



## *In silico* analysis of polymorphic microsatellites in penaeid shrimp and construction of a free-access database

Patrícia Domingues de Freitas, Allysson Pontes Pinheiro, Thiago Buosi Silva and Pedro Manoel Galetti Junior

*Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.*

### Abstract

We performed an *in silico* analysis of all microsatellites so far described for penaeid shrimp and for which the polymorphic behavior has previously been analyzed. The objective of the study was to evaluate the structural characteristics of these microsatellites and identifying patterns which allow the characterization of the nature of these sequences in the penaeid genome. All data were compiled in a free-access database specially constructed for this study. Three hundred non-monomucleotide polymorphic microsatellite loci described for 12 shrimp species belonging to the family Penaeidae were analyzed and simple and compound microsatellites with di-, tri-, tetra-, penta- and hexanucleotide motifs were found. Dinucleotides and trinucleotides were the most frequent motifs among both the simple and the compound microsatellites. Although a certain bias related to different microsatellite isolation methodology could not be discarded, it is possible that part of this microsatellite abundance reflects some degree of conservation of microsatellite motifs among the different species. There was a pronounced motif variability within and between species, indicating high differentiation dynamism of these repetitions in this animal group. This study not only sheds light on the structure of the microsatellites present in the penaeid shrimp genome but also resulted in the free-access Penaeid Shrimp Microsatellite Database (available at <http://www.shrimp.ufscar.br>) which may be very useful for optimizing the use of these microsatellites.

*Key words:* Crustacea, microsatellite structure, Penaeidae, SSR.

Received: September 19, 2006; Accepted: May 7, 2007.

Over the last few decades, microsatellite sequences have been widely used in many animal groups, including shrimp, and principally in economically important species (Wolfus *et al.*, 1997; Bierne *et al.*, 2000; Xu *et al.*, 2001). Due to their high level of allelic variation and codominant nature, microsatellite loci have been extremely useful for the determination of genetic diversity, stock discrimination, identification of lineages and individuals, the establishment of pedigrees, the development of breeding programs, linkage map studies and the identification of loci related to commercially important characteristics (Moore *et al.*, 1999; Ozaki *et al.*, 2000).

In the past half decade, different research groups have been striving to characterize microsatellite markers in the genome of a number of penaeid shrimp species (Cruz *et al.*, 2002; Meehan *et al.*, 2003; Pérez *et al.*, 2005). However, the validation of these sequences has generally not been easy. The penaeid genome seems to contain very long

microsatellite repetitions, hindering the cloning and sequencing of microsatellites containing both flanking regions. A few comparisons demonstrate that the microsatellites found in penaeids (Tassanakajon *et al.*, 1998; Cruz *et al.*, 2002) are approximately twice the size of the sequences found in other animal groups (Estoup *et al.*, 1993; Brooker *et al.*, 1994), being four or five times larger in some cases.

Currently, there is a significant number of polymorphic microsatellite loci characterized for different penaeid species. This important source of markers is available in public databanks and/or published in indexed journals that may be accessed online. We performed an *in silico* analysis of the polymorphic microsatellite loci described for penaeids up to June 2006, the aim being to characterize the structural patterns of these sequences and compile a free-access database containing the information needed for their optimized.

The microsatellite nucleotide sequences analyzed were accessed using their GenBank (NCBI) ID codes and/or via an electronic search of the journals available online. All the sequences were converted to the FASTA

format and submitted to analysis using the Tandem Repeats Finder (TRF) program version 3.21 (Benson, 1999) for the identification of the repeats and localization of the flanking regions. The TRF program generated data files describing the motif regions and respective flanking regions. These data were compared with those described for the referred locus in the scientific literature and deposited in a free-access Penaeid Shrimp Microsatellite Database, the address of which is given below in the Internet resources section.

This database allows researchers to perform different searches and generate specific reports based on eight pre-established criteria: (1) author; (2) species; (3) ID code; (4) microsatellite type (simple or compound); (5) number of bases in the motif; (6) type of bases in the motif, *i.e.* AT-rich, CT-rich, etc.; (7) number and size of alleles per locus; and (8) results generated by the TRF program during the analysis, including any inconsistencies with the literature concerning the locus in question. Among the main observed inconsistencies in the sequences analyzed were the presence of microsatellites which were not flanked by one or both of the described primers, partially or totally absent repeats in the sequence described, microsatellites described with different motifs than those found by the TRF program and microsatellites with more than one GenBank ID or for which the ID numbers were different to those described in the paper in which it was cited.

