of known genes) were identified. Microsatellite sequences were characterized as perfect repeats, imperfect repeats and compound repeats. One hundred and seven loci (95.65%) identified in this study contained perfect repeat motifs, five loci (4.39 %) contained imperfect repeats and one locus contained compound repeat.

Forty-three loci (37.7%) contained dinucleotide repeats, 44 loci (38.5%) contained trinucleotide repeats, 22 loci (19.3%) contained tetranucleotide repeat motifs and four loci (3.5%) contained for pentanucleotide repeats. Among dinucleotide repeat type, TC/CT (25%), TG/GT (21%), AC/CA (21%), and GA/AG (18%) were the most abundant whereas TA/AT (9%), and GC (2%) were present at lower abundance. The most abundant trinucleotide motifs were TCC (16%), TTA (11%), AAT (9%), AAG (9%), GAT (7%), followed by ATT, TAT, TGA, GCA, CTA each were 4.5% whereas CCT, GCA, CAT, CGG, CTG, GAG, GCT, ACC, CCA, TAA, TTG were rare, each with 2.2%. The most abundant tetranucleotide repeats were (TTTC (27%), CTTT (4.5%), AAAT (9%), TAAA (4.5%), AAAC (9%) ,CAAA (4.5%), TTGG (9%), AAAG (9%) whereas TGAG, TTTG, TTTA, TGGA, CCTT, were evenly distributed, each with 4.5%. Four pentanucleotide repeat types, TTCAA, AAAAG, AAAAC, and ATTAG were present each with 25%. The length of core sequences ranged from 3 (penta-) to 31 (di-) repeats (Appendix Table 1).

EST sequences of known genes indicated that 32 microsatellite loci were found in the 3′ UTR region, 7 loci in the 5′ UTR region and 13 loci in the open reading frame (ORF) of known genes, e.g., vitellogenin, myosin light chain, troponin, and parvalbumin (Appendix Table 1).

2. Design of Primers

Primers were synthesized for thirty-nine loci (34%) from 113 loci identified from the EST sequences. Four loci were non-amplifiable and five loci showed un-specific amplification. Of 30 loci, sixteen loci contained dinucleotide repeats, twelve loci contained trinucleotide repeats, one locus contained tetranucleotide repeat motifs and one locus contained pentanucleotide. Annealing temperatures for these primers ranged from 50 to 60°C.

3. Primer analysis

3.1 Characterization of EST-linked microsatellite primers

Thirty loci were tested on genomic DNA from 30 unrelated fish sampled from hatchery population. Fourteen microsatellite loci were polymorphic. BLAST results showed that all microsatellite loci identified in this study did not have significant similarity with sequences from the previous efforts by Sukmanomon *et al*. (2003). Eleven of the *C. macrocephalus* microsatellites contained dinucleotide repeat motif, two loci contained trinucleotide repeat motifs and one locus contained a pentanucleotide repeat motif. GenBank Accession number of primers, primer sequence, repeat sequence, annealing temperature (T_A), number of alleles, allele size range, observed and expected heterozygosities and *P*-value of polymorphic loci were shown in Table 1 and the genotypes of 30 fishes were shown in Figure 4. The number of alleles ranged from 2 to 15. The observed and expected heterozygosities ranged from 0.4667 to 0.9333 and from 0.4271 to 0.8819 respectively. Significant deviations from Hardy-Weinberg expectations were found at loci *Cmac*5 and *Cmac*7 after Bonferroni correction (adjusted *P*-value <0.003). The heterozygosity excess at these

loci could have resulted from breeding practice in the hatchery. There was no significant linkage disequilibrium among all pairs of loci**.**

3.2 Cross-species Amplification

Fourteen polymorphic loci were tested in four other species of catfish including the African catfish *C. gariepinus*, the striped catfish *Pangasius hypophthalmus*, the Black-ear catfish *P. larnaudii*, and the Mekong giant catfish *Pangasianodon gigas*. Ten loci were found to amplify in other species. Cross-species amplifications of ten polymorphic loci were observed in African catfish. The numbers of polymorphic loci detected in pangasiid catfishes were five in the striped catfish and the black ear catfish, and four in the Mekong giant catfish (Table 2).

Table 1 Characterization of 14 polymorphic microsatellite loci for *Clarias macrocephalus*, including GenBank accession number, primer sequences, annealing temperature (T_a), repeat sequence, number of alleles, size range (bp), observed heterozygosity (H_o), expected heterozygosity (H_e), and *P*-value: *indicates significant deviation from HWE after Bonferroni correction.

Locus	GenBank Accession	Gene ID	Primer sequences 5'-3'	T_a (°C)	Repeat sequence	No. of alleles	Size range (bp)	H_o	H_e	*P*
*Cmac*1	EU179738	Histone deacetylase 6	F: AGTTCGGGGAGGAGATTC R:ATGGAGACGAGTGTTCAGTG	53	$(TG)_{20}$	15	243-279	0.8333	0.8740	0.0113
*Cmac*2	EU179739	Unknown	F: GAGCAATCAGCAGTGGAG R: AGGCAACAGTGAAACAGC	58	$(CT)_{14}$	8	240-260	0.6333	0.7379	0.1711
*Cmac*3	EU179740	Unknown	F: TCCTGATGACCTGATGATTC R: CTATATGGTCTCGCTGCT	56	$(TC)_{17}$	13	185-227	0.7667	0.8819	0.1057
*Cmac*4	EU179741	Myosin light chain 1	F: AGTGGGGCAGTAACAAGG R: ATAAAGACTCCAGG AAAGG	54	$(AC)_{11}$	8	191-223	0.5667	0.5503	0.0608
*Cmac*5	EU179742	Unknown	F: GAGATGACGTGTGTAGCAC R: GACCTGACTTTCAGGAAGC	53	$(ATT)_9$	6	226-240	0.9000	0.7661	0.0001*
*Cmac*6	EU179743	Unknown	F: AGGGGGAGACTGACGAGC R: GGGCACAGGCATCAGGAC	60	$(TC)_4TT(TC)_9$	7	244-260	0.7333	0.7480	0.3835
*Cmac*7	EU179744	phosphatidylethanolamine binding protein	F: CCCATCAGAGCAGACAAG R: GAATACATGATCCTGCCC	57	$(TC)_{23}$	12	203-229	0.9333	0.8593	0.0002*
*Cmac*8	EU179745	*Polypterus bichir*, clone -164C2	F: TGACCCAGTTATGCTGGG R: CTCAACCATTTGGAAGAGG	57	$(AAAAC)_3$	4	260-272	0.6667	0.6249	0.2315
*Cmac*9	EU179746	Unknown	F: CAGACCGAGTCTGTGCG R: CAGAGCACACTGATGTGG	57	$(CCT)_5$	5	220-244	0.7000	0.6350	0.4000
*Cmac*10	EU179747	Unknown	F: GGACGTTGATGCAGATGG R: CTTCGAGGTCATCGGCC	56	$(GA)_8$	4	194-202	0.7000	0.6571	0.2027
*Cmac*11	EU179753	receptor for activated protein kinase C homolog	F: CTGATCACCAATCGACTG R: GTTATTACTGGGCCAGAG	53	$(AC)_5$	2	159-161	0.4667	0.4271	0.6871
*Cmac*12	EU179748	Unknown	F: GGTGCGATAAAGGTGACCC R: CCGGGCGAGTAGGAACAC	58	$(TA)_6$	14	144-182	0.7000	0.8249	0.0062
*Cmac*13	EU179749	Unknown	F: GCTCTGGTTTCCATAGCAAC R: GATTGTCCTCTAGTGCTGC	55	$(GA)_6$	3	233-241	0.7667	0.6689	0.1741
*Cmac*14	EU179750	Unknown	F: TAGAGATCCCATCCTGGC R: CGACACAACCTTCTGCTG	55	$(GA)_6$	13	256-290	0.6667	0.8723	0.0247

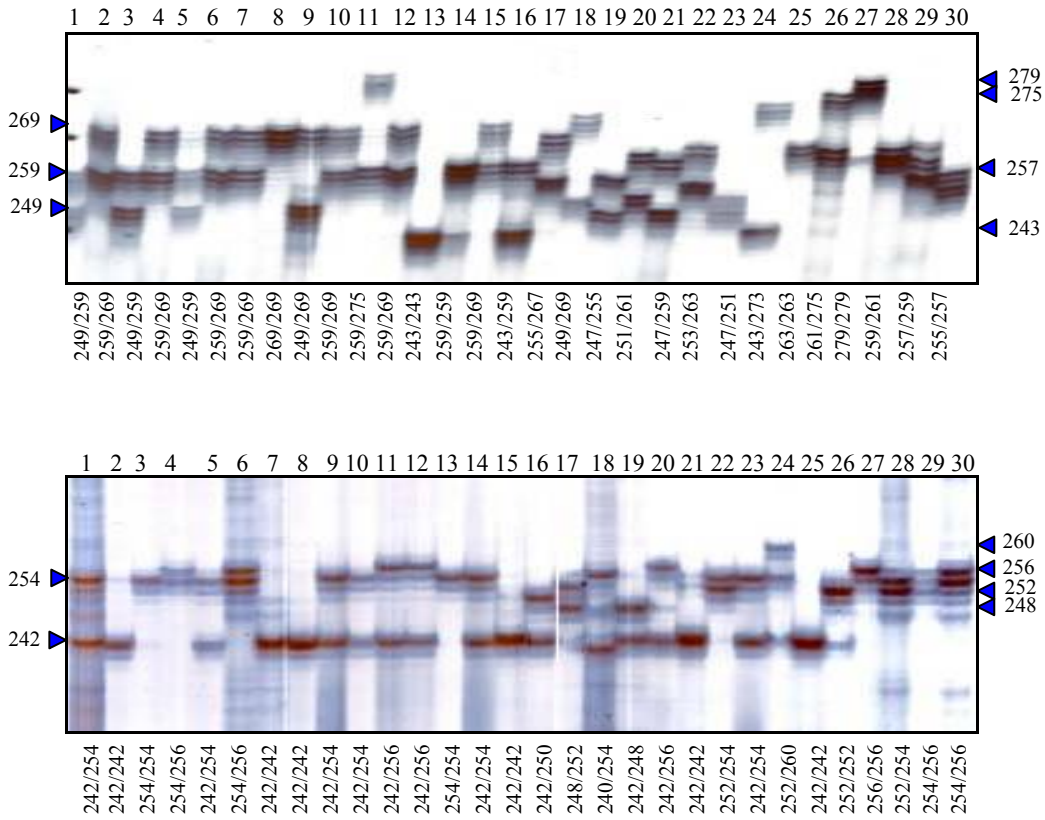


Figure 4 Genotypes of 30 unrelated fish at four loci, *Cmac*1, *Cmac*2, *Cmac*3, and *Cmac*4.

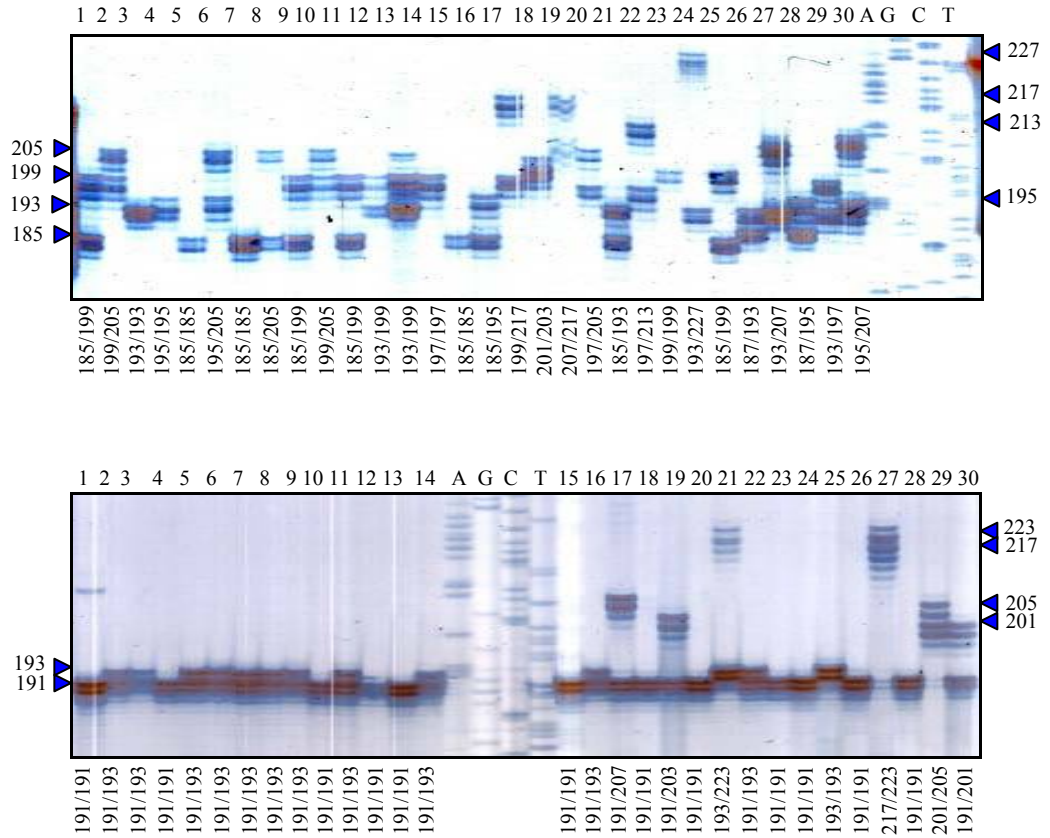


Figure 4 (Continued).

Table 2 Cross-species amplification and allele size range (bp) of 10 ESTs-linked microsatellite loci from *C. macrocephalus* in four other catfishes (sample size = 5), including the African catfish *C. gariepinus*, striped catfish *Pangasius hypophthalmus*, the black-ear catfish *P. larnaudii* and the Mekong giant catfish *Pangasianodon gigas*.

Locus	Species			
	C. gariepinus	*P. hypophthalmus*	*P. larnaudii*	*P. gigas*
*Cmac*1	246-254(3)	242-250(4)	226-244(4)	-
*Cmac*2	215-221(2)	-	-	-
*Cmac*3	197-213(3)	-	183(1)	-
*Cmac*4	188-198(2)	-	188(1)	-
*Cmac*5	211(1)	213(1)	-	213(1)
*Cmac*6	256-268(2)	244-260(3)	256(1)	244-260(2)
*Cmac*8	264-272(2)	264-272(2)	266(1)	264-272(2)
*Cmac*10	197-203(2)	197(1)	-	179-197(2)
*Cmac*11	162(1)	-	-	-
*Cmac*13	236-274(3)	-	-	-

(-) is the number of alleles.

3.5 Mendelian Inheritance and Segregation Analysis

Fourteen EST-linked microsatellite loci were tested for Medelian inheritance in two full-sib families of twenty normal diploid progeny per family. Results of Chi-square Goodness of fit test revealed that Twelve loci segregated in agreement with the Mendelian expectations, whereas two loci, *Cmac*4 and *Cmac*13 showed significant segregation distortion in one of two families with chi-square values of 8.01 and 12.55 (df = 3, $P_{0.05}$ = 7.82) respectively (Table 3). Microsatellite genotyping at each locus is shown in Figure 5.

Segregation of all forty-one microsatellite loci developed from genomic library conformed to Mendelian expectations (S. Poompuang unpublished data).

4. Production of gynogenetic diploids

Approximately 3,000 eggs were obtained for each female. Average hatching rate of gynogenetic diploids was 60%. No evidence of male contribution from UV-inactivated striped catfish sperm was found among ten families of gynogenetic diploid walking catfish. Two families (A-1 and A-2) were randomly chosen for half-tetrad analysis.

5. Estimating Gene-centromere distance

5.1 Genotyping

Because some DNA samples were degraded, only 47 gynogenetic diploid progeny each from two families were used for estimating gene-centromere distances. Results showed that the females were heterozygous in two families or either at 11 EST-linked microsatellite loci and at 33 microsatellite loci derived from genomic library. The genotypes of progeny from two families were shown in Figure 6.

5.2 The proportion of heterozygote progeny (*y*)

The frequencies of two homozygous classes were significant different in Family 2 at locus *Cmac*4, *Cma*-13, *Cma*-42 and *Cma*-47 and in both families at locus *Cma*-33. The *y* values of these loci were then estimated by

$$Y = \frac{\text{No. of heterozygotes}}{\text{No. of heterozygotes} + 2(\text{No. of the most common class of homozygotes})}$$

as described by Thorgaard *et al*. (1983).

Genotyping of 44 loci revealed that the proportion of heterozygotes (*y*) ranged from 0.043 (locus *Cmac*14) to 0.936 (loci *Cmac*10 and *Cma*-36). Twenty-six loci (59%) showed high levels of microsatellite-centromere recombination with

frequencies greater than 0.67. Three loci (*Cmac*10, *Cma*-13 and *Cma*-36) displayed recombination frequencies greater than 0.9.

5.3 Gene-centromere map distances

The gene-centromere map distance ranged from 2.15 cM for *Cmac*14 to 46.8 cM for *Cmac*10 and *Cma*-36. The average distance was 33.1 cM. The maternal genotypes, numbers of heterozygotes and each homozygote classes, proportions of heterozygotes and gene-centromere distances are shown in Table 4.

Four loci, *Cmac*14, *Cma*-29, *Cma*-51 and *Cma*-55 were closely linked to their centromeres as indicated by gene-centromere distances of 2.15, 6.0, 5.9 and 5.3 cM, respectively. Three loci, *Cmac*10, *Cma*-13 and *Cma*-36 were located at 45.2 and 46.8 cM. Twenty-six loci were located from the centromeres at distances ranged from 33.5 to 44.7 cM. Eleven loci were located at distances ranged from 12.9 to 32 cM with an average distance of 20 cM.

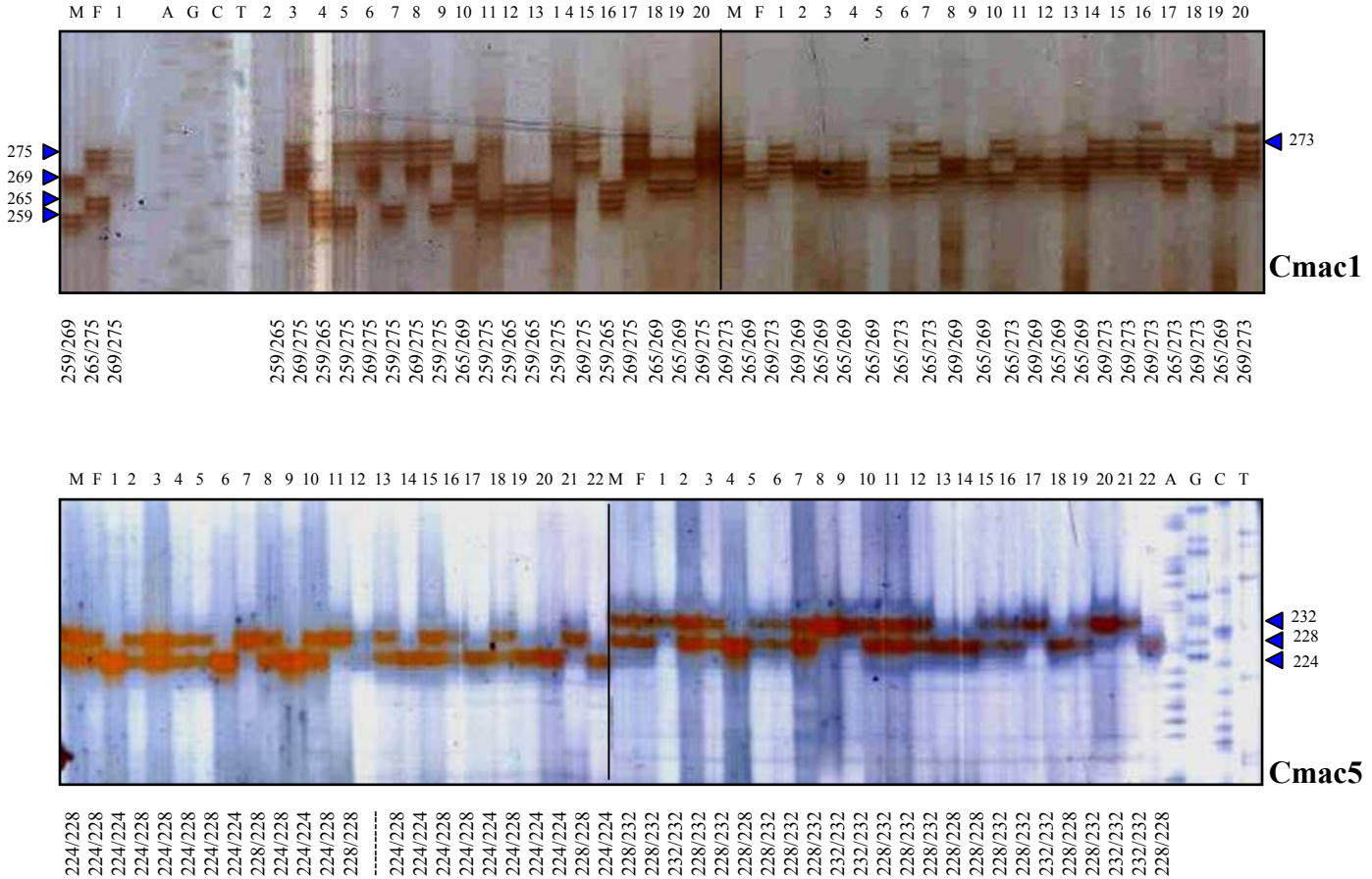


Figure 5 Mendelian segregation at four EST-linked microsatellite loci (*Cmac*1, *Cmac*5, *Cmac*7, and *Cmac*9) of two normal diploid families (C-1 and C-2).

