Population structure of the giant tiger shrimp *Penaeus monodon* in Bangladesh based on variation in microsatellites and immune-related genes

M. M. Mahbub Alam\textsuperscript{a,b} and Snæbjörn Pálsson\textsuperscript{a}

\textsuperscript{a}Department of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland; \textsuperscript{b}Faculty of Fisheries, Sylhet Agricultural University, Sylhet, Bangladesh

**ABSTRACT**

*Penaeus monodon* is the most economically important penaeid shrimp in the Indo-West Pacific region for both aquaculture and wild capture. This study evaluated the population structure of *P. monodon* in Bangladesh and searched for signs of natural selection in immunological genes. Wild *P. monodon* sampled at four locations off Bangladesh were studied using 10 microsatellite markers and 14 single nucleotide polymorphic sites (SNPs), including six unannotated regions and eight immunological genes (seven C-type lectin and one HLA3), of which variations were associated with tolerance to pathogens and survival in aquaculture. Three genetically distinguishable populations were observed, one in the eastern mangrove forest in the Sundarban–Barguna coast (south Bangladesh), the second in the Sundarban–Middle ground close to the delta of the river Ganges and the third at St Martin’s Island, southeastern Bangladesh. \(G’’\) ST ranged from 0.042 to 0.093, and significant \(F_{ST}\) ranged from 0.004 to 0.011. A large proportion of individuals caught in Middle ground and at St Martin’s Island showed a genetic origin that was distinct from the other individuals at those sites, possibly from unsampled southern regions in the Bay of Bengal. Variation and the \(F_{ST}\) values for the SNPs from C-type lectin genes and HLA3 did not differ from the other markers studied, indicating no selective effects at these loci in the natural populations. Sustainable management of *P. monodon* should consider the population differentiation described in this study. Further assessment of signs of selection in aquaculture stocks could be useful for successful breeding of the species.

**Introduction**

The giant tiger shrimp, *Penaeus monodon* Fabricius, 1798, is an important penaeid shrimp for aquaculture and marine capture in the Indo-West Pacific region (Motoh 1985; Kumar et al. 2007; You et al. 2008; Azad et al. 2009; Kamal & Khan 2009; Quader 2010; Mandal et al. 2012; Waqairatu et al. 2012; Vaseeharan et al. 2013; Debnath et al. 2015; Alam et al. 2016). In Bangladesh marine waters, it is the most targeted out of 36 species of shrimps caught, due to its high economic value (Quader 2010; Department of Fisheries Bangladesh 2014). Shrimp aquaculture in Bangladesh has been gradually expanding since the late 1980s (Debnath et al. 2015) and *P. monodon* is now the dominant cultured marine species, with most of the shrimp farms located in the southwestern region (Debnath et al. 2015). The species is cultured using different methods, ranging from extensive rearing of natural postlarvae in rice fields to semi-intensive rearing of natural/hatchery-produced postlarvae in aquaculture farms (Hoq et al. 2001; Azad et al. 2009; Debnath et al. 2015). Hatcheries, mainly located in the southeastern region, use both natural gravid females and farmed *P. monodon* for postlarvae production.

Diseases, particularly white spot syndrome virus (WSSV), are highly detrimental to shrimp aquaculture production (Escobedo-Bonilla et al. 2008; Hayes et al. 2010; Stentiford et al. 2012; Debnath et al. 2015). The management of shrimp aquaculture, with its vulnerability to viral diseases and the availability of high-quality gravid females for postlarvae production from the natural stock, is important for long-term sustainability of the shrimp industry in Bangladesh (Walker & Mohan 2009; Debnath et al. 2015). Broods from the deep zone of Bangladesh marine waters have been found to have higher reproductive performance and a lower incidence of WSSV infection than the broods caught from the shallow zone (Debnath et al. 2014); thus, the natural origin of the broods may matter for the successful breeding of the species.
The marine diversity of shrimp in Bangladesh is threatened by overexploitation (e.g. postlarvae collection, the use of banned fishing gear), coastal and marine pollution, the impact of coastal aquaculture, natural disasters (e.g. cyclones, sea level rise) (Quader 2010; Department of Fisheries Bangladesh 2013) and the escape of aquaculture shrimps (Kumar et al. 2007). The Department of Fisheries (DOF) in Bangladesh imposed a ban on wild prawn larvae fishing in September 2000 to minimize the potential negative effects on diversity from high levels of bycatch of non-target species caught incidentally (Department of Fisheries Bangladesh 2002; Ahmed & Troell 2010). To overcome the threats to shrimp biodiversity and to manage wild *P. monodon* populations in a sustainable way, basic information on genetic variation and population structure is of great importance as it can provide guidance in restoration efforts and for the sustainable harvest of local populations (Willette et al. 2014). Limited molecular genetic information is available for Bangladesh *P. monodon*, except for a recent mitogenomic study which described high variation and a panmictic population along the coast (Alam et al. 2016).

