Molecular Marine Biology and Biotechnology (1996) 5(1), 71–83

Molecular analysis of a RAPD marker (B20) reveals two microsatellites and differential mRNA expression in *Penaeus vannamei*

Denise K. Garcia, Arun K. Dhar, and Acacia Alcivar-Warren
Department of Comparative Medicine
Tufts University School of Veterinary Medicine
200 Westboro Road
North Grafton, Massachussetts
01536, U.S.A.

Abstract

We previously reported a population-specific DNA fragment (B20) in *Penaeus vannamei* shrimp, fragment found using the randomly amplified polymorphic DNA (RAPD) procedure, that was present in Population 2 but not in Populations 1 and 4. The specific objectives of this study were to clone and sequence this genetic marker, determine if all or part of this cloned sequence could be found in any of the other populations in which this marker could not be amplified, and examine if this marker represents a functional gene by examining the steady-state levels of mRNA expression using Northern blot hybridization. Sequence information of the 1259-bp B20 clone revealed two microsatellites and two candidate open reading frames. Although the entire B20 sequence could only be amplified in Population 2 (from Ecuador), Population 3 (a hybrid of Populations 1 and 2), and a few individuals from wild Ecuadorian shrimp samples, portions of the B20 DNA could be amplified in individuals from Populations 1, 2, 3, candidate Population 4, and wild Ecuadorian samples. These microsatellites vary in size between populations and families. Northern blot hybridization analysis using radiolabeled B20 probe detected two mRNA transcripts of approximately 1.5 and 2.0 kb. Expression data throughout development indicated that these transcripts were present at low levels in nauplii from two of the three crosses examined using broodstocks of Population 1. Higher levels were observed in postlarvae (PL) 6, PL8, and PL10 in one of the three crosses. Individuals from all crosses showed higher levels of expression in the juvenile tail muscle. The mRNA transcript levels were undetected in zoea 3, PL2, and PL4 stages of development and broodstock tail muscle. The levels of expression of B20 mRNA transcripts varied significantly between Populations 1, 2, 3, 4, and wild Ecuadorian individuals as well as between families and within individuals representative of seven families from Population 1. In summary, the B20 clone revealed the presence of two microsatellites that vary in size between populations. These microsatellites will be useful for estimating genetic diversity within and between populations, identifying family-specific markers, and mapping loci responsible for economically important traits in penaeid shrimp. The mRNA levels detected by the B20 clone showed differential expression during development, and the pattern of expression was influenced by the genetic background of the parental crosses used.

*Correspondence should be sent to this author.
© 1996 Blackwell Science, Inc.

Introduction

It has been shown in many livestock species that the analyses of genomes and the identification of genetic markers will enable us to understand the genetic control of economically important traits (Lande and Thompson, 1990; Moore et al., 1992; Slettan et al., 1993). These same principles can be applied to the commercially important *Penaeus vannamei* shrimp. Genetic markers have numerous applications in the study of shrimp breeding programs including identification of populations and maternal lines and analysis of genetic diversity (Sunden and Davis, 1991; Garcia et al., 1994, and references therein). Some of these markers could be used to map loci responsible for diseases or other economically important traits. However, the usefulness of genetic linkage analyses for mapping loci responsible for quantitative traits in domestic species depends on the availability of a sufficient number of highly polymorphic marker loci that are evenly distributed throughout the genome (Fries and Ruddle, 1986).

The first step in finding genetic markers for *P. vannamei* shrimp is to establish a breeding program. The United States has established the Gulf Coast Research Laboratory (GCRL) Consortium, which manages the Marine Shrimp Farming Program (MSFP) (Dill et al., 1994). The MSFP is developing a highly sophisticated breeding program using *P. vannamei* populations known to differ in growth and disease susceptibility (Gjedrem and Fimland, 1995; Pruder et al., 1995; A.A. Alcivar-Warren, unpublished data). As a Consortium member, our objective is to monitor the levels of genetic diversity in these populations. To accomplish this, we have used various molecular genetic techniques including restriction fragment length polymorphism (RFLP) of mitochondrial genes and RAPD techniques (Alcivar-Warren et al., 1994a, 1994b; Garcia et al., 1994). During the course of our investigation we found a RAPD marker specific to Population 2 (Ecuador) (Garcia et al., 1994). This genetic marker (B20) was not present in families of Population 1 and in the candidate Population 4 (both from Mexico) (Garcia et al., 1994).

To date, little is known about the shrimp genome; therefore we decided to further characterize this genetic marker at the molecular level. Our specific objectives were to clone and sequence this genetic marker; determine if all or part of this cloned sequence could be found in any of the other populations in which this marker could not be amplified; examine if this marker represents a functional gene or genes by examining the steady-state levels of mRNA expression in different stages of development; analyze the mRNA expression of this marker in tail muscle samples obtained from individuals of Populations 1, 2, 3, candidate Population 4, and wild Ecuadorian females; and compare mRNA expression of this marker in tail muscle samples obtained from individuals of the Population 1 families.

Results and Discussion

Cloning and sequencing of the B20 RAPD marker

The B20 RAPD marker was cloned into pCR Script vector (Stratagene) and sequenced. Two different clones (B20-2 and B20-19) were sequenced following the dideoxy chain termination method. Internal primers were synthesized for both cistrons to complete the sequence. The B20-19 clone was 1259 bp and the B20-2 clone was 1261 bp long. There were three nucleotide differences between the two clones examined (Figure 1). Compared with the B20-19 clone, the B20-2 clone had two additional A's in the long stretch of A's at nucleotide position 206, and this clone also had a transition from C to T at nucleotide 1237. We used the 1259-bp B20-19 clone for further characterization as well as for polymerase chain reaction (PCR) and hybridization analysis.

Both B20-2 and B20-19 clones contained two tandemly repeat sequences (microsatellites) within their sequence. Presence of such microsatellite se-

```
   1  ggacccttac acactgcagt ataatcgcat tactaaactg taacacatta
  51  tatgtagcag tttctatgcg tcagtatctc cctaacatag cgttgcttgt
 101  agtgatccca actccatttt tcagaatcca ctaaggcgac ctaaaaaatc
 151  tgttatatta tgggtaaacc tctaaaaaat ggaagaaaaa aacattacct
 201  ttgacaaaaa aaaaaaaaaa aaaaaatcgc tggttacttt tgcgaatttc
                          xx
 251  ataatttttaa cagctagatt gatatttata ttcgctagat ttagcggaaa
 301  catcgccatg atggcaacag tggtctcgta aataagaaat attactaacc
 351  caatatcgaa tctttatttt ttttttcttc tttctttctt tttctttctt
 401  tgtaactata tttgttttta actttctttc tttctttctt tgttgctttc
 451  taacttttatt tctttctaac tttctttctt tctaactttc tttctttctt
 501  tgtttgtttc tttttcgtct tcttctgaga tcggttaaca taattcgatt
 551  ccgcaacaca ctataccttg tagaaacaga ttcgttcctg aaattcccga
 601  attcctggta tctcgaatga cgtcaagctt ataatcaata actttgatga
 651  atattaaatg agaatgtctc ctggataaca ctgattagat atcacggttt
 701  ACCTTATATA ATTCGAAGCC TTTGATACGA AATATGGTGC AACAGGACTT
 751  GAAGTTGCCT TGTACAGTAT ATACAGAAGA ACCTCATTTA TCATTATTGT
 801  tatcattatt attatcattt tttcattatc attcttaatc ttttatgttg
 851  gtgatgatga tgatgaagag aaggaggagg aagaagaaga agaagaagaa
 901  gaggaagagg aagtggaaga ggaggaggtc gaggaggatt atgtgacgat
 951  gttgcagtga taatgaagat ttatgatgat aatggtgata attaagataa
1001  TAATAAGAGT AATAAAGATA ACAGCAACGA AGTGAAACTA TCTGTATGAG
1051  tacccaagaa tcgcccttgt acaaaaaaat ttacccttct tcttttatcg
1101  atttacacag gtgcatctcg tgtacattta tatcaaacgg acattaaaat
1151  aatcagattc tgacattgca cccaggatta acagggggaaa aagtctatta
1201  ggtttatata atggatttgt cagtcatatg ggtatccttg tagtacacac
                                                t
1251  tgtaagggt
```