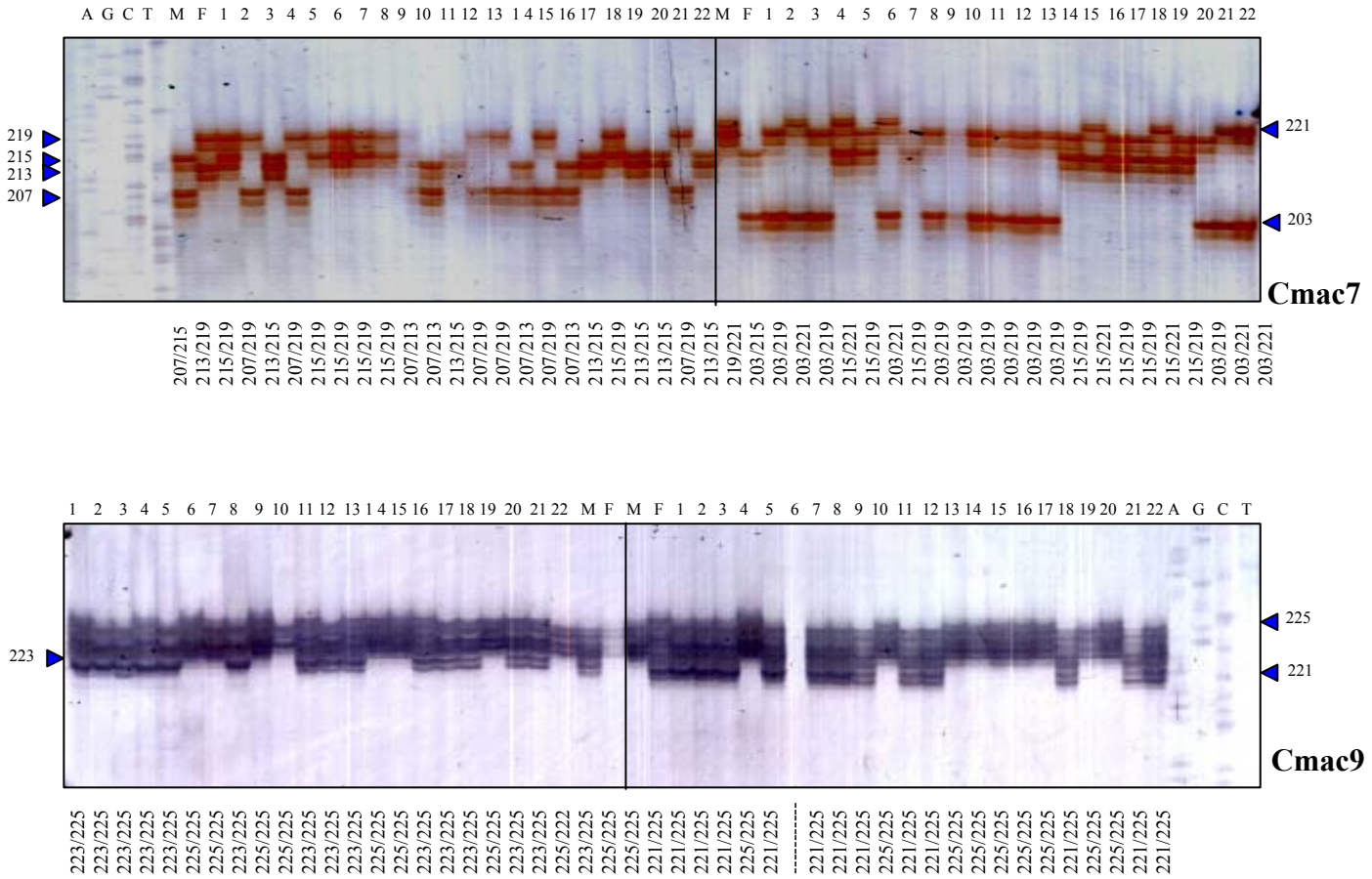


Figure 5 (Continued)

Table 3 Mendelian segregation at fourteen EST-linked microsatellite loci of two control families (C-1 and C-2).

Locus	Family	Genotype		Progeny genotype		χ^2
		Male	Female	Genotype	Number	
*Cmac*1	C-1	259/269	265/275	259/265	5	
				259/275	5	
				265/269	4	
				269/275	6	2.00
	C-2	269/273	265/269	265/269	7	
				265/273	4	
				269/269	3	
				269/273	4	1.00
*Cmac*2	C-1	241/247	241/255	241/241	6	
				241/255	9	
				241/247	4	
				247/255	3	2.55
	C-2	251/255	253/255	251/253	7	
				251/255	0	
				253/255	8	
				255/255	7	5.64
*Cmac*3	C-1	195/205	185/193	185/195	4	
				185/205	5	
				193/195	4	
				193/205	9	2.00
	C-2	181/187	183/199	181/183	2	
				181/199	9	
				183/187	4	
				187/199	6	3.65
*Cmac*4	C-1	187/187	187/201	187/187	10	
				187/201	12	0.04
	C-2	199/203	187/199	187/199	2	
				199/199	11	
				187/203	7	
				199/203	2	8.01*
*Cmac*5	C-1	224/228	224/228	224/224	8	
				224/228	10	
				228/228	3	2.68

Table 3 (Continued)

		Genotype		Progeny genotype		χ^2
		Male	Female	Genotype	Number	
*Cmac*5	C-2	228/232	228/232	228/228	5	
				228/232	11	
				232/232	6	
						1.94
*Cmac*6	C-1	256/262	250/256	250/256	8	
				256/256	6	
				250/262	5	
				256/262	2	
						2.43
	C-2	254/262	246/262	246/254	7	
				254/262	3	
				246/262	4	
				262/262	8	
						1.82
*Cmac*7	C-1	207/215	213/219	207/213	4	
				207/219	6	
				213/215	6	
				215/219	6	
						0.18
	C-2	219/221	203/215	203/219	10	
				215/219	6	
				203/221	3	
				215/221	3	
						7.27
*Cmac*8	C-1	265/269	265/267	265/265	11	
				265/267	3	
				265/269	5	
				267/269	3	
						5.99
	C-2	265/269	265/271	265/265	4	
		224/228	224/228	265/271	9	
				265/269	2	
				269/271	7	
						3.64
*Cmac*9	C-1	223/225	225/225	223/225	14	
				225/225	8	
						1.14
	C-2	225/225	221/227	221/225	12	
				225/227	9	
						0.19

Table 3 (Continued).

		Genotype		Progeny genotype		
		Male	Female	Genotype	Number	χ^2
*Cmac*10	C-1	173/175	175/177	173/175	4	
				173/177	6	
				175/175	6	
				175/177	6	
						0.18
	C-2	169/169	175/175	169/175	22	
						0.00
*Cmac*11	C-1	158/160	158/160	158/158	5	
				158/160	13	
				160/160	4	
						5.20
	C-2	158/158	160/160	158/160	22	
						0.00
*Cmac*12	C-1	142/148	144/176	142/144	5	
				142/176	6	
				144/148	5	
				148/176	6	
						0.00
	C-2	148/158	144/144	144/148	9	
				144/158	13	
						0.41
*Cmac*13	C-1	236/238	234/236	234/236	7	
				236/236	2	
				234/238	7	
				236/238	6	
						2.00
	C-2	234/238	236/238	234/236	5	
				234/238	3	
				236/238	13	
				238/238	1	
						12.55*
*Cmac*14	C-1	280/280	258/284	258/280	12	
				280/284	9	
						0.19
	C-2	274/280	272/280	272/274	5	
				274/280	3	
				272/280	7	
				280/280	6	
						0..91

* indicates significant deviation from χ^2 (p-value< 0.05, df = 1, χ^2= 3.84 and p-value<0.05, df = 3, χ^2= 7.82).

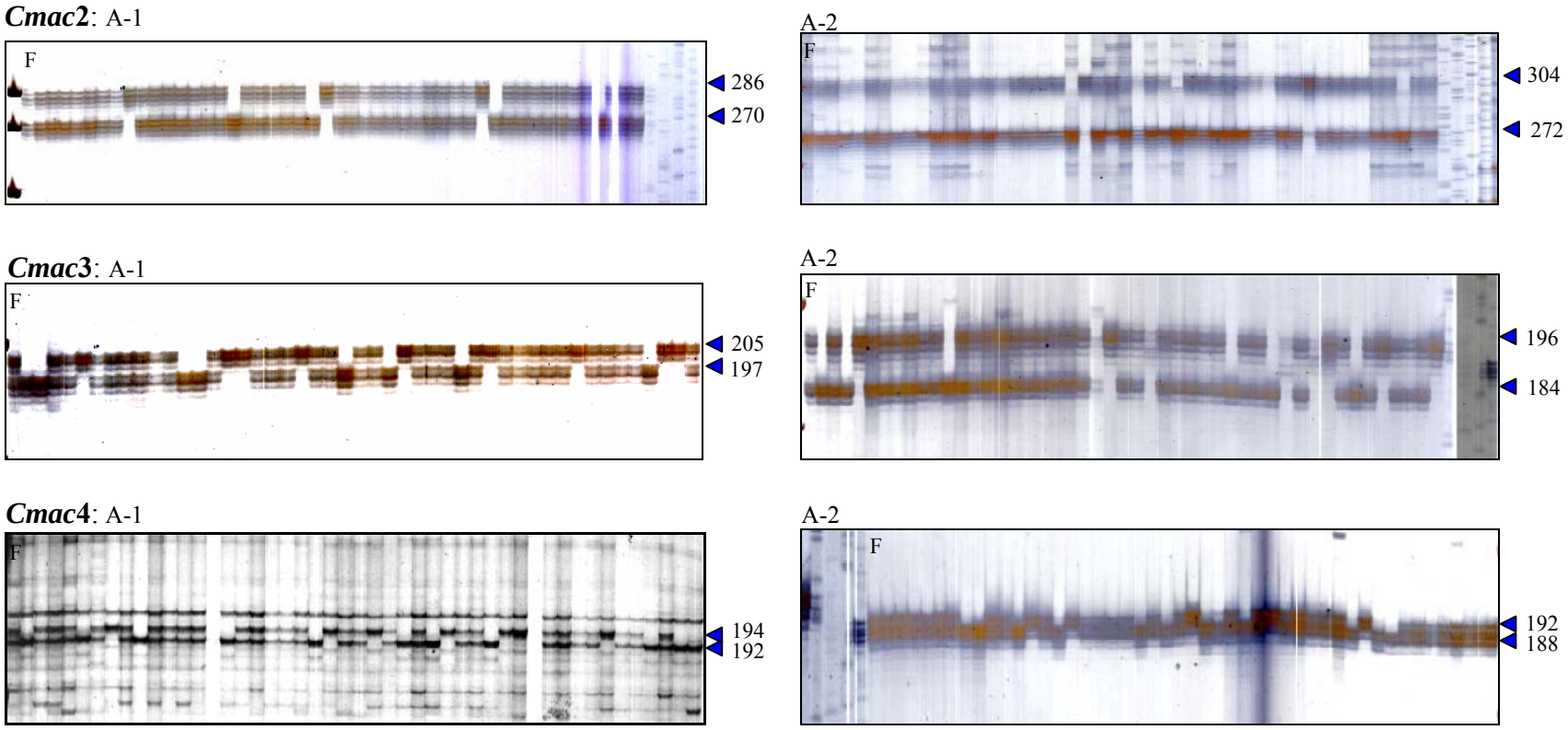


Figure 6 Progeny genotypes at three loci in two gynogenetic diploid families (A-1 and A-2).

Table 4 Progeny genotypes at forty-four loci in gynogenetic diploid walking catfish

Locus	Family	Maternal genotype	Progeny genotypes[a]			Proportion of heterozygotes (*y*)	Gene-centromere distance (cM)
			11[a]	12	22[a]		
*Cmac*1	A-1	270/286	4	40	3	0.851	
	A-2	272/304	3	42	2	0.893	
		Total	7	82	5	0.872	43.60
*Cmac*2	A-1	242/256	4	36	6	0.783	
	A-2	242/254	4	32	11	0.681	
		Total	8	68	17	0.731	36.60
*Cmac*3	A-1	197/205	8	30	9	0.638	
	A-2	184/196	6	33	8	0.702	
		Total	14	63	17	0.670	33.50
*Cmac*4	A-1	192/194	11	27	7	0.600	
	A-2	188/192	12	32	3	0.571	
		Total	23	59	10	0.641	32.05
*Cmac*5	A-1	226/242	3	44	0	0.936	
	A-2	226/242	6	38	3	0.809	
		Total	9	82	3	0.872	43.60
*Cmac*6	A-1	243/257	7	39	1	0.830	
	A-2	253/257	7	36	4	0.766	
		Total	14	75	5	0.798	39.90
*Cmac*7	A-1	212/220	15	17	15	0.362	
	A-2	207/229	21	12	13	0.261	
		Total	36	29	28	0.312	15.60
*Cmac*8	A-1	261/269	5	36	6	0.766	
	A-2	261/265	4	33	8	0.733	
		Total	9	69	14	0.750	37.50
*Cmac*10	A-1	194/198	3	44	0	0.936	
	A-2	197/199	3	41	3	0.872	
		Total	6	85	3	0.904	45.20
*Cmac*13	A-2	236/238	8	36	3	0.766	38.30
*Cmac*14	A-1	278/286	27	2	18	0.043	2.15
Cma-8	A-1	204/254	15	13	19	0.277	
	A-2	248/252	19	19	9	0.404	
		Total	34	32	28	0.340	17.00
Cma-13	A-1	234/250	1	43	3	0.915	
	A-2	234/248	0	42	5	0.808	
		Total	1	85	8	0.904	45.20
Cma-14	A-1	187/197	8	30	9	0.638	
	A-2	189/197	6	34	7	0.723	
		Total	14	64	16	0.681	34.05
Cma-17	A-1	226/230	2	37	5	0.841	42.05
Cma-20	A-1	242/246	1	42	4	0.894	
	A-2	220/228	5	36	6	0.766	
		Total	6	78	10	0.830	41.50

Table 4 (Continued)

Locus	Family	Maternal genotype	Progeny genotypes[a]			Proportion of heterozygotes (*y*)	Gene-centromere distance (cM)
			11[a]	12	22[a]		
Cma-21	A-1	179/183	2	43	2	0.915	
	A-2	178/186	4	39	4	0.830	
		Total	6	82	6	0.872	43.60
Cma-28	A-1	195/221	5	39	3	0.830	
	A-2	195/221	4	40	2	0.870	
		Total	9	79	5	0.849	42.50
Cma-29	A-1	213/249	17	3	26	0.065	
	A-2	213/241	19	8	19	0.174	
		Total	36	11	45	0.120	6.00
Cma-31	A-1	223/249	10	31	6	0.660	
	A-2	223/225	10	28	8	0.609	
		Total	20	59	14	0.634	31.70
Cma-32	A-1	492/526	9	22	16	0.468	
	A-2	492/576	9	29	9	0.617	
		Total	18	51	25	0.543	27.15
Cma-33	A-1	191/213	0	37	10	0.649	
	A-2	191/213	0	38	9	0.679	
		Total	0	75	19	0.664	-*
Cma-34	A-1	217/245	8	36	3	0.766	
	A-2	217/245	10	32	4	0.696	
		Total	18	68	7	0.731	36.55
Cma-35	A-1	261/275	13	14	20	0.298	
	A-2	247/261	19	14	14	0.298	
		Total	32	28	34	0.298	14.90
Cma-36	A-2	218/250	0	44	3	0.936	46.80
Cma-37	A-1	208/230	7	35	5	0.745	
	A-2	208/248	8	36	3	0.766	
			15	71	8	0.755	37.75
Cma-38	A-1	164/190	12	21	14	0.447	22.35
Cma-40	A-1	241/247	5	36	5	0.783	
	A-2	249/253	7	34	6	0.723	
		Total	12	70	11	0.753	37.65
Cma-41	A-1	564/588	13	23	11	0.489	24.45
Cma-42	A-2	247/251	23	16	8	0.258	12.90
Cma-43	A-1	242/252	8	32	6	0.696	34.80
Cma-44	A-1	232/254	5	34	8	0.723	36.20
Cma-45	A-1	237/271	15	15	17	0.319	
	A-2	241/325	14	20	13	0.426	
		Total	29	35	30	0.372	18.60
Cma-46	A-1	219/227	9	34	4	0.723	36.20

Table 4 (Continued)

Locus	Family	Maternal genotype	Progeny genotypes[a]			Proportion of heterozygotes (*y*)	Gene-centromere distance (cM)
			11[a]	12	22[a]		
Cma-47	A-1	206/214	11	24	12	0.511	
	A-2	186/206	15	25	5	0.455	
		Total	26	49	17	0.533	26.65
Cma-49	A-1	305/315	2	40	5	0.851	42.55
Cma-50	A-2	235/239	6	35	5	0.761	38.05
Cma-51	A-1	232/238	19	8	20	0.170	
	A-2	238/250	23	3	21	0.021	
		Total	42	11	41	0.117	5.90
Cma-52	A-1	261/283	8	36	3	0.766	
	A-2	261/275	5	36	6	0.766	
		Total	13	72	9	0.766	38.30
Cma-53	A-1	232/254	3	42	2	0.894	44.70
Cma-54	A-1	217/259	7	38	2	0.809	
	A-2	249/259	5	40	2	0.851	
		Total	12	78	4	0.830	41.50
Cma-55	A-1	246/254	26	5	16	0.106	
	A-2	242/262	24	5	18	0.106	
		Total	50	10	34	0.106	5.30
Cma-56	A-1	280/292	3	40	3	0.870	
	A-2	280/300	1	39	6	0.848	
		Total	4	79	9	0.859	42.95
Cma-57	A-1	231/251	3	36	8	0.766	38.30

[a]11 and 22 are designations for homozygotes for the first and second maternal allele respectively, and 12 are heterozygotes.
*gene-centromere distance at this locus was not included because of segregation distortion.

5.5 Linkage Analysis (Joint Segregation)

Linkage groups were determined using the JoinMap program based upon the test for independence with a minimum LOD score of 3.0. Results of the gene-centromere orders for six linkage groups were consistent when using LOD scores of 3, 4 and 5 (Figure 7). Two marker loci were detected on the same chromosomes in six linkage groups. The order *centromere-marker* 1-*marker* 2, which the two markers were on the same chromosome arm was observed in five linkage groups. One linkage group displayed the order *marker locus* 1-*centromere-marker locus* 2, which indicated that the centromere was flanked by the two markers. In linkage group 1, *Cma*-44 and *Cma*-57 were closely linked and located at 36.2 and 38.3 cM from the centromere. Group 2 consisted of two closely linked loci, *Cma*-28 and *Cma*-54 with distances from the centromere of 41.5 and 42.5 cM respectively. In Group 3, two closely linked loci *Cmac*3 and *Cma*-14 were located 33.5 and 34.0 cM from the centromere. In Group 4, two loci *Cma*-35 and *Cma*-47 were located 14.9 and 26.7 cM relative to the centromere. Group 5 contained two closely linked loci, *Cma*-21 and *Cmac*10 at 43.6 and 45.2 cM distal to the centromere. In Group 6, *Cma*-51 and *Cma*-45 were on the different chromosome arms with distances from the centromere of 5.9 and 18.4 cM respectively.

Joint segregation of loci *Cma*-31 and *Cma*-34 allowed the placement of centromere, *Cma*-31 on Linkage Group (LG) 3 of the AFLP map (Poompuang and Na-Nakorn, 2004). The centromere was placed at 5.3 cM above marker locus *Cma*-55 or at 15.6 cm below the AFLP locus *TAGCAG*10. Two loci, *Cma*-31 and *Cma*-34 were located on the same chromosome arm at 31.7, and 36.7 cM from the centromere. Two loci *Cma*-34 and *Cma*-55 were previously mapped onto this linkage group (Figure 8) (S. Poompuang, unpublished data).

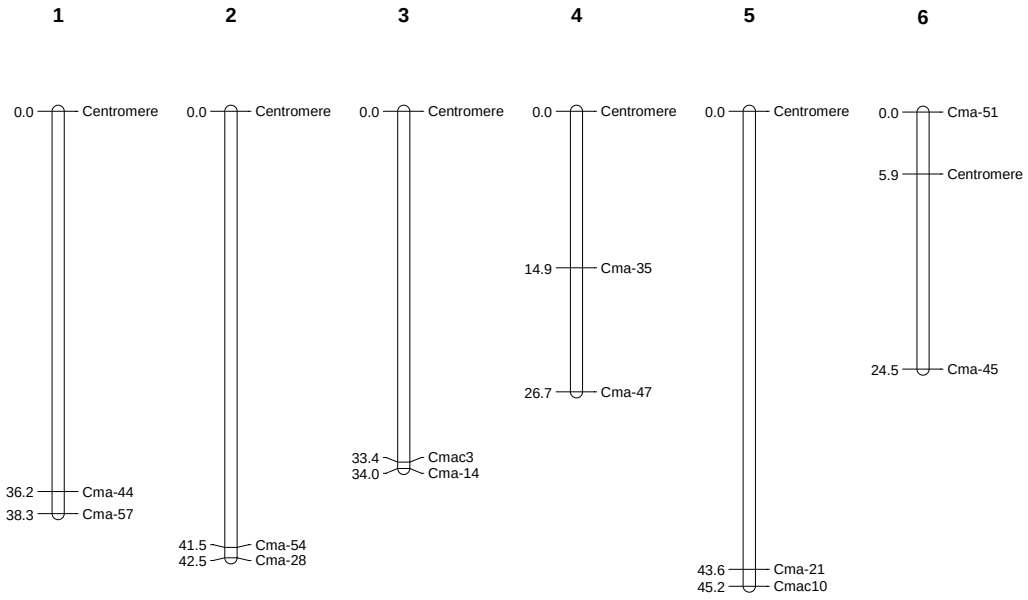


Figure 7 Six groups of the microsatellite–centromere map. The centromere is the starting point and numbers present the map distances between centromere and each locus.