Genetic improvement through related biotechnological applications is of great importance for sustainable development in aquaculture (Gjedrem et al. 2012; Mandal et al. 2012; Debnath et al. 2015), and genetic variation is essential for artificial selection programmes that aim to enhance ecologically or economically important traits and to identify disease-resistant traits for aquaculture (Moss et al. 2012; Maralit et al. 2014). Furthermore, patterns of genetic variation carry signatures of population structure, indicating areas of recruitment and the effects of natural selection at linked loci, and, thus, provide insight into the biology of the species which could be of importance for management (e.g. Carvalho & Hauser 1994).

*Penaeus monodon* has been studied for sustainable aquaculture production (Kumar et al. 2007; Mandal et al. 2012) and has been a subject of numerous studies on genetic variation to conserve its genetic diversity through management of wild populations in the Indo-West Pacific region (Benzie et al. 2002; You et al. 2008; Mandal et al. 2012; Waqairatu et al. 2012; Alam et al. 2016). Population differentiation and the identification of stocks and nonindigenous *P. monodon* have been studied by microsatellites (Pan et al. 2004). Based on 10 microsatellite loci, significant genetic structure was identified among eight *P. monodon* populations from the Bay of Bengal, south of Bangladesh, with the highest genetic diversity at Andaman Island sites (Mandal et al. 2012). High levels of polymorphism were also reported from Malaysian waters, with highly significant heterozygote deficiencies indicating probable inbreeding in the populations or, alternatively, admixture of breeding units (Aziz et al. 2011). Further study, utilizing genomic DNA such as microsatellites, is warranted to examine population structure of *P. monodon* in Bangladesh waters.

Several nuclear markers are linked to resistance to WSSV infection in penaeids, including C-type lectins, which effectively participate in the initial step of pathogen recognition, as well as in various antimicrobial effector functions (Song et al. 2010; Zeng et al. 2013; Wang et al. 2014). Eight SNPs, including a linkage to C-type lectin, showed genome-wide associations in *P. monodon* (Robinson et al. 2014), and individuals surviving more than 84 h following a WSSV infection showed higher levels of expression of C-type lectin (Jeswin et al. 2013). If WSSV is a problem in wild populations of *P. monodon*, variation in genes causing resistance to the virus may be reduced within populations and increased among populations due to natural selection (e.g. Luikart et al. 2003; Wan et al. 2004). Identification of SNPs contributing to the susceptibility to common diseases could be of importance for husbandry, stock enhancement and improved aquaculture production (Debnath et al. 2014, 2015).

The aims of the present study were: (1) to examine genetic patterns of *P. monodon* in Bangladesh waters of the Bay of Bengal to identify putative management units, and (2) to assess variation and the possible effect of natural selection in genes that have been associated with survival and resistance to microbial pathogens in aquaculture.
Materials and methods

Sample collection and DNA extraction

A total of 100 wild-origin *P. monodon* were collected and preserved in 96% ethanol from four locations along the Bangladeshi coastline (Figure 1) during the period of December 2012 to September 2013. Artisanal fishermen collected 20 subadult specimens (average weight of 40 g) from the Sundarban mangrove forest (SB). Samples of adults, with an average weight of 100 g, were collected from the Barguna coast (BC, 10 individuals), Middle ground (one of four fishing grounds) of the Bay of Bengal (MB, 40 individuals) and the coast of St Martin’s Island (SM, 30 individuals) by fish/shrimp trawlers. Total genomic DNA was extracted from the pleopod tissue using standard phenol–chloroform extraction (Maniatis et al. 1982).

