

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE ANIMAL**

**ISOLAMENTO E CARACTERIZAÇÃO DE
MICROSSATÉLITES EM *AEGLA LONGIROSTRI*
(CRUSTACEA, DECAPODA, ANOMURA)**

DISSERTAÇÃO DE MESTRADO

Paula Angélica Roratto

Santa Maria, RS, Brasil

2007

**ISOLAMENTO E CARACTERIZAÇÃO DE
MICROSSATÉLITES EM *AEGLA LONGIROSTRI*
(CRUSTACEA, DECAPODA, ANOMURA)**

por

Paula Angélica Roratto

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Biodiversidade Animal, Área de Concentração em Biologia Evolutiva, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Ciências Biológicas.**

Orientadora: Prof^a. Dr^a. Marlise L. Bartholomei-Santos

Santa Maria, RS, Brasil

2007

**Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Curso de Pós-Graduação em Ciências Biológicas –
Biodiversidade Animal**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**ISOLAMENTO E CARACTERIZAÇÃO DE MARCADORES MICROSSATÉLITE
EM *Aegla longirostri* (CRUSTACEA, DECAPODA, ANOMURA)**

elaborada por
Paula Angélica Roratto

como requisito parcial para obtenção do grau de
Mestre em Ciências Biológicas – Biodiversidade Animal

COMISSÃO EXAMINADORA:


Marlise Ladvat Bartholomei Santos
(Presidente/Orientador)


Andréa Mara Macedo, Dra. (UFMG)


Thales Renato Ochotorena De Freitas, Dr. (UFRGS)

Santa Maria, 27 de fevereiro de 2007.

AGRADECIMENTOS

A realização de um projeto de mestrado e elaboração da dissertação é um produto coletivo, embora o trabalho, redação, responsabilidade e *stress* sejam predominantemente individual. Várias pessoas contribuíram para que este trabalho se concretizasse, às quais expresso minha gratidão.

À minha orientadora, Prof^a Marlise, que desde o nosso primeiro encontro me ensinou a “amar” os microssatélites, despertando meu interesse e entusiasmo pela pesquisa. Meu trabalho de mestrado foi objeto de dupla orientação da Prof^a Marlise e do Prof Sandro, a quem agradeço imensamente a confiança, a amizade e a oportunidade de ter participado deste projeto, que foi o primeiro em conjunto de ambos!

Aos co-orientadores do meu projeto, Prof Élgion Loreto e prof Lia Rejane Silveira Reiniger, pela análise crítica inicial e pelas infinitas sugestões e esclarecimentos que vieram a contribuir com o resultado final deste trabalho.

Um agradecimento especial ao Prof Élgion Loreto e a Prof^a Lenira Sepel, por generosamente disponibilizarem o espaço e a infraestrutura que nos faltavam, sem os quais este trabalho jamais teria sido realizado; além da disponibilidade em esclarecer dúvidas e sugerir soluções nos momentos de dificuldades, e também pelos momentos de descontração compartilhados. Aproveito para estender o agradecimento aos demais integrantes do LabDros: Ana Lúcia, Brenda, Carol, Felipe, Gabriel, Juline, Karina, Luciéle, Mauro, Nina, Paloma e Ronaldo, pela amizade e por muitas vezes terem dividido seus recursos comigo; além da Adriana e do Newton que desde a graduação muito me ensinaram e sempre me incentivaram.

Aos meus eternos colegas divergianos: Alexsandra, Alice, Darine, João Vitor, Josmael, Stela e Talis, por toda AJUDA e APOIO, AMIZADE e principalmente por tornarem a convivência e o trabalho muito prazerosos e divertidos todos os dias! À Darine, em especial, com quem dividi as tarefas e as responsabilidades deste trabalho.

Aos companheiros do LaCrust: André, Bianca, Beto, Daniel, Gláucia, Joele, Luciane e Marcelo; obrigada pela amizade e pela troca de experiências. Reforço um agradecimento especial a Lu, minha *personal trainer* de dissertação, pelos elogios e críticas construtivas à minha redação.

Aos meus pais, Arlindo e Romilda, que tiveram a sabedoria de me educar tendo como prioridade o estudo e sempre me incentivaram à vida acadêmica.

A toda minha pequena grande família: irmãos e irmã, cunhado(as) e sobrinhos(as), que mesmo sem compreender direito o universo do meu trabalho, torceram pelo meu sucesso e sempre me forneceram um porto seguro diante das incertezas e dificuldades.

Ao meu companheiro Leandro, que nos últimos dois anos pacientemente me ouviu falar de eglídeos e microssatélites, de clonagem a corridas eletroforéticas intermináveis... e além disso teve que aprender a conviver com a minha ausência e aturar meu estado de stress e cansaço nos momentos de maiores dificuldades. Obrigada pelo carinho, atenção, compreensão, companheirismo e apoio que dedicaste a mim, sempre!

Às minhas companheiras de apartamento Carol, Carine, Nice e Simone, e à minha prima e colega de profissão Manoela, por sempre me incentivarem e acreditarem no meu sucesso.

Ao Paulo, secretário da pós-graduação, por estar sempre disposto a ajudar, esclarecer dúvidas e nos alertar sobre nossos compromissos, e à Marilene – a “Mari” funcionária do Departamento de Biologia, por toda ajuda, amizade e incentivo.