Three hundred non-monomucleotide polymorphic microsatellite loci described for 12 shrimp species belonging to the family Penaeidae were analyzed: *Farfantepenaeus notialis* (Robainas *et al.*, 2002; Robainas *et al.*, 2003), *Farfantepenaeus paulensis* and *Farfantepenaeus subtilis* (Maggioni and Rogers 2002), *Fenneropenaeus chinensis* (Wang *et al.*, 2005), *Fenneropenaeus merguensis* (Wanna *et al.*, 2004), *Litopenaeus schmitti* (Espinosa *et al.*, 2001; Maggioni *et al.*, 2003; Borrel *et al.*, 2004), *Litopenaeus setiferus* (Ball *et al.*, 1998), *Litopenaeus stylirostris* (Vonau *et al.*, 1999; Bierne *et al.*, 2000), *Litopenaeus vannamei* (Wolfus *et al.*, 1997; Cruz *et al.*, 2002; Meehan *et al.*, 2003; Pérez *et al.*, 2005), *Marsupenaeus japonicus* (Moore *et al.*, 1999), *Penaeus esculentus* (Meadows *et al.*, 2003) and *Penaeus monodon* (Xu *et al.*, 1999; Pongsomboon *et al.*, 2000; Brooker *et al.*, 2000; Xu *et al.*, 2001; Wuthisuthimethavee *et al.*, 2003; Pan *et al.*, 2004). Cross-species amplification results are practically inexistent, being limited to a few loci described for *L. vannamei* that were also efficient in *Rimapenaeus byrdi* (= *Trachypenaeus byrdi*) and *L. stylirostris* (Pérez *et al.* 2005).

Simple and compound microsatellites (Garza and Desmarais, 2000) with di-, tri-, tetra-, penta- and hexanucleotide motifs were found. Almost 52% of the total microsatellites analyzed contained only one type of motif and were characterized as simple microsatellites. In a few cases, simple microsatellites presented interruptions be-

tween their repetition units and were thus classified as imperfect microsatellites. Approximately 48% were microsatellites composed of more than one motif-type, sometimes interrupted by divergent bases. Approximately 50% of the loci studied were described for *L. vannamei* and about 24% for *P. monodon*, both of which are considered to be the most commercially important marine shrimp in the world. Over 25% of the loci have been described for the other species, with, to date, only a single characterized and validated locus for the shrimp *F. merguensis* (Table 1).

Dinucleotide sequences (interrupted or uninterrupted) were the predominant motif, present in 56% of the total microsatellites analyzed. Nearly 58% of the simple and 55% of the compound microsatellites contained dinucleotides, TC (19%) and GT (17%) being the most abundant motifs. Although a certain bias related to different microsatellite isolation methodology could not be discarded, it is possible that part of this abundance reflects some conservation of microsatellites motifs among the different species. Trinucleotide sequences also appeared at high frequencies and represented 27% of the total simple and 39% of the total compound microsatellites, and although a wide diversity of motifs has been described for this specific class of microsatellites the most frequent motifs were GAA (13%) for the simple trinucleotide microsatellites and ATT (7%) for the compound trinucleotide microsatellites. Tetra-, penta- and hexanucleotides were observed at lower frequencies, although tetranucleotides appeared in more than 30% of compound microsatellites.

The same microsatellite distribution trend verified for the set of all penaeid species was also observed for the species analyzed separately, although the trinucleotide microsatellites in *P. esculentus* were predominant. Dinucleotide repetitions appeared in 64% of the total microsatellites described for *L. vannamei* and in 45% for *P. monodon*, the two penaeid species with the largest number of known loci. However, the motifs TA (24%) in *L. vannamei* and TC and CA that appeared in equal proportion (6%) in *P. monodon* were the most abundant, differing from those observed when the set of all species is considered. The tetranucleotides were also abundant in the compound microsatellites of both species, corroborating the results that considered the entire penaeid set.

Despite the observed exceptions, similar to those reported for other crustaceans and insects (Robainas *et al.*, 2003), microsatellites containing dinucleotide repetitions seem to be the most frequent among penaeid shrimp, followed by microsatellites containing trinucleotides. Moreover, even though a few motifs of all the analyzed classes may appear at greater abundance, there is a pronounced variability within and between species, indicating the high differentiation dynamism of these repetitions within this animal group.

Thus, besides contributing to a better understanding of the microsatellite structure in the penaeid shrimp

**Table 1** - Abundance and distribution of penaeid shrimp simple (S) and compound (C) polymorphic microsatellites among 300 microsatellite loci available from published work and/or genetic databases. Motif and size (di-, tri-, tetra-, penta- and hexanucleotide) repeats. N = number of described loci.