Figure 1. DNA sequence of the B20 population-specific marker. This marker was cloned into pCR Script and transformed into *E. coli* XL1-Blue cells before sequencing. Items in boldface indicate the differences between the two clones (B20-19 amplified from individual 2.12 and B20-2 amplified from individual EL1). The two x's that are in boldface indicate that two a's have been inserted here in the B20-2 clone. Nucleotides that are underlined are the microsatellites. Nucleotides that are in italics are the approximate open reading frames with a 95% confidence level using the program TestCode (GCG). Capitalized nucleotides are areas of indecision in which the program cannot tell with 95% confidence if it is an open reading frame or not. This sequence has been deposited in GenBank (accession number U35314).

quences indicates the nuclear origin of B20 RAPD marker. The two microsatellites observed in the B20 sequence vary in their base composition as well as length of tandem repeats. The first microsatellite contained a tetranucleotide repeat "CTTT" starting at nucleotide 361 and extending for 166 nucleotides (Figure 1). The second microsatellite contained a contiguous stretch of a trinucleotide repeat "GAA" starting at nucleotide 853 and extending for 86 bp (Figure 1). Such repetitive DNA sequences are known to constitute a substantial portion of the eukaryotic genome (Charlesworth et al., 1994). The first microsatellite "CTTT" had a tandem repeat of $T(CTTT)_1(ATTT)_1(T)_5(CTTT)_4TTCTTC(CTTT)_1CTA$ $ACTATATTT(CTTT)_1CTAA(CTTT)_5CTTC(CTTT)_1$- $CTAA(CTTT)_1(ATTT)_1CTTT_1(CTAA)_1(CTTT)_3$- $(CTAA)_1(CTTT)_7TTCGTCTTCTTCT$. The second microsatellite "GAA" had a repeat of $(GAT)_4(GAA)_1$ $GA(GAA)_1G(GAA)_1GGAG(GAA)_7GAG(GAA)_1GAG$ $(GAA)_1GTG(GAA)_1GAGGAGGAGGTCGAGGAGGA$. As observed, these microsatellites or microsatellite regions are not perfect repeats. Different genetic mechanisms, namely replication slippage, rolling circle mechanism, unequal exchange, and mutation by substitution, have been proposed to explain the anomalies in tandem array length and composition (Schlotterer and Tautz, 1992; Charlesworth et al., 1994).

The B20 sequence was aligned to sequences in GenBank using the GCG (Genetics Computer Group) analysis software version 7.3. The FastA program was used to find the sequences that are most similar using the method of Pearson and Lipman (1988). All the first 40 matches aligned to the 166-bp "CTTT" microsatellite region. These included matches to human sequence tag sites (STS) on chromosomes 1, 2, 6, 8, 9, and 14, mouse liver phosphofructokinase gene, nontranscribed spacer region of the human ribosomal DNA, retina-specific protein, rat tyrosine aminotransferase gene, mouse serum amyloid P component, human metallothionein IV gene, and human β-actin-related pseudogene. When the "CTTT" microsatellite region was removed from the B20 sequence and FastA was used to align with the sequences in GenBank, 35 of the best 40 matches were to the 86-bp "GAA" microsatellite region, while the other five matches were to regions outside the microsatellites. The two microsatellites were then taken out, and FastA was run again to look for the best-aligned sequences (Table 1). Four of the best 40 matches aligned to more than 500 bp of the B20 sequence, which may make them a more likely match (Table 1). In the remaining 36 sequences, the relatively small number of base pairs (26–396 bp) that were similar to the B20 sequence would probably not make them a likely match (Table 1).

The four maximum aligned matches included two nuclear DNA-encoded genes (ATPase-1 gene of *Plasmodium falciparum,* the human malaria parasite, and the elongation factor 1-α gene of *Licopersicon esculentum,* tomato) (Krishna et al., 1993; Shewmaker et al., 1990) and two mitochondrial DNA (mtDNA)–encoded genes (cytochrome *c* oxidase subunit I gene of *Saccharomyces douglassii,* the yeast, and the ribosomal protein *S3* gene of *Prototheca wickerhamii,* the chlorophyte algae) (Tian et al., 1993; Wolff et al., 1994) (Table 1).

The ATPase-1 gene of *P. falciparum* had only about 52.5% sequence homology to B20, unlike most of the Na^+/K^+ ATPase genes, which are highly conserved (79% amino acid sequence homology observed in the ATPase genes from *Xenopus* to mammals and at least 70% homology to the invertebrate brine shrimp, *Artemia*) (Krishna et al., 1993). In addition, the *P. falciparum* ATPase probe revealed a single transcript of 5.65–6.3 kb in Northern blot analysis, whereas the B20 probe detected two mRNA transcripts of approximately 1.5 and 2.0 kb (see below for details). This ruled out the possibility that B20 sequence may represent the family of cation ATPase-like gene. We also observed that the region of the ATPase-1 gene showing homology with B20 sequence encompasses a segment which is highly AT-rich (> 80%). Interestingly, the B20 sequence, even without the microsatellites, contained 67% AT. This probably explains why B20 sequences had homology with a rather unlikely gene such as ATPase-1.

The region of the elongation factor 1-α gene homologous to the B20 sequence includes untranslated sequences, upstream of the initiation codon ATG (Shewmaker et al., 1990). This untranslated region, like B20 sequence, was also found to be rich in AT (> 75%). The predicted mRNA size (~3.0 kb) of the elongation factor 1-α gene was much larger than the mRNA transcripts detected by the B20 probe, and therefore this gene too would be an unlikely match.

In view of these dissimilarities of B20 sequences with the two maximum aligned nuclear DNA sequences, one may speculate that the B20 sequences might represent mtDNA genes. However, the presence of repetitive DNA in the B20 sequence established its nuclear origin, indicating that RAPD polymorphism essentially reflects nuclear DNA variability rather than mtDNA polymorphism. Also.