Aos meus colegas de mestrado: Aline, Gisele, Fernanda, João Vitor, Lizélia, Manuela, Rafaelle e Ricardo; pela amizade, momentos e experiências compartilhados em nossa caminhada.

Além de todas estas e outras pessoas que me apoiaram e que de alguma forma contribuíram com o meu trabalho, registro também minha gratidão a Universidade Federal de Santa Maria, por ter me acolhido nestes dois anos, tendo sido praticamente minha segunda casa.

Ao PPG Biodiversidade Animal pela oportunidade de desenvolver meu projeto e pela infraestrutura proporcionada.

À CAPES, instituição que contribuiu financeiramente para a realização deste trabalho.

Muito obrigada!

There is no freshwater Crustacea at all like
Aegla anywhere else in the world.

(Schmitt, 1942)

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Biodiversidade Animal
Universidade Federal de Santa Maria

ISOLAMENTO E CARACTERIZAÇÃO DE MICROSSATÉLITES EM *AEGLA LONGIROSTRI* (CRUSTACEA, DECAPODA, ANOMURA)

AUTORA: PAULA ANGÉLICA RORATTO

ORIENTADORA: MARLISE LADVOCAT BARTHOLOMEI-SANTOS

Data e Local da Defesa: Santa Maria, 27 de fevereiro de 2007.

Aeglidae é a única família de crustáceos anomuros cujos representantes vivos são habitantes exclusivos de água doce. A origem marinha do grupo é evidenciada pelos registros fósseis. Entretanto, o gênero *Aegla* conquistou o ambiente dulcícola na região temperada do sul da América do Sul onde é endêmico, amplamente distribuído e diversificado, constituindo-se de aproximadamente 70 espécies descritas. Apesar da riqueza de espécies, os eglídeos têm sido vítimas de progressiva perda de habitat e declínio populacional devido à degradação dos ecossistemas límnicos sul-americanos. Visando possibilitar trabalhos futuros voltados à avaliação da diversidade do gênero, este estudo propôs o desenvolvimento de marcadores microssatélites para *Aegla longirostri*. Devido ao seu alto grau de polimorfismo, estes marcadores têm sido amplamente utilizados para avaliação de estrutura populacional, dinâmica e relações entre e dentro de populações dos mais variados organismos. Foram obtidos 42 locos verdadeiros de microssatélites com a construção de uma biblioteca genômica enriquecida com microssatélites (CA)_n para *A. longirostri*, além de 17 seqüências quiméricas oriundas de artefatos da técnica. As repetições encontradas variaram de três a mais de 40 unidades. Desse total, 13 permitiram a projeção de *primers*. Condições ideais de amplificação foram estabelecidas para os locos AICA135 e AIGA138 até o momento, os quais estão sendo testados em 15 indivíduos de uma população da região central do estado do Rio Grande do Sul, onde esta espécie é endêmica, para avaliação do polimorfismo. Testes de amplificação seguem sendo feitos para os demais locos desenvolvidos. Estes marcadores são promissores para investigação de estruturação genética em populações desta espécie. Além disso, estes locos apresentaram amplificação cruzada positiva quando testados para a espécie *Aegla uruguayana*. Para as seqüências quiméricas obtidas, as quais correspondem a 29% do total capturado, sugeriu-se um mecanismo para explicar sua causa e avaliaram-se as implicações deste evento em construções de biblioteca genômica enriquecida com microssatélites.

Palavras-chave: caranguejo de água doce; diversidade genética; seqüência quimérica

ABSTRACT

Master Dissertation
Post-Graduation in Animal Biodiversity
Universidade Federal de Santa Maria

ISOLATION AND CHARACTERIZATION OF MICROSATELLITE MARKERS FROM *Aegla Longirostri* (CRUSTACEA, DECAPODA, ANOMURA)

AUTHOR: PAULA ANGÉLICA RORATTO
ADVISER: MARLISE LADVOCAT BARTHOLOMEI-SANTOS

Aeglidae is the only freshwater anomuran family. Fossil evidences relate a marine origin for aeglids. However, genus *Aegla* invaded the freshwater environment of southern South America, becoming endemic and widespread in that Neotropical region. Besides, the group is widely diversified, consisting of approximately 70 recognized species. Despite the species richness, aeglids have experienced progressive habitat loss and reduction in their populations due to the steam environments alterations. In order to enable future works on *Aegla* diversity, this study proposed to develop microsatellite markers for *Aegla longirostri*. Because of their high polymorphism, such molecular markers have been widely used for establishing genetic structure, dynamic and relationship within and among populations. A total of 42 true microsatellite loci was isolated from the library enriched for (CA)_n repeats, besides 17 chimeric sequences attained from technical artifacts. The repeats ranged from 3 to more than forty units. However, primer pairs could be developed for thirteen sequences. Successful PCR amplifications were obtained for loci AICA135 and AIGA138 up to the moment. This microsatellite loci have been screening in 15 individuals of a population from the central region of Rio Grande do Sul State, Brazil, where that species is endemic, in order to evaluate polymorphism. The other developed primer pairs are still undergoing optimization. These markers are promising for investigating population genetic structure for this species. Besides, they revealed cross-species amplification when tested with *Aegla uruguayana*. Chimeric PCR products attained in this study, which correspond to 29% of the captured sequences, were object of a deep study about artifacts in microsatellite isolation. Moreover, it was proposed a mechanism to explain their formation and implications.