	Simple/compound microsatellites						Most frequent motifs			
	N	Total S/C	Di-S/C	Tri-S/C	Tetra-S/C	Penta-S/C	Hexa-S/C	Di-S/C	Tri-S/C	Tetra-S/C
All penaeids	300	154/146	89/81	42/57	11/46	10/5	2/9	GT/TC	GAA/ATT	AAAAG/CTTT
<i>Farfantepenaeus</i>										
<i>F. notialis</i>	9	5/4	5/3	-/1	-/1	-/1	-/-	GA/GT	-/CTT	-/CATA
<i>F. paulensis</i>	6	4/2	4/2	-/-	-/1	-/-	-/-	CA/CA, GT, GA, CT	-/-	CAAA/-
<i>F. subtilis</i>	3	1/2	1/2	-/-	-/-	-/-	-/-	GT/CA	-/-	-/-
<i>Fenneropenaeus</i>										
<i>F. chinensis</i>	9	7/2	5/1	2/1	-/-	-/-	-/-	TG/CA, TA	TTA, TAA/TTA, ATT	-/-
<i>F. merguensis</i>	1	1/-	1/-	-/-	-/-	-/-	-/-	CT/-	-/-	-/-
<i>Litopenaeus</i>										
<i>L. schmitti</i>	5	2/3	2/2	-/1	-/1	-/-	-/-	GT/CT, GA, GT	-/CCT	-/CTTT, CATT, CATG
<i>L. setiferus</i>	6	6/-	6/-	-/-	-/-	-/-	-/-	GT/-	-/-	-/-
<i>L. stylirostris</i>	3	3/-	3/-	-/-	-/-	-/-	-/-	TG/-	-/TTC, TCC	AAAAG, AAAAC/-
<i>L. vannamei</i>	164	73/91	27/78	25/39	9/20	10/4	2/5	TA/TC	TGG/ATT	AAAT/CTTT
<i>Marsupenaeus</i>										
<i>M. japonicus</i>	12	9/3	4/1	5/2	-/-	-/-	-/-	CA/TA, TG	CTT, GAA/CTG, CTT, GAA	-/-
<i>Penaeus</i>										
<i>P. esculentus</i>	9	6/3	-/-	5/3	1/-	-/-	-/-	-/-	TGA/GAG, AAG, TCA, ACT, ATG, CTG	CATA/-
<i>P. monodon</i>	73	37/36	31/2	5/10	1/23	-/-	-/4	TC, CA/CA	CAT/CAT	CAGT/ACA

genome, our analysis and the microsatellite database supporting it may be very useful for further genetic studies in shrimp and for a greater optimization of the use of these microsatellites.

## Acknowledgments

This work was supported by the Brazilian agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## References

- Ball AO, Leonard S and Chapman RW (1998) Characterization of (GT)<sub>n</sub> microsatellites from native white shrimp (*Penaeus setiferus*). Mol Ecol 7:1247-1263.
- Bierne N, Beuzart I, Bonhomme V and Bédier E (2000) Microsatellite-associated heterosis in hatchery-propagated stocks of the shrimp *Penaeus stylirostris*. Aquaculture 184:203-219.
- Borrell Y, Espinosa G, Romo J, Blanco G, Vázquez E and Sánchez JA (2004) DNA microsatellite variability and genetic differentiation among natural populations of the Cuba white shrimp *Litopenaeus schmitti*. Mar Biol 144:327-333.
- Brooker AL, Cook D, Bentzen P, Wright JM and Doyle RW (1994) Organization of microsatellites differs between mammals and cold-water teleost fishes. Can J Fish Aquat Sci 51:1959-1966.
- Brooker AL, Benzie JAH, Blair D and Versini JJ (2000) Population structure of the giant tiger prawn *Penaeus monodon* in Australian waters determined using microsatellite markers. Mar Biol 136:149-157.
- Cruz P, Mejia-Ruiz CH, Perez-Henriquez R and Ibarra AM (2002) Isolation and characterization of microsatellites in Pacific White shrimp *Penaeus (Litopenaeus) vannamei*. Mol Ecol Notes 2:239-241.
- Espinosa GL, Jager M, García-Machado E, Pichs YB, Rodríguez NC, Barcia AR and Deutsch J (2001) Microsatellites from white shrimp *Litopenaeus schmitti* (Crustacea, Decapoda). Biotechnol Appl 18:232-234.
- Estoup A, Solignac M, Harry M and Cornut JM (1993) Characterization of GT and CT microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nucleic Acids Res 21:1427-1431.
- Garza JC and Desmarais E (2000) Derivation of a simple microsatellite locus from a compound ancestor in the genus *Mus*. Mamm Genome 11:1117-1122.
- Maggioni R and Rogers AD (2002) Microsatellite primers for the three Western Atlantic *Farfantepenaeus* prawn species. Mol Ecol 2:51-53.
- Maggioni R, Rogers AD and Maclean N (2003) Population structure of *Litopenaeus schmitti* (Decapoda, Penaeidae) from Brazilian coast identified using six polymorphic microsatellite loci. Mol Ecol 12:3213-3217.
- Meadows JRS, Ward RD, Grewe PM, Dierens LM and Lehnert SA (2003) Characterization of 23 tri- and tetranucleotide microsatellite loci in the Brown Tiger Prawn, *Penaeus esculentus*. Mol Ecol Notes 3:454-456.