Polymerase chain reaction (PCR) and genotyping

Ten polymorphic microsatellite loci (Pan et al. 2004) and 14 unlinked SNPs, six unannotated, seven from C-type lectins and one from HLA3 (Baranski et al. 2014), were genotyped (see supplementary material Tables SI and SII for details of microsatellites and SNPs, respectively). Amplification of microsatellite loci was performed in a final volume of 10 µl, which included 50 ng DNA, 0.2 mM dNTP, 0.1% Tween 20, 1× standard Taq Buffer (New England Biolabs), 0.5 mg Bovine Serum Albumin, 0.5 U Taq Polymerase and 0.34 mM each of the forward and reverse primers. The amplification protocol for all microsatellite loci included an initial denaturation at 95°C for 5 min, 35 cycles of denaturation at 95°C for 30 s, annealing at 52–56°C for 45 s and extension at 72°C for 1 min, and a final extension at 72°C for 10 min. Two microlitres of the PCR products (diluted 2×), with 0.1% GS500 (250) LIZ size standard and 100% formamide mix were run in an Applied Biosystems 3500xl Genetic Analyser. Amplification of SNP loci was done using qPCR (7500 Real Time PCR system, Applied Biosystems) in a final volume of 10 µl, which contained 100 ng DNA, 5.0 µl KASP master mix (2×) and 0.14 µl assay mix (72×) (LGIC Limited), following the manufacturer’s instructions. The amplification protocol for qPCR consisted of four stages: (I) activation at 94°C for 15 min, (II) 10 cycles including denaturation at 94°C for 20 s and annealing at 61–55°C for 1 min with a decrement of −0.6°C per cycle, (III) denaturation at 26 cycles of 94°C for 20 s and annealing at 55°C for 1 min, and (IV) post-PCR read at 30°C for 1 min. A second post-PCR read was performed by three additional cycles, at 94°C for 20 s and 57°C for 1 min. Microsatellite allele sizes at each locus were determined from electrophogram images, using the GeneMapper software (version 4.1, Applied Biosystems). SNP genotypes were determined from allelic discrimination plots of the qPCR analysis for each of the SNP loci.

Data analysis

Microsatellites were inspected by looking for signs of stuttering and large allele dropout, and the frequency of null alleles was determined using MICRO-CHECKER (Van Oosterhout et al. 2006) and GENEPOP (Rousset 2008). Deviation from Hardy–Weinberg equilibrium, summarized with the inbreeding coefficient (Fis), and genotypic disequilibrium were analysed by performing exact tests in GENEPOP (Rousset 2008), using a Markov Chain Monte Carlo (MCMC) approach with default parameters. Overall significance values across loci were obtained with Fisher’s combined probability method, implemented in GENEPOP. Basic statistics (including allelic richness, observed heterozygosity and expected heterozygosity) of the data sets (i.e. microsatellite, non-annotated SNPs, annotated SNPs and all loci) were summarized with the HIERFSTAT package (Goudet 2005) in R (R Core Team 2015). Identification of candidate loci under natural selection for all loci (microsatellites and SNPs) was done using BayeScan (Foll 2012), with the probability for the test statistic adjusted due to multiple testing using the FDR method (q-values) in the package BOA (Smith 2007) in R (R Core Team 2015).

The number of population units (K value) and the multivariate ordination of the different genotypes for the combined data set of microsatellites and SNPs were investigated using Discriminant Analysis of Principal Components (DAPC) with the ADEGENET package (Jombart et al. 2015) in R (R Core Team 2015). Variation among populations was summarized by calculating 

\[ F_{ST} \]

for microsatellites with or without null alleles and the corresponding allele size-based statistics for microsatellites (R<sub>ST</sub>), and for the combined data set of microsatellites (without null alleles) and SNPs, using GENEPOP. Sequential Bonferroni correction was applied to adjust the significance level due to multiple tests (Rice 1989). Considering that 

\[ F_{ST} \]

is dependent on heterozygosity within populations (Meirmans & Hedrick 2011), two unbiased methods of population comparisons, G<sub>ST</sub> (Meirmans & Hedrick 2011) and D<sub>ST</sub> (Jost 2008), were calculated, in addition to 

\[ F_{ST} \]

based on the microsatellite and SNP variation with the MMOD package.
Wilcoxon tests were used to compare the heterozygosities and the $F_{ST}$ between the SNPs from the C-type lectin genes, HLA3 and the unannotated SNPs. Genetic distances ($G''_{ST}$) between populations were visualized in a multidimensional scale plot (MDSP) utilizing the R package MASS (Venables & Ripley 2002) and the fit to the multidimensional scaling distances to the original distances (the stress) was estimated according to Kruskal (1964). Correlations between genetic distances for all loci ($F_{ST}$) and geographic distances were tested using the Mantel (1967) test implemented in the package APE (Paradis 2012) in R (R Core Team 2015).

Results

The average null allele frequency for the microsatellites was $0.05 \pm 0.02$ (Tables I and SIII) and was found to be significant at one locus in SB and BC, two in MG and three in SM, ranging from 0.08 to 0.23 (Table SIII). No signs of stuttering and large allele dropout were observed in the microsatellites. Four pairwise comparisons of the allelic variation at different loci showed indication of linkage disequilibrium ($P = 0.004–0.027$), but they were not significantly different from zero, when correction for multiple tests was applied (Rice 1989). On average, a deviation from a random association of alleles within populations was observed for all populations, except for BC, with an overall $F_{IS} = 0.07 \pm 0.13$ (Table I; see Table SIII for deviation at each locus within each population). The inbreeding coefficient was nominally larger for the microsatellites ($F_{IS} = 0.11 \pm 0.07$), with seven of 10 loci showing deficiency of heterozygotes, than for both the unannotated SNPs ($F_{IS} = 0.08 \pm 0.17$) and the immune-related SNPs ($F_{IS} = 0.00 \pm 0.13$) (Tables I and SIII).