Key words: freshwater crab; genetic diversity; chimeric sequence

SUMÁRIO

1 INTRODUÇÃO	09
1.1 Revisão Bibliográfica	09
1.2 Objetivos	16
1.2.1 Objetivo Geral	16
1.2.1 Objetivos Específicos	17
2 ARTIGO 1	18
2.1 Abstract	20
2.2 Introduction	20
2.3 Material and Methods	22
2.4 Results and Discussion	23
2.5 References	34
3 ARTIGO 2	37
Abstract	39
Introduction	39
Material and Methods	41
Results	41
Discussion	42
The mechanism implicated in chimeric microsatellite loci	42
How to detect chimeric sequences?	45
Minimizing the PCR recombination	46
Frequency of recombinant formation	46
Concluding remarks	47
References	48
4 DISCUSSÃO	57
5 CONCLUSÃO	62
6 PERSPECTIVAS	63
7 REFERÊNCIAS BIBLIOGRÁFICAS	64
ANEXO A – Descrição da Espécie	70

1 INTRODUÇÃO

1.1 Revisão bibliográfica

Aeglidae é uma excelente família candidata a estudos moleculares por aspectos ecológicos, sistemáticos e biogeográficos exclusivos que instigam os pesquisadores e exigem mais do que as análises morfológicas podem inferir, principalmente pela falta de compartilhamento de caracteres derivados em relação ao grande número de espécies existentes (Pérez-Losada et al., 2002a). Compreende o único grupo ecológico de anomuros habitantes de água doce, em geral encontrados embaixo de pedras e folhas do substrato de rios e riachos, podendo ocorrer também em lagos e cavernas (Bond-Buckup e Buckup, 1994).

Sistematicamente, Aeglidae pertence à superfamília Galattheoidea, juntamente com as famílias Galatheidae, Chirostylidae, Porcellanidae e Kiwaidae (Bond-Buckup, 2003; Macpherson et al., 2005). Entretanto, características morfológicas relevantes destacam os eglídeos dos demais Galattheoidea: os pleópodos dos machos são vestigiais em aeglídeos e bem desenvolvidos nos demais; a eclosão dos ovos ocorre em estágio pós-larval enquanto que os ovos dos outros galateídeos eclodem em fase de zoea; brânquias do tipo tricobrânquias são exclusivas dos Aeglidae entre os Galattheoidea, ao passo que as outras famílias são filobranqueadas; ausência da linea anomurica em aeglídeos, sendo que a mesma é evidente em outros galateídeos (Martin e Abele, 1988). Além disso, Tudge e Scheltinga (2002), por comparação da morfologia espermática, e diversas reconstruções filogenéticas (Pérez-Losada et al. 2002b; Morrison et al., 2002; Ahyong e O'Meally, 2004; Porter et al., 2005; MacPherson et al., 2005;) sugerem que Aeglidae deve representar uma superfamília distinta entre os Anomura. Apesar dessas constatações, os autores continuam adotando apenas o taxa de família.

Ao contrário da posição filogenética contraditória do grupo, o contexto biogeográfico dos eglídeos é bem estabelecido. As espécies atuais são endêmicas às áreas temperada e subtropical do sul da América do Sul, incluindo o Brasil, Uruguai, Argentina, Paraguai, Bolívia e Chile. O município de Franca, no estado de São Paulo (~20°S) caracteriza o limite norte de ocorrência do gênero *Aegla* Leach,

1820; ao passo que ao sul, espécimes são encontrados até a Ilha de Madre Díos, Província de Última Esperanza (~50° S), Chile (Bond-Buckup e Buckup, 1994).

As duas espécies fósseis, descritas até o momento, são provenientes de sedimentos marinhos do Cretáceo, *Haumuriaegla glaessneri* Feldmann (1984) encontradas na Nova Zelândia e *Protoaegla minuscula* Feldmann et al. (1998), proveniente do sul do México.

A origem marinha dos eglídeos e o posterior estabelecimento e dispersão no habitat de água doce da América do Sul são amplamente aceitos e fundamentados nas evidências fósseis e análises filogenéticas (Martin e Abele, 1986; Pérez e Losada et al., 2002a e 2002b). Entretanto, a forma como a invasão ao continente e posterior dispersão teriam ocorrido, têm instigado diversos pesquisadores.

Padrões de dispersão (Ortmann, 1902), registros fósseis (Feldmann, 1986) e uma análise filogenética baseada em seqüências de DNA mitocondrial (Pérez-Losada et al., 2002a) atribuem origem pacífica (ocidental) aos eglídeos, com posterior dispersão para o oriente. Por outro lado, análise de padrões morfológicos e biométricos (Schmitt, 1942; Ringuélet, 1949) e o estudo panbiogeográfico realizado por Morrone e Lopretto (1994) priorizam a origem atlântica do grupo.

A fim de proporcionar uma hipótese filogenética robusta para o gênero *Aegla*, Pérez-Losada et al. (2004) utilizaram metodologias-padrão de filogenética heurística, considerando o maior número possível de espécies e subespécies de *Aegla* (58 das 63 descritas) além de seis espécies não descritas. As filogenias obtidas agruparam essas espécies em cinco clados com correlação geográfica. Os resultados dão suporte à origem pacífica do gênero e os autores destacam ainda que a distribuição geográfica atual não reflete o padrão de drenagem dos sistemas hídricos onde eles ocorrem, sugerindo que a radiação do grupo é mais antiga que as configurações atuais de escoamento dos rios.