Table 1. Sequence homology of the B20 clone (excluding the microsatellite regions) with the sequences in GenBank.[a]

Organism	Gene	Region of nucleotide sequences similar to B20	% Homology
Plasmodium falciparum	ATPase I gene	3052–3727	52.5% in 697 bp
Lycopersicon esculentum	Gene for elongation factor 1-α	484–1126	53.1% in 680 bp
Saccharomyces douglassii	Cytochrome c oxidase subunit 1 gene	2570–3170	51.3% in 628 bp
Prototheca wickerhamii	Complete mitochondrial genome	45924–46423	52.5% in 528 bp
Dictyostelium discoideum	Adenylyl cyclase germination protein	2528–2910	54.8% in 396 bp
Saccharomyces cerevisiae	70kb region of right arm of chrom.2	39582–39952	52.3% in 384 bp
Plasmodium falciparum	RNA polymerase III largest subunit	4992–5300	53.7% in 313 bp
Plasmodium falciparum	High mobility group-like protein PS16	349–639	56.6% in 302 bp
Drosophila melanogaster	18s, 5.8s, and 28s, 2s ribosomal RNA	10516–10780	55.0% in 282 bp
Dictyostelium discoideum	P8A7 gene, transmembrane protein	570–834	56.7% in 277 bp
Plasmodium falciparum	pfmdrl, multiple drug resistance gene	1898–2136	59.1% in 242 bp
Plasmodium falciparum	pfmdrl, multiple drug resistance gene	2397–2634	59.1% in 242 bp
Plasmodium falciparum	pfmdrl gene	1900–2027	59.1% in 237 bp
Caenorhabditis elegans	Cosmid K02F3	19784–20008	56.6% in 235 bp
Kluyveromyces lactis	mtDNA COl and ATPase 8	5146–5320	61.4% in 184 bp
Plasmodium falciparum	Topo II gene for topoisomerase II	271–434	57.7% in 163 bp
Saccharomyces cerevisiae	Chromosome 6 open reading frame	285–429	59.9% in 157 bp
*Podospora anserina**	Complete mitochondrial genome	17851–17985	58.5% in 142 bp
Plasmodium falciparum	pfmdr 1 gene associated sequence	1900–2027	66.7% in 129 bp
Petunia hybrida	Chalcone synthase gene	1052–1162	61.2% in 116 bp
Trypanosoma brucei brucei	maxicircle genes, 12s rRNA, 9s rRNA	15992–16106	56.1% in 114 bp
Dictyostelium discoideum	TFIID mRNA	143–253	63.2% in 114 bp
Caenorhabditis elegans	Cosmid T05G5,	35282–35379	55.4% in 112 bp
Dictyostelium discoideum	rap1 gene GTP-GDP binding protein	304–410	60.6% in 109 bp
Plasmodium falciparum	RNA polymerase II large subunit	5127–5231	62.0% in 108 bp
Plasmodium falciparum	Tat-binding protein	1492–1596	63.2% in 106 bp
Saccharomyces cerevisiae	DNA for chromosome V	2087–2192	60.0% in 105 bp
Tetrahymena thermophila	Ribosomal protein L21 gene	823–902	63.1% in 84 bp
Saccharomyces cerevisiae	car80 gene	2749–2826	60.2% in 83 bp
Plasmodium falciparum	DNA polymerase α gene	2360–2436	75.0% in 80 bp
Vaccinia virus	Complete genome	122009–122047	69.2% in 78 bp
Saccharomyces cerevisiae	DNA of chromosome, XI, right arm	1802–1872	70.3% in 74 bp
Homo sapiens	CD1 R2 gene for MHC-related antigen	2586–2647	66.2% in 68 bp
Caenorhabditis elegans	Glycerol 3-phosphate dehydrogenase	17379–17436	65.5% in 58 bp
Plasmodium falciparum	ATPase I gene	3682–3736	81.8% in 55 bp
Plasmodium falciparum	S antigen gene	3704–3756	71.2% in 52 bp
Homo sapiens	Cosmid MMDA chromosome 19q13. 3	30888–30938	76.0% in 50 bp
Caenorhabditis elegans	Cosmid ZK637	9540–9570	83.3% in 30 bp
Drosophila melanogaster	Abd-A gene	55833–55859	88.5% in 26 bp

*This sequence appeared twice in FastA due to its two different accession numbers in GenBank.

[a]Only the sequences of the first 40 matches which show greater than 50% homology are shown.

it has been shown that the mtDNA of brine shrimp *Artemia*, a closely related species to *P. vannamei* for which the entire mtDNA sequence is known, has highly compact organization with maximum economy (Valverde et al., 1994). However, there were many instances in which mtDNA or mtDNA-like sequences were found in the nuclear genome of organisms (Fukuda et al., 1985; Ossorio et al., 1991; Zullo et al., 1991; Shay and Werbin, 1992). Like the B20 sequence, many of these mtDNA-like inserted sequences were flanked by repetitive elements (Ossorio et al., 1991; Zullo et al., 1991). In fact, repetitive sequences were shown to be a favorable site for integration of foreign DNA (Kato et al., 1986).

The region of the complete mtDNA from *Prototheca wickerhamii* that aligned 52.5% to the B20

sequence included the ribosomal RNA small-subunit gene *srn* (Wolff et al., 1994). However, the transcript size (~4.5 kb) of this gene was much larger than either of the two B20 transcripts. The mtDNA of *P. wickerhamii*, like most mtDNAs, has a high AT content (74.2%) which is very close to the AT content (67%) of the B20 sequence. Again, this would probably explain the reason for similarity of the B20 sequence to this *srn* gene.

The sequences of the mitochondrial cytochrome oxidase subunit I gene *COI* from *Saccharomyces douglassi* that aligned to the B20 DNA lie within the noncoding intronic 2 region of the *COI* gene (Tian et al., 1993). This region contained 85% AT, which probably explains the similarity to the B20 sequence. Although the expected *COI* transcript for the yeast gene is approximately 1.6 kb and one of the B20 transcripts is approximately 1.5 kb, we have been able to rule out the possibility that B20 mRNA is similar to mitochondrial *COI* mRNA by Northern blot hybridization analysis (A.A. Alcivar-Warren et al., unpublished data; see mRNA expression results below). The data therefore suggested that the B20 sequence represented unidentified genes. Isolation of cDNA clones using the B20 probe and subsequent characterization of the genes would help to resolve this issue.

The GCG analysis software was also used to look for open reading frames (ORFs) with the program TestCode. TestCode works by plotting the nonrandomness at every third base at a 95% confidence level using statistics developed by Fickett (1982). Two potential ORFs were identified using the whole B20 sequence (Figure 1). One ORF starts at approximately base pair 227, extends about 135 bp, and ends at the beginning of the "CTTT" microsatellite. The other ORF starts at approximately base pair 801, extends about 200 bp, and completely encompasses the "GAA" microsatellite. When either the "CTTT" microsatellite or both the microsatellites were taken out, the ORFs were at approximately the same nucleotide positions as before. TestCode also predicted the areas of indecision in which the program cannot tell with 95% confidence if it is an ORF or not (Figure 1). All other areas are not ORFs at a 95% confidence level.

There were three polyadenylation signals (AATAAA/ATTAAA) (Kan and Moran, 1995) within this sequence. The first signal was at nucleotide 653, after the first ORF and after the first microsatellite. The other two signals were found after the second ORF at nucleotides 1011 and 1143 (see Figure 1). There were also 12 potential TATA sites and one CCAAT potential binding factor throughout the entire B20 sequence. Seven of the 12 TATA sites were located before either of the ORFs and microsatellites. One was found within the first microsatellite, one was within the first ORF, and three were after both ORFs and microsatellites. The CCAAT site was found toward the end of the first ORF, at nucleotide 350. The presence of such signals, coupled with the ability of the B20 clone to detect two mRNA transcripts, supports the view that the B20 sequence represents potential coding genes. In addition, we found many possible initiation signals (ATA and ATG) and termination signals (TAG, TAA, and TGA) in the B20 sequence; however, we were unable to identify the specific signals for transcription of the B20 genes. This information could be obtained only if further sequences upstream and downstream of the current B20 clone were available.