Genetic diversity

The average overall observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$) for all samples, based on both microsatellites and SNPs, were: $H_O = 0.57 \pm 0.24$ and $H_E = 0.62 \pm 0.28$, and were higher for the microsatellites ($H_O = 0.83 \pm 0.05$ and $H_E = 0.93 \pm 0.04$) than for both the unannotated SNPs ($H_O = 0.39 \pm 0.10$ and $H_E = 0.40 \pm 0.10$) (Table I). Variation, summarized with the different statistics, varied among locations (Tables I, SIII and SIV). The average overall allelic richness ($A_R$) for both microsatellites and SNPs was similar among the sampling locations, microsatellites ranging from 10.6 in BC to 10.9 in SM, unannotated SNPs ranging from 1.90 in SB to 2.00 in BC, and immune-related SNPs ranging from 1.94 in MG to 2.00 in BC (Table I). For the microsatellites, the number of private alleles and the average number of alleles per locus were smallest in BC (8 and 11.6, respectively) and largest in MG (37 and 22.4) (Table I; see Tables SI and SIV for allele sizes and frequency of private alleles, respectively). Heterozygosity and the $F_{ST}$ of the SNPs from those of the C-type lectin genes and the HLA3 did not differ from those of the unannotated SNPs (Wilcoxon test, $P > 0.05$).

Table I. Average of summary statistics for four populations of *Penaeus monodon* from Bangladesh based on 10 microsatellite, six unannotated, seven C-type lectin and one HLA3 loci.

| Locus name | Parameters | SB       | BC       | MG       | SM       | Overall    |
|------------|------------|----------|----------|----------|----------|------------|
| Ten microsatellite loci | NullA$_G$-$GP$ | 0.066    | 0.028    | 0.049    | 0.060    | 0.051 ± 0.017 |
|            | $A_R$      | 10.859   | 10.572   | 10.714   | 10.931   | 10.769 ± 1.869 |
|            | $H_O$      | 0.813    | 0.856    | 0.828    | 0.815    | 0.828 ± 0.053 |
|            | $H_E$      | 0.936    | 0.938    | 0.923    | 0.932    | 0.932 ± 0.043 |
|            | $F_{IS}$   | 0.132*   | 0.089ms  | 0.098*   | 0.123*   | 0.111* ± 0.066 |
|            | Ap         | 15       | 8        | 37       | 24       | –          |
|            | An         | 17.100   | 11.600   | 22.400   | 20.900   | –          |
| Six SNPs, unannotated | $A_R$      | 1.900    | 1.999    | 1.978    | 1.980    | 1.964 ± 0.085 |
|            | $H_O$      | 0.417    | 0.367    | 0.412    | 0.373    | 0.392 ± 0.141 |
|            | $H_E$      | 0.396    | 0.428    | 0.420    | 0.425    | 0.417 ± 0.121 |
|            | $F_{IS}$   | −0.052   | 0.137    | 0.019    | 0.142    | 0.076 ± 0.171 |
| Eight SNPs; C-type lectin and HLA3 | $A_R$      | 1.982    | 1.996    | 1.940    | 1.985    | 1.976 ± 0.059 |
|            | $H_O$      | 0.407    | 0.443    | 0.327    | 0.385    | 0.390 ± 0.096 |
|            | $H_E$      | 0.385    | 0.427    | 0.392    | 0.376    | 0.395 ± 0.100 |
|            | $F_{IS}$   | −0.076   | −0.042   | 0.154    | −0.031   | 0.000 ± 0.131 |
|            | $H_O$      | 0.578    | 0.596    | 0.557    | 0.561    | 0.573 ± 0.238 |
|            | $H_E$      | 0.617    | 0.640    | 0.620    | 0.620    | 0.624 ± 0.279 |
|            | $F_{IS}$   | 0.017*   | 0.057ns  | 0.097*   | 0.076*   | 0.065* ± 0.125 |

Detailed statistics for each marker are given in Table SIII. $A_R$ is allelic richness for microsatellite with the expected number considering the smallest sample size of 10 individuals. Letters for sampling locations correspond to Figure 1. NullA$_G$-$GP$, null allele frequency estimated in GENEPOP; $A_R$, allelic richness; $H_O$, observed heterozygosity; $H_E$, expected heterozygosity; $F_{IS}$, inbreeding coefficient; Ap, number of private allele only found in the respective populations; An, mean number of alleles per locus; ns, not significant.