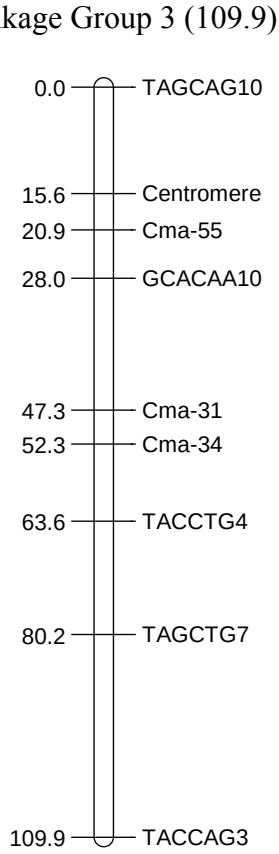


Figure 8 Location of centromere on Linkage Group 3 of the AFLP linkage map.

DISCUSSION

Characterization of EST-linked microsatellite markers

Development of microsatellite markers using the traditional approach which involves creating genomic library, hybridization with the oligonucleotide probes and sequencing of selected clones is expensive and time-consuming. In walking catfish, for instance, a very small number of microsatellites have been developed using this strategy. Forty-six microsatellite loci were identified, including four loci were reported by Na-Nakorn *et al*. (1999), 12 loci (*Cma*-6 to *Cma*-27) were developed by Sukmanomon *et al*. (2003) and 30 loci (*Cma*-28 to *Cma*-57) were unpublished (S. Poompuang unpublished data).

Expressed sequence tag (EST) techniques provide a fast and more efficient approach for developing microsatellites in many species of plant and animal. Because microsatellites can be searched for in the EST databases, the approach helps shortening time and reducing cost for marker development. The abundance of microsatellite containing ESTs in fish genome varied across species. For example, Serapion *et al*. (2004) reported that the frequency of microsatellites found in ESTs of channel catfish was relatively high, i.e., 11.2% (4,855 out of 43,033 ESTs) contained microsatellites. The frequencies of microsatellite containing ESTs, however, were much lower in Atlantic salmon (Vasemagi *et al*., 2005) and turbot (Chen *et al*., 2007) at 2% and 2.2% respectively. Despite a relatively small number of EST sequences (2,029) for walking catfish, 5.57% (113) of ESTs were found to contain microsatellites.

Dinucleotide and trinucleotide repeats were the most abundant in walking catfish ESTs, of which 43 loci (37.7%) contained dinucleotide repeats, and 44 loci (38.5%) contained trinucleotide repeats, followed by 22 loci (19.3%) and four loci (3.5%) of tetranucleotide and pentanucleotide repeats respectively. It should be noted that the frequencies of all nucleotide repeat types will change with new EST information of walking catfish. Distribution of repeat types in microsatellite

containing ESTs of channel catfish was different, the frequency of dinucleotide repeats was much higher at 71.7% followed by 11.2% and 19.2%, respectively, for tri- and tetra- repeats (Serapion *et al*., 2004). The abundance of tri-nucleotide repeats in walking catfish genes can be explained by the suppression of non-trimeric repeats in coding regions due to the risk of frameshift mutations that may occur when those microsatellites alternate in size of one unit (Metzgar *et al.,* 2000).

EST sequences of known genes indicated that 32 microsatellite loci (62%) were found in the 3′ UTR region, 7 loci in the 5′ UTR region and 13 loci in the open reading frame (ORF) of known genes. It has been reported that microsatellites are more abundant in the 3′ ends of cDNA clones which contain untranslated regions (Scott *et al*., 2000; Thiel *et al*., 2003). Due to the process of cDNA generation (oligo-dT priming) there is a preferential selection of UTR within 3′-ESTs, therefore being more variable than 5′-ESTs.

Primers were synthesized for thirty-nine loci (59%) from 113 loci identified from the EST sequences because there were not enough flanking regions for some of those microsatellite loci. Analysis of primers indicated that four loci (10.25%) failed to amplify products. Possible explanations for the non-amplifiable loci could be that primers extend across a splice site or the presence of large introns in the genomic sequence (Theil *et al*., 2003). Five loci showed unspecific amplification which could have resulted from low stringency of the PCR reaction. Many microsatellites found in the ESTs of walking catfish were located at the ends of the ESTsequences, therefore, additional sequence data are needed for designing PCR primers. For those located at the 3′ or 5′ ends of the cDNA clones, their flanking sequences cannot be obtained unless the corresponding genomic sequences are available (Serapion *et al*., 2004).

The usefulness of EST-microsatellites depends on their polymorphism as well as cross-species transferability. It has been suggested that EST derived microsatellites were less polymorphic than those derived from genomic sequences because the conservation of ESTs in coding sequences may limit their polymorphism (Scott *et al*., 2000). Previous studies in fish, however, have shown higher number of alleles per

locus (7.7 vs. 4.9) and polymorphism information content (PIC) (72.7% vs. 54.0%) of EST-derived microsatellites than that of anonymous microsatellites respectively in common carp (Yue *et al*., 2004) and rainbow trout (Coulibaly *et al*., 2005). In this study, characterization of EST-derived microsatellites showed similar levels of polymorphism to those of microsatellites developed from genomic library of walking catfish (Sukmanomon *et al*., 2003). The number of alleles ranged from 2 to 13 (7.67) and 2 to 15 (8.14) alleles per locus and the averaged observed heterozygosities were 0.692 and 0.71 for Type II and type I microsatellites respectively. But the levels of heterozygosity of type I microsatellites were significantly higher than that of type II markers.

Generally, it is expected that the EST-derived microsatellites are more transferable than anonymous microsatellites. The transferability of the EST-linked microsatellites makes them useful for comparative mapping. The EST-derived microsatellites from rainbow trout showed cross-species amplification with nine other salmonids, 85% in other six *Oncorhynchus*, and 79% could be transferred to *Salvelinus*, *Salmo*, and *Thymallus* (Coulibaly *et al*., 2005). The EST-linked microsatellite markers developed for walking catfish will be useful for other clariid and pangasiid catfishes due to the high rate of cross-species amplification of these markers. Cross-species amplification results demonstrated that *Cmac*6 and *Cmac*8 amplified in all species indicating that these two loci were mostly conserved for family clariidae and family pangasiidae. Three loci, *Cmac*2, *Cmac*11, and *Cmac*13 amplified only in walking catfish and African catfish (*C. gariepinus*). Cross-species amplification of EST-linked microsatellite loci also confirmed close evolutionary relationship between African catfish and walking catfish. However, the use of these microsatellites for species identification requires verification in a larger sample size.

Despite their properties as ideal genetic markers, it has been suggested that microsatellites found within genes are not appropriate for population genetic studies because they are not selectively neutral. Selection on these loci might influence the estimation of population genetic parameters which is needed to reveal gene flow, bottleneck events, effective population size, and population origin of individuals

(Pashley *et al*., 2006). From conservation genetics perspective, local adaptation is a key component to preserving an organism's survival and reproduction within a particular environment. Neutral molecular marker data, however, may not correctly reveal local adaptation of populations as allele frequency changes at adaptive loci can occur much faster than changes in neutral loci. Alternatively, the application of molecular markers that are linked to adaptive genes will provide important tool for measurement of adaptive differentiation between populations (Vasemägi and Primmer, 2005). The EST-linked microsatellites have been applied in population genomics approach to identify functionally important adaptive genes, for instance, in populations of anadromous species such as Atlantic salmon and rainbow trout which were affected by dam construction. Although microsatellite containing ESTs have been identified in this study, the number of available loci is too small for their application in population genomics study in walking catfish.

Microsatellite-centromere mapping

Characterization of EST-linked microsatellites in this study indicated that most of the identified microsatellites were informative and suitable for gene-centromere mapping in walking catfish.

Medelian inheritance and segregation analysis

Segregation distortion is a common phenomenon in genetic mapping studies. It is defined as deviation of the observed genotypic frequencies from their Mendelian expected values. The presence of deleterious alleles which results in reduced viability of progeny and errors in marker genotyping are two major factors that may cause deviation from Medelian segregation ratio. It has been reported that the levels of segregation distortion varied significantly among different marker types. The estimation of recombination fractions between co-dominant markers was less affected by segregation distortion than dominant markers. Because the locations of RFLP and microsatellite markers are more evenly distributed throughout the genome thus they are more stable whereas the AFLP and RAPD markers tend to cluster near the

centromeres. Although segregation distortion is commonly found in mapping populations of plants and animals, the phenomenon occurred less frequent in diploid gynogenetic fish. Nevertheless, segregation distortion was apparent in turbot with up to 33% of microsatellite loci showed deviation from Mendelian inheritance (Martinez *et al.,* 2008). Distortions on the same set of markers, however, were not detected in the one- and 10 day post hatching, suggesting that expression of recessive deleterious alleles in turbot was stage-specific. Use of newly hatched larvae, therefore, could prevent the detection of the expected segregation distortion. In this study, all microsatellite loci exhibited Medelian segregations, indicating that they were suitable for gene-centromere mapping.

Gene-centromere distances

Gene-centromere mapping based on half-tetrad analysis is possible for loci at which the female is heterozygous. By determining the genotype distribution of gynogenetic progeny, recombination frequencies between genes and their centromere can be estimated. In the absence of crossover between genes and centromeres, all progeny will be homozygous at those loci. If there are crossovers, the genotype of the progeny will be heterozygous. Female walking catfish of two families were genotyped at 55 microsatellite loci, 14 of EST-linked and 41 of genomic-derived markers. Results indicated that the females were heterozygous in two families or either at 11 EST-linked microsatellite loci and at 33 microsatellite loci derived from genomic library. A total of 44 loci (78.5%) were mapped to their centromeres whereas the remaining 11 loci were not applicable in this experiment. Genotyping of additional females and their families will increase the likelihood of finding other females heterozygous at the remaining loci, and would allow mapping of these loci to their centromeres.

The unequal numbers of homozygotes were evident at four loci *Cmac*4, *Cma*-13, *Cma*-42 and *Cma*-47 in one family and at locus *Cma*33 in both families. Such differences were also reported in previous study by Na-Nakorn *et al*. (2006). Differential survival of two homozygous classes due to deleterious recessive alleles

linked to the markers and small size of the family (n = 47) may have contributed to this difference. Locus *Cmac*4 was microsatellite-containing myosin light chain gene whereas loci *Cma*13, *Cma*33, *Cma*42 and *Cma*47 were anonymous markers. The myosin light chain gene encodes a skeletal muscle protein which involved in growth and development of fish (Xu *et al*., 1999). The presence of alternate alleles of myosin light chain may affect expression of fitness traits of homozygous gynogens. However, there was no evidence for differential survival of homozygous gynogens of rainbow trout (Allendorf *et al*., 1986), pink salmon (Lindner *et al*., 2000) and channel catfish (Liu *et al*., 1992). Allendorf *et al*., (1986) suggested that the presence of duplicated loci could prevent the detection of deleterious alleles in species of tetraploid origin such as salmonids.

The proportion of heterozygotes (*y*) in walking catfish gynogenetics ranged from 0.043 to 0.936. Twenty-six loci (59%) showed high levels of microsatellite-centromere recombination with *y* values greater than 0.67 which is the maximum expected value of independent segregation, assuming no chiasma interference (Thorgaard *et al*., 1983). Three loci *Cmac*10, *Cma*-13 and *Cma*-36 (22%) had *y* values greater than 0.9. These loci with the *y* values near 1.0 were the result of reduced recombination on the telomeric region of the chromosome due to interference. The high values of heterozygote proportion of most loci confirmed strong chiasma interference in walking catfish. Chiasma interference has been reported to be common in fish species, e.g., zebrafish (Johnson *et al*., 1996), medaka (Naruse and Shima, 1989), channel catfish (Liu *et al*., 1992), rainbow trout (Thorgaard *et al*., 1983), tilapia (Kocher *et al*., 1998), pink salmon (Lindner *et al*., 2000), Japanese eel (Nomura *et al*., 2006), yellow croaker (Li *et al*., 2007) and turbot (Martinez *et al*., 2008). In addition, complete interference was detected in walking catfish gynogens (Na-Nakorn *et al*., 2006).

Gene-centromere distances ranged from 2.15 cM of *Cmac*14 to 46.8 cM of *Cma*-36. Four loci, *Cmac*14, *Cma*-29, *Cma*-51 and *Cma*-55 were closely linked to their centromere as indicated by gene-centromere distances of 2.15, 6, 5.9 and 5.3 cM, respectively. Three loci, *Cmac*10, *Cma*-13 and *Cma*-36 with highest recombination

frequencies, were located telomerically on different chromosomes with an average distance of 46.0 cM. Twenty-six loci showing high frequencies of microsatellite-centromere recombination greater than 0.67 were located from the centromeres at distances ranged from 33.5 to 44.7 cM. Gene-centromere distances of the remaining 11 loci ranged from 12.9 to 32 cM with average distance of 20 cM. The recombination frequency data suggested that microsatellites were almost randomly distributed within the genome of walking catfish and supported the use of these markers for constructing a linkage map. Differential distributions of various markers along the chromosomes were observed in pink salmon as revealed by different average *y* values of marker loci (Lindner *et al*., 2000). For example, AFLP loci with mean *y* value of 0.40 were distributed differently from other classes of marker loci including allozymes, paired interspersed nuclear elements (PINEs), and microsatellites which displayed similar *y* values ranged from 0.63 to 0.69. Many studies have shown that AFLPs tend to cluster in centromeric regions of the chromosomes.

The order of centromere and genes on the chromosome can be simply determined by linkage analysis between marker loci. For two marker loci, *A* and *B*, there are three possible gene orders. When the centromere flanks the markers, the order *centromere-A-B*, or *centromere-B-A*, can be assumed, and the two genes are located on the same chromosome arm. When the centromere is flanked by two markers as displayed by the order *A-centromere-B*, the two genes are located on the different chromosome arms. Linkage analysis resulted in five linkage groups which markers were on the same chromosome arm and one linkage group which marker loci were located on the different chromosome arms.

In species with a well-defined linkage map, the location of centromeres can be determined by selecting a set of marker loci from each linkage group and genotyping gynogenetic diploid progeny. When the position of centromere is established, the consistency of recombination frequencies of marker loci and their position on the map would be tested to confirm the location of centromere on the chromosome. In turbot, 79 microsatellite loci selected from 242 loci across 26 linkage groups were used to

locate centromeres, of which centromeres were placed on 22 linkage groups (Martinez *et al*., 2008). By using the gene-centromere recombination data in this study, it was possible to place the centromere as well as two other markers, *Cma*-31 and *Cma*-47 onto Linkage Group 3 of the AFLP linkage map of walking catfish. The correct position of the centromere was confirmed by the consistency of previous mapping data and gene-centromere recombination frequencies for *Cma*-34 and *Cma*-55 which have been placed on the AFLP map.

CONCLUSION

This work has, to some extent, increased genome information of walking catfish by identifying type I microsatellite markers from the existing EST databases. Further, this study presents new gene-mapping data for the expansion of the female linkage map by mapping the identified EST-associated microsatellites in relation to their centromeres. By searching the current EST database which consists of 2,029 sequences, 113 microsatellites were found within functional genes from liver and muscle tissues as well as genes of unknown function. Less than 10% of the loci were polymorphic and applicable for genetic mapping. Design of PCR primers was not possible for many identified genes because most of the microsatellites found within these genes were located near the 3′ or 5′ ends so that there were not enough flanking regions. Nevertheless, if needed, the primers can be designed from their corresponding genomic DNA sequences. Some of the designed primers were non-amplifiable due to the presence of large introns in the gene sequences. In this case, primers should be redesigned using genomic DNA information.

One objective of the genome research for aquaculture species is to identify genes underlying quantitative traits. A well-defined linkage map is, therefore, needed for mapping of QTLs. Due to the limited EST sequence information for walking catfish, the number of polymorphic microsatellite loci, however, is too small to increase marker density on the map to be useful for mapping of quantitative trait loci or positional cloning of these loci. Therefore, large EST databases are needed for walking catfish to provide resources for isolation of gene-linked microsatellite markers. It is expected that with sufficient number of EST-derived microsatellites, a highly detailed linkage map will be developed for walking catfish.

This study has demonstrated that gene-centromere mapping was an effective way for construction of female linkage map because gynogenesis can be easily practiced in walking catfish. For species with many chromosomes, linkages between markers gene are difficult to detect but all genes can be mapped in relation to their centromeres. A total of 44 microsatellite markers including 11 EST-linked and 33

DNA-derived loci were mapped to the centromeres. In addition, a centromere as well as two marker loci was placed on Linkage Group 3 of the previous map based on AFLP markers. In the future, when new sets of markers are available, further expansion of the existing map will be practical by genotyping of the polymorphic markers in the same mapping families. Determining the position of centromere on the linkage map is useful for studies of their meiotic function or comparative mapping.

Although the number of polymorphic microsatellite loci identified in walking catfish were relatively small, these loci exhibited high level of cross-species amplification in other clariid and pangasiid catfishes, e.g., African catfish, striped catfish, black-ear catfish and Mekong giant catfish. It has been reported in other eukaryotic organisms that gene-derived microsatellites are more transferable than microsatellites developed from genomic DNA due to the conservation nature of functional genes. Many studies suggested that gene-related microsatellites are not appropriate for traditional population genetic investigations because they are considered non-neutral markers. For conservation studies, using neutral marker data, however, may not correctly assume adaptive population differentiation between populations. On the other hand, gene-associated markers provide alternative tool for identifying adaptive genetic diversity in natural populations.

This work will serve as a starting point towards increasing genetic marker information for improving genetic map. As the research field of conservation genetics has grown rapidly, it is expected that gene-derived markers will receive more attentions as markers of choice. For walking catfish, the availability of a large numbers of gene-related microsatellite markers will provide the first step for adaptive genetic investigations in this species.

LITERTURE CITED

Adam, M. D., J.M. Kelly, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, M. R. Kerlarvage, W. R. McCombie and J. C.Venter. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**. 252: 1651-1656.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. **J. Mol. Biol**. 215: 403-410.

Aliah, R. S., S. Sato and N. Taniguchi. 2000. An evaluation of genetic variability in Nishikigoi, *Cyprinus carpio*, stock from Niigata Prefecture based on microsatellite DNA markers. **Suisan Zoshoku.** 48: 25–31.

Allendorf, F. W. and R. F. Leary. 1984. Heteroygosity in gynogenetic diploids and triploids estimated by gene-centromere recombination rates. **Aquaculture**. 43: 413-420.

Allendorf, F. W., J. E. Seeb, K. L. Knudsen, G. H. Thorgaard and R. F. Leary. 1986. Gene-centromere mapping of 25 loci in rainbow trout. **J. Hered**. 77: 307-312.

Anderson, L. K., A. Reeves, L. M. Webb and T. Ashley. 1999. Distribution of crossing over on mouse synaptonemal complexs using immonofluorescent localization of MLH1 protein. **Genetics**. 151: 1569-1579.

Chen, S. L., H. Y. Ma, Y. Jiang, X. L. Liao and L. Meng. 2007. Isolation and characterization of polymorphic microsatellite loci from an EST library of turbot (*Scophthalmus maximus*) and cross-species amplification. **Molecular Ecology Notes**. 7: 848-850.

Copenhaver, G. P., K. C. Keith and D. Preuss. 2000. Tetrad analysis in higher plants. A Budding Technology. **Plant Physiol**. 124: 7-16.

Coulibaly, I. K.Gharbi, R. G. Danzmann , J. Yao and C.E. Rexroad. 2005. Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. **Animal Genetics**. 36: 309-315.

Chistiakov, D. A., B. Hellemans and F. A. M. Volckaert. 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. **Aquaculture**. 255: 1-29.

Da, Y., V. L. Jarrell, T. Wang, R. L. Fernando, M. B. Wheeler and H. A. Lewin. 1995. Multilocus Analysis for Gene-Centromere Mapping Using First Polar Bodies and Secondary Oocytes. **Genetics.** 139: 1091-1097.

Danzmann, R. G. and K. Gharbi. 2001. Gene mapping in fishes: a means to an end. **Genetica** 111: 3-23.

Donsakul, T. and W. Magtoon. 1989. **A chromosome study on two species of Clariid catfishes: *Clarias batrachus* and *Clarias macrocephalus*, from Thailand**. In: Proceeding . 27th Kasetsart University Conference January 30 – February 1. 1989 pp. 421- 428.

Ebersold, W. T. and R. P. Levine. 1958. A genetic analysis of linkage group I of *Chlamydomonas reinhardi*. **Molecular and general genetics**. 90: 74-82.

Excoffier, L., G. Laval and S. Schneider. 2005. **Arlequin ver. 3.0: An integrated software package for population genetics data analysis**. Computational and Molecular Population Genetics Lab, University of Berne, Switzerland.

Gestl, E. E., E. J. Kauffman, J. L. Moore and K. C. Cheng. 1997. New conditions for half-tetrad embryos in the zebrafish (*Danio rerio*). **The J. Hered**. 88(1): 76-79.

Hussain, M. G., B. J. McAndrew, D. J. Penman and P. Sodsuk. 1994. **Estimating gene-centromere recombination frequencies in gynogenetic diploids of *Oreochromis niloticus* L., using allozymes, skin colour and a putative sex-determination locus (SDL-2)**. In: Beaumont A.R. (ed). Genetics and Evolution of Aquatic Organisms. Chapman and Hall, New York, pp. 502-509.

Johnson, S. L., M. A. Gates, M. Johnson, W. S. Talbot, S. Horne, K. Baik, S. Rude, J. R. Wong and J. H. Postlethwait. 1996. Centromere-Linkage Analysis and Consolidation of the Zebrafish Genetic Map. **Genetics**. 142: 1277-1288.

Kamonrat, W. 1996. **Spatial genetic structure of Thai silver barb *Puntius gonionotus* (Breeker) population in Thailand**. Ph. D. Thesis, Dalhousie University, Dalhousie, Canada.

Kauffman, E. J., E. E. Gestl, D. J. Kim, C. Walker, J. M. Hite, G. Yan, P. K. Rogan, S. L. Johnson and K. C. Cheng. 1995. Microsatellite-Centromere Mapping in Zebrafish (*Danio rerio*). **Genomics**. 30: 337-341.

Kirby, L. T. 1992. **DNA Fingerprinting: An Introduction**. W.H. Freeman and Company, New York.

