THE mtDNA CONTROL REGION OF THE BARBA-RUÇA SHRIMP _ARTEMESIA LONGINARIS_ (DECAPODA: PENAEIDAE) AND ITS POTENTIAL USE AS A MARKER FOR POPULATION ANALYSIS

DUMONT, LUIZ FELIPE Cestari\textsuperscript{1,2,3}, HWANG, GYU-LIN\textsuperscript{1} & MACLEAN, NORMAN\textsuperscript{1}.

1 - School of Biological Science, University of Southampton, Basset Crescent East, Southampton, SO16 7PX, UK.
2 - Universidade Federal do Rio Grande – Instituto de Oceanografia
Caixa Postal 474 – Rio Grande – RS – CEP 96201-900 - Brazil
3 – felipe_dumont@hotmail.com

ABSTRACT

The shrimp _Artemesia longinaris_ is endemic from Southwestern Atlantic and is commercially exploited from Argentina (Puerto Rawson – 21°37’S) to Southeastern Brazil (Rio de Janeiro – 43°00’S). Molecular markers, such as the mtDNA control region, (CR) have been used to elucidate the population structure of penaeid shrimps worldwide. The suitability of mtDNA CR of the barba-ruça shrimp as a molecular marker at a population level was tested and a novel set of primers to amplify this region has been designed. Primers were rooted in the flanking genes of the CR that showed the same order (12S at 5’ and ITRNA\textsubscript{Ile} at 3’ extreme) as reported for other penaeid shrimps. The CR of _A. longinaris_ was 990 bp long, presenting two hypervariable regions at the 5’ and 3’ extremes (more variable), and a central one with less polymorphism. In addition, an internal primer set to amplify approximately 800 bp of 5’ extreme of CR, including the hypervariable region I, is provided to help resolving population structure. Comparison of the CR with cytochrome oxdase I (COI) sequences showed that the former gene presents higher polymorphism. Nucleotide diversity estimated for CR was low (\(\pi=0.017\)), and haplotype diversity high (\(Hd=0.92\)), but both fall within the values suggested for the family. Preliminary F\textsubscript{st} values suggest that populations inhabiting extremes of distribution show less genetic interchange. Briefly, we were able to confirm the suitability of CR hypervariable regions of _A. longinaris_ as a molecular marker to resolve the population structure of _A. longinaris_. identification.

KEY WORDS: _Artemesia longinaris_, control region, population structure, mtDNA, stock

INTRODUCTION

_Artemesia longinaris_ is an endemic penaeid shrimp, distributed from Argentina (Puerto Rawson – 21°37’S) to Southeastern Brazil (Rio de Janeiro – 43°00’S) (D’Incao 1999). Its life cycle does not demand an estuarine phase, but the juveniles migrate to shallow marine waters during grow out phase and females move to deeper waters for spawning (Boschi 1969, Castillo et al. 2007). The Penaeidae family includes several commercially important shrimp species, such as _Farfantepeneaus paulensis_, _Farfantepeneaus brasiliensis_ and _Litopenaeus schmitti_. However, the genus _Artemesia_ is monospecific and includes only the cold water species _Artemesia longinaris_ (D’Incao, 1999). This species is distributed along the coastal waters (rarely further than 30 m) of Southwestern Atlantic, presenting a clear preference for soft bottoms, as well as a higher affinity for cold and salty water masses (Boschi 1969, Fransozo et al. 2004, Dumont, 2005, Dumont & D’Incao 2008).

Unlike many other penaeid fisheries (e.g. _Farfantepeneaus duorarum_, _F. paulensis_, _F. brasiliensis_), landings of _A. longinensis_ have been recently increased (Navarrete et al. 1995, D’Incao et al. 2002.), despite of a great interannual variability that mainly results from environmental effects. Landings of _A. longinensis_ in Brazil have grown since the 1990’s achieving a maximum value of 7,424 tons in 2003 (Unival 2004). To date, it is the second most profitable shrimp fishery in Southern and

CONTROL REGION OF _A. LONGINARIS_
Southeastern Brazil (Pezzuto 2001). Since stock structure is vital for a proper management strategy, molecular studies have been used to identify particular populations or stocks (Benzie 2000).