*Significant at $P < 0.001$. 
**Detecting markers under selection**

Variation at all markers, except for the microsatellite PM4868, was in compliance with the neutral expectation obtained with BayeScan, q values (adjusted \( P \)) ranging from 0.098 to 0.809 (Table SV). The alpha value for PM4868 was negative (−1.520; 95% CI: −2.926−0.307) with q = 0.016, indicating purifying or balancing selection. Since the microsatellite marker (PM4868) seems to be affected by selection that could, thus, affect the estimation of population divergence, it was omitted from the downstream analyses. The average \( F_{ST} \) value for all markers, excluding PM4868, was 0.012 (sd = 0.002) and was considerably higher than for PM4868 (\( F_{ST} = 0.003 \)) (Table SV).

**Clustering of individuals and population differentiation**

The DAPC analysis without prior information determined five clusters (\( K = 5 \), stat = 199.63; Figure 2a), two with individuals solely from the southerly locations at MG and SM and three with individuals from all locations, but in different frequencies (Figure 2c). The DAPC analyses with prior information of the sampling sites revealed separation among the four locations with the largest overlap between SB and MG (Figure 2b; see Figure 2d for the proportion of individuals in the clusters, representing four sampling sites, and Figure 3 for the posterior probabilities of samples from four locations). All population comparisons between sampling locations, based on nine microsatellite loci and all loci (nine microsatellites and 14 SNPs), were significant, except for the comparisons of SB with BC and MG. For microsatellite loci, significant \( F_{ST} \) ranged from 0.002 to 0.010, and for all loci slightly larger differentiations were obtained (significant \( F_{ST} \) ranged from 0.004 to 0.011) (Table II). By adjusting the allele frequencies with the null alleles, a slight reduction in \( F_{ST} \) was observed for comparisons of SM with BC and SB, which were lowered by 10% but still remained significant (\( P < 0.01 \)). The \( F_{ST} \) between SM and MG was lowered to 0.0004, but remained significant (\( P < 0.05 \)). Analysis of the microsatellites with \( F_{ST} \) and \( R_{ST} \) gave similar results; little or no differentiation was observed between SB and BC or MG, but the values from the other comparisons were larger and more pronounced for the \( R_{ST} \), especially between BC and MG (Table II). Where \( R_{ST} \) was larger than zero, values were considerably larger than the \( F_{ST} \), and the ratio of \( R_{ST}/F_{ST} \) ranged from 2.7 to 15.4, indicating that allele sizes were more similar within populations than among populations. The interpretation of \( R_{ST} \) should be made with caution as it is dependent on the mutational model. Similar patterns are observed by adding the SNP data to the analyses. By considering

---

**Figure 2.** Discriminant analysis of principal components (DAPC) of Bangladesh *Penaeus monodon*: (a) without prior information, (b) with prior information, (c) proportion of individuals in a (clusters: 1, 2, 3, 4 and 5) sampled from four locations and (d) proportion of individuals from four locations in b (SB, BC, MG and SM, see Figure 1). Shadings (from black to white) represent the four locations SB, BC, MG and SM, respectively.
the dependency of $F_{ST}$ on heterozygosity within a population, the $G_{ST}$ and $D_{ST}$ statistics showed larger differentiation, as expected (Table II). Similar patterns were observed with the pairwise $G_{ST}$ and $D_{ST}$ values for all loci, as revealed by high correlation ($r = 0.998$) between them, but the former was about two times larger. $G_{ST}$ and $D_{ST}$ ranged from 0.042 to 0.093 and from 0.022 to 0.053, respectively (Table II), and the largest differentiation was observed between BC and SM ($G_{ST} = 0.093$ and $D_{ST} = 0.053$; Figure 4), followed by the differentiation between SB and SM, and between BC and MG (Table II; Figure 4). Less differentiation was observed between MG and SM, SB and BC, and SB and MG ($G_{ST}$ and $D_{ST}$ ranged from 0.042 to 0.055 and from 0.022 to 0.030, respectively) (Table II). The fit of the distances in the multidimensional scaling was excellent, with a low stress value of 0.0002. Based on the Mantel test, differentiation between populations ($F_{ST}$) was not correlated with the geographic distances separating the sample sites ($r = -0.13$, $P = 0.83$).