Baseados no trabalho de Lundberg et al. (1998) que trata de uma extensa revisão da Paleodrenagem e Paleogeografia da América do Sul, Pérez-Losada et al. (2004) esclarecem que certas incongruências entre os clados de suas filogenias podem ser atribuídas a eventos geológicos como a formação da Serra do Mar e uma transgressão marinha do Atlântico (há cerca de 11 milhões de anos), que teriam proporcionado isolamentos em refúgios e posterior recolonização.

Esta hipótese justifica o fato de que a maioria das espécies de *Aegla* apresenta áreas de ocorrência restritas, habitando apenas uma bacia ou geralmente

bacias de drenagens adjacentes, como enfatizado por Bond-Buckup e Buckup (1994) e Morrone (1996), além da ampla diversificação do gênero no sul da América do Sul.

Atualmente o gênero *Aegla* é composto por 63 espécies e subespécies descritas (Bond-Buckup e Buckup, 1994; Bond-Buckup, 2003). No entanto, alguns trabalhos mais recentes apontam para a existência de novas espécies (Jara et al., 2003; Pérez-Losada et al., 2004) e o *status* de outras têm sido questionado (Pérez-Losada et al., 2002a).

Eventos geológicos contribuíram para a diversificação do gênero *Aegla* no passado. Entretanto, os eglídeos atualmente estão sujeitos a um outro tipo de isolamento, causado pela perda de habitat que restringe cada vez mais as populações naturais desse crustáceo às nascentes, o que caracteriza o grupo como exigente em qualidade de habitat e vulnerável a alterações ambientais (Bond-Buckup, 2003).

Segundo Jara (1996 apud PÉREZ-LOSADA et al., 2002c) e Bahamond et al. (1998), as 19 espécies e subespécies de *Aegla* que ocorrem no Chile têm sofrido redução progressiva em suas populações devido à degradação ambiental nos últimos 25 anos.

Pérez-Losada et al. (2002c) avaliaram o *status* de conservação para as espécies chilenas de *Aegla* com o objetivo de priorizar áreas de conservação. Os autores confirmaram a extinção de *Aegla conceptionensis* e *Aegla expansa*, enquadraram três espécies na categoria “criticamente em perigo” e classificaram seis espécies como vulneráveis. Em todos os casos, o grau de ameaça está relacionado às severas alterações ambientais e diminuição do nível de água nos rios, atribuídos à exploração florestal, urbanização, ocupação dos recursos hídricos para irrigação e conseqüente poluição dos mesmos com detritos orgânicos e compostos químicos provenientes da agricultura, como já vinha sendo demonstrado em diversos trabalhos (veja referências em Pérez-Losada et al. 2002c).

Não há estudos específicos sobre este tipo de impacto para populações de *Aegla* no Brasil, apesar da diversificada fauna de eglídeos descrita. Entretanto, alterações ambientais têm causado o desaparecimento de populações de *Aegla* em localidades onde havia registro (G. Bond-Buckup comunicação pessoal). Somente no estado do Rio Grande do Sul são encontradas 22 das 35 espécies de *Aegla* descritas para o Brasil (Bond-Buckup, 2003). Entretanto, quatro espécies de

eglídeos (*Aegla grisella*, *Aegla inermis*, *Aegla obstipa* e *Aegla violacea*) estão incluídas no Livro Vermelho da fauna ameaçada de extinção no estado na categoria de espécies vulneráveis (Bond-Buckup et al., 2003).

Os eglídeos são componentes fundamentais das cadeias tróficas, sendo considerados como elos nas mesmas, pois apresentam hábito alimentar onívoro, incluindo larvas de simulídeos hematófagos popularmente conhecidos como “borrachudos” (Magni e Py-Daniel, 1989), além de utilizarem amplamente o tecido vegetal disponível no ambiente (Bueno e Bond-Buckup, 2004; Castro-Souza e Bond-Buckup, 2004). Também servem como recurso alimentar para aves, rãs e peixes, especialmente para salmonídeos como *Salmo gairdnerii* e *Percichthys trucha* (Arenas, 1974).

Apesar da relevância dos eglídeos para os ecossistemas aquáticos e da riqueza de espécies do gênero (Bond-Buckup, 2003), apenas dois estudos de estrutura populacional utilizando marcadores moleculares foram dedicados a espécies de *Aegla* até o momento.

D’Amato e Corach (1997) publicaram o primeiro estudo de estrutura populacional de *Aegla* utilizando a técnica de RAPD (*Random Amplification Polimorphic DNA*). Quatro populações de *Aegla jujuyana* do noroeste da Argentina foram analisadas em um raio de 120 Km, sem a existência de barreira geográfica entre as mesmas. Os resultados permitiram detectar a ocorrência de fluxo gênico entre os indivíduos dos pontos analisados.

Recentemente, um estudo baseado na conformação do DNA heteroduplex foi dedicado a *Aegla longirostri* e caracterizou perfis eletroforéticos distintos para populações da bacia Uruguiaia (oeste) e Atlântica (leste), demonstrando o efeito da Serra do Pinhal (Figura 1) como barreira geográfica para a espécie (Santos et al., 2007).