The use of mtDNA has proved to be a useful tool for population genetic studies of many terrestrial and aquatic organisms (Avise 1994). Two portions of mtDNA can be especially useful at population level, the cytochrome oxydase I (COI) and the control region (CR) (Caccone et al. 1996). The control region, which is also known as the AT-rich region, does not code for a functional gene; therefore, it is under fewer functional and structural constraints, leading to a high average substitution rate (Saccone et al. 1987). It is usually the fastest evolving region in the mitochondrial genome of either vertebrates or invertebrates, and consequently more sensitive than protein loci as a marker of phylogeographic structuring of many organisms (Caccone et al. 1996; Avise 2000, Roldán et al. 2009, Wang et al. 2009).

Due to these characteristics, the CR is of special utility for phylogeographic analysis over extreme microevolutionary time scales, which often show an exceptionally fast pace of nucleotide substitution and high level of intraspecific polymorphism (McMillan & Palumbi 1995). Since no information on the CR of *A. longinaris* is available, summed to the fact that the increasing interest in this resource requires the identification of possible unit stocks, the aim of this work is to explore the potential use the of control region of *A. longinaris* as a molecular marker in order to characterize phylogeographic patterns of this species.

**MATERIAL & METHODS**

**Sampling strategy, tissue preservation and DNA extraction**

Samples were obtained from fishery across the distribution area of *A. longinaris*, from Argentina to Rio de Janeiro in spring and were arbitrarily divided in three different stocks: Rio de Janeiro (RJ- 21°37’S), Convergence Zone (CZ- 32°00’S and 26°54’S) and Mar del Plata (AR- 37°56’S) (Figure1). Three individuals from each region were used to assess the suitability of CR as a molecular marker. It is important to point out that the aim of this investigation was only to test the suitability of CR as a marker for population analysis. Therefore, the results obtained must not be used to infer population structure, but only to confirm the suitability of this region through the comparison with similar investigations.
Figure 1. Southwestern Atlantic Ocean and distribution range of *A. longinaris* (black solid line). Black dots indicate sites where samples were obtained; from Mar del Plata (Argentina, AR) Rio Grande do Sul (Brazil, CZ) and Macaé (Brazil, RJ).

Muscle samples from pereiopods and abdomen were removed, fixed in 95% ethanol and stored at 4°C. DNA was extracted using a phenol-chloroform-isoamyl alcohol (25:24:1) extraction of sodium dodecysulfate (SDS) and proteinase K digested (Sambrook & Maniatis 1989). DNA was isolated by 100% ethanol precipitation and visualized by gel electrophoresis to check the quality and amount obtained.
Design of mitochondrial control region primers

Universal control region primers and other primers designed to amplify the partial or the entire control region of penaeid shrimps were tested with no success (Chu et al. 2003, MacMillen-Jackson & Bert 2003, Grabowski et al. 2004). Non-amplification or yield of a large number of nonspecific fragments and unsuccessfully optimization of amplification conditions led to the necessity of designing a specific primer set for *A. longinaris*.

The first set of oligonucleotide primers were designed on more conserved genes that flank the penaeid shrimps control region (Figure 2) and were based on a consensus alignment of *Penaeus monodon* (GenBank accession number NC002184, Wilson et al. 2000), *Marsupenaeus japonicus* (GenBank accession number AP006346, Yamauchi et al. 2005) and *Penaeus notialis* (GenBank accession number X84350, Garcia-Machado et al. 1999). The next primer sets used were designed based on *A. longinaris* sequences.