Table II. Coefficients of population differentiation among four sampling locations along the Bangladesh coastline, based on nine microsatellites, six unannotated, seven C-type lectin and one HLA3 loci.

| Pop pairs | Microsatellites | Microsatellites and SNPs |
|-----------|-----------------|--------------------------|
|           | $F_{ST}$ | $R_{ST}$ | $F_{ST}$ | $G_{ST}$ | $D_{ST}$ | $F_{ST}$ | $G_{ST}$ | $D_{ST}$ |
| SB–BC     | 0.000   | −0.008  | 0.015ns  | 0.051   | 0.029   |            |           |           |
| SB–MG     | 0.000   | −0.004  | −0.004ns | 0.005   | 0.030   |            |           |           |
| SB–SM     | 0.005***| 0.014   | 0.010*** | 0.092   | 0.051   | 0.007***  | 0.081     | 0.046     |
| BC–MG     | 0.007***| 0.082   | 0.017*** | 0.093   | 0.053   | 0.010**   | 0.093     | 0.053     |
| BC–SM     | 0.010** | 0.032   | 0.009*   | 0.093   | 0.053   | 0.002*    | 0.042     | 0.022     |
| MG–SM     | 0.002*  | 0.023   | 0.004**  | 0.042   | 0.022   |            |           |           |

Table II. Coefficients of population differentiation among four sampling locations along the Bangladesh coastline, based on nine microsatellites, six unannotated, seven C-type lectin and one HLA3 loci.

Letters for sampling locations correspond to Figure 1. ns, not significant after sequential Bonferroni correction.

*Significant at 0.01 < $P < 0.05$.
**Significant at 0.001 < $P < 0.01$.
***Significant at $P < 0.001$.

Figure 3. Assignment plot: posterior probabilities of *Penaeus monodon* samples representing four sampling locations in Bangladesh (SB, BC, MG and SM). Letters for sampling locations correspond to Figure 1.

Figure 4. Multidimensional scale plot of four *Penaeus monodon* populations along Bangladesh coastline, based on pairwise $G_{ST}$. Letters for sampling locations correspond to Figure 1.

Discussion

Microsatellites and SNP variation, based on both DAPC and population differentiations ($G_{ST}$ and $D_{ST}$), revealed three distinct populations in *Peneaus monodon* sampled from four locations along the Bangladesh coastline of the Bay of Bengal: one in the east in the mangrove forest of the Sundarban–Barguna coast (south Bangladesh), the second at the Sundarban–Middle ground in the Bay of Bengal, and the third at St Martin’s Island in southeastern Bangladesh. The population at St Martin’s Island in south Bangladesh was differentiated from the other samples, and the population sampled at Middle ground differed from the one at the Barguna coast. The population samples from both Middle ground and the Barguna coast had a certain level of similarity with Sundarban samples, indicating an ongoing gene flow among these three locations, which may have resulted via the large mangrove forest, Sundarban, which is a major nursery ground for shrimps (Hoq 2007; Ghosh et al. 2015). As most of the aquaculture farms are located in the southwestern region of Bangladesh, the possible escape from aquaculture might have reduced the differentiation among Sundarban, Barguna coast and Middle ground. A large proportion of individuals caught in Middle ground and St Martin’s Island shared a genetic origin that was distinct from the other individuals, indicating a possible influx from unsampled southern regions. Previous studies of microsatellite patterns in *P. monodon* from the Bay of Bengal (Mandal et al. 2012) and Malaysia (Aziz et al. 2011) revealed a genetic structure where divergence reflected geographic distance. Distinct genetic differentiation was also reported
among *P. monodon* populations from the Gulf of Thailand (Tassanakajon et al. 1998) and Australia (Li et al. 2007). Here, we did not detect an association with distance, but neighbouring populations sampled only 95 km apart were found to be differentiated, pointing to a limited connection between the regions. Genetic variation estimated using microsatellites was high, and characterized by heterozygote deficiency and a large number of private alleles. Similar genetic variability was reported in *P. monodon* sampled from India (Mandal et al. 2012), Malaysia (Aziz et al. 2011) and the Indo-West Pacific region (You et al. 2008). In general, most marine decapods, including penaeids, have shown high microsatellite diversity (Benzie 2000), reflecting large population sizes. High mitogenomic variation was also reported in Bangladesh *P. monodon*, where haplotype diversity of 520 bp of the control region was close to one (Alam et al. 2016). The average overall observed heterozygosity based on microsatellites and SNPs was slightly lower than the expected heterozygosity, which indicated a deficiency of heterozygotes. This pattern has been reported in previous genetic studies on *P. monodon* samples from India (Mandal et al. 2012) and the Indo-Pacific region (You et al. 2008). Seven of 10 microsatellite markers showed significant deviation from HWE with positive F<sub>S</sub> values, indicating a lack of heterozygotes. One probable reason for the deficiency could be mutations in the primer binding sites that might result in null alleles (e.g. García et al. 1995; Aziz et al. 2011). Alternatively, such deviations could result from the sampling design or admixture of differentiated populations, as suggested for *P. monodon* in India (Mandal et al. 2012), due to different cohorts (Christiansen 1988), even breeding at different times of the year. Null allele frequencies below 0.20 have a limited effect on general population structure analyses and analyses of F<sub>ST</sub> with and without markers which showed a frequency of null alleles in the range of 0.1–0.2 did not affect the results (Chapuis & Estoup 2007). It is unlikely that the deficiency of heterozygotes is due to inbreeding, as the high variation in both marker types presented here and the high mitogenomic variation in *P. monodon* in Bangladesh (Alam et al. 2016) suggested a large population size.