B20 DNA amplification

Because the B20 marker could not be amplified by RAPD in Populations 1 and 4, we sought to determine if these populations contained the B20 sequence at all using Southern blot hybridizations. However, efforts using different blotting protocols, blocking agents, and labeling procedures (Westneat et al., 1988; Alcivar et al., 1989) did not give us clear fingerprints, although the control hybridization with mtDNA *COI* probe gave the expected product sizes (data not shown). Therefore, we used six internal sequencing primers to determine if they could amplify B20 DNA from all the populations, including Populations 1 and 4 in which the B20 RAPD marker was absent. From the sequence data, the expected sizes, including the primers, were 214 bp for the first combination of primers (B20F and B202RB; no microsatellite, from position 149 to 361), 336 bp for the second set of primers (B202FB and B203R; encompassing microsatellite 1, from position 344 to 678), and 306 bp for the third set of primers (B203F and B204R; encompassing microsatellite 2, from position 659 to 965).

The PCR-amplified products were run on a 3% agarose gel and were present in all individuals examined (Figure 2). This suggests that the absence of the 1259-bp B20 marker in Populations 1 and 4 was due to a mismatch in the priming sites of the B20 primer in these two populations. The presence of the B20 marker in individuals from Population 3 was expected in that this population is a hybrid of Populations 1 and 2. We also observed that parts of the B20 sequence were amplified in wild Ecuadorian samples. This was expected, as we have pre-

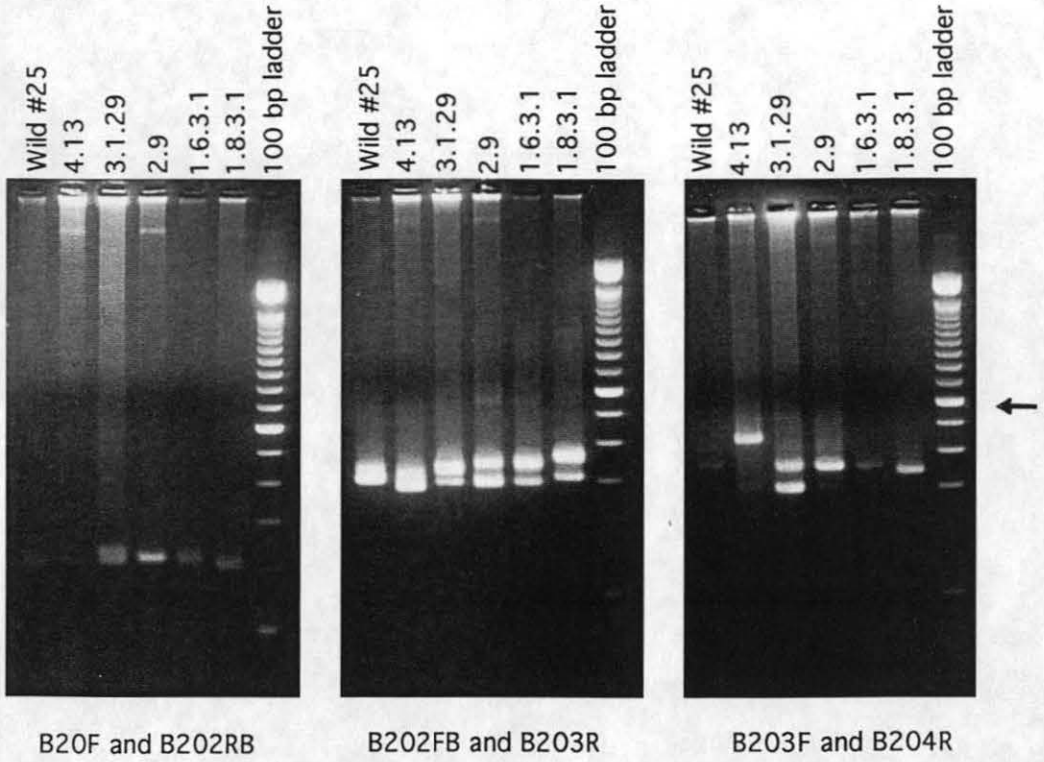

Figure 2. Different combinations of the sequencing primers were used to amplify portions of the B20 marker in individuals from all populations. These individuals were 1.8.3.1 and 1.6.3.1 from Population 1, 2.9 from Population 2, 3.1.29 from Population 3, 4.13 from Population 4, and a wild-caught individual from Ecuador, 25. The primer combinations were B20F and B202RB (214 bp expected), B202FB and B203R (336 bp expected), and B203F and B204R (306 bp expected). The forward and reverse primer sequences as well as the positions of the primers in the B20 sequence are indicated in the Experimental Procedures. The PCR cycle conditions were as follows: 93°C for one minute, 52°C for one minute 30 seconds, and 72°C for one minute, for a total of 40 cycles. Reactions were run on a 3% agarose gel. These results were repeated twice. Arrow indicates the 600-bp fragment of the molecular weight marker (Gibco, BRL).

viously found that 17% of these wild Ecuadorian individuals amplified the 1259-bp B20 marker (data not shown).

Variable-sized bands were amplified in the individuals from the four populations examined (Figure 2). This was most likely when the primers flanking the microsatellites were used and suggested the presence of different alleles. Length variability observed in microsatellite sequences arises due to unequal crossing over by sister chromatid exchange, replication slippage, or rolling circle amplification and mutation (Stephan and Cho, 1994). Computer model studies have determined that unequal crossing over is probably not the case for microsatellites because these loci are too short and are usually distributed over the entire genome (Levinson and Gutman, 1987; Stephan and Cho, 1994). Rather, the consensus is that the length variation in microsatellites is due to replication slippage (Levinson and Gutman, 1987; Schlotterer and Tautz, 1992; Stephan and Cho, 1994).

The results presented here are important for our future research in order to estimate genetic diversity within and between populations, to identify family-specific markers, and to map loci responsible for economically important traits in penaeid shrimp using the microsatellite technique.

Expression of B20 mRNA during development of P. vannamei

To determine if the B20 marker represented any functional genes, the steady-state levels of mRNA

expression were examined using Northern blot analysis. Total RNA samples from different stages of development were obtained from offspring of three combinations of crosses as detailed in the Experimental procedures. Autoradiograms of the Northern blot analysis showed two B20 mRNA transcripts of approximately 1.5 (1.5–1.7) kb and 2.0 (2.0–2.2) kb in juvenile tail muscle samples from all three crosses (Figure 3). We were not able to detect B20 mRNA transcripts in zoea 3, PL2, PL4, PL6, PL8, and PL10 samples from the L × L cross, but low levels of the larger transcript were observed in the nauplii sample obtained from this cross. For the H × L cross, high levels of the two transcripts were found in stages PL6, PL8, and PL10 samples and very low levels in the nauplii stage, but transcripts were not detected in all other stages (zoea 3, PL2, and broodstocks). Results from the H × H cross indicated that the transcripts were only found in a juvenile individual and not in zoea 3, PL2, PL4, PL6, PL8, and PL10 samples (data not shown). In general, variable amounts for each of the B20 transcripts were observed in the juvenile tail muscle samples analyzed from the three crosses, but equal levels of the two mRNA transcripts were observed in the PL samples from the H × L cross. The broodstock tail muscle samples from the L × L, H × L, and H × H crosses did not show any B20 mRNA transcripts. This was probably due to the smaller amount of total RNA present in the gel (see the ethidium bromide picture in Figure 3). A similar situation was observed for zoea 3 from the L × L cross and the PL2 individual from the H × L cross. Once again, we believe that the B20 probe is of nuclear origin because when these same Northern blots were rehybridized with a shrimp-specific *COI* mtDNA probe, different patterns of expression were observed (A.A. Alcivar-Warren et al., unpublished results).