The forward primer (12Sa-F) was rooted in the beginning (5’ end of 12S) of the small subunit ribosomal RNA gene (SSU rRNA – 12S, position 14209 of *P. monodon* mtDNA) and the reverse (ILE3a-R) in the isoleucine transfer RNA (tRNAIle3, position 7 of *P. monodon* mtDNA) gene (Figure 2). The second set of primers (12Sb-F and ILE3b-R) was positioned in the middle of the 12S gene (position 14596 of *P. monodon* mtDNA) and inside the CR (position 15945 of *P. monodon* mtDNA). A third set of primers (12Sc-F and ILE3c-R) was positioned in the 5’ flank of CR (position 14988 of *P. monodon* mtDNA) and inside the CR (position 15310 of *P. monodon* mtDNA). Primer-dimmer, secondary structure and G+C percentage were checked by using PRIMER3 software online (Rozen & Skaletsky 2000).

Figure 2. Primer positions used for partial amplification of SSU rRNA (12S) mtDNA and the entire control region of *A. longinaris*. Diagram also shows the position of flanking genes surrounding control region.

The first set of primers successfully amplified a fragment of 1.8 kb, including approximately 90% of the 12S gene and the entire CR. Two more primer walking steps have been performed to obtain more reliable sequences from the entire fragment. In order to establish a variation reference point, we amplified a 0.7kb fragment from COI region (HCO/LCO) (Folmer et al. 1994). Additionally, an interspecific comparison of 12S was also performed to check whether the fragment amplified produced the same traits observed for similar penaeid shrimp species. Primer oligonucleotides and their position in *P. monodon* mtDNA entire genome (GenBank accession number NC002184; Wilson et al. 2000) are provided in table 1.
Table 1. Primer sets used for *A. longinaris* mtDNA amplification, including oligonucleotide sequences, gene location and primer position based on *P. monodon* sequences. *nr= position not reported

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Oligonucleotide sequences</th>
<th>Gene location</th>
<th>Primer position (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12Sa-F</td>
<td>5'-AGCGACGCGCAGTGATGTTACAT- 3'</td>
<td>SSU rRNA</td>
<td>14209</td>
</tr>
<tr>
<td>ILE3a-R</td>
<td>5'-GATACTTCTTTCAGGCAGN- 3'</td>
<td>tRNA-Ile</td>
<td>7</td>
</tr>
<tr>
<td>12Sb-F</td>
<td>5'-GTGTAACAGGGTATCTACTATC- 3'</td>
<td>SSU rRNA</td>
<td>14596</td>
</tr>
<tr>
<td>ILE3b-R</td>
<td>5'-GAGGGTGCCGCAAGAACAAA- 3'</td>
<td>CR</td>
<td>15945</td>
</tr>
<tr>
<td>12Sc-F</td>
<td>5'-GAACTCAAGCGCAGATAAAC- 3'</td>
<td>SSU rRNA</td>
<td>14988</td>
</tr>
<tr>
<td>ILE3c-R</td>
<td>5'-GTTGTATCGATTTAGTATTT- 3'</td>
<td>CR</td>
<td>15310</td>
</tr>
<tr>
<td>HCO</td>
<td>5'-TAAACTTCAGGTGACCAAA-3'</td>
<td>COI</td>
<td>nr</td>
</tr>
<tr>
<td>LCO</td>
<td>5'-GTCACAAAAATCGGGTGACCATAAAT- 3'</td>
<td>COI</td>
<td>nr</td>
</tr>
</tbody>
</table>

Amplification, purification of PCR products and sequencing

Amplification reactions were conducted using an Applied Biosystems PCR machine and each 20 µl of PCR contained approximately 5 ng of DNA template, 3.5 µl of 10x buffer (with 15 mM of MgCl$_2$), 0.2 µl of dNTP (20 mM), 0.5 µl of MgCl$_2$ (15 mM), 3.0 µl of Q-solution (Qiagen), 0.2 µl of each primer (100 pM), 0.5 µl of Taq DNA polymerase (5U/µl) and sterile HPLC-grade water. Ten cycles were carried out under the following conditions: 1x94°C for 1 minute, 94°C for 10 seconds, 56°C for 1 min (12Sa-F/ILE3a-R), 68°C for 1:50 min and finally 5 min at 68°C (10x). Twenty five cycles more were carried out adding 10 sec of extension time at each cycle. This strategy was adopted to compensate the loss of synthesizing ability of polymerase through the cycles, when amplifying long products.