Variation at the SNPs from the unannotated regions, and the C-type lectin genes and HLA3 did not differ, indicating no signs of selective effects at SNPs, e.g. due to pathogens at these loci in the natural populations. A single microsatellite marker with an unknown genetic location was found to be affected by natural selection and was excluded from further analysis. Lower pathogenic infection was observed in gravid shrimps from the deeper zone of Bangladesh marine waters (Debnath et al. 2014) and the harmful pathogenic effects may be restricted to aquaculture stocks. A further study should be done to assess the putative impact of selection in these genes, both by analysing larger numbers of SNPs and by assessing variation in aquaculture stocks. This could be valuable for aquaculture by allowing breeding of disease resistant individuals.

To conclude, three genetically distinguishable populations with high genetic diversity were observed along the Bangladesh coastline as described above, which can be considered as separate management units. The Sundarban mangrove forest region can be considered as an important nursery ground and could be managed as a separate unit, as postlarvae from both the Barguna coast and Middle ground of the Bay of Bengal share this nursery ground. The southernmost population (St Martin’s Island) should be considered as a separate unit for management. The study did not identify any signs of selective effects on wild *P. monodon* populations that could have resulted, for example, from severe viral diseases such as WSSV. This could suggest that the negative effects of pathogens are mainly restricted to conditions in aquaculture or that more extensive sampling is needed to detect the signs of selection.

Acknowledgements

This study was funded by the United Nations University–Fisheries Training Programme (UNU-FTP), Iceland. The first author was supported under a doctoral scholarship by the UNU-FTP, Iceland. The authors are grateful to UNU-FTP, Iceland, for its support and the Department of Fisheries, Bangladesh, for its assistance during shrimp sample collection. Thanks to Kristen Marie Westfall for her assistance with reading the microsatellites and comments on the manuscript. The authors are grateful to reviewers and editors for their valuable comments which improved the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ahmed N, Troell M. 2010. Fishing for prawn larvae in Bangladesh: an important coastal livelihood causing negative effects on the environment. Ambio 39:20–29. doi:10.1007/s13280-009-0002-y

Alam MMM, Westfall KM, Pálsson S. 2016. Mitogenomic variation of Bangladesh *Penaeus monodon* (Decapoda, Penaeidae) and reassessment of its phylogeography in
Mandal A, Rao D, Karuppaiah D, Gopalakrishnan A, Pozhoth J, Chellamma Y, et al. 2012. Population genetic structure of P. monodon, in relation to monsoon current patterns in Southwest, East and Andaman coastal waters of India. Gene 491:149–57. doi:10.1016/j.gene.2011.10.002

Maniatis T, Fritsch EF, Sambrook J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 545 pages.

Man tel NA. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–20.

Maralit BA, Ventolero MFH, Maningas MBB, Amar EC, Santos MD. 2014. Subtracted transcriptome profile of tiger shrimp (P. monodon) that survived WSSV challenge. Dataset Papers in Science 2014:807806. 11 pages. doi:10.1155/2014/807806

Meirmans PG, Hedrick PW. 2011. Assessing population structure: FST and related measures. Molecular Ecology Resources 11:5–18. doi:10.1111/j.1755-0998.2010.02927.x

Moss SM, Moss DR, Arce SM, Lightner DV, Lotz JM. 2012. The role of selective breeding and biosecurity in the prevention of disease in penaeid shrimp aquaculture. Journal of Invertebrate Pathology 110:247–50. doi:10.1016/j.jip.2012.01.013

Motoh H. 1985. Biology and ecology of P. monodon. In: Taki Y, Primavera JH, Llobrera JA, editors. Proceedings of the First International Conference on the Culture of Panaeid Prawns/Shrimps. Iloilo City, Philippines, p 27–36.

Pan Y, Chou H, You E, Yu H. 2004. Isolation and characterization of 23 polymorphic microsatellite markers for diversity and stock analysis in tiger shrimp (P. monodon). Molecular Ecology Notes 4:345–47. doi:10.1111/j.1471-8286.2004.00692.x

Paradis E. 2012. Analysis of Phylogenetics and Evolution with R. New York: Springer. 386 pages.