Kocher, T. D., W-J. Lee, H. Sobolewska, D. Penman and B. McAndrew. 1998. A genetic linkage map of cichild fish, the tilapia (*Oreochromis niloticus*). **Genetics** 148: 1225-1232.

Li, Q. and A. Kijima. 2005. Segregation of Microsatellite Alleles in Gynogenetic Diploid Pacific Ablone (*Haliotis discus hannai*). **Marine Biotechnology**. 7: 669-676.

Li, Y., M. Cai, Z. Wang, W. Guo, X. Liu, X. Wang and Y. Ning. 2008. Microsatellite-centromere mapping in large yellow croaker (*Pseudosciaena crocea*) using gynogenetic diploid families. **Marine Biotechnology**. 10: 83-90.

Lindner, K. R., J. E. Seeb, C. Habicht, K. L. Knudsen, E. Kretschmer, D. J. Reedy, P. Spruell and F. W. Allendorf. 2000. Gene centromere mapping of 312 loci in pink salmon by half-tetrad analysis. **Genome**. 43: 538-549.

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

Liu, Z., A. Karsi and R. A. Dunham. 1999. Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. **Marine Biotechnology**. 1: 437-447.

Martínez, P., M. Hermida, B. G. Pardo, C. Fernández, J. Castro, R. M. Cal, J. A. Álvarez-Dios, A. Gómez-Tato and C. Bouza. 2008. Centromere-linkage in the turbot (*Scophthalmus maximus*) through half-tetrad analysis in diploid meio-gynogenetics. **Aquaculture**. 280(1-4): 81-88.

Metzgar, D. and J. Bytof.. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. **Genome Res**. 10(1): 72-80.

Mohideen, M. P. K., J .L. Moore and K. C. Cheng. 2000. Centromere-linked microsatellite marks for linkage groups 3, 4, 6, 7, 13 and 20 of Zebrafish (*Danio rerio*). **Genomics.** 67: 102-106.

Naruse, K. and A. Shima. 1989. Linkage relationships of gene loci in the medaka, *Oryzias latipes* (Pisces: Oryziatidae), determined by backcross and gynogenesis. **Biochem. Genet**. 27:183–198.

Na-Nakorn, U., N. Tanigushi, E. Nugroho, S. Sedi and W. Kamonrat. 1999. Isolation and characterization of microsatellite loci of *Clarias macrocephalus* and their application to genetic diversity study. **Fisheries Science**. 64(4): 520-526.

Na-Nakorn, U., S. Sripairoj , T. Ngamsiri , S. Poompuang and N. Taniguchi. 2006. Gene-centromere Mapping at Four Microsatellite and Five Isoyme Loci in *Clarias mcrocephalus* Based on Meiotic Gynogenetic Offspring. **Aquaculture Science**. 54(4): 505-513.

Na-Nakorn, U., W. Luengpromporn and R. A. Dunham. 1994. **Gynogenesis induction of *Clarias macrocephalus* eggs using irradiated sperm of *Pangasius sutchi*.** PP.536-539, In Proceedings of the Third Asian Fisheries Forum, 26-30 October, 1992. Singapore.

Na-Nakorn, U., W. Rangsin and S. Witchasunkul. 1993. Suitable conditions for induction of gynogenesis in the catfish, *Clarias macrocephalus* using sperm of *Pangasius sutchi*. **Aquaculture**. 118, 53-62.

Neff, B. D. and M. R. Gross. 2001. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. **Evolution**. 55(9): 17-33.

Ng, S. H. S., G. Carlo, I. E. Artieri, R. C. Bosdet, R. G. Danzmann, W. S. Davidson, M. M. Ferguson, C. D. Fjell, B. Hoyheim and S. J. M. Jones. 2005. A physical map of genome of Atlantic salmon, *Salmo salar*. **Genomics**. 86: 396-404.

Nomura, K., K. Morishima, H. Tanaka, T. Unuma, K. Okuzawa, H. Ohta and K. Arai. 2006. Microsatellite-centromere mapping in the Japanese eel (*Anguilla japonica*) by half-tetrad analysis using induced triploid families. **Aquaculture**. 257: 53-67.

O'Brien, S. J. 1991. Mammalian genome mapping: lessons and prospects. **Current Opinion in Genetics and Development**. 1: 105-11.

Panprommin, D., S. Poompuang and P. Srisapoome. 2007. Characterization of expressed sequence tags from liver and muscle tissues of walking catfish, *Clarias macrocephalus*. **Aquaculture**. 276 (1-4): 60-68.

Pashley, C. H., J. R.Ellis, D. E. Mccauley and J. M. Burke. 2006. EST databases as a source for molecular markers: lessons from *Helianthus*. **J. Hered**. 97(4): 381–388.

Poompuang, S., and S. Sukmanomon 2003. **Conservation of microsatellite loci between *Clarias spp*. and *Pangasius hypophthalmus* and identification of AFLP markers in *C. macrocephalus* and *C. gariepinus***. Proceedings of 41st Kasetsart University Annual Conference, 3-7 February 2003.

Poompuang, S., and U. Na-Nakorn. 2004. A preliminary genetic map of walking catfish (*Clarias macrocephalus*). **Aquaculture**. 232:195-203.

Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. **J. Hered**. 86: 248-249.

Rice, W. R. 1989. Analyzing tables of statistical tests. **Evolution**. 43(1): 223-225.

Sato, T., S. Yokomizo, M. Matsuda, S. Hamaguchi and M. Sakaizumi. 2001. Gene-centromere mapping of medaka sex chromosomes using triploid hybrids between *Oryzias latipes* and *O. luzonensis*. **Genetica**. 111: 71-75.

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

Seki, S. and N. Taniguchi. 1989. The frequency of second division segregation of some isozyme loci in gynogenetic diploid of ayu, *Plecoglossus altivelis*. **Fish Genetics and Breeding Science**. 14: 43-48.

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

Siemon, H. S., A. Chang, G. D. Brown, B.F. Koop and W.S. Davidson. 2005. TypeI microsatellite markers from Atlantic salmon (*Salmo salar*) expressed sequence tags. **Molecular Ecology Notes.** 5(4): 762-766**.**

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package JOINMAP. **Plant J**. 3: 739-744.

Streisinger, G., F. Singer, C. Walker, D. Knauber and N. Dower. 1986. Segregation analysis and gene-centromere distances in zebrafish. **Genetics**. 112: 311-319.

Sukmanomon, S., S. Poompuang and M. Nakajima. 2003. Isolation and characterization of microsatellites in the Asian walking catfish *Clarias macrocephalus*. **Molecular Ecology Notes**. 3: 350-351.

Taggart, J. B., R. A. Hynes., P. A. Prodohl and A. Ferguson. 1992. A simplified protocol for routine total DNA isolation from Salmonid fishes. **J. Fish Biology**. 40: 963-965.

Taniguchi, N., A. Kijima and J. Fukai. 1987. High heterozygosity at *GPI-1* in gynogenetic diploids and triploids of ayu, *Plecoglossus altivelis*. **Nippon suisan Gakkaishi**. 53: 717-720. (in Japanese, with English tables)

Thiel, T., W. Michalek, R. K. Varshney and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene- derived SSR-markers in barley *(Hordeum vulgare* L.). **Theoretical and Applied Genetics**. 106: 414-22.

Thorgaard, G. H., F. W. Allendorf and K. L. Knudsen. 1983. Gene-Centromere Mapping in Rainbow trout : High interference over long map distances. **Genetics**. 103: 771-783.

Varshney, R. K., A. Graner and M. E. Sorrells. 2005. Genic microsatellite markers in plants: features and applications. **Biotechnology**. 23(1): 48-55.

Vasemagi, A and C. R. Primmer. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. **Molecular Ecology**. 14: 3623–3642.

Vasemagi, A., J. Nilsson and C. R. Primmer. 2005. Seventy-five EST-linked Atlantic salmon (*Salmo salar* L.) microsatellite markers and their cross-amplification in five salmonid species**. Molecular Ecology Notes**. 5: 282-288.

Voorrips, R. E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. **J. Hered**. 93 (1): 77-78.

Xu, Y., J. He, H. L. Tian, C. H. Chan, J. Liao, T. Yan, T. J. Lam and Z. Gong. 1999. Fast Skeletal Muscle-Specific Expression of a Zebrafish Myosin Light Chain 2 Gene and Characterization of Its Promoter by Direct Injection into Skeletal Muscle. **DNA and Cell Biology**. 18(1): 85-95.

Yeh, F. C., R. C. Yang and T. Boyle. 2000. **POPGENE version 1.32 Microsoft windows-based freeware for population genetic analysis**. University of Alberta and Center for International Forestry Reseaech, Alberta, Canada.

Yue, G. H., B. Kovacs and L. Orban. 2003. Microsatellites from *Clarias batrachus* and their polymorphismin seven additional catfish species. **Molecular Ecology Notes**. 3: 465-468.

Yue, G. H , M.Y. Ho, L. Orban and J. Komen. 2004. Microsatellite within genes and ESTs of common carp and their applicability in silver crucian carp. **Aquaculture**. 234: 85-98.

Zhao, H. and T. P. Speed. 1998. Statistical analysis of half-tetrads. **Genetics**. 150: 473-485.

Lahrech Z., K. Chiharu, M. Kagayaki, M. Tatsunari, S. Setsuo and A. Katsutochi. 2007. Genetic verification of induced gynogenesis and microsatellite-centromere mapping in Barfin flounder *Verasaper moseri*. **Aquaculture**. 272: S115-S124.

Appendix

Appendix Table 1 Microsatellite sequences searched from EST database of liver and muscle tissue using "Modified Sputnik II" program.

Seq.ID	Gene ID	Position	EST length	SSR length	SSRs_Seq
LS103	Unknown	-	604	12	$(TG)_6$
LS169-1	Unknown	-	842	12	$(TC)_6$
LS169-2	Unknown	-	842	62	$(CT)_{31}$
LS270*	Unknown	-	859	12	$(CGA)_4$
LS313*	Unknown	-	836	28	$(CT)_{14}$
LS631	putative secreted protein 4	3′UTR	876	12	$(AAG)_4$
LS663	Unknown	-	762	18	$(AC)_9$
LS1092	Unknown	-	302	40	$(CA)_{20}$
LS1123	Unknown	-	869	12	$(TGA)_4$
LS1138-1	Unknown	-	672	12	$(TTTC)_3$
LS1138-2	Unknown	-	672	12	$(TTTC)_3$
LS1138-3*	Unknown	-	672	12	$(CCTT)_3$
LS1138-4	Unknown	-	672	21	$(CAT)_7$
LS1121	Unknown	-	401	12	$(TAT)_4$
LS2219	Unknown	-	351	20	$(TC)_{10}$
LS2235	Unknown	-	854	15	$(TTCAA)_3$
LS2366	Unknown	-	858	12	$(CGG)_4$
LS2461*	Unknown	-	347	46	$(GT)_{23}$
LS114-1	Unknown	-	1494	14	$(GT)_7$
LS114-2	Unknown	-	1494	12	$(AAT)_4$
LS114-3	Unknown	-	1494	12	$(GCA)_4$
LS679	Vitellogenin	3′UTR	811	44	$(TC)_{22}$
LS2220	Vitellogenin	5′UTR	865	12	$(CTG)_4$
LS100016*	Vitellogenin	ORF	683	12	$(GCA)_4$
LS182	peroxiredoxin 4	3′UTR	654	12	$(CA)_6$
LS327*	dexamethasone-induced protein	3′UTR	865	15	$(GAT)_5$
LS353*	glucose transporter 2	3′UTR	684	12	$(TAT)_4$
LS652	cellular retinoic acid binding protein	3′UTR	612	18	$(TA)_9$
LS1062	Fibrinogen, B beta polypeptide	3′UTR	690	36	$(AC)_{18}$
LS1110-1	Anserinase	3′UTR	866	36	$(TC)_{18}$
LS1110-2	Anserinase	3′UTR	866	38	$(AC)_{19}$

Appendix Table 1 (Continued)

Seq.ID	Gene ID	Position	EST length	SSR length	SSRs_Seq
LS2127-1	mouse DNA sequence from clone RP23-189P1 on chromosome 11	5′UTR	906	15	$(AAAAG)_3$
LS2127-2	mouse DNA sequence from clone RP23-189P1 on chromosome 11	ORF	906	12	$(AAAT)_3$
LS2159-1*	histone deacetylase 6	3′UTR	577	18	$(GA)_9$
LS2159-2*	histone deacetylase 6	3′UTR	577	40	$(TG)_{20}$
LS2276	12S ribosomal RNA gene	ORF	977	12	$(TGAG)_3$
LS2286	Ribosomal protein L22	5′UTR	446	12	$(AAG)_4$
LS2329	hypothetical protein LOC393488	3′UTR	312	12	$(TTTC)_3$
LS2382	Ictacalcin	3′UTR	334	12	$(TTGG)_3$
LS100018	heat shock 90kDa protein 1 beta isoform b	ORF	649	12	$(GAG)_4$
LS191*	NADH dehydrogenase subunit 2	3′UTR	509	12	$(CTA)_4$
LS244	Unknown	-	814	12	$(AAT)_4$
LS360-1*	Polypterus bichir, clone - 164C2	3′UTR	814	15	$(AAAAC)_3$
LS360-2	Polypterus bichir, clone - 164C2	3′UTR	814	12	$(AAAC)_3$
LS376*	Unknown	-	782	15	$(CCT)_5$
LS411*	Unknown	-	445	16	$(GA)_8$
LS428-1	Unknown	-	810	12	$(TAAA)_3$
LS428-2	Unknown	-	810	21	$(AAT)_7$
LS474-1	Unknown	-	725	12	$(GCT)_4$
LS474-2*	Unknown	-	725	15	$(TTA)_5$
LS492*	Unknown	-	711	27	$(ATT)_9$
LS493*	Unknown	-	420	12	$(GAT)_4$
LS510-1	Unknown	-	447	27	$(ATT)_9$
LS510-2	Unknown	-	447	28	$(TG)_7C(TG)_6$
LS510-3*	Unknown	-	447	12	$(TTTG)_3$
MS105	Unknown	-	591	54	$(TG)_{27}$
MS113	Unknown	-	417	28	$(TTTA)_7$
MS136	Unknown	-	524	57	$(TAA)_{19}$
MS147*	Unknown	-	309	14	$(CA)_7$
MS551-1*	Unknown	-	329	15	$(ACC)_5$
MS551-2	Unknown	-	329	14	$(AC)_7$
MS556-1	Unknown	-	860	14	$(AAAG)_{11}$
MS556-2	Unknown	-	860	16	$(GT)_8$
MS556-3	Unknown	-	860	16	$(GA)_8$

Appendix Table 1 (Continued)

Seq.ID	Gene ID	Position	EST length	SSR length	SSRs_Seq
MS561	Unknown	-	284	21	$(TGA)_7$
MS591*	Unknown	-	369	34	$(TC)_{17}$
MS1129	Unknown	-	318	33	$(TTG)_{11}$
MS2191	Unknown	-	782	36	$(GAT)_{12}$
MS2205-1	Unknown	-	392	12	$(TC)_6$
MS2220	Unknown	-	767	12	$(CCA)_4$
MS2242	Unknown	-	959	12	$(CTTT)_3$
MS2259	Unknown	-	301	40	$(ATTAG)_8$
MS2432	Unknown	-	346	28	$(GA)_{14}$
MS613*	myosin light chain 1	3′UTR	824	22	$(AC)_{11}$
MS91116*	myosin heavy chain	5′UTR	618	12	$(AAAC)_3$
MS995-1	Schistosoma japonicum SJCHGC01887 protein	3′UTR	555	22	$(GA)_{11}$
MS995-2	Schistosoma japonicum SJCHGC01887 protein	3′UTR	555	12	$(CAAA)_3$
MS2147	Danio rerio zgc:66286, mRNA	3′UTR	599	12	$(TTTC)_3$
MS2178	elongation factor 1 alpha	3′UTR	371	15	$(TTA)_5$
MS552*	slow troponin T 2 (sTnT2)	3′UTR	705	12	$(TC)_6$
MS2350-2*	Synaptotagmin binding, cytoplasmic RNA interacting protein	3′UTR	396	12	$(TCC)_4$
MS125	Zebrafish DNA sequence from clone CH211-285O8 in linkage group 19	3′UTR	382	12	$(TTA)_4$
MS1048	Tnnt3b protein (troponin T3b)	ORF	463	12	$(GA)_6$
MS2230	NADH dehydrogenase (ubiquinone) 1 beta subcomplex	3′UTR	478	18	$(AG)_9$
MS2289	18S rRNA gene (partial), 5.8S rRNA gene, 28S rRNA gene (partial)	ORF	458	12	$(GC)_6$
MS2328	Ribosomal protein L14	5′UTR	849	22	$(TG)_{11}$
MS2369	Mus musculus RIKEN cDNA 1810014F10 gene	3′UTR	426	12	$(TGGA)_3$
MS2414	GTH alpha mRNA for gonadotropin alpha-subunit	5′UTR	845	12	$(TTA)_4$
MS2506	Ictacalcin	3′UTR	307	12	$(TTGG)_3$
MS2509*	hypothetical protein LOC550569	3′UTR	510	34	$(TG)_{17}$
MS9310	Pavalbumin	5′UTR	563	28	$(AAAG)_7$
MS106*	Actin	ORF	680	12	$(TCC)_4$
MS532*	skeletal muscle actin	ORF	889	12	$(TCC)_4$
MS2153*	Danio rerio zgc:66286, mRNA	3′UTR	597	12	$(TTTC)_3$
MS227	actin mutant	ORF	565	12	$(TCC)_4$
MS625	Unknown	-	325	14	$(TA)_7$
MS702*	receptor for activated protein kinase C homolog	3′UTR	745	22	$(TC)_6(AC)_5$
MS713*	Unknown	-	819	46	$(TC)_{23}$

Appendix Table 1 (Continued)

Seq.ID	Gene ID	Position	EST length	SSR length	SSRs_Seq
MS748*	Unknown	-	228	12	$(TA)_6$
MS792*	Unknown	-	713	12	$(AAG)_4$
MS805*	Unknown	-	681	12	$(GA)_6$
MS825*	Unknown	-	308	12	$(TTTC)_3$
MS887	Unknown	-	233	18	$(AT)_4T(AT)_4$
MS921*	Unknown	-	347	28	$(TC)_4TT(TC)_9$
MS942*	phosphoglucose isomerase-2 (pgi-2 gene)	3′UTR	288	18	$(AC)_5AT(AC)_3$
MS954*	Unknown	-	815	42	$(AAT)_{14}$
MS1138	Unknown	-	382	12	$(TTA)_4$
MS1179	actin mutant	ORF	557	12	$(TCC)_4$
MS1207	Unknown	-	454	18	$(CTA)_3A(CTA)_3$
MS1244-2*	Troponin T	ORF	485	12	$(AAG)_4$
MS1271*	actin	ORF	820	12	$(TCC)_4$
MS9587	Troponin T	3′UTR	585	12	$(AAAT)_3$
MS91007	actin	ORF	688	12	$(TCC)_4$

* indicates sequences which primers were designed.

Appendix Table 2 Forty-one microsatellite primers developed from genomic DNA including GenBank accession number, primer sequences, repeat sequence, annealing temperature (T_a) and size range (bp).