Annealing temperature used to amplify 1.3 kb with the second primer set (12Sb-F/ILE3b-R), was 48.5°C for 1 minute and an elongation time of 1:30 minutes at 68°C, also adding 10 seconds of extension time at each cycle. Last primer walking step amplified approximately 0.3kb at the 5’ extremity of CR (annealing temperature 62°C), in such a way that the entire CR was entirely covered.

Internal primers (12Sc-F/ILE3b-R) were used to amplify approximately 0.8kb (709bp were effectively analyzed) of CR, including the hypervariable region at the 5’extreme, in attempt to test the suitability of CR as a genetic marker for *A. longinaris*. The annealing temperature used with this set of primers was 67°C. A total of nine individuals were sequenced for intraspecific comparison of CR polymorphism and compared to sequences obtained from the cytochrome oxydase I gene (700bp fragment length, 544bp were effectively analyzed) (Table 1). Annealing temperature used to amplify COI region was 60°C.

The PCR products were checked for correct size and amount of DNA on 1.0% agarose gel. Products were then purified by using PCR purification kit (Qiaquick™ PCR Purification Kit, Qiagen, Inc.) to remove excess nucleotides and primers as well as concentrate PCR products. The purified double-stranded amplification products were used as template DNA in sequencing reactions. Cycle-sequencing was performed using the ABI Prism BigDye Ready Mix (Applied Biosystems) and all PCR products were sequenced in both forward and reverse directions.

Sequence alignment, phylogenetic and phylogeographic analyses

Homologous nucleotide sequences from all samples were aligned using the program CLUSTAL W included in BIOEDIT (Hall 1999) and refined when necessary. Control region and flanking genes sequences were reported for the first time for *A. longinaris* (Accession number EU400382), as well as the COI sequence (Accession number EU400383) and have been submitted to GenBank database (National Center of Biotechnology Information). Estimates of nucleotide diversity ($\pi$), haplotype diversity ($Hd$) and preliminary $Fst$ distances between CR sequences were estimated by using Arlequim 3.1 (Schneider et al. 2006). Kimura 2-parameter was used in the interspecific comparison with 12S. Substitution rate as well as the nucleotide composition were estimated through the software MEGA (Tamura et al. 2007).
RESULTS

Interespecific comparison of 12S and CR

We were able to sequence approximately 90% of the 3' extremity of the 12S gene (764 of 852bp, 400bp were effectively analyzed) as well as the entire CR (990bp). Comparison of A. longinaris 12S with those of P. monodon and M. japonicus resulted in 2.5% (10/400bp) of polymorphic sites (S%), an average number of differences (k) of 7.66 (±2.55) and nucleotide diversity (π) of 0.0196 (±0.00004).

Nucleotide composition of the 12S gene was 36.43% (A), 35.44% (T), 17.84% (C) and 10.30% (G). Kimura 2-parameter, based on 12S alignment between three different Penaeidae genera (Penaeus, Marsupenaeus and Artemesia), varied from 0.197 (Marsupenaeus and Penaeus) to 0.256 (Artemesia and Penaeus).

Interspecific alignment of the entire CR sequence with P. monodon and M. japonicus showed three hypervariable regions, two at extremes and one central. The hypervariable region (I) extends from the 5' extreme for approximately 300bp. The hypervariable region I is followed by a more conserved region, that stretches for approximately 100bp. This conserved area is followed by a central hypervariable domain (II), which presents lower variation rates. The central domain is followed by another short conserved stretch. At the end of this conserved region there is another hypervariable region (III), similar to the hypervariable region I in terms of variation rates, that ends concurrently with the control region (Figure 3).

![Figure 3](image-url) Variation in nucleotide diversity (Pi) along the 990bp of penaeid control region. Three hypervariable regions were observed (HI, HII and HIII), agreeing with previous penaeid shrimps CR sequenced. Comparison was made between three different Penaeid genera (Penaeus, Artemesia and Marsupenaeus).