We conclude from these results that the steady-state levels of B20 mRNA were expressed differently throughout development. This differential expression appears to be influenced by the genetic background of the parental crosses. Therefore, it appears that B20 mRNA encompasses functional genes, and the transcription, initiation, and termination signals

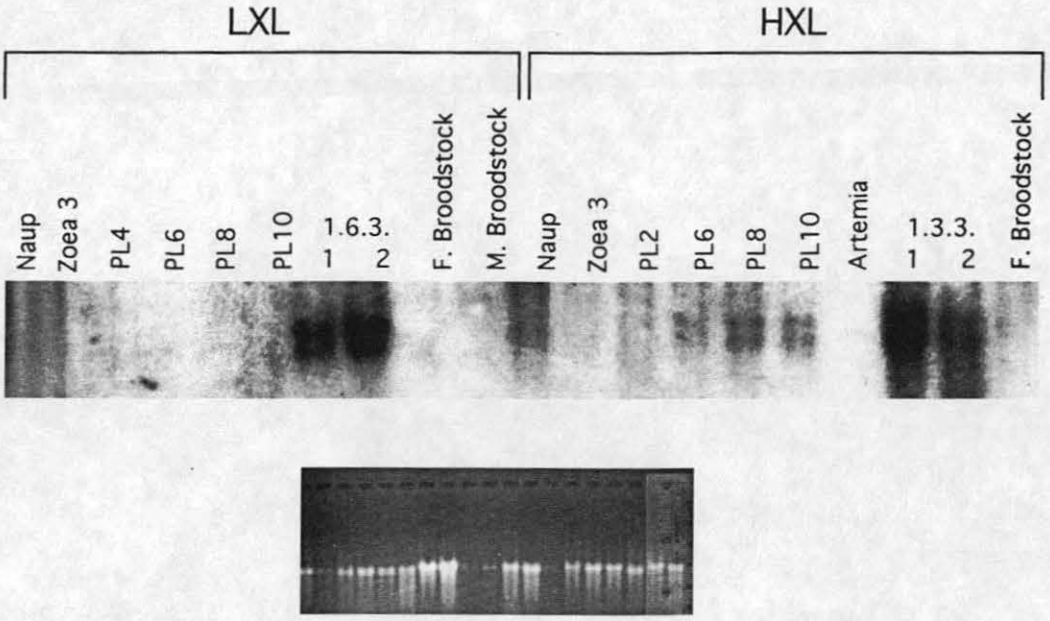

Figure 3. Northern blot analysis of B20 mRNA expression at different stages of development of *P. vannamei* shrimp. Offspring were obtained from three crosses of broodstocks from three families of Population 1. The crosses were H × H, L × L, and H × L. The broodstocks used to perform these crosses were 1.3 female × 1.8 male (H × H), 1.6 female × 1.6 male (L × L), and 1.3 female × 1.6 male (H × L). Aliquots of total RNA were loaded in a 1% agarose/formaldehyde gel, electrophoresed, and blotted onto a nitrocellulose (Nitropure, MSI, Westboro, MA) membrane. Hybridization conditions were as detailed in the Experimental Procedures. Blots were hybridized using the 1259-bp B20 insert labeled with ^{32}P and the random priming procedure. The ethidium bromide picture shows the relative amount of total RNA loaded in each lane.

must lie upstream or downstream of the sequenced B20 clone.

Differential expression of B20 mRNA in P. vannamei *populations*

B20 mRNAs were expressed at variable levels in tail muscle of individuals examined from Population 1 (Figure 4). Even though the amplified DNA fragment was only found in Population 2, we were unable to detect strong hybridization signals for these transcripts in Population 2 (Figure 4). This finding was not unexpected as polymorphisms in RAPD amplifications arise due to mutation, insertion, or deletion of one or more nucleotides in the priming site, and the presence or absence of a DNA band does not bear any correlation with the mRNA expression for those genes. Of 24 samples examined from two families of Population 1, 12 showed high levels of B20 mRNA transcripts (Figure 4). In addition, results from a different Northern blot showed that the B20 mRNAs were highly expressed in wild Ecuadorian female broodstocks, were undetectable in Population 3, and were very low in Population 4 (data not shown). The lack of expression in Population 2 (Ecuador) and 3 (hybrid of Populations 1 and 2) and yet the strong expression in the wild Ecuadorian samples could be due to the fact that Population 2 is seventh-generation inbred and Population 3 was made using males of Population 2. These results with Population 2 were not unique to B20 in that the same results were observed when an *actin* gene probe was hybridized to a Northern blot containing aliquots of these samples (Alcivar-Warren et al., 1994a).

Variability in B20 mRNA expression between and within families of Population 1

As Population 1 individuals showed higher levels of B20 mRNA expression than the other populations, we decided to evaluate B20 expression across the families of Population 1. It was interesting to observe that the steady-state levels of B20 mRNA

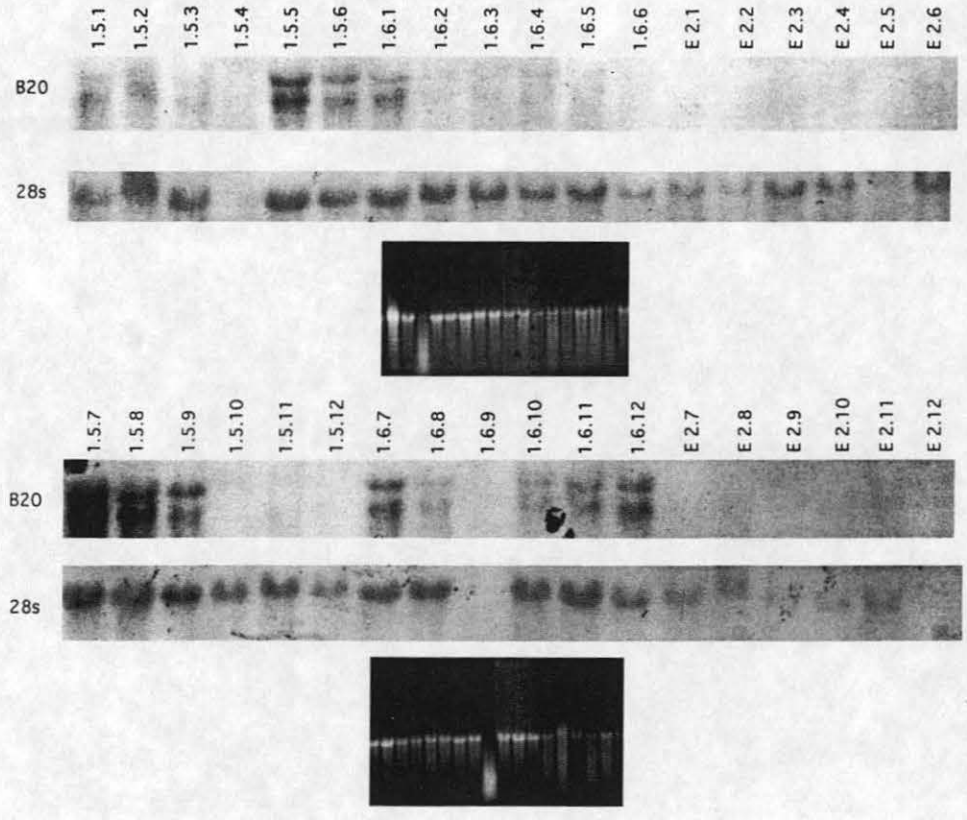

Figure 4. Autoradiogram of B20 mRNA expression in individuals from Populations 1 and 2. RNA was loaded in a 1% agarose/formaldehyde gel. Blots were first hybridized to the 1259-bp B20 insert and then to the 28S amplified fragment. The DNA was labeled with α-^{32}P-dCTP and the random priming procedure. Hybridization conditions were as detailed in the Experimental Procedures. The ethidium bromide picture shows the relative amount of total RNA loaded in each lane.