- Meehan D, Xu Z, Zuniga G and Alcivar-Warren A (2003) High frequency and large number of polymorphic microsatellites in cultured shrimp, *Penaeus (Litopenaeus) vannamei* (Crustacea, Decapoda). *Mar Biotechnol* 5:311-330.
- Moore SS, Whan V, Davis GP, Byrne K, Hetzel DJS and Preston N (1999) The development and application of genetic markers for the Kuruma prawn *Penaeus japonicus*. *Aquaculture* 173:19-32.
- Ozaki A, Sakamoto T, Khoo S, Nakamura K, Coimbra MR, Akutsu T and Okamoto N (2000) Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Onchorhynchus mykiss*) *Mol Genet Genom* 265:23-31.
- Pan Y-W, Chou H-H, You E-M and Yu H-T (2004) Isolation and Characterization of 23 polymorphic microsatellite markers for diversity and stock analysis in tiger shrimp (*Penaeus monodon*). *Mol Ecol Notes* 4:345-347.
- Pérez F, Ortiz J, Zhinula M, Gonzabay C, Calderón J and Volckaert FAMJ (2005) Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus byrdi*. *Mar Biotechnol* 7:554-569.
- Pongsomboon S, Whan V, Moore SS and Tassanakajon A (2000) Characterization of tri- and tetranucleotide microsatellites in the black tiger prawn, *Penaeus monodon*. *ScienceAsia* 26:1-8.
- Robainas A, Espinosa G and Machado EG (2003) Characterization of the pink shrimp *Farfantepenaeus notialis* (Crustacea, Decapoda) microsatellite DNA. *Rev Invest Mar* 24:161-164.
- Robainas A, Monnerot M, Solignac M, Dennenbouy N, Espinosa G and Machado EG (2002) Microsatellite loci from the pink shrimp *Farfantepenaeus notialis* (Crustacea, Decapoda). *Mol Ecol Notes* 2:344-345.
- Tassanakajon A, Tiptawonnukul A, Supungul P, Rimphanitchayakit V, Cook D, Jarayabhand P, Klinbunga S and Boonsaeng V (1998) Isolation and characterization of microsatellite markers in the black tiger prawn *Penaeus monodon*. *Mol Mar Biol Biotechnol* 7:55-61.
- Vonau V, Ohresser M, Bierne N, Delsert C, Beuzart I, Bedier E and Bonhomme F (1999) Three polymorphic microsatellites in the shrimp *Penaeus stylirostris*. *Anim Genet* 30:234-235.
- Wang H, Li F and Xiang J (2005) Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. *Aquaculture* 249:107-114.
- Wanna W, Rolland J-L, Bonhomme F and Phongdara A (2004) Population genetic structure of *Penaeus merguensis* in Thailand based on nuclear DNA variation. *J Exp Mar Biol Ecol* 311:63-78.
- Wolfus GM, Garcia DK and Alcivar-Warren A (1997) Application of the microsatellite technique for analysing genetic diversity breeding programs. *Aquaculture* 152:35-47.
- Wuthisuthimethavee S, Lumubol P, Vanavichit A and Tragonrungs S (2003) Development of microsatellite markers in black tiger shrimp. *Aquaculture* 224:39-50.
- Xu Z, Dhar AK, Wyrzykowski J and Alcivar-Warren A (1999) Identification of abundant and informative microsatellites from shrimp (*Penaeus monodon*). *Anim Genet* 30:150-156.
- Xu Z, Primavera JH, de la Pena LD, Pettit P, Belak J and Alcivar-Warren A (2001) Genetic diversity of wild and culture Black Tiger shrimp (*Penaeus monodon*) in the Philippines using microsatellites. *Aquaculture* 199:13-40.

## Internet Resources

- Benson G (1999) Tandem repeats finder: A program to analyze DNA sequence. *Nucleic Acids Res* 27:573-580. Available at <http://tandem.bu.edu/trf/trf.submit.options.html>.
- National Center for Biotechnology Information (NCBI) Online, GenBank, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide> (June 30, 2006).
- Penaeid Shrimp Microsatellite Database, available at <http://www.shrimp.ufscar.br>.

*Associate Editor: Fabio de Melo Sene*