Intraspecific comparison of control region (CR) and cytochrome oxydase (COI)

Intrapopulation analysis showed that the percentage of polymorphic sites (S%) estimated for CR was 6.9% (49/709bp), haplotype diversity (Hd) found was 0.92 (0.83-0.98), nucleotide diversity (π) 0.017 (±0.0009) and the average number of differences (k) 12.09 (±0.367). The nucleotide composition was 42.49% (A), 45.61% (T), 7.79% (C) and 4.11% (G).

The COI gene showed a lower percentage of polymorphic sites (1.28% or 7/544bp), haplotype
diversity ($Hd=0.38<0.77>0.94$), nucleotide diversity ($\pi=0.006\pm0.000001$) as well as lower average number of differences ($k= 3.55\pm0.83$) than CR (Table 2). The nucleotide composition was 28.31% (A), 30.51% (T), 22.06% (C) and 19.12% (G).

Table 2. Summary of intrapopulational comparison between control region (CR) and cytochrome oxydase (COI) and gene, containing fragment length (bp), percentage of polymorphic sites ($S\%$), haplotype diversity ($Hd$), nucleotide diversity ($\pi$) and A+T composition obtained from Artemesia longinaris. Estimates were based on nine individuals sampled along distribution area.

<table>
<thead>
<tr>
<th>Amplified fragment</th>
<th>CR</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment length (bp)</td>
<td>709</td>
<td>544</td>
</tr>
<tr>
<td>Percentage of polymorphic sites ($S%$)</td>
<td>6.91</td>
<td>1.28</td>
</tr>
<tr>
<td>Haplotype diversity ($Hd$)</td>
<td>0.92</td>
<td>0.77</td>
</tr>
<tr>
<td>Nucleotide diversity ($\pi$)</td>
<td>0.017</td>
<td>0.006</td>
</tr>
<tr>
<td>Nucleotide composition (A+T%)</td>
<td>88.1</td>
<td>58.82</td>
</tr>
</tbody>
</table>

Since CR showed more suitability to resolve population structure than 12S and COI, preliminary $Fst$ distances were estimated based on this region. Results revealed that populations inhabiting the Convergence Zone (CZ) and Argentinean waters are more closely related ($Fst=0.02$). Highest $Fst$ value was observed between the populations inhabiting the extremes of distribution area ($Fst= 0.043$). Additionally, the $Fst$ distance between CZ and RJ population was 0.023.

DISCUSSION

Interespecific comparison of 12S

The metazoan mitochondrial genome is a circular, double-stranded DNA molecule that is highly variable in DNA sequence but conservative in gene content and order (Wolstenholme 1992). Sequencing of the CR and part of the adjacent flanking genes confirmed this hypothesis, since gene size and order are identical to those reported for $P. monodon$ (Wilson et al. 2000), $P. notialis$ (Garcia-Machado et al. 1999) and $M. japonicus$ (Yamauchi et al. 2005). The same gene order was also observed for the lobster *Panulirus argus* (Diniz et al. 2005).

The nucleotide composition, observed in partial sequence from 12S gene of *A. longinaris*, was in agreement with values observed for other penaeid mtDNA, such as the giant tiger shrimp *P. monodon* (A+T= 70.6%; G+C= 29%) (Wilson et al. 2000). The bias in favor of A+T has also been observed for other arthropodan mtDNA sequences (Simon et al. 1994), including penaeids (Benzie 2000), which is in accordance with the data here obtained.

Comparison between 12S and CR sequences polymorphism clearly demonstrates that the second greatly increases the number of informative characters and the resolution power in population analysis (Chu et al. 2003). On the other hand, the 12S gene shows a good potential to be used in phylogenetic investigations, at family and species level, as suggested by previous investigations (Taylor et al. 1996). Similar Kimura 2-distances were observed when using 16S and COI genes to elucidate penaeids phylogeny (Lavery et al. 2004). For instance, distance estimated between the genera *Litopenaeus* and *Melicertus* was 0.135, while the average distance estimated within the subgenus was 0.111 (Lavery et al. 2004). Similar distances were observed in comparison with previous results (e.g. Lavery et al. 2004) and reduced polymorphism suggests that there is a lower potential use of 12S sequences to elucidate phylogeographic relationships in crustaceans. It is important to point out that *A. longinaris* is an endemic species from Southwestern Atlantic and has never been included in penaeid shrimp phylogenetic analysis. Therefore, the information presented here suggests that 12S gene is suitable for this kind of investigation, help resolving the phylogeny of penaeid group which is still controversial to date (Lavery et al. 2004).
Intraspecific comparison of control region and cytochrome oxydase

The mtDNA is widely known as a polymorphic marker (Avise 1994); however, the control region can be even more variable and therefore, has been successfully used to detect population structure in many different phyla, from mammals (Lau et al. 1998; Nagata et al. 1998) to invertebrates (Duran et al. 2004). Sequencing of AT-rich region has also been widely used as an effective marker for population studies in crustaceans as whole (Diniz et al. 2005), but especially in shrimp studies (see Benzie, 2000 for a review; Chu et al. 2003, McMillen-Jackson & Bert 2003, Grabowski et al. 2004).

The percentage of polymorphic sites observed among nine specimens (three different sites analyzed based on CR sequencing) of A. longinaris alignment was 6.91, which is close to the lower limit values suggested by Benzie (2000) for penaeid shrimps (2.2% to 33%). It is important to point out that low sampling numbers may have affected this estimate, but nevertheless a very low percentage of polymorphic sites is present in A. longinaris. Nucleotide diversity was also low ($\pi=0.017$) and subject to sampling numbers, but also within the values suggested for the group, that range from 0.058 in P. monodon from Indonesia to 0.002 in Western Australia (Benzie 2000). Conversely, haplotype diversity was high (0.83-0.92>0.98) but within the estimated in recent investigations on mtDNA control region for penaeid shrimps (McMillen-Jackson & Bert 2003). Low nucleotide diversity combined to high haplotype diversity has frequently been attributed to an expansion after a period of small effective population size, retaining new mutations (Avise et al. 1984).

The A-T composition is known to be high in the CR, as observed for A. longinaris (87.1% A+T) and other penaeid shrimps such as P. monodon (81.5% A+T). The A+T percentage values observed for A. longinaris control region were within values observed in insect genomes (86% - 96% A+T), but higher than those in some crustaceans such as Artemia franciscana (68% A+T) and Daphnia pullex (67.1%) (Wilson et al. 2000).

The COI region showed lower polymorphism than CR when an intraspecific comparison was performed. This observation is in agreement with previous investigations concerning comparison of variation levels between these two genes in penaeids (Chu et al. 2003). Despite lower polymorphic levels observed in COI gene, it may still be used to elucidate A. longinaris phylogeography, since variation observed was not much lower than that of CR.

In conclusion, the present investigation provides a novel set of primers to amplify the entire CR of A. longinaris, as well as internal primers capable of resolving population structure of this species. Additionally, the preliminary Fst distances showed that AR and CZ populations are more closely related, which is supported by the shorter geographical distance separating them, as well as the circulation pattern in near shore Southwestern Atlantic (Piola et al. 2004).

ACKNOWLEDGEMENTS

Authors acknowledge Secretaria de Ciência e Tecnologia do Estado do Rio Grande do Sul for providing financial support to project Camarões Oceânicos and to CAPES for the studentship provided during PDEE at University of Southampton. We would also like to thank Dr. Enrique Boschi (INIDEP), Dr. Marcelo Vianna and Dr. Karina Keunecke (UFRJ) for helping in collections of individuals from distribution limits (Argentina and Rio de Janeiro).

REFERENCES


DUMONT, LFC & F D’INCAO, 2008. Distribution and abundance of the argentinean (Artemesia longinaris) and red (Pleoticus muelleri) prawns (Decapoda-Penaeoidea) in Southern Brazil during the commercial double-rig trawl fishery season. Nauplius, 16(2): 83-94.