Quader O. 2010. Coastal and marine biodiversity of Bangladesh (Bay of Bengal). In: Proceedings of International Conference on Environmental Aspects of Bangladesh (ICEAB10), Fukuoka, Japan, p 83–86.

R Core Team. 2015. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Rice WR. 1989. Analyzing tables of statistical tests. Evolution 43:223–25. doi:10.2307/2409177

Robinson NA, Gopikrishna G, Baransi K, Katneni VK, Shekhar MS, Shanmugakarthik J, et al. 2014. QTL for white spot syndrome virus resistance and the sex-determining locus in the Indian black tiger shrimp (P. monodon). BMC Genomics 15(1):731. 21 pages. doi:10.1186/1471-2164-15-731

Rousset F. 2008. GENEPOP’007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8:103–06. doi:10.1111/j.1471-8286.2007.01931.x

Smith BJ. 2007. BOA: an R package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software 21(11):1–37.

Song X, Zhang H, Zhao J, Wang L, Qiu L, Mu C, et al. 2010. An immune responsive multidomain galectin from bay scallop A. irradians. Fish & Shellfish Immunology 28(2):326–32. doi:10.1016/j.fsi.2009.11.016

Stenttford GD, Neil DM, Peeler EJ, Shields JD, Small HJ, Flegel TW, et al. 2012. Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. Journal of Invertebrate Pathology 110:141–57. doi:10.1016/j.jip.2012.03.013

Tassanakajon A, Tiptawonnukul A, Supungul P, Rimphanitchayakit V, Cook D, Jarayabhand P, et al. 1998. Isolation and characterization of microsatellite markers in the black tiger prawn P. monodon. Molecular Marine Biology and Biotechnology 7:55–61.

Van Oosterhout C, Weetman D, Hutchinson WF. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Molecular Ecology Notes 6:255–56. doi:10.1111/j.1471-8286.2006.01082.x

Vaseeharan B, Rajakamaran P, Jayaseelan D, Vincent AY. 2013. Molecular markers and their application in genetic diversity of penaeid shrimp. Aquaculture International 21:219–41. doi:10.1007/s10499-012-9582-9

Venables WN, Ripley BD. 2002. Random and mixed effects. In: Venables WN, Ripley BD, editors. Modern Applied Statistics with S. New York: Springer, p 271–300.

Walker PJ, Mohan CV. 2009. Viral disease emergence in shrimp aquaculture: origins, impact and the effectiveness of health management strategies. Reviews in Aquaculture 1:125–54. doi:10.1111/j.1753-5131.2009.00107.x

Wan Q, Wu H, Fujihara T, Fang S. 2004. Which genetic marker for which conservation genetics issue? Electrophoresis 25:2165–76. doi:10.1002/elps.200305922

Wang X, Xu J, Zhao X, Vasta GR, Wang J. 2014. A shrimp C-type lectin inhibits proliferation of the hemolymph microbiota by maintaining the expression of antimicrobial peptides. Journal of Biological Chemistry 289(17):11779–90. doi:10.1074/jbc.M114.552307

Waqairatu SS, Dierens L, Cowley JA, Dixon TJ, Johnson KN, Barnes AC, Li Y. 2012. Genetic analysis of black tiger shrimp (P. monodon) across its natural distribution range reveals more recent colonization of Fiji and other South Pacific Islands. Ecology and Evolution 2(8):2057–71. doi:10.1002/ece3.316

Willette DA, Allendorf FW, Barber PH, Barshis DJ, Carpenter KE, Crandall ED, et al. 2014. So, you want to use next-generation sequencing in marine systems? Insight from the Pan-Pacific Advanced Studies Institute. Bulletin of Marine Science 90(1):79–122. doi:10.5343/bms.2013.1008

Winter D, Green P, Kamvar Z, Gosselin T. 2015. Package ‘MMOD’: modern measures of population differentiation. https://github.com/dwinter/mmod. Computer program.

You EM, Chiu TS, Liu KF, Tassanakajon A, Klinbunga S, Tino, et al. 2008. Microsatellite and mitochondrial haplotype diversity reveals population differentiation in the tiger shrimp (P. monodon) in the Indo-Pacific region. Animal Genetics 39:267–77. doi:10.1111/j.1365-2052.2008.01724.x

Zeng D, Chen X, Xie D, Zhao Y, Yang C, Li Y, et al. 2013. Transcriptome analysis of Pacific white shrimp (L. vannamei) hepatopancreas in response to Taura Syndrome Virus (TSV) experimental infection. PLoS One 8(2):e57515. 8 pages. doi:10.1371/journal.pone.0057515