varied considerably between and within families of Population 1 (Figure 5). As shown in Figure 5, individuals from families 1.4, 1.6, and 1.11 expressed higher levels of B20 mRNAs than individuals from families 1.9 and 1.10. In addition, individuals from all these families showed high within-family variability (Figure 5 and other unpublished data). It remains to be determined if these variable levels of expression between individuals of the families of Population 1 are correlated with any economically important trait like growth performance or disease resistance and susceptibility. Because the B20 genes could not be identified precisely from the sequence, future efforts will be aimed at constructing cDNA clones that will allow identification of the genes and study of their function and regulation during growth and development.

Experimental Procedures

Isolation of the population-specific RAPD marker

To isolate and characterize the population-specific B20 RAPD marker, we cut the DNA fragment (amplified from Population 2 individual 2.12) out of a 1.2% agarose gel run in 1 × TAE (40 mM Tris, 20 mM glacial acetic acid, and 2 mM EDTA). The excised DNA band was eluted using a Spin-X column (Costar). The DNA was precipitated using one-tenth volume of 3 M sodium acetate (pH 5.2) and two volumes of ethanol and stored overnight at −20°C. The DNA was then centrifuged for 30 minutes at 4°C, washed with 70% ethanol, and dried. The pellet was resuspended in 10 μl of sterile water. Because the PCR does not always yield blunt ends for cloning efforts, the eluted DNA from the Spin-

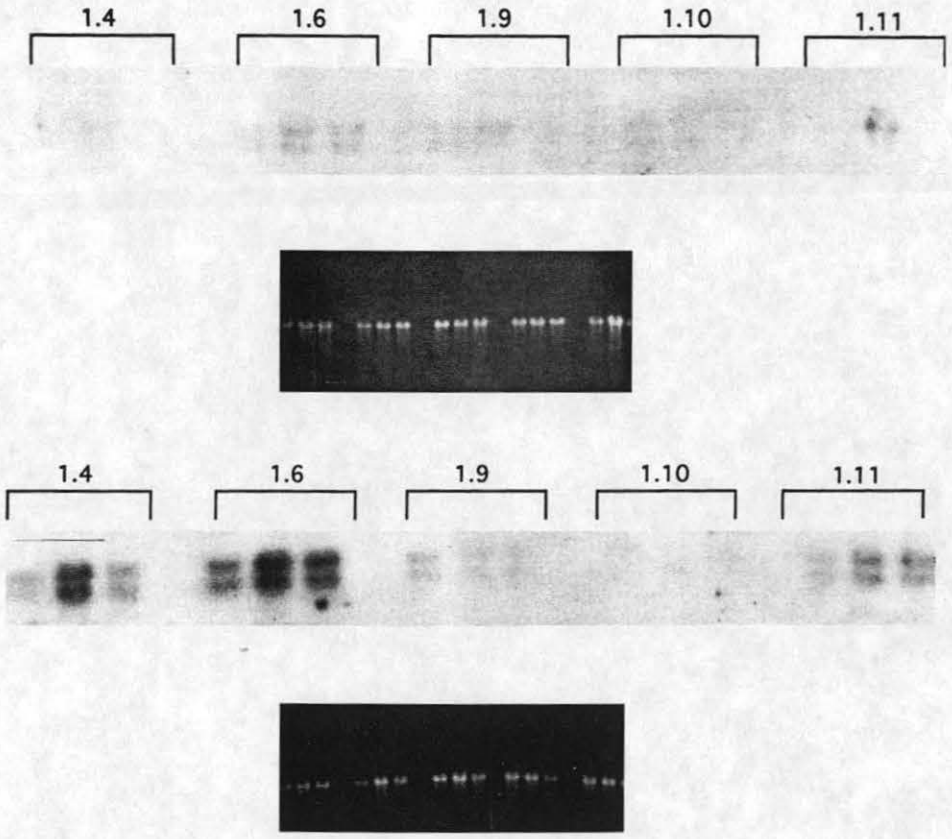

Figure 5. Northern blot analysis of B20 mRNA expression in individuals from families of Population 1. Three individuals are included within each family (1.4, 1.6, 1.9, 1.10, and 1.11) in the top and bottom panels. Individuals within these families are numbers 1–3 for the top autoradiogram and numbers 4–6 for the bottom autoradiogram. RNA was loaded in a 1% agarose/formaldehyde gel. Hybridization conditions were as detailed in the Experimental Procedures. Blots were hybridized to the 1259-bp B20 insert labeled with ^{32}P and the random priming procedure. The ethidium bromide picture shows the relative amount of total RNA loaded in each lane.

X column was treated with T4 DNA polymerase in the following reaction conditions: 50 mM Tris-Cl, pH 8.0, 5 mM $MgCl_2$, 5 mM DTT, 100 μM 4 dNTP mix, 50 μg/ml bovine serum albumin (BSA), and 10 units T4 DNA polymerase in a total reaction volume of 50 μl. The reaction mixture was incubated at 11°C for 20 minutes and then stopped by adding 2 μl of 0.5 M EDTA. The resulting blunt-ended DNA was then purified using a PCR Select II column (5 Prime–3 Prime) following the manufacturer's instructions. The 50-μl elute was subsequently dried and resuspended in 6.5 μL of water for cloning.

Cloning and sequencing the population-specific DNA fragment

The 6.5-μl PCR product was ligated into pCR Script vector and transformed into *Escherichia coli* XL1-Blue cells using the pCR Script cloning kit (Stratagene) and following the manufacturer's instructions. Recombinant clones were also grown in 2 ml of Lauria broth containing 50 μg/ml ampicillin at 37°C overnight. Plasmid DNA was extracted following alkali lysis methods (Ausubel et al., 1993) and run on a 1%, 1× TAE gel at 80 V for three to four hours with nonrecombinant plasmid DNA as a control for size. The potential recombinant clones were then digested with the restriction enzymes *Bam*HI and *Sac*I (Promega) and run on a 1%, 1× TAE gel at 80 V for three to four hours using the digested vector and the PCR-amplified B20 DNA as size controls.