Locus	GenBank Accession	Primer sequence 5′-3′	Repeat sequence	T_a (°C)	Size range (bp)
*Cma-6**	AY185607	GGG CAC TAA GGG GTC GCT CTC GGG GCT TCT GGG ACA TCC TCT	$(GT)_5(GA)_8$	50	232-239
*Cma-8**	AY185608	CCG TGA TAC AAC TGT GAC T CGG TGC ACT GAA AGG	$(CA)_{26}$	55	212-286
*Cma-13**	AY185610	TGA GGG GAG GCA GGA G TGT TTC TCA CTC TTG GCA TTT	$(CA)_{42}$	55	224-248
*Cma-14**	AY185611	TAG AGA CAT TTA CGC TTC A TGC AAA GGC TAA TCA A	$(GT)_{34}$	50	178-212
*Cma-17**	AY185612	CGC CAT TGT TGT GAT AAA G GAT GAA GAT AAA AGC GAA GGA	$(CA)_{27}$	55	184-250
*Cma-19**	AY185613	ATC AGA GCC CTT TCA TCA CC CGT GCG AGT TCC CAG AG	$(GA)_8$	50	131-136
*Cma-20**	AY185614	TGT AAA CCA GGG CTG ATT TGT TG AAA CGG TGG GGA CAA ACT GTA TC	$(GT)_{26}$	60	213-255
*Cma-21**	AY185615	CTC GCT TAA AGG CAA GTT CAC TC CGC CAT ATA GCC ATA GAG GTG TG	$(CT)_9$	60	178-200
*Cma-22**	AY185616	TGT GTA CGA GTG TGT TTC TCA GTG CAG TTA CAC ACT CAC GCA AAT CAG	$(GT)_{10}$	50	184-216
*Cma-23**	AY 185617	GCA ACC CAA CAA TGT GTT TAC AG TAT GAA AAA TAA CGT GCC TAC CG	$(AAT)_6GAG(AAT)_5N_{13}(CA)_3GC(CA)_3$	55	262-278
*Cma-24**	AY 185618	GCC TCT AAT CCG TCG TAA AAA TG TCA TGG GTA AAT CTG GCA GAA AG	$(CA)_5CG(CA)_2CG(CA)_7CG(CA)_4$	55	236-254

Appendix Table 2 (Continued)

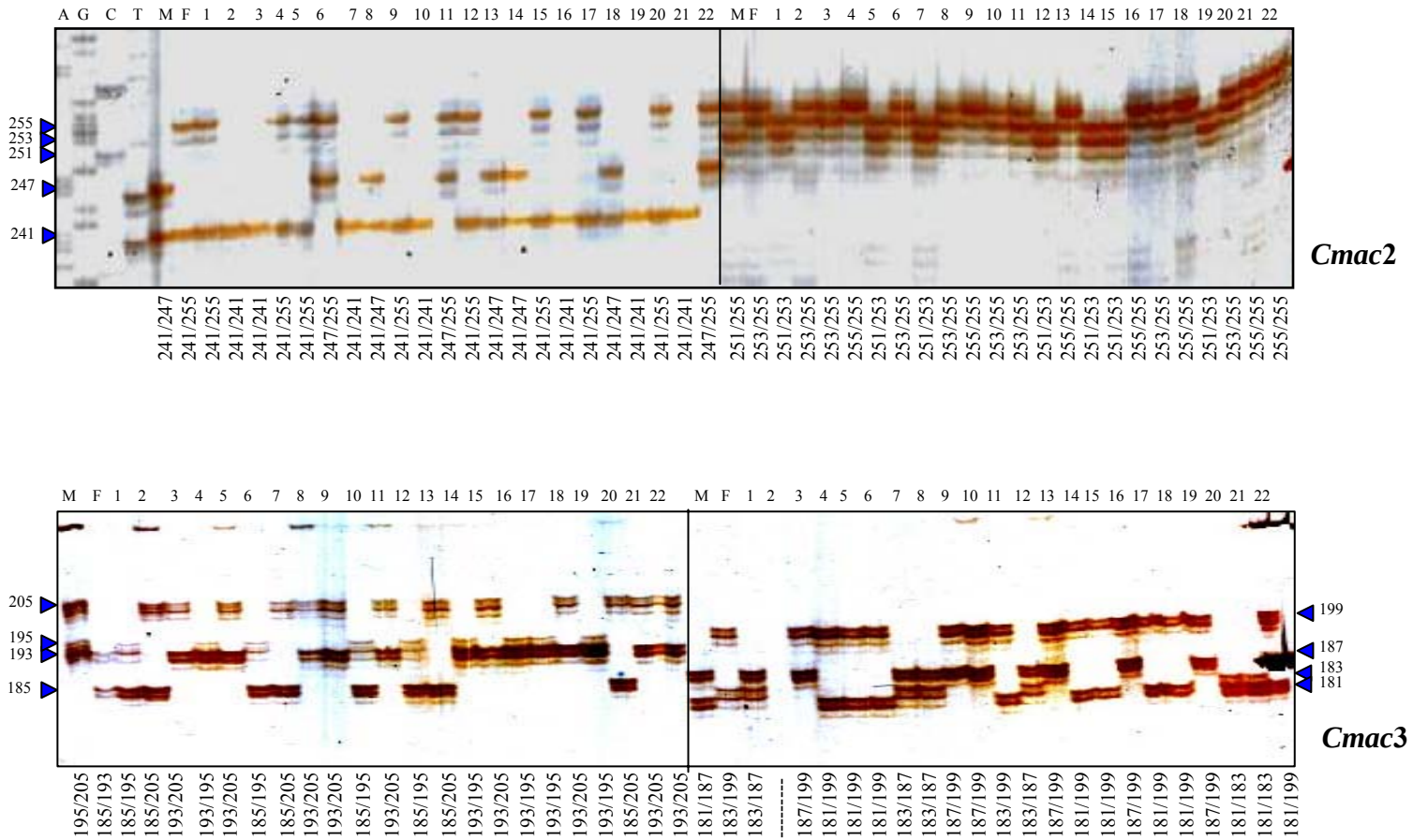
Locus	GenBank Accession	Primer sequence 5′-3′	Repeat sequence	T_a (°C)	Size range (bp)
*Cma-28**	AY357077	GAG GAA GAC GGA AAG GGA AGA G TCC GAA ACA AAG AGC GTA GTG TG	$(GT)_{43}$	60	151-249
*Cma-29**	AY357078	GAA GTC ATG TGG AAC GAA TGA GG TTT TAC ACT GGG CAA ACT GAA AC	$(GT)_4CT(GT)_4AT(GT)_{19}GC(GT)_{12}$ $(GA)_{15}GTAT(GT)_4$	60	249-427
*Cma-30**	AY357079	TCA ATC ATT TCC TCC TGA TGG AC CTCA ATC AAG CCC TGC AAT CTG	$(CT)_{16}GT(CT)_5GT(CT)_4AT(CT)_3$	60	311-500
*Cma-31**	AY357080	AGA TTC TCC CTC CCT GCT TGG AGA ATT TGC CTC ACC CGA AAT G	$(CT)_{17}$	60	200-250
*Cma-32**	AY357081	ACA ATT CAG CTT CCG TCG ATG AG AGA CCC TGT GGA TGA AGG TCT G	$(CT)_3CC(CT)_7GT(CT)_6CCGT(CT)_4TT\ (CT)_3$ $GT(CT)_5GT(CT)_7GT(CT)_7GT(CT)_8GT(CT)_4$ $CC(CT)_4CCGT(CT)_4GT(CT)_3GT(CT)_5GT$ $(CT)_9CC(CT)_6GT(CT)_7GT(CT)_6GT(CT)_5GT$ $(CT)_5GT(CT)_3$	65	427-500
*Cma-33**	AY357082	CCA TGC CTC AGA CCT CTC TAC TG TAT TGA TGC GGG AAT AAA TGA CG	$(GT)_{41}$	65	200-311
*Cma-34**	AY357083	CGA AAA ATG AAG CCC AAT GTA TG AAG CCT CCC AAC ATA ACA ACA CC	$(GT)_{34}$	62	180-249
*Cma-35**	AY357084	CAA AAC GGC ACT ATG GAG GAA C TCT GAC GTG TTA GTT TGT GAC AGC	$(GA)_{13}(GT)_{29}$	65	180-249
*Cma-36**	AY357085	ATA GAA GCA GCC CAG ACC TCA GT TCA AGT CAG ACG TTT CAG GTC AC	$(GT)_{23}$	68	180-249
*Cma-37**	AY357086	GAC CTT GTT AGG GGA CAA GTG TG GAT TCG CTA GCT GAC CTC ATC CT	$(GA)_{14}N_{71}(GT)_{29}$	65	200-311

Appendix Table 2 (Continued)

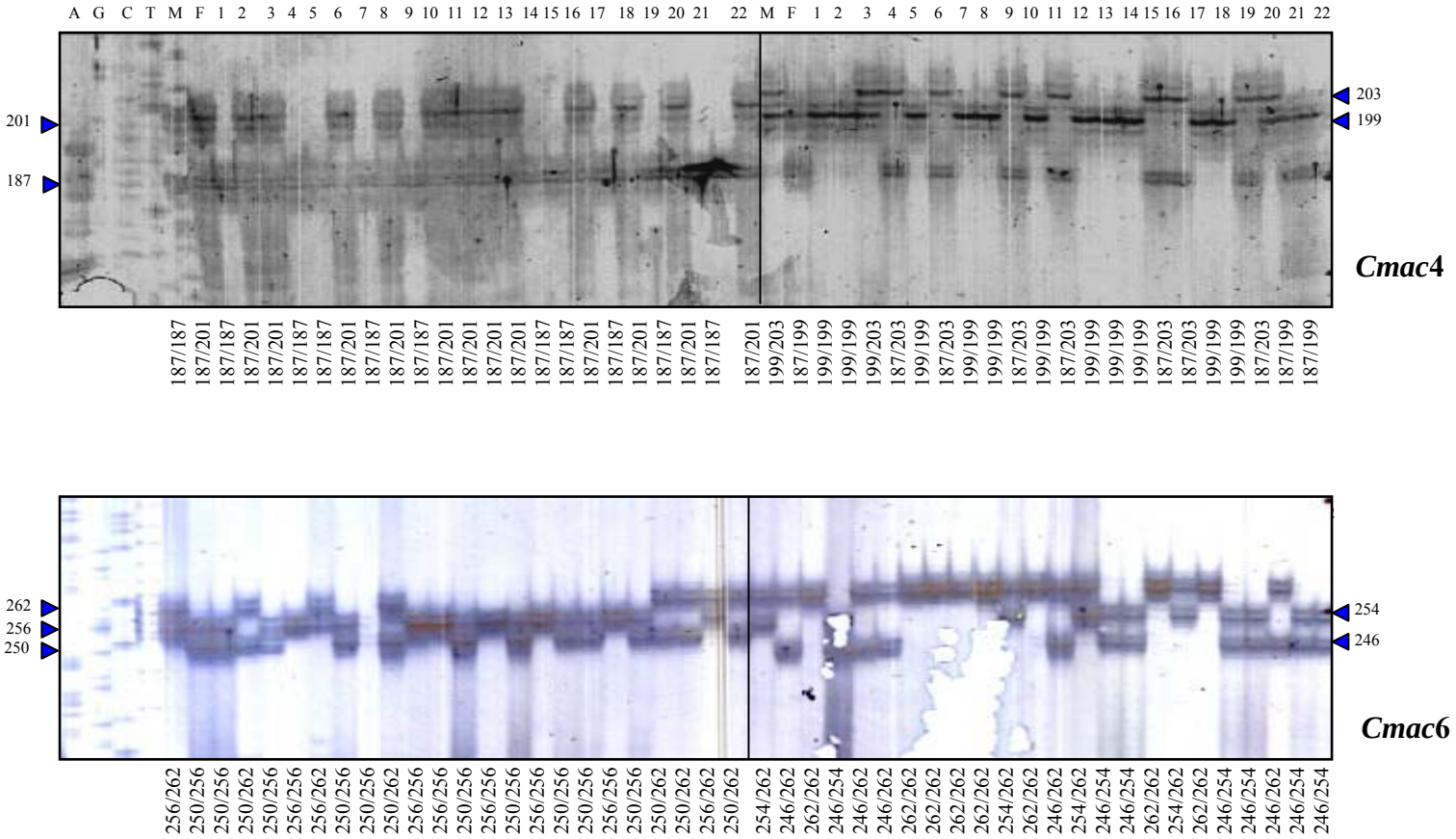
Locus	GenBank Accession	Primer sequence 5′-3′	Repeat sequence	T_a (°C)	Size range (bp)
*Cma-38**	AY357087	GGG TAT GAG GTC GAG TCT CTG TG GAA GAA GCT GCG AAT CTG ACG	$(CT)_6CA(CT)_3CA(CT)_5CA(CT)_5CA(CT)_5$ $CAGG(CT)_3CA(CT)_5CG(CT)_5CA$ $(CT)_2TT(C)\ TT(CT)_2CA(CT)_6CA(CT)_5CA(CT)_8$ $CG(CT)_5G(CT)_6CAGG(CT)_3CA(CT)_5CA(CT)_7$	60	151-533
*Cma-39**	AY357088	CCA TGT GGA ATG CTG TGA ACC AGC AAT CTC GCA AAA ACT CAG C	$(GA)_2CA(GA)GG(GA)_4GAA(GA)_3GGGA$ $(GA)_9GAAA(GA)_{16}G(CA)_{36}$	60	311-500
*Cma-40**	AY357089	AGG GGT TGG TTC CAT AAA TTG AC TGC TGG CAA CCT AAA TAA AAG CA	$(GT)_{21}$	65	200-311
*Cma-41**	AY357090	CGT CGT GCA ATA CAC ACA TTT TG GGT GCA GAA AAG TAA AGT TGA GCA G	$(CA)_{11}CC(CA)_7GA(CA)_4$	60	533-729
*Cma-42**	AY357091	CGA GAT CAA AGA ACC GAG TCA C ATT GGG GCC AGA ATG ACT TAT TG	$(GT)_{34}$	55	220-249
*Cma-43**	AY357092	CAG TCA ATT CCG AAC TCT TGC TC GAT CTT GTG TTT CCC TTT CAT GG	$(CT)_{21}$	55	220-249
*Cma-44**	AY357093	GTG GCT TAGTTG TTG GCT TGA TG TGC CTT TCT TTC AGT CTC TCC TC	$(CA)_{32}$	60	200-249
*Cma-45**	AY357094	GCC AGT CTC TCC CCT CCT TAT TC CAG AGA CAG GGA CAG AGA GCA AC	$(CT)_{13}$	55	220-427
*Cma-46**	AY357095	GCT CTC GTC ATC AGC CAG TAG AG ATG CAA GGC TTC ACT TTC TCA AG	$(GT)_{20}(GA)_7$	55	200-249
*Cma-47**	AY357096	TTG TTT CTC CAG CAC ACA CCT G GTG AGT CGT GTC CAA GTT GGT G	$(CA)_{25}$	60	180-220
*Cma-48**	AY357097	TTG TGT GTA AGA TCG GAC CAT TG CGA GTG ATC ATA AAG AGA GAG AGC	$(CT)_{17}(CA)_8$	60	220-280

Appendix Table 2 (Continued)

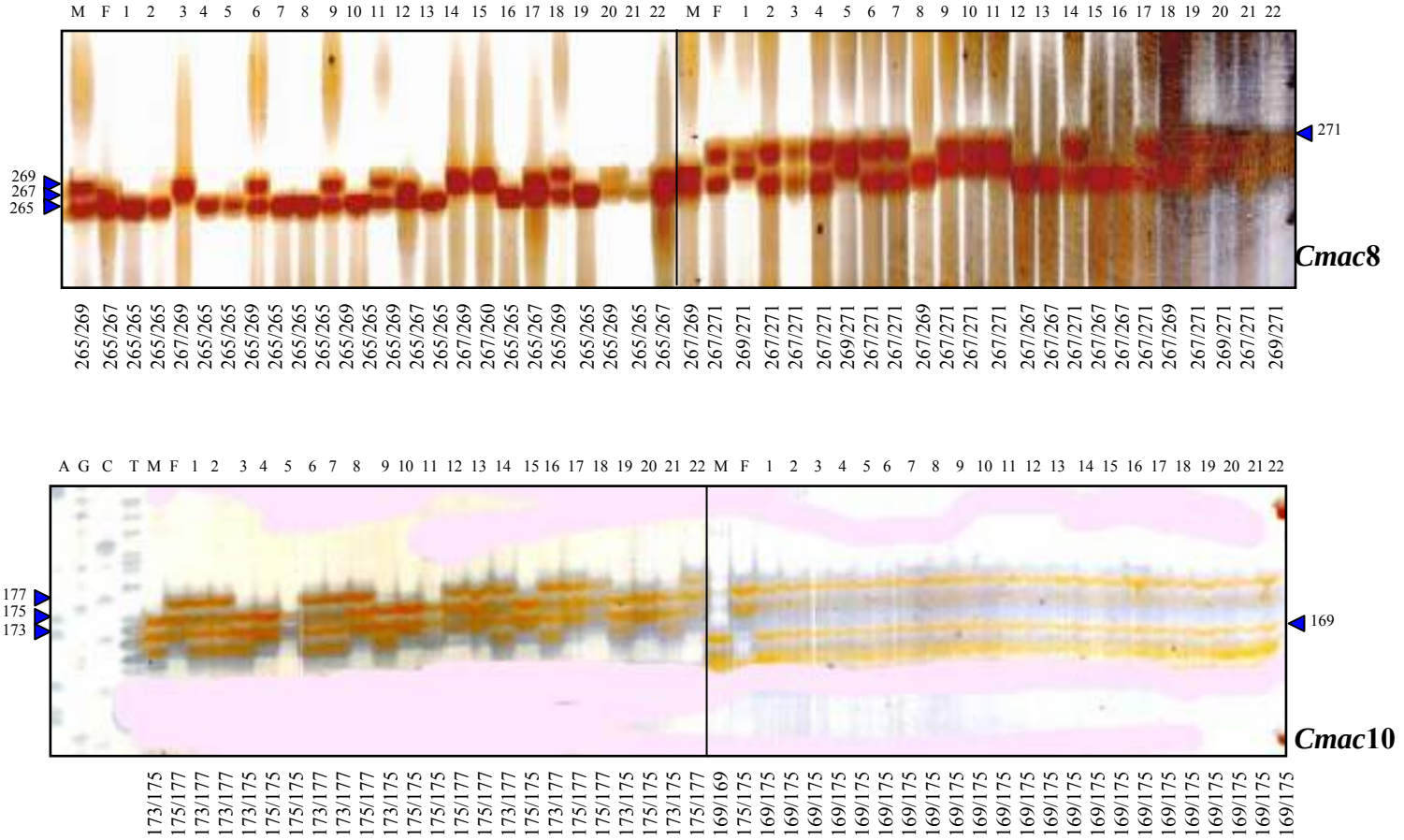
Locus	GenBank Accession	Primer sequence 5′-3′	Repeat sequence	T_a (°C)	Size range (bp)
*Cma-49**	AY357098	CGT ACC CCT ACT CGA ATT GAA GG ATC CTA ATC AAT GGG GCT GCT C	$(GT)_{30}$	55	249-311
*Cma-50**	AY357099	GCA ACA CAC ACC TCT GGT TTC AC TTC ATC TGC TGG GAT TTA CCT CA	$(CA)_{21}$	65	220-311
*Cma-51**	AY357100	ATA TGG GGT TAG GAA GCG AGA GC AGT GGG AGG AAC ACA TGG AAT G	$(GA)_{19}N_{108}(CA)_8TG(CA)_4C(CA)_4$	60	249-311
*Cma-52**	AY357101	AGC AAG GAA AGC TGG AGA TGT G CCC ACT TAC TGG GTC TCA CTG G	$(CA)_{17}$	60	249-311
*Cma-53**	AY357102	GAG AGA TGG AGG AAG TGT GAT GTG TGG CCA TGT CTA CAC AGG TCA G	$(GT)_{40}$	62	240-311
*Cma-54**	AY357103	CAT CTA AAT GGA ATG GAA TGT GC TCG GTC GTT CCT CTG TGT TTA TC	$(GT)_{26}$	55	200-260
*Cma-55**	AY357104	TAT GGA ACA AGG CAT GGA GTA CC TTT CCT CCA ACA GTC CAA AGA CA	$(CA)_9GA(CA)_4GA(CA)_{25}AAGA(CA)_3(CT)_{10}$	60	230-280
*Cma-56**	AY357105	TGC GAT GTC TTA TTC AAT CCA CAG TCT ATT CAG GAA GCG GCT ACG ATG	$(CT)_7(CA)_2CC(CA)_{21}$	60	249-311
*Cma-57**	AY357106	GTG GCT TAG TTG TTG GCT TGA TG TGC CTT TCT TTC AGT CTC TCC TC	$(CA)_{11}C(GA)_2(CA)_{32}$	62	200-311



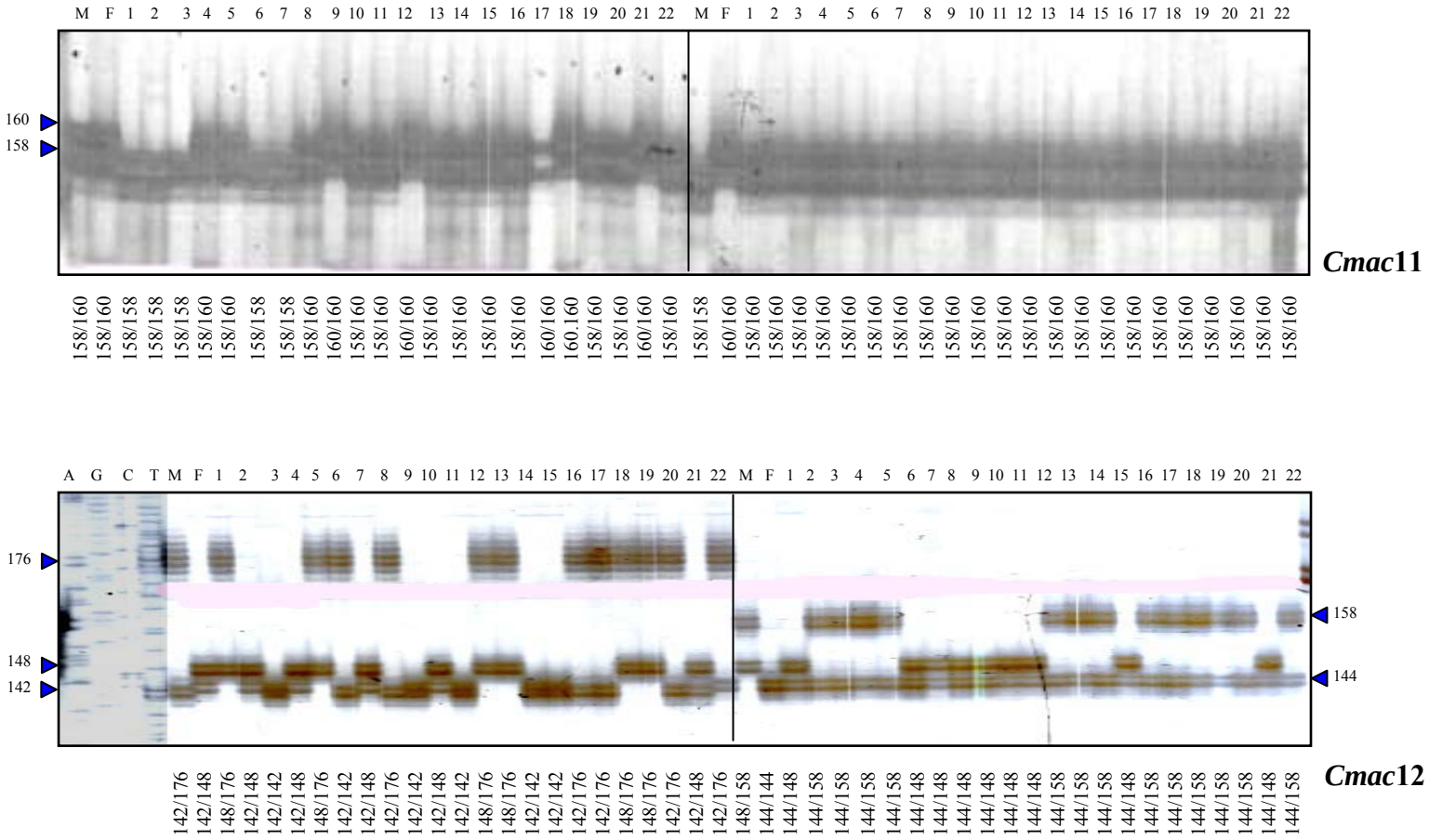
Appendix Figure 1 Medelian segregation at 10 microsatellite loci in two diploid families



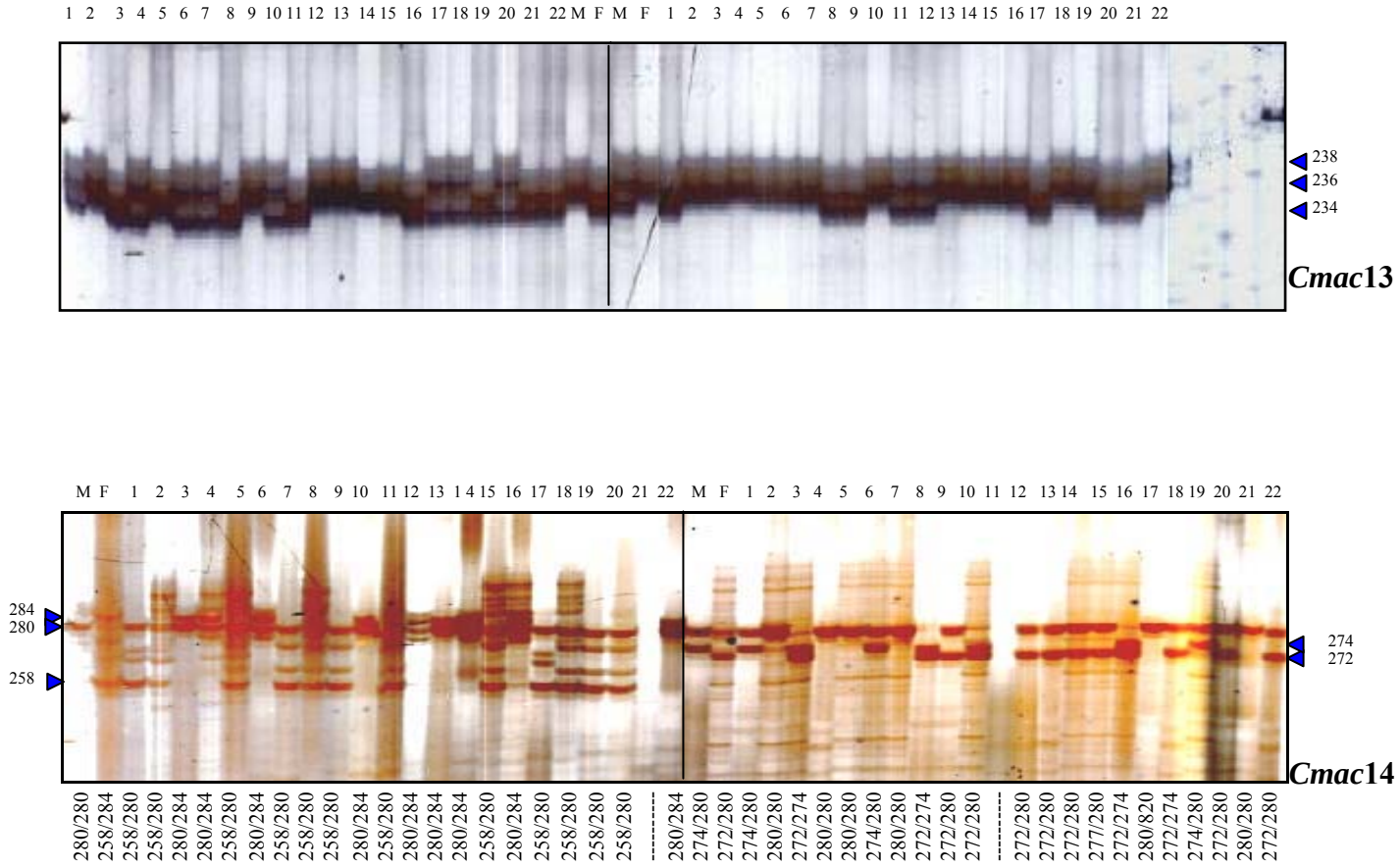
Appendix Figure 1 (Continued)



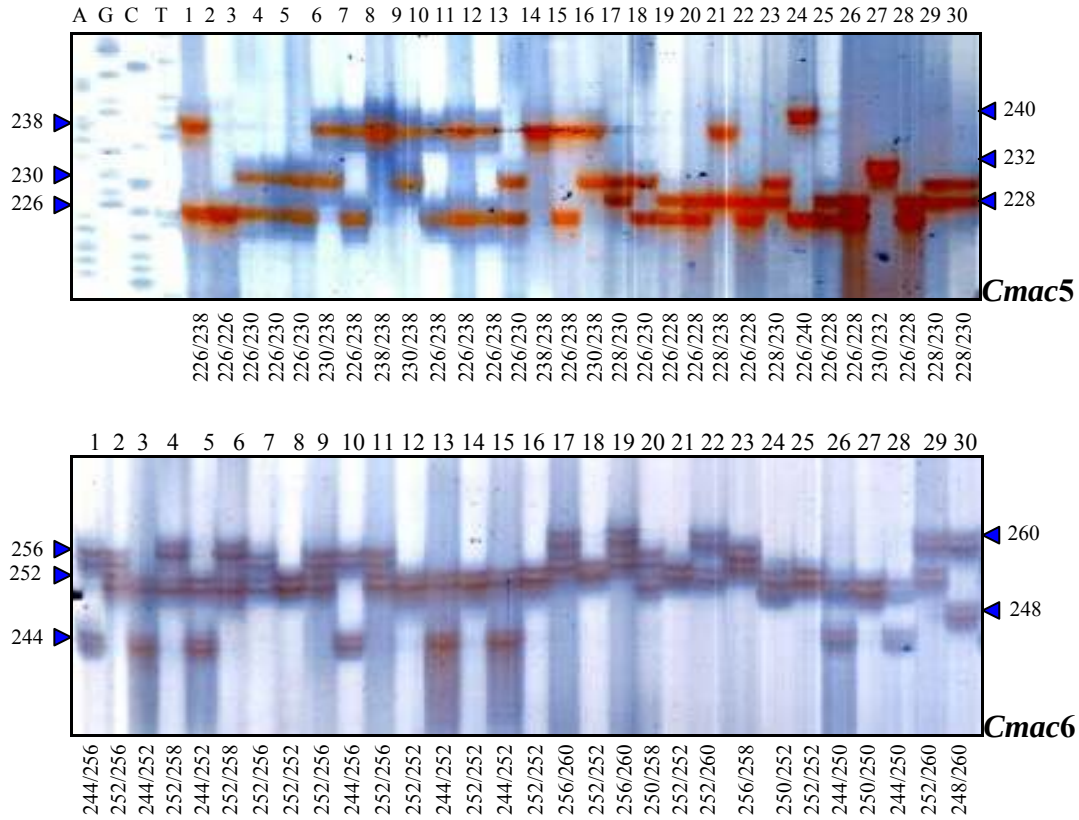
Appendix Figure 1 (Continued)



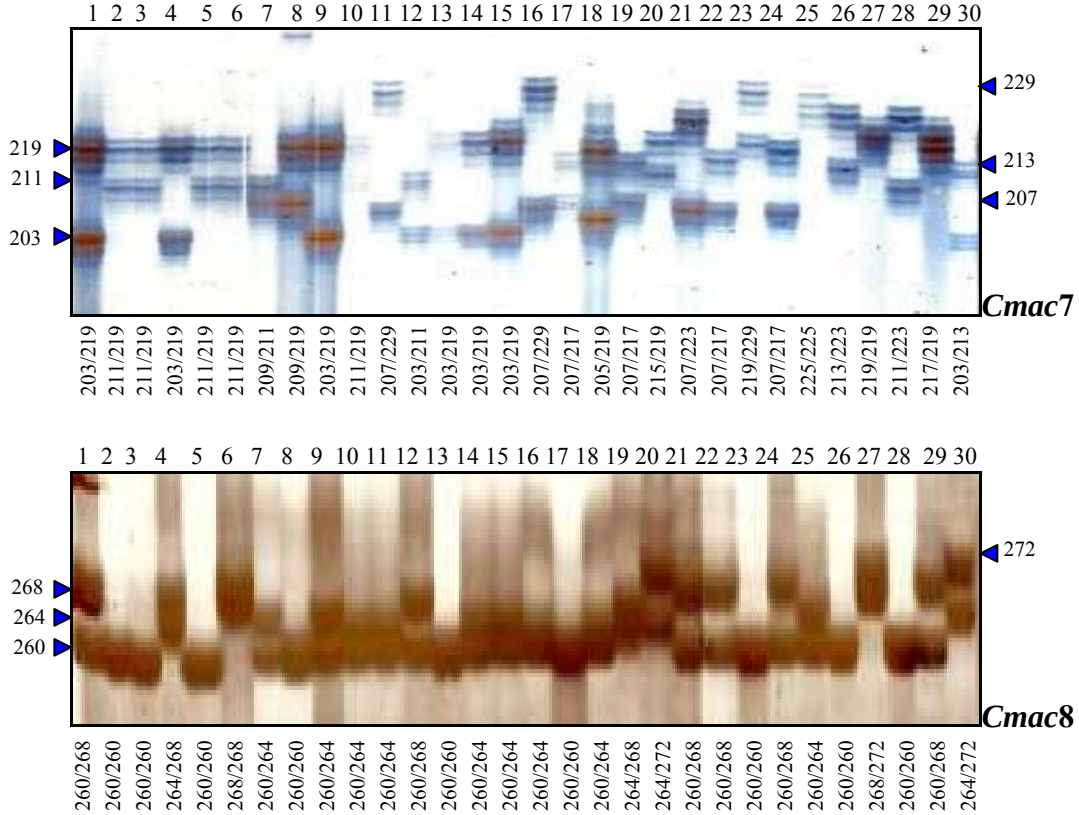
Appendix Figure 1 (Continued)



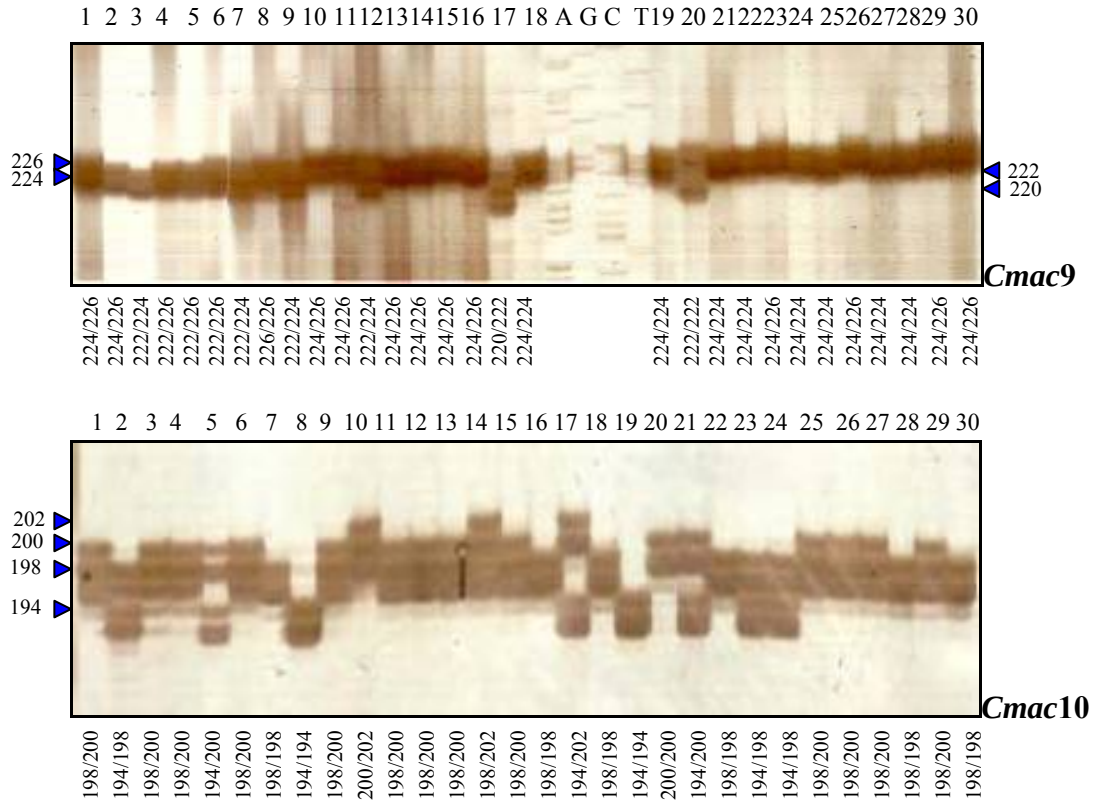
Appendix Figure 1 (Continued)



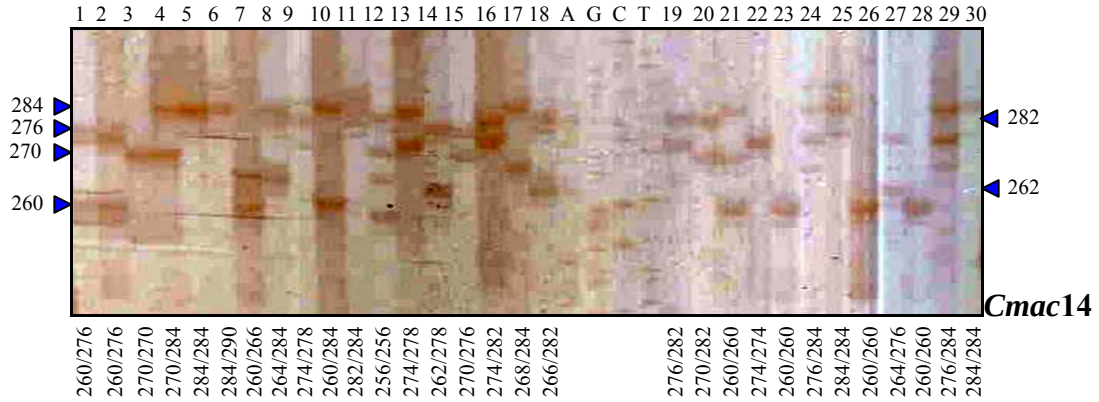
Appendix Figure 2 Genotypes of 30 unrelated fish at seven loci.



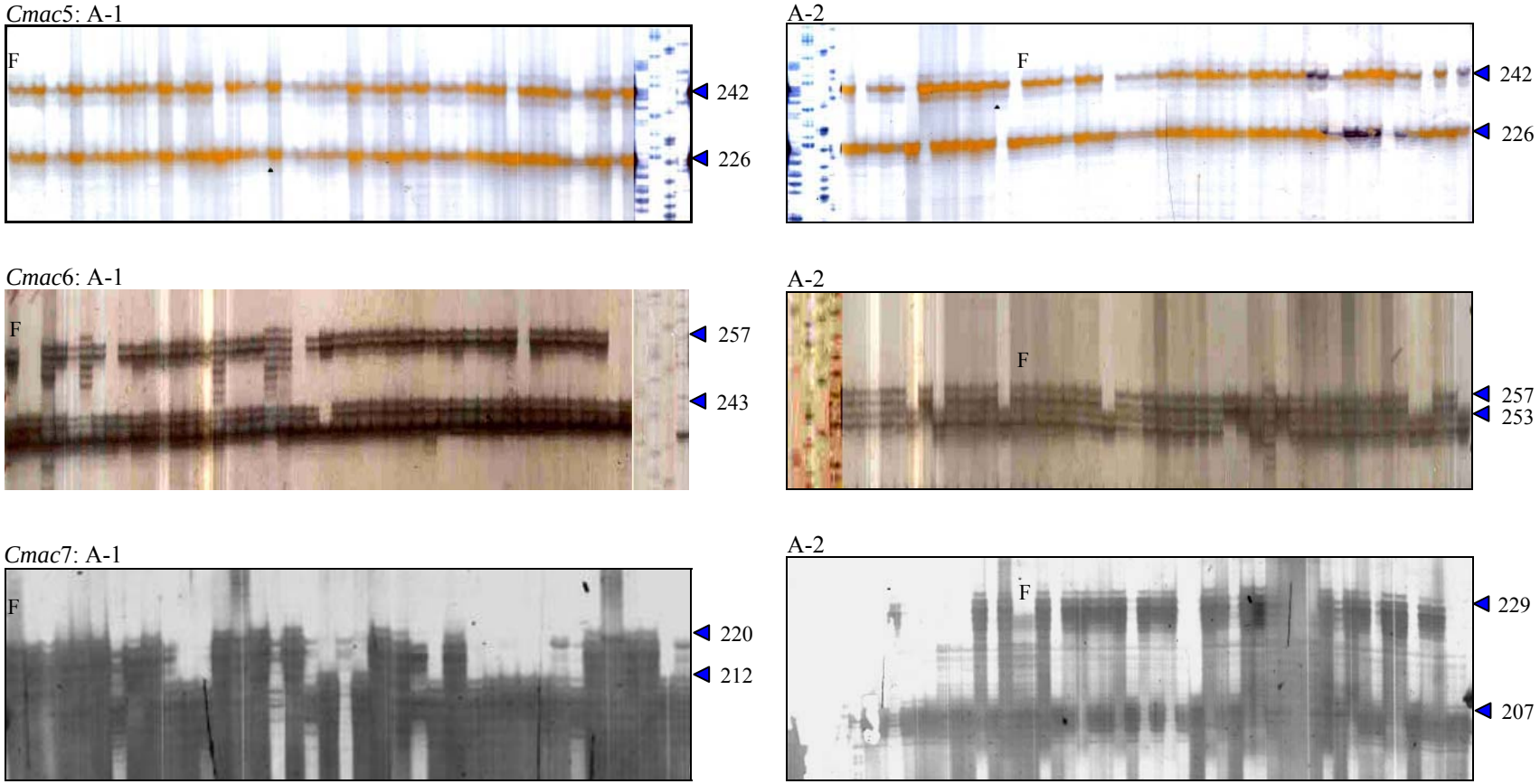
Appendix Figure 2 (Continued)



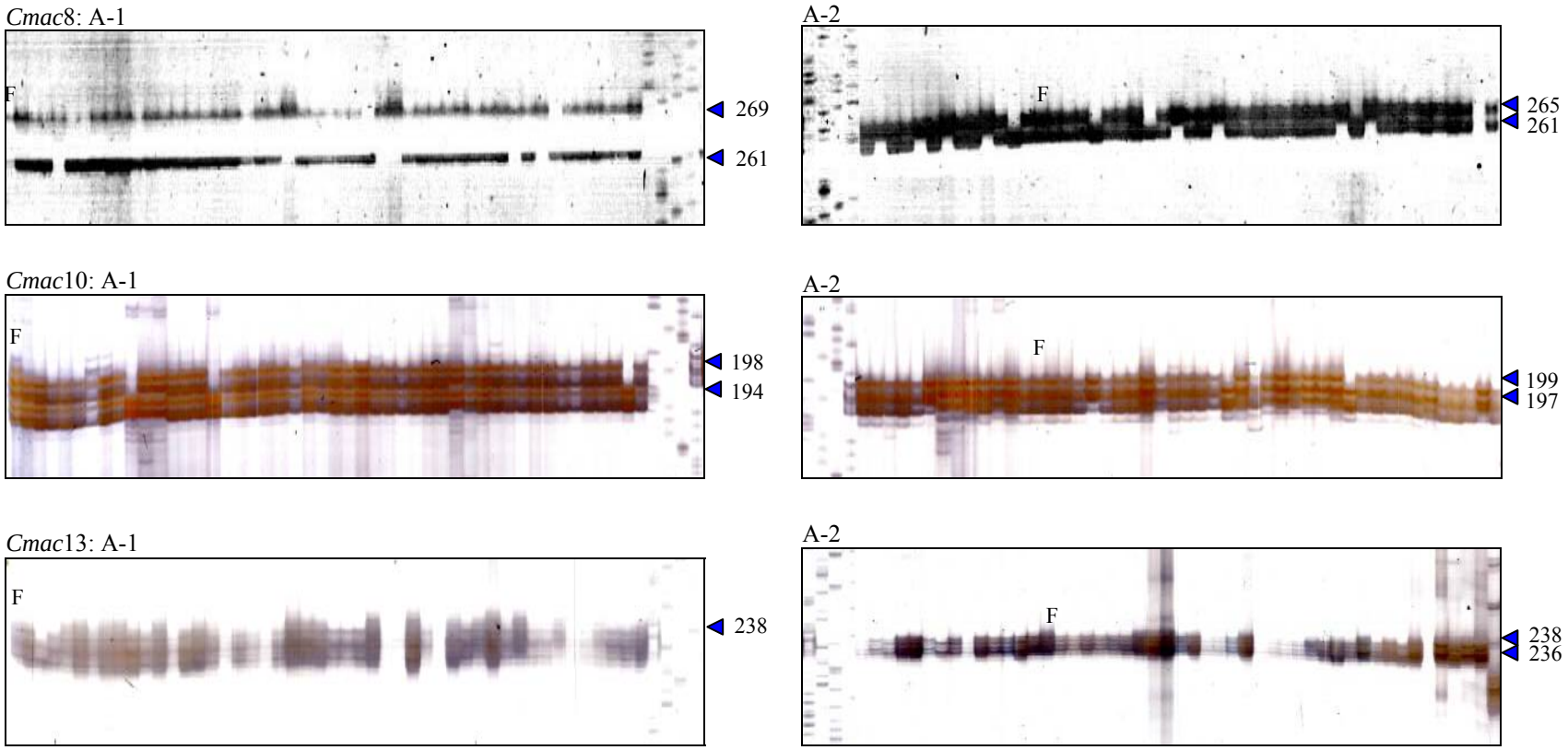
Appendix Figure 2 (Continued)



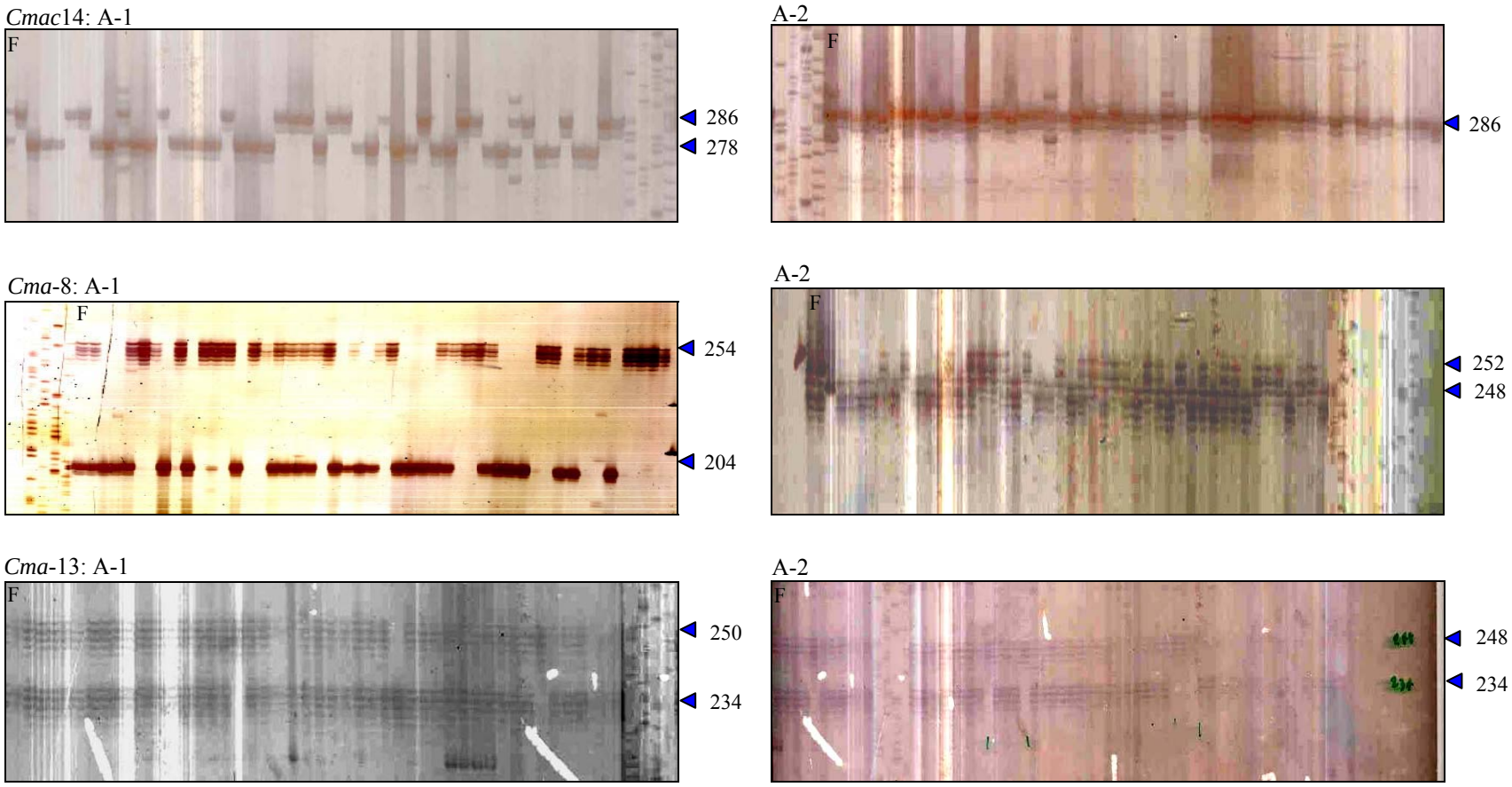
Appendix Figure 2 (Continued)



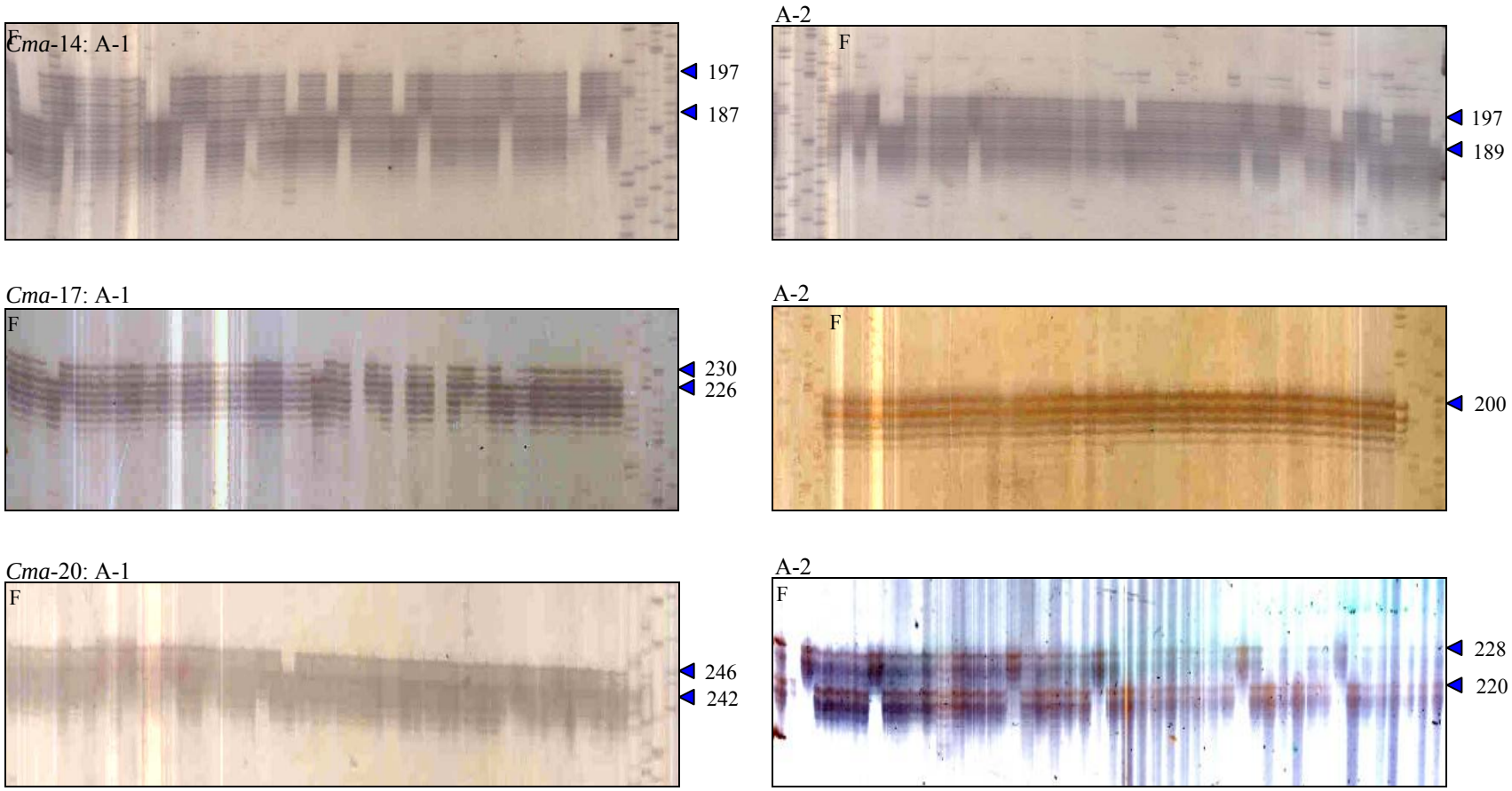
Appendix Figure 3 Progeny genotypes at 40 loci in two gynogenetic diploid families (A-1 and A-2)



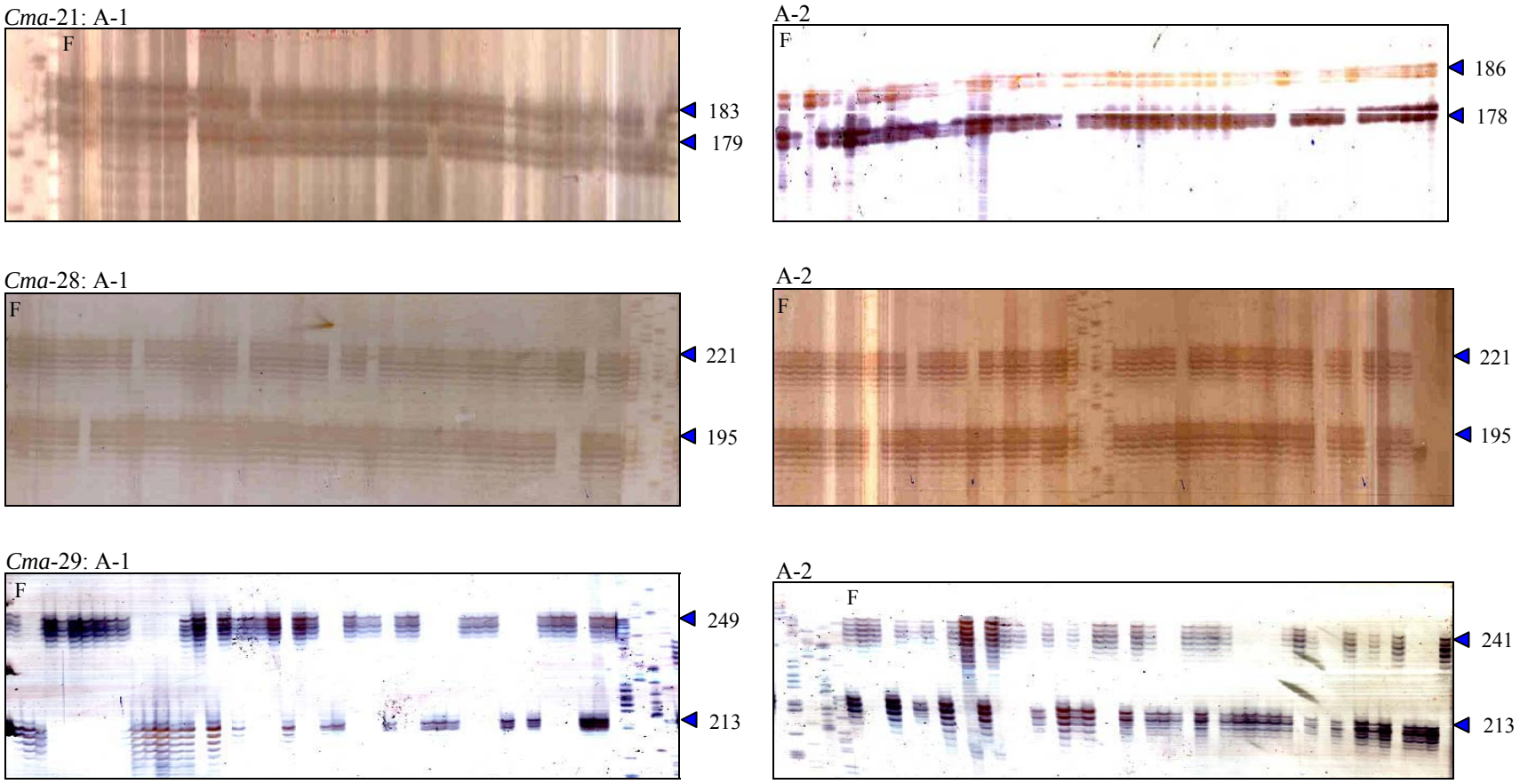
Appendix Figure 3 (Continued)



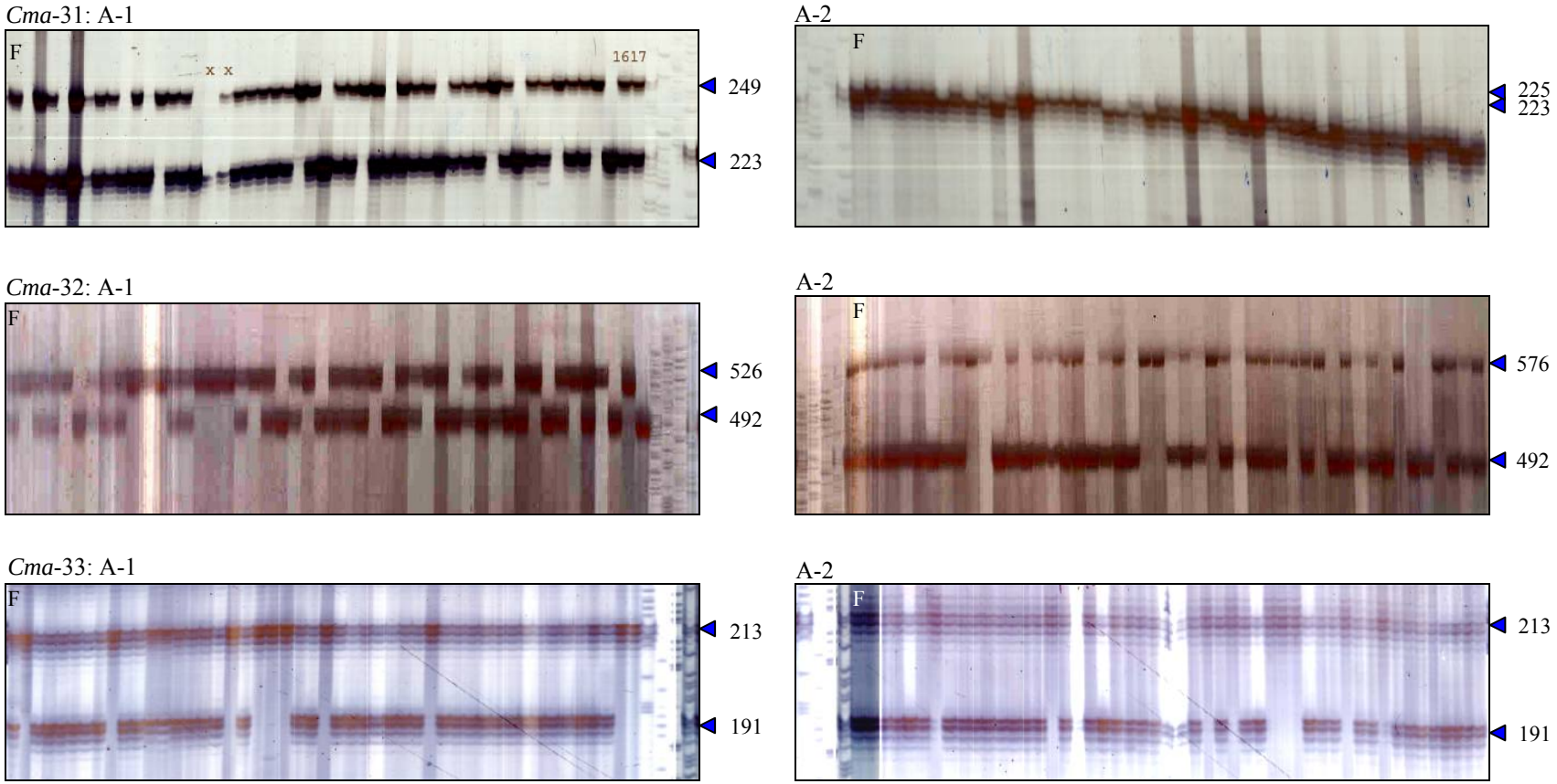
Appendix Figure 3 (Continued)



Appendix Figure 3 (Continued)

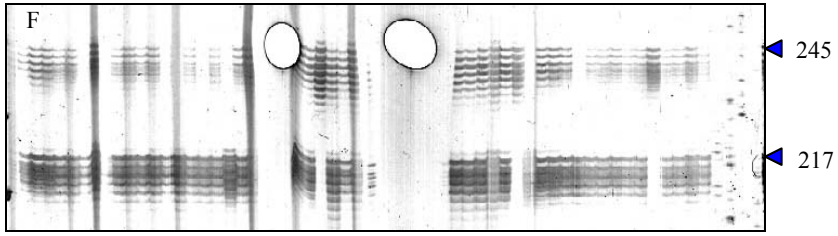


Appendix Figure 3 (Continued)

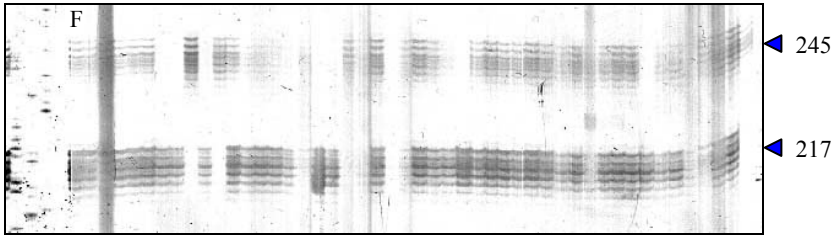


Appendix Figure 3 (Continued)

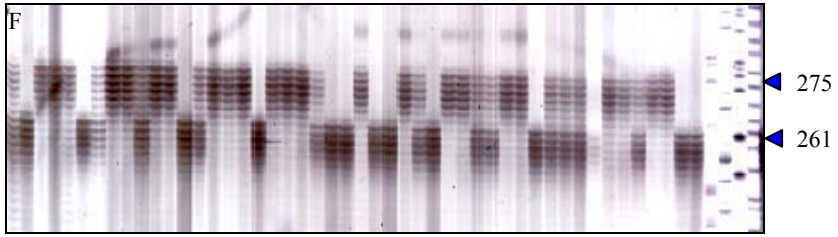
Cma-34: A-1



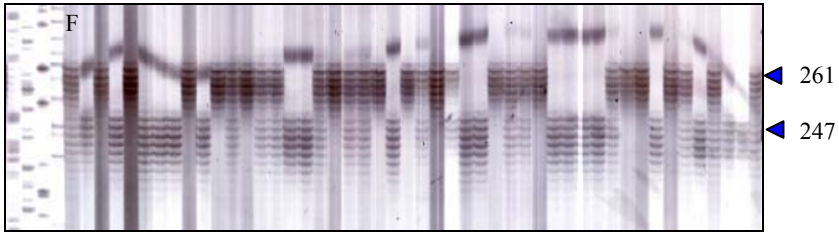
A-2



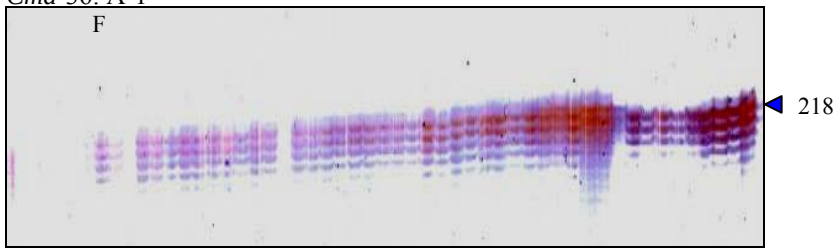
Cma-35: A-1



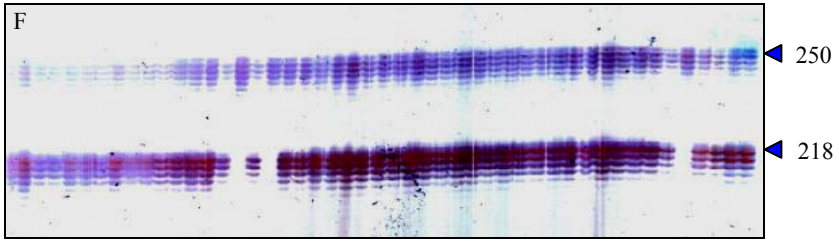
A-2



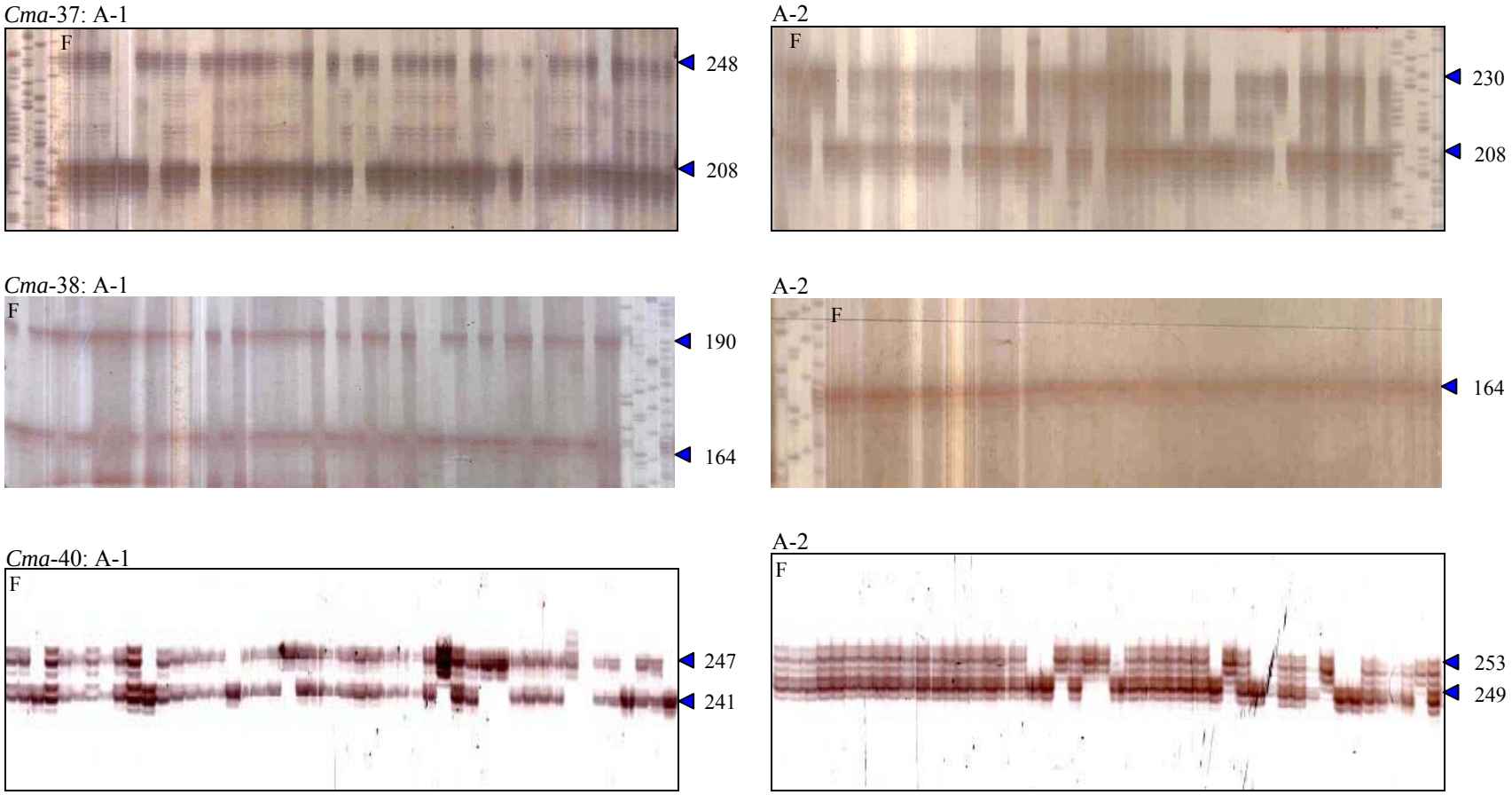
Cma-36: A-1



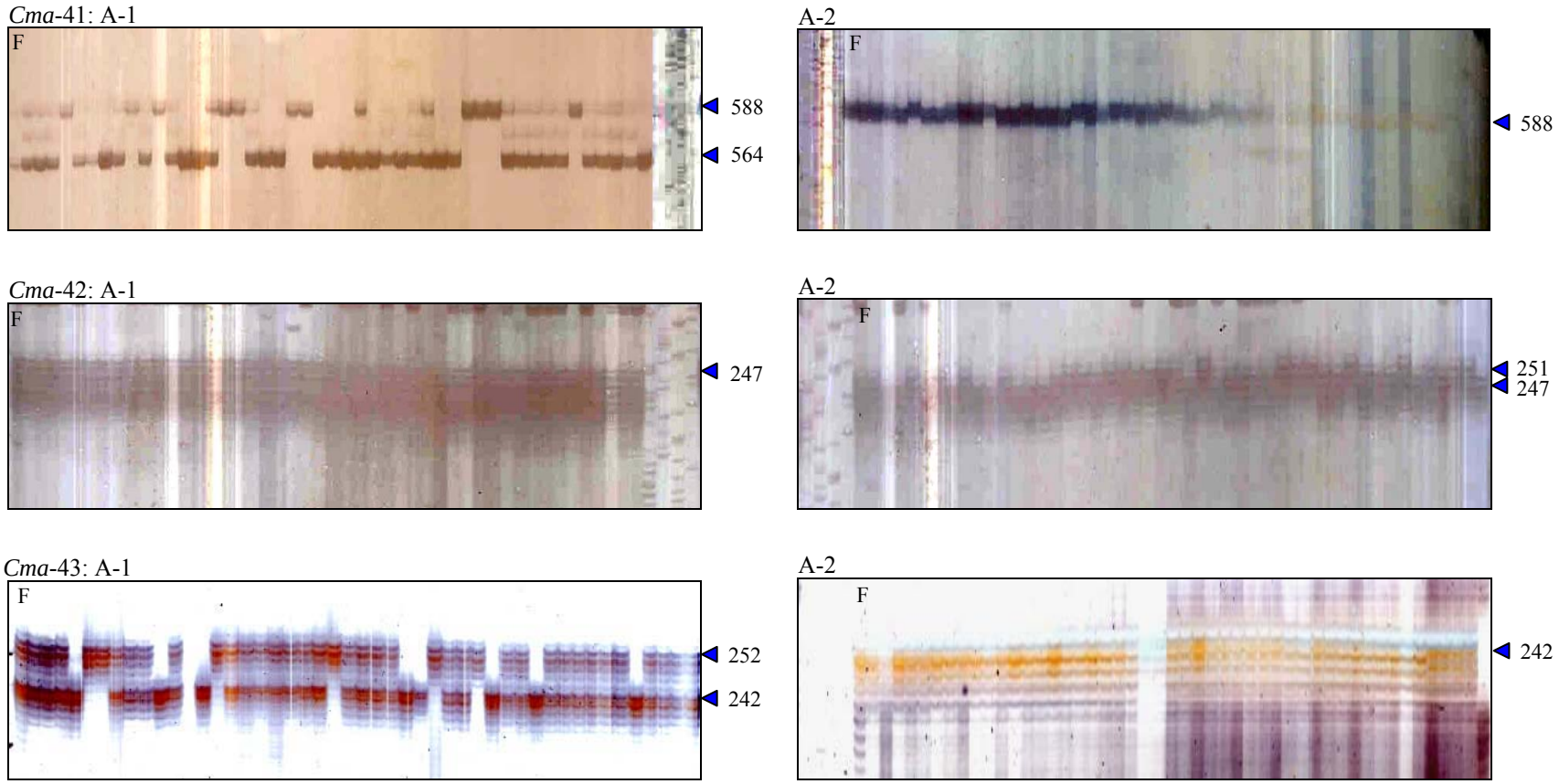
A-2



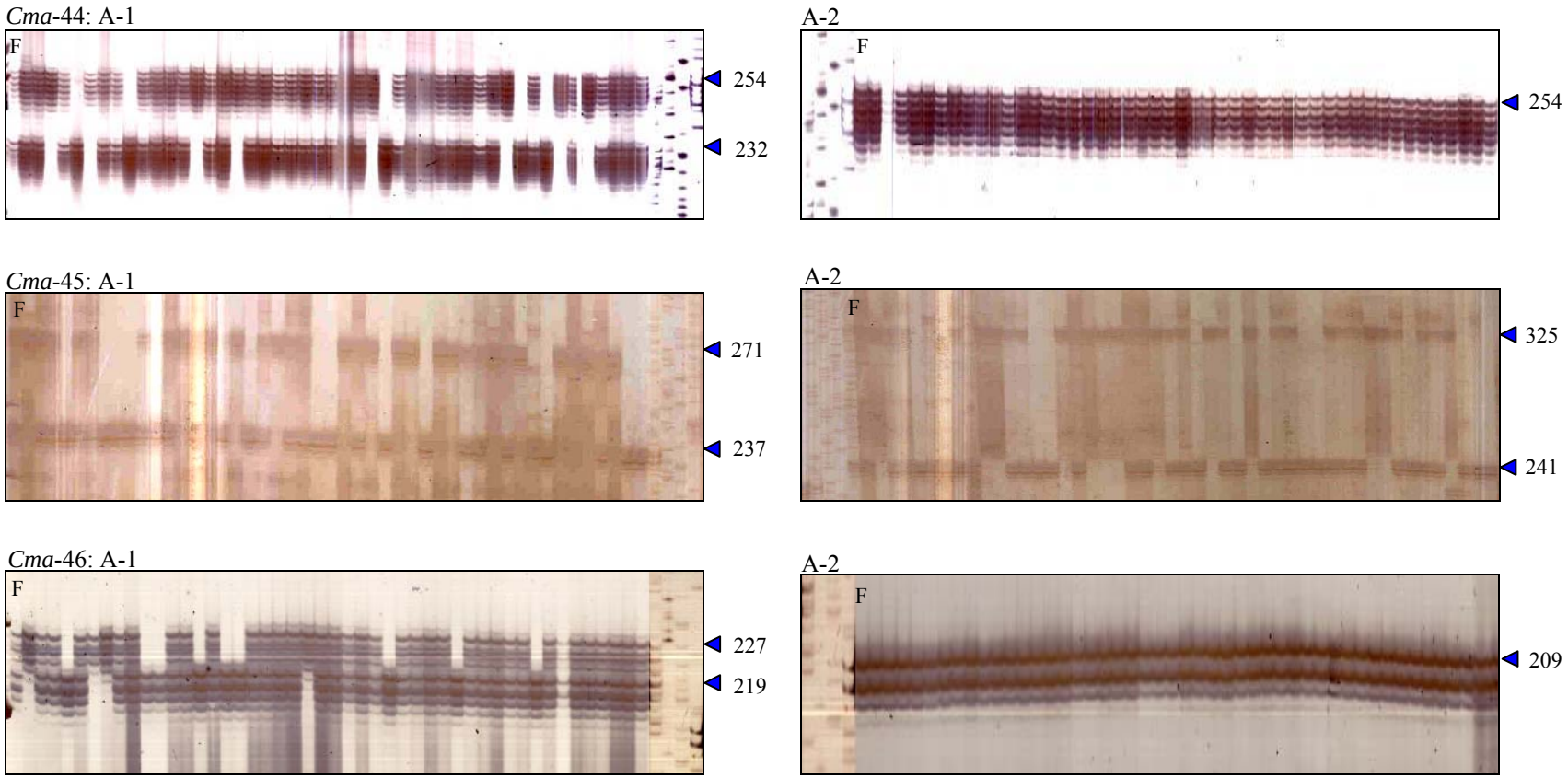
Appendix Figure 3 (Continued)



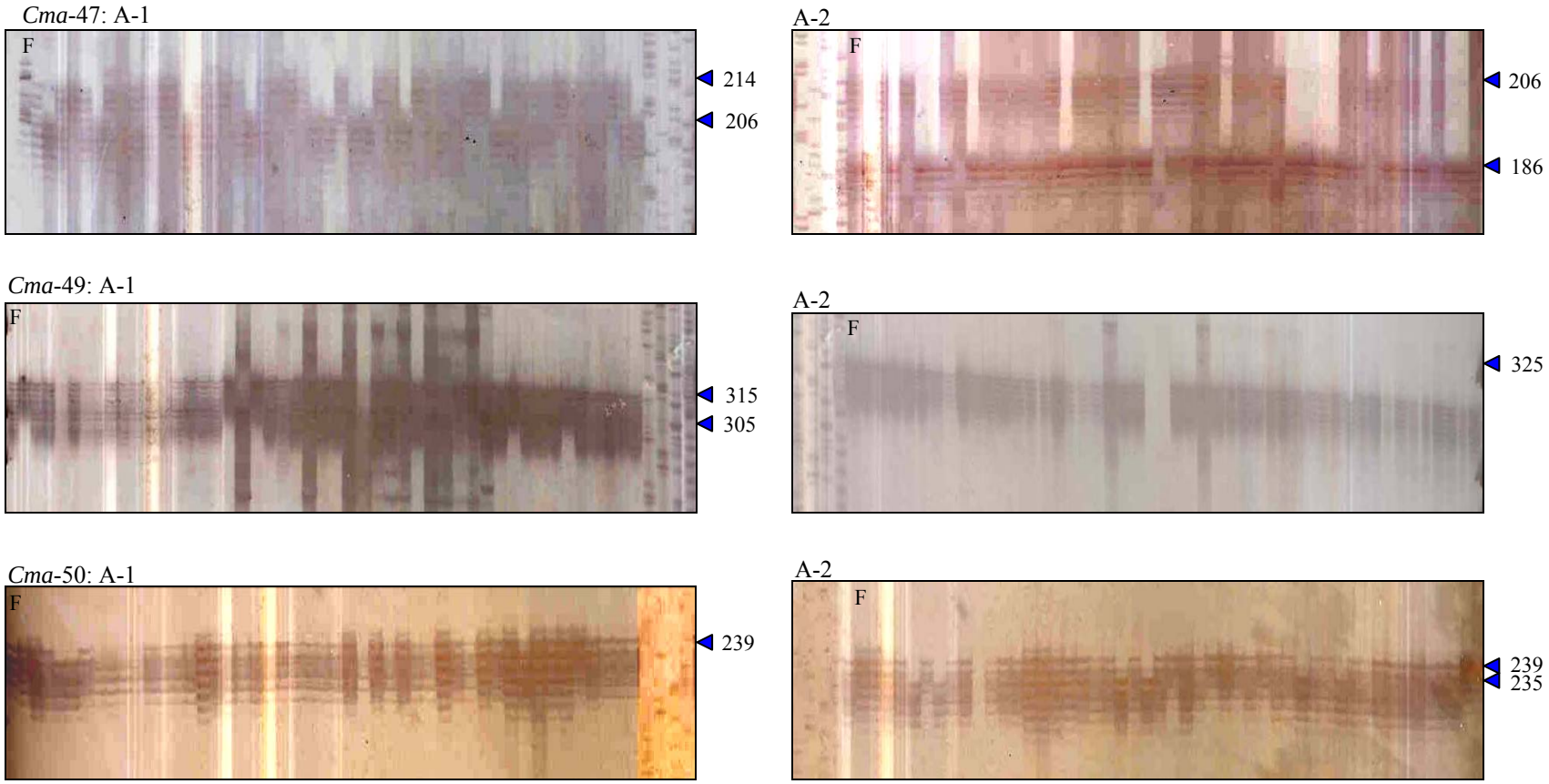
Appendix Figure 3 (Continued)



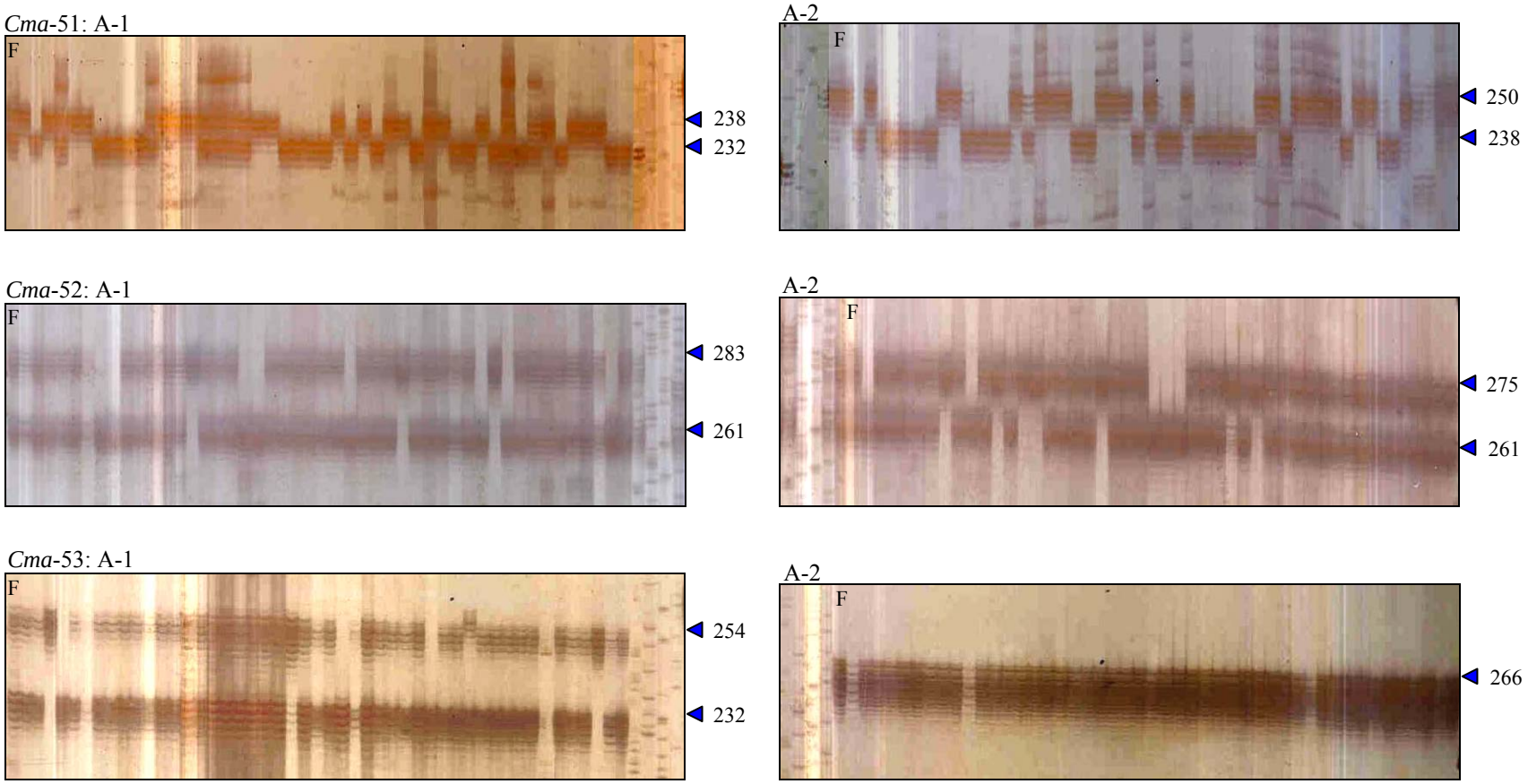
Appendix Figure 3 (Continued)



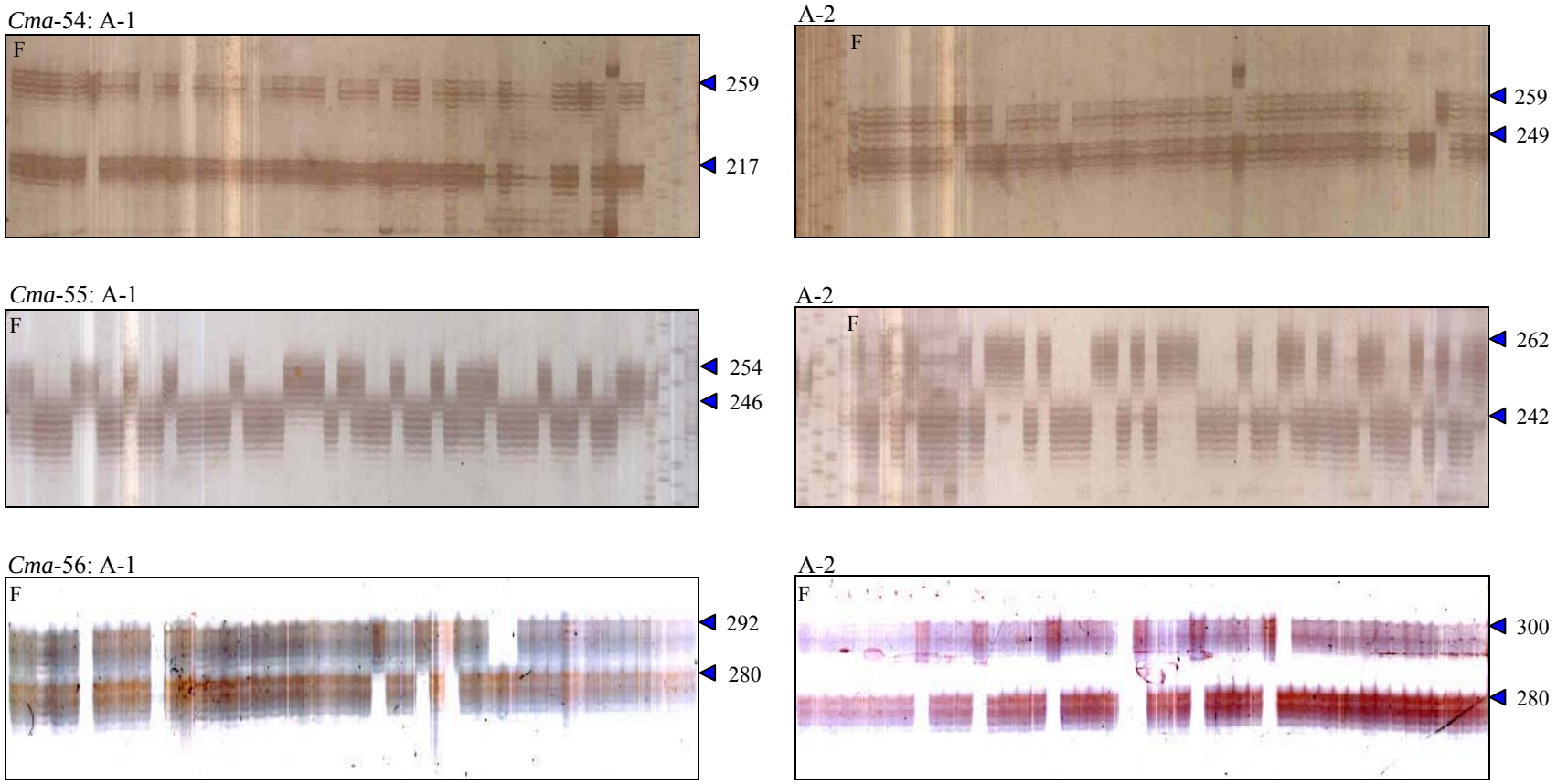
Appendix Figure 3 (Continued)



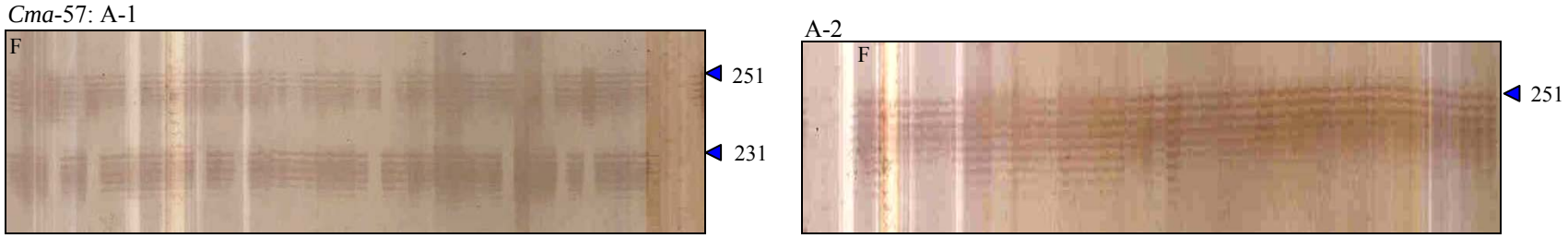
Appendix Figure 3 (Continued)



Appendix Figure 3 (Continued)



Appendix Figure 3 (Continued)



Appendix Figure 3 (Continued)

CURRICULUM VITAE

NAME : Ms. Chantapim Sukkorntong

BIRTH DATE : Jan 8, 1982

BIRTH PLACE : Pattanee, Thailand

EDUCATION :	**YEAR**	**INSTITUTE**	**DEGREE/DIPLOMA**
	2004	Kasetsart University	B. S. (Fisheries)

POSITION/TITLE : -

WORK PLACE : -

SCHOLARSHIP/AWARD : -