O estudo de Pérez-Losada et al. (2004) destaca-se como o mais fundamentado dedicado à origem, biogeografia e relações filogenéticas de espécies de *Aegla*. Além de descrever um quadro geral sobre o gênero, expôs algumas incongruências a nível de espécie a serem esclarecidas. Os espécimes de *A. longirostri* empregados naquele estudo, por exemplo, foram coletados em diferentes rios da mesma rede hidrográfica (Rio Taquari-Guaiba) no estado do Rio Grande do Sul e representaram um grupo monofilético nas análises de Máxima Verossimilhança e Bayesiana. Entretanto, a filogenia gerada por Máxima Parcimônia revelou um

status parafilético à espécie. Somado a isso, os achados de Santos et al. (2007) para populações desta espécie na região central do estado (citado anteriormente), refletem a necessidade de um aprofundamento no conhecimento do grupo para melhor compreender sua história evolutiva.

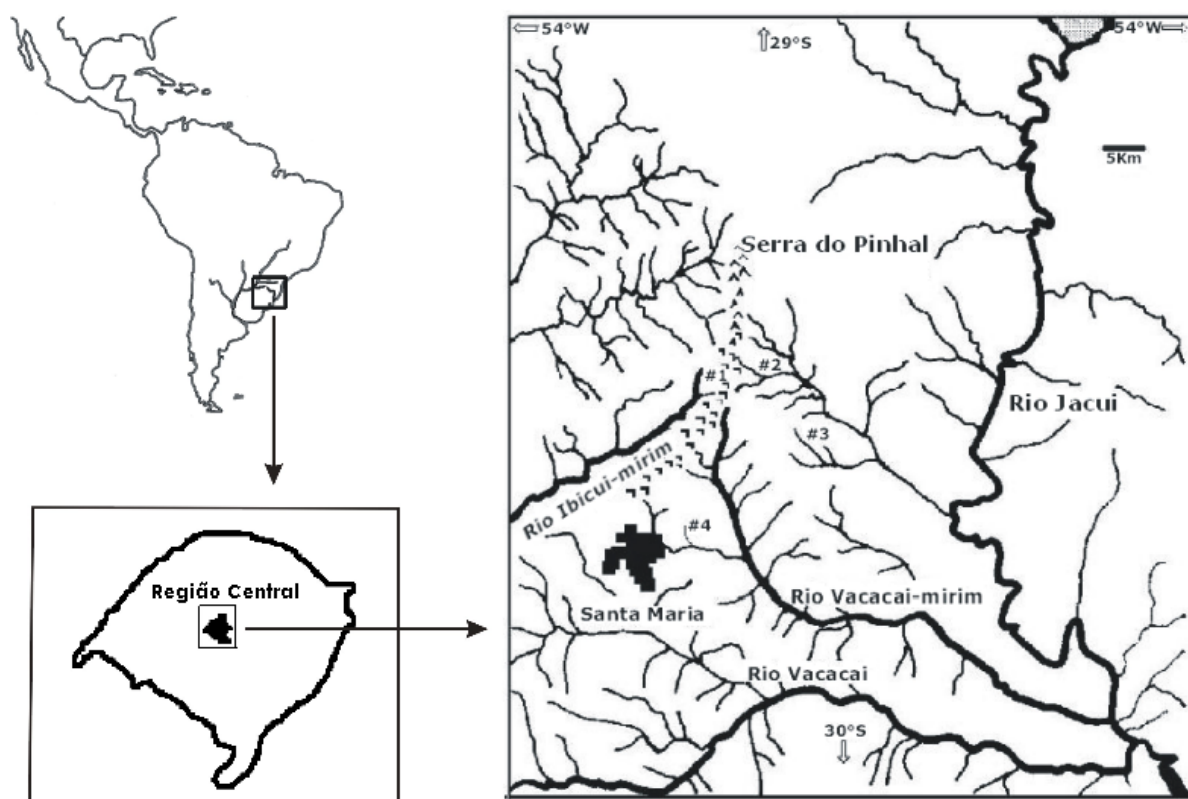


Figura 1 - Mapa da região central do estado do Rio Grande do Sul, tendo como referência o município de Santa Maria. O símbolo sustenido representa quatro pontos de coleta de *Aegla longirostri*, cujas populações foram analisadas no estudo de Santos et al. (2007).

Os microssatélites são as ferramentas indicadas para este tipo de estudo. Constituem-se de seqüências de DNA de 2 a 6 pares de bases repetidos em *tandem*, cuja principal característica é a variabilidade de alelos descrita para cada loco. Os inúmeros alelos em um determinado loco de microssatélite são diferenciados pela variação no número de unidades repetitivas presentes na

seqüência. Por exemplo, um loco de microsatélite pode ter sido inicialmente isolado a partir do genoma de um indivíduo e apresentar a repetição (CA)₁₀. O mesmo loco, analisado em outros indivíduos da mesma espécie, pode revelar alelos (CA)₈, (CA)₁₁, (CA)₁₅, (CA)₁₉ etc.

Repetições (CA) em uma fita do DNA subentendem a presença da repetição (GT) na outra fita. Portanto, os motivos CA ou AC, bem como GT ou TG são considerados equivalentes (Jin et al., 1994). Além de repetições simples, diferentes motivos repetitivos dispostos de forma adjacente constituem locos de microsatélites compostos, como por exemplo, CTCTCTCTAGTAGTAGT. Estes marcadores também são classificados quanto à integridade da repetição. Um microsatélite pode ser **Perfeito**, quando o mesmo motivo repetitivo está presente do início ao fim, sem interrupções (CACACACACA); **Imperfeito**, quando uma ou mais unidades repetitivas apresentam uma base que altera o padrão (CATCATCATCAGCATCAT); ou **Interrompidos**, quando há presença de alguns pares de base que diferem do padrão repetitivo (ATATAT**CAGG**ATATAT) (Goldstein e Schlötterer, 1999).

Microsatélites ocorrem, em grande parte, em regiões não-codificantes do genoma. Entretanto, funções biológicas atribuídas a estas seqüências vêm sendo evidenciadas em estudos que mostram locos de microsatélites associados a regiões promotoras de genes, atuando na expressão gênica como *enhancer* ou reguladores, além de estarem relacionadas a doenças genéticas (veja uma revisão em Kashi e Soller, 1999).

O mecanismo mais aceito para explicar a natureza hipervariável destas seqüências reside no deslizamento das fitas de DNA durante o processo de replicação – *Slip strand mispairing*. Fitas de DNA repetitivo dissociadas eventualmente poderiam reanelar de forma desalinhada, formando uma alça não pareada em uma delas. O pareamento errôneo proporciona perda ou ganho de unidades repetitivas, se a alça se formar na fita molde ou nascente, respectivamente. Passando despercebida pelo sistema de reparo celular, esta mutação passa a constituir um novo alelo a este loco de microsatélite (Eisen, 1999).

Além do seu caráter polimórfico, locos de microsatélite apresentam segregação segundo o padrão de herança mendeliana co-dominante. Dessa forma, são extremamente úteis para identificação individual, estudos populacionais e esclarecimento de questões biológicas relacionadas à taxonomia, reprodução, história evolutiva entre outros aspectos (Chambers e MacAvoy, 2000).

Estes marcadores genéticos são mais informativos que marcadores protéticos como isozimas (Estoup et al., 1998) ou dominantes como RAPD (Yan et al., 2005), consagrando-se como o método mais comumente utilizado para acessar estrutura genética populacional em fina escala.

Os microssatélites estão presentes nos mais variados tipos de organismos. Entre os Crustacea, camarões e caranguejos verdadeiros (*Brachyura*) são os grupos que mais possuem microssatélites descritos. Em termos de aplicação destes marcadores, destacam-se trabalhos envolvendo espécies de relevante papel ecológico como o *krill* (*Euphausiacea*), base da cadeia alimentar em ecossistemas polares (Ostellari et al., 2000); e espécies de interesse econômico, em geral camarões, priorizando o manejo e a exploração sustentável (Imgrund et al., 1997; Benzie, 2000; Xu et al., 2001; Robainas et al., 2002; Goyard et al., 2003). Além disso, Toonen (2004) e Bilodeau et al. (2005) utilizaram marcadores microssatélites para investigar a múltipla paternidade em ovos dos decápodos *Petrolisthes cinctipes* (*Anomura*) e *Callichirus islagrande* (*Thalassinidea*), respectivamente, e encontraram fêmeas multifecundadas.

Espécies de água doce em geral são pouco visadas em estudos de estrutura populacional e biologia da conservação, embora não sejam menos afetadas e ameaçadas por mudanças ambientais e climáticas (Abell, 2002). Embora poucos, a maioria dos estudos sobre estrutura populacional de organismos de água doce utilizando microssatélites, refere-se a peixes (Parker et al., 1999; Lucentini et al., 2006; Fessehayé et al., 2006). Este marcador tem se mostrado eficiente também para estudos populacionais de invertebrados de importância clínica, a maioria moluscos (Meunier et al., 2001; Gow et al., 2004) e econômica (Geist e Kuehn, 2005; Chand et al., 2005; Charoentawee et al., 2006;).

Em contrapartida, apesar da ampla aplicabilidade, os microssatélites não são facilmente isolados e, em geral, os locos descritos para uma espécie dificilmente são transferíveis a outras, a menos que as espécies sejam intimamente relacionadas (Scribner e Pearce, 2000 apud Zane et al., 2002).

Até o presente momento não há nenhum trabalho ou seqüência publicados com microssatélites para Aeglidae. O único *Anomura* que possui microssatélites descrito é o *Petrolisthes cinctipes* (Toonen, 2004) da família Porcellanidae. Entretanto, o fato das relações filogenéticas entre Aeglidae e Porcellanidae serem

contraditórias é um forte indício de que o marcador microssatélite descrito para uma família pouco provavelmente possa ser aplicado à outra.

Por outro lado, marcadores microssatélites desenvolvidos para uma espécie de *Aegla* têm grandes chances de apresentarem amplificação positiva para as demais espécies do gênero, pelo seu caráter significativamente monofilético descrito anteriormente. *Aegla longirostri* foi selecionada neste estudo para ser a primeira espécie de *Aegla* a ter microssatélites descritos, pela sua ocorrência estratégica na região central do estado do Rio Grande do Sul, onde o trabalho foi desenvolvido; pela sua distribuição na região, associada a diferentes bacias hidrográficas; e pelos aspectos intrigantes revelados pelo estudo de Santos et al. (2007) a diferentes populações, merecedor de aprofundamento.

Embora o processo de isolamento dos microssatélites seja trabalhoso, a obtenção das seqüências é garantia de resultados após serem aplicadas em análises populacionais.

1.2 Objetivo

1.2.1 Objetivo Geral

O objetivo deste trabalho foi o de isolar marcadores microssatélites a partir do genoma de *Aegla longirostri* (ver descrição da espécie no anexo 1), caracterizar os locos obtidos e selecionar seqüências que desempenhem um padrão polimórfico para que possam ser utilizadas em estudos futuros de estrutura populacional, entre outros, com esta e demais espécies do gênero *Aegla* para as quais estes marcadores amplificarem.

1.2.2 Objetivos Específicos

- Isolar locos de microssatélites de repetição (CA)_n a partir do genoma de *Aegla longirostri*, através da metodologia de bibliotecas genômicas enriquecidas com microssatélites por captura;
- Classificar os locos obtidos quanto a sua composição e número de repetição, selecionando os locos que se mostrarem apropriados para a projeção de pares de *primers*;
- Otimizar as condições de amplificação por PCR para os pares de *primers* estabelecidos;
- Investigar a possibilidade de utilização dos locos identificados em populações de outras espécies de *Aegla* (por exemplo, *Aegla uruguayana*);
- Avaliar a metodologia utilizada sob o ponto de vista das limitações da técnica, baseado nos artefatos obtidos neste estudo, bem como suas implicações na construção de bibliotecas genômicas enriquecidas com microssatélites.

Isolation and Characterization of microsatellite markers
in *Aegla longirostri* (Crustacea, Decapoda, Anomura).

**Isolation and characterization of microsatellite markers in *Aegla longirostri*
(Crustacea, Decapoda, Anomura)**

Paula Angélica Roratto¹, Darine Buchmann², Sandro Santos^{1,2}, Marlise Ladvoat
Bartholomei-Santos^{1,2*}

¹Curso de Pós-graduação em Biodiversidade Animal, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria, 97105-900 – Santa Maria, RS, Brazil

²Departamento de Biologia, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria, 97105-900 – Santa Maria, RS, Brazil

*Corresponding author:

marliselbs@gmail.com

Laboratório de Diversidade Animal

Departamento de Biologia, sala 3129

Universidade Federal de Santa Maria

Av Roraima 1000 Camobi

97105-900 Santa Maria, RS, Brazil

Phone: +55 55 32208339 (Ramal 31); Fax: +55 55 3220 8465 (Ramal 21)

2.1 Abstract

Genus *Aegla* is the only anomuran crab living in freshwater and the group is endemic to temperate South America. This study developed polymorphic microsatellite markers for *A. longirostri*, inhabitant of hydrographic basins from central region of Rio Grande do Sul state, Brazil. The microsatellite enriched library yielded 13 appropriated sequences for primer design. For two of the 9 loci tested up to now the ideal conditions for PCR amplification were determined and allelic variability was found in a 15-individuals sample of *A. longirostri*. Besides, positive amplification was attained when applying these primer pairs for an *A. uruguayana* sample as template. Allele numbers and observed heterozygosities will be determined for these loci and other else which present reproducible amplification and polymorphism. These markers will be useful for investigating population genetic structure of this species and other *Aegla* sp with successful cross-amplification.

Keywords: freshwater crab, SSR (Simple Sequence Repeats), molecular marker.

2.2 Introduction

Family Aegliidae consists of the only freshwater anomuran crabs nowadays. The single existing genus *Aegla* had its origin from marine environment (Feldmann, 1984; Feldmann et al., 1998), invaded continental water and showed a wide dispersion and diversification throughout temperate South America, having 63 recognized species (Bond-Buckup and Buckup, 1994). Distribution and endemism are well reported for aeglids, as well as the progressive and worrying population declining as result of anthropogenic effects during the last decades (Bond-Buckup and Buckup, 1994; Pérez-Losada et al., 2002a).

These crustaceans can be found in streams, living under rocks or fallen leaves, and are extremely strict to unspoiled habitats (Bond-Buckup and Buckup, 1994). Despite their ecological significance, little is known about their genetic population structure.

The first study investigating the genetic structure of aeglids used RAPD (Random Amplified Polymorphic DNA) to assay four Argentinean populations of *Aegla jujuyana* within a radius of 120 Km. Results showed slight differences, especially for the most distant populations, and significant gene flow among them (D'Amato and Corach, 1997a).

Recently, the polymorphism of a pentanucleotide microsatellite of the U1 snRNA gene complex was assessed through heteroduplex DNA assay in four populations of *A. longirostri*, isolated by a geographic barrier. It was demonstrated significant differentiation among populations of opposite sides of the barrier (Santos et al., 2007).

The fact of no morphological differences occur among several populations of *A. longirostri* distributed in different geographic basins, in despite of some of them are isolated for millions of years, makes this species a very instigating group for studying its genetic structure (Santos et al., 2007). The use of a powerful molecular marker will be useful to detect intraspecific genetic structure in these species.

Microsatellite markers, because of their high polymorphism, have proved to be the fittest molecular tool for establishing genetic structure, dynamic and relationships within and among populations. They have been widely used in many species nowadays, but they were described only for one marine anomuran species, *Petrolisthes cinctipes* (Toonen, 2004). For crustaceans in general, most works are dedicated to true crabs – Brachyura (e.g. Gopurenko et al., 2002; Hänfling and Weetman, 2003; Jensen and Bentzen, 2004; Steven et al., 2005; Tepolt et al., 2006) – or shrimps from super-family Penaeoidea (e.g. Pongsomboon et al., 2000; Robainas et al., 2002; Maggioni and Rogers, 2002; Meehan et al., 2003; Wang et al., 2005) and other species of economic importance.

In this work we constructed a microsatellite-enriched genomic library for *A. longirostri* and characterized the yielded loci to be evaluated in studies of population genetics.

2.3 Material and Methods

Live *A. longirostri* specimens (n = 15) were collected by hand from a population of a Vacacai-Mirim River tributary. The genomic DNA was extracted with phenol-chloroform according to Bitencourt et al. (2006).

Microsatellites were developed following Refseth et al. (1997). DNA of one sample was digested with Taq I, fragmented DNA was ligated to an adapter formed by the annealing of two oligonucleotides (Taq 20 Mer: 5'-ATGAAGCCTTGGTACTGGAT-3' and TaqI 22 Mer: 5'-CGATCCAGTACCAAGGCTTCAT 3') and then hybridized with a biotinylated probe [(CA)₈]. DNA fragments hybridized with CA-probe were selectively separated using streptavidin-coated beads (Dynabeads® M-280 Streptavidin, Dynal) in the presence of a magnetic field. Captured DNA was then eluted from the beads, amplified using the oligonucleotide TaqI 20 Mer as primer and finally cloned into a pCR® 2.1 vector (TA Cloning kit, Invitrogen). See Figure 1 for a schematic representation.

Positive clones were screened by amplification using M13 primers and sequenced using the DYEnamic ET Dye Terminator Cycle Sequencing kit for MegaBACE (Amersham Bioscience). Sequences were analyzed and aligned using Staden Package software (Staden et al., 1998). PCR primers were designed for sequences containing more than ten repeats and enough flanking regions, using primarily the program OLIGO (Rychlin and Rhoads, 1989) and then the program PRIMER3 (Rozen and Skaletsky, 2000).

Primer pairs that generated the PCR products with the expected molecular weight were tested in 15 *A. longirostri* samples. PCR amplifications were conducted in a 25 µl reaction volume containing ~40 ng of template DNA, 100 mM Tris-HCl, 50mM KCl, 0.1 mM of each dNTP, 1 U Taq DNA polymerase (Cenbiot), MgCl₂ and primers as specified on Table 4, with the following cycling program: initial denaturation at 94°C for 5 min followed by 30 cycles of 1min denaturation at 94°C, 45s primer-specific annealing temperature (Table 4) and 45 s extension at 72°C, besides a final extension step at 72°C for 5 min. The amplification products were separated by electrophoresis in 6% polyacrylamide gels and visualized by traditional silver staining.

2.4 Results and discussion

The first attempt to construct a microsatellite-enriched genomic library for *A. longirostri* yielded 62 positive clones. Thirty two were sequenced and, except three sequences without repeat, all the others presented inserts lacking a flanking region and ending with the repeat (CA)_n, as a result of internal priming of the biotinylated probe during the PCR. Despite the impossibility to characterize them as molecular markers, these isolated loci showed repeats ranging from nine to 38 units, some were compounds, and the microsatellite motif (GA)_n was also found (Table 1). Some strategies were adopted to avoid carrying oligonucleotide probes with the recovered DNA, such as reducing the probe concentration from 0.4 μM to 0.3 μM and the elution temperature from 90°C to 80°C, besides an additional elution step.

The second microsatellite library attained 68 positive clones. A total of 61 clones was sequenced and approximately 90% of them contained microsatellite with more than 10 repeats. Seventeen clones were chimeric PCR products (Roratto et al., 2007) and 15 did not contain sufficient flanking regions for primer design. Besides, it was not obtained a consensus sequence for nine clones (Table 3).

Primer pairs could be designed for 13 microsatellite loci (Table 4). For loci AICA105, ALCA119 and AICA128, whose primers were designed by program OLIGO, exhaustive attempts to optimize PCR conditions did not yield the expected amplification product. A possible reason for unsuccessful amplifications could be ascribed to the high rate of chimeric clones attained. These loci could represent PCR artifacts not detected in sequence alignment by the absence of related shuffled clones (see Roratto et al., 2007).

The choice of primer design program was also considered as a cause of unfruitful primer pairs. The second batch of six primer pairs was designed by PRIMER3. Ideal amplification conditions were quickly well established for loci AICA135 and AIGA138 up to now.

Preliminary analysis of these loci in polyacrylamide gel has demonstrated polymorphism in a sample of 15 *A. longirostri* individuals. The electrophoresis conditions have been optimized to evaluate allele diversity. Locus AIGA138 has a (GA)₁₇ repeat, that make easy the interpretation of polyacrilamide gel bands for allele identification. However, locus AICA135 represent a complex sequence with more

than one repeat, which could represent allele homoplasy in such analysis. Although this event should harm polymorphism analysis, complex sequences as locus AICA135 are useful for microsatellite evolution assays.

The other loci AICA108, AICA140, AICA145 and AICA159, did not present reproducible PCR products at the moment. Despite two primer pairs developed from PRIMER3 have worked, it is not possible to affirm that the applied program is the justification for the unsuccessful primers.

Recently, additional sequencing reactions allowed designation of primer pairs for loci AICA99, AIGA120, AICA131 and AICA136, which will be evaluated as well.

Perfect isolated microsatellite loci (CA)_n ranged from 3 repetitive unit (AICA162) to more than forty (e. g. AICA94, AICA128 and AICA158). It was highlighted the occurrence of GA/TC motif associated with