The B20-19 clone, amplified from individual 2.12, was then sequenced using protocol VI of the fmol sequencing kit (Promega). The reaction conditions were as suggested by the manufacturer, and the cycle conditions were as follows: 95°C for two minutes, then 95°C for 30 seconds, 42°C for 30 seconds, and 70°C for one minute, for 30 cycles. Sequencing gels were run following standard laboratory procedures. All samples were run at least three times, and the autoradiograms were read by two different people in order to confirm the sequence. Initially, the pUC universal −20 forward primer and M13 reverse were used for sequencing. Internal primers were then synthesized for both the cistrons from the ends of the first sequences. Altogether, four internal primers were synthesized to complete the sequence of the 1259 bp in both directions. The four forward and reverse primers used were B20F 5′-tctgttatattatgggtaaacctct-3′ (starts at position 149); B202FB 5′-ctaacccaatatcgaatc-3′ (starts at position 344); B203F 5′-gagaatgtctcctggataac-3′ (starts at position 659); B204F 5′-gttgcagtgataatgaag-3′ (starts at position 948); B20R 5′-agaggtttacccataatataacaga-3′ (starts at position 173); B202RB 5′-gattcgatattgggttag-3′ (starts at position 361); B203R 5′-gttatccaggagacattctc-3′ (starts at position 678); B204R 5′-cttcattatcactgcaac-3′ (starts at position 965). The clone B20-2, amplified from individual EL1 of Population 2, was also sequenced using the above primers.

Sequence analysis was done using software GCG version 7.3 (Genetics Computer Group) and analyzed using the programs FastA and TestCode. The clone B20-19 was used for Southern and Northern hybridizations.

Animals and nucleic acid extractions

Animals from three specific-pathogen-free (SPF) populations (1, 2, and 3), one candidate SPF population (4), and high-health-status (HHS) stocks were used in this study. Lotz et al. (1995) described the terms SPF and HHS to indicate the organization and flow of shrimp in an SPF industry. SPF applies to F_1 juveniles in the founder populations and postlarvae that are rigorously screened for a standard set of pathogens with the best available diagnostic techniques. The SPF shrimp are produced in the Nucleus Breeding Center (NBC) of the U.S. Marine Shrimp Farming Program Consortium (MSFP) (Wyban et al., 1993; Carr et al., 1994). HHS is used only to describe stocks downstream from the NBC, which are less intensively monitored for the standard set of pathogens (Lotz et al., 1995).

Total nucleic acids were extracted from tail muscle of the following individuals: 48 from Population 1 (initially from Sinaloa, Mexico, third generation), 12 from Population 2 (originally from Ecuador, seventh generation), 9 from Population 3 (hybrid, Population 1 × 2), 8 from candidate Population 4 (originated from wild-caught postlarvae, Oaxaca, Mexico), and 6 wild-caught female Ecuadorian broodstocks. Detailed descriptions of how these populations were developed by the consortium have been reported elsewhere (Garcia et al., 1994, and references therein). RNA samples were also obtained from the following developmental stages: nauplii stages 3/4 (Naup), zoea stage 3 (Zoea 3), postlarvae stages 2, 4, 6, 8, and 10 (PL2–PL10), juveniles, and broodstocks from three combinations of crosses between families of Population 1. These crosses were H × H (1.3 × 1.8), L × L (1.6 × 1.6), and H × L (1.3 × 1.6), where H is high growth and L is low growth. These crosses were made on the basis of growth perfor-

mance at harvest, as reported by the MSFP (Technical Report, 1992).

We used a guanidine isothiocyanate (GT) procedure to obtain total DNA and RNA from all shrimp samples (Alcivar et al., 1989). The RNA extraction protocol was exactly as published, and modifications to the DNA extraction protocol have been reported (Garcia et al., 1994).

DNA amplifications

Six of the sequencing primers were used to amplify portions of the B20 genes from DNA in individuals from all populations. Different combinations of primers were used to amplify different regions of the B20 sequence within these individuals. These primer combinations were B20F and B202RB, B202FB and B203R, and B203F and B204R. The individuals used were 1.8.3.1 and 1.6.3.1 from Population 1, 2.9 from Population 2, 3.1.29 from Population 3, 4.13 from Population 4, and a wild-caught individual from Ecuador, 25. The reaction conditions were as follows: 1 μg of DNA was used in a 50-μl PCR reaction containing 1 × buffer (Promega), 0.2 mM dNTPs (Promega), 50 ng of each primer (Operon), 2 mM $MgCl_2$, and 5 units *Taq* DNA polymerase (Promega) overlaid with 50 μl of paraffin oil. The PCR cycle conditions were as follows: 93°C for one minute, 52°C for one minute 30 seconds, and 72°C for one minute, for a total of 40 cycles. Samples were loaded on a 3% agarose 1 × TAE gel containing 0.3 μg/ml ethidium bromide and electrophoresed at 80 V for six hours using 0.5 μg of 100-bp ladder (Gibco, BRL) for a molecular weight standard. Gels were then photographed.

Agarose gel electrophoresis, Northern and Southern blot hybridizations, and probe preparation

RNA samples were electrophoresed in a 1% agarose/formaldehyde gels and blotted onto nitrocellulose (MSI, Westboro, MA) (Alcivar et al., 1989). Northern blot hybridizations were performed with formamide-based hybridization solutions as indicated by the suppliers (5 Prime–3 Prime), and posthybridization washes were performed at 50°C using the medium-stringency washing protocol (5 Prime–3 Prime). Blots were first hybridized using the 1259-bp B20 fragment as a probe and then stripped using the moderate-stripping treatment of Ausubel et al. (1993). To confirm proper blotting and hybridization techniques, the blots were rehybridized using a PCR-amplified 28S gene. The primers used to amplify this gene were 28SF 5′-gcatatcaataagcggaggaaaag-3′ and 28SR 5′-ggtccgtgtttcaagacg-3′ (Silberman and Walsh, 1992). The reaction was as follows: 50 ng of each primer, 1 × Perkin Elmer reaction buffer, 0.2 mM dNTPs, 5 units of *Taq* DNA polymerase (Perkin Elmer), and 100 ng of DNA in a total of 100 μl. The cycle conditions were 94°C for one minute, 55°C for one minute, and 72°C for one minute, all for 40 cycles. The band was then cut out of a 1% agarose gel and purified as described above. All probes were labeled using either random priming or nick translation (Gibco, BRL) with α-^{32}P-dCTP (Amersham) and specific activities of 7×10^7 to 1×10^{10}/50 μl of probe.

Southern hybridizations were performed following two protocols, using different blocking agents and labeling probes by random priming or by nick translation (Westneat et al., 1988; Alcivar et al., 1989). Either the entire 1259-bp insert or different portions (131, 394, or 734 Dde I-digested fragments) of the B20 clone were used for probe preparations. For control hybridizations we probed the same membranes with a *P. vannamei* mtDNA COI probe (Garcia et al., 1994).

Acknowledgments

We thank Ms. Maura Faggart, Mr. Eugene Davidov, and Ms. Julie Gonsalves for extracting most of the DNA and RNA samples used in this study and Ms. Laurel Rhodes for helping with the developmental sample collection. We would also like to thank Dr. Will Carr, Mr. Jim Sweeney, Mr. Jim Swingle, and all of the technical staff at the Oceanic Institute in Makapuu Point for their able assistance with sample collection. We are thankful to Mr. David Garriques and Mr. Gorky Arevalo, from Granjas Marinas El Rosario S.A., Salinas, Ecuador, for access to the wild Ecuadorian samples. This work was supported by grant 95-38808-1424 from the U.S. Department of Agriculture to the U.S. Marine Shrimp Farming Program Gulf Coast Research Laboratory Consortium. Partial support for this project was received from NOAA National Sea Grant College Program Office, Department of Commerce, under grant NA90-AA-D-SG480, Woods Hole Oceanographic Institution Sea Grant Project No. R/A-28-PD-New Initiatives Program. The views expressed herein are those of the authors and do not necessary reflect the views of NOAA.

References

Alcivar, A.A., Hake, L.E., Millette, C.F., Trasler, J.M., and Hecht, N.B. (1989). Mitochondrial gene expression in male germ cells of the mouse. *Dev Biol* 135:263–271.

Alcivar-Warren, A.A., Garcia, D.K., Faggart, M.A., and Rich, C. (1994a). Evaluation of genetic diversity in *Penaeus vannamei* shrimp using molecular genetic techniques. *USMSFP 10th Anniversary Review.* GCRL Special Publication No. 1, pp. 27–34.

Alcivar-Warren, A.A., Faggart, M.A., Nevin, P.A., Rich, C., Wyban, J.A., Ebert, K.M., and Sweeney, J.N. (1994b). Mitochondrial DNA as a marker for genetic diversity of Penaeid shrimp: polymorphisms in the cytochrome oxidase I gene of SPF *P. vannamei* shrimp. *Proceedings of the World Aquaculture Society Meeting, Book of Abstracts.* p.28. Jan. 14–18, New Orleans, LA, U.S.A.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1993). *Current Protocols in Molecular Biology.* Waltham, MA: John Wiley & Sons, Inc.

Carr, W.H., Sweeney, J.N., and Swingle, J. (1994). The Oceanic Institute's SPF shrimp breeding program: preparations and infrastructure. *USMSFP 10th Anniversary Review,* GCRL Special Publication No. 1, pp. 47–54.

Charlesworth, B., Snegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–220.

Dill, J., McIlwain, T., Rowland, W.C., and Pruder, G. (1994). The Gulf Coast research laboratory consortium's U.S. marine shrimp farming program. *USMSFP 10th Anniversary Review.* GCRL Special Publication No. 1, pp. 7–24.

Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res* 10:5303–5318.

Fries, R., and Ruddle, F.H. (1986). Gene mapping in domestic animals. In: St. John, J. (ed.), *10th Beltsville Symposium in Agriculture Research: Biotechnology for Solving Agricultural Problems,* U.S. Department of Agriculture, Beltsville, MD, Vol. 10, pp. 19–37.

Fukuda, M., Wakasugi, S., Tsuzuki, T. Nomiyama, H., Shimada, K., and Muyata, T. (1985). Mitochondrial DNA-like sequences in the human nuclear genome. Characterization and implications in the evolution of mitochondrial DNA. *J Mol Biol* 186:257–266.

Garcia, D.K., Faggart, M.A., Rhoades, L., Wyban, J.A., Carr, W.H., Ebert, K.M., Sweeney, J.N., and Alcivar-Warren, A. (1994). Genetic diversity of cultured *Penaeus vannamei* shrimp using three molecular genetic techniques. *Mol Mar Biol Biotechnol* 3:270–280.

Gjedrem, T., and Fimland, E. (1995). Potential benefits from high health and genetically improved shrimp stocks. In: Browdy, C.L., and Hopkins, J.S. (eds.). *Swimming Through Troubled Waters.* Proceedings of the special session on shrimp farming, Aquaculture '95. Baton Rouge, LA: World Aquaculture Society, pp. 60–65.

Kan, J.L.C., and Moran, R.G. (1995). Analysis of a mouse gene encoding three steps of purine synthesis reveals use of an intronic polyadenylation signal without alternative exon usage. *J Biol Chem* 270:1823–1832.

Kato, S., Anderson, R.A., and Camerini-Otero, R.D. (1986). Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L. cell genome. *Mol Cell Biol* 6:1787–1795.

Krishna, S., Cowan, G., Meade, J.C., Wells, R.A., Stringer, J.R., and Robson, K.J. (1993). A family of cation ATPase-like molecules from *Plasmodium falciparum. J Cell Biol* 120:385–398.

Lande, R., and Thompson, R. (1990). Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.

Levinson, G., and Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203–221.

Lotz, J.M., Browdy, C.L., Carr, W.H., Frelier, P.P., and Lightner, D.V. (1995). USMSFP suggested procedures and guidelines for assuring the specific pathogen status of shrimp broodstocks and seed. In: Browdy, C.L., and Hopkins, J.S. (eds.). *Swimming Through Troubled Waters.* Proceedings of the special session on shrimp farming, Feb. 1–4, San Diego, CA. Baton Rouge, LA: World Aquaculture Society, pp. 66–75.

Moore, S.S., Barendse, W., Berger, K.T., Armitage, S.M., and Hetzel, D.J.S. (1992). Bovine and ovine DNA microsatellites from the EMBL and GENBANK databases. *Anim Genet* 23:463–467.

Ossorio, P.N., Sibley, L.D., and Boothroyd, J.C. (1991). Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii. J Mol Biol* 222:525–536.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence analysis. *Proc Natl Acad Sci USA* 85:2444–2448.

Pruder, G.D., Brown, C.L., Sweeney, J.N., and Carr, W.H. (1995). High health shrimp systems: seed supply—theory and practice. In: Browdy, C.L. and Hopkins, J.S. (eds.). *Swimming Through Troubled Waters,* Proceedings of the special session on shrimp farming, Aquaculture '95. Baton Rouge, LA: World Aquaculture Society.

Schlotterer, C., and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20:211–215.

Shay, J.W., and Werbin, H. (1992). New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutat Res* 275:227–235.

Shewmaker, C.K., Ridge, N.P., Pokalsky, A.R., Rose, R.E., and Hiatt, W.R. (1990). Nucleotide sequence of an EF-1α genomic clone from tomato. *Nucleic Acids Res* 18:4276.

Silberman, J.D., and Walsh, P.J. (1992). Species identification of spiny lobster phyllosome larvae via ribosomal DNA analysis. *Mol Mar Biol Biotechnol* 1:195–205.

Slettan, A., Olsaker, I., and Lie, Ø. (1993). Isolation and characterization of variable (GT)n repetitive sequences from Atlantic salmon, *Salmo salar* L. *Anim Genet* 24:195–197.

Stephan, W., and Cho, S. (1994). Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* 136:333–341.

Sunden, S.L.F., and Davis, S.K. (1991). Evaluation of genetic variation in a domestic population of *Penaeus vannamei* (Boone): a comparison with three natural populations. *Aquaculture* 97:131–142.

Tian, G.L., Michel, F., Macadre, C., and Lazowska, J. (1993). Sequence of the mitochondrial gene encoding subunit I of cytochrome oxidase in *Saccharomyces douglassi. Gene* 124:153–163.

Valverde, J.R., Batuecas, B., Moratilla, C., Marco, R., and Garesse, R. (1994). The complete mitochondrial DNA sequence of the crustacean *Artemia fransiscana. J Mol Evol* 39: 400–408.

Westneat, D.F., Noon, W.A., Reeve, H.K., and Aquadro, C.F. (1988). Improved hybridization conditions for DNA 'fingerprints' probed with M13. *Nucleic Acids Res* 16:4161.

Wolff, G., Plante, I., Lang, B.F., Kuck, V., and Burger, G. (1994).

Complete sequence of the mitochondrial DNA of the chlorophyte algae *Prototheca wickerhamii*. *J Mol Biol* 237: 75–86.

Wyban, J.A., Swingle, J.S., Sweeney, J.N., and Pruder, G.D. (1993). Specific pathogen-free *Penaeus vannamei*. *World Aquacult* 24:39–45.

Zullo, S., Sieu, L.C., Slightom, J.L., Hadler, H.I., and Eisenstadt, J.M. (1991). Mitochondrial D-loop sequences are integrated in the rat nuclear genome. *J Mol Biol* 221: 1223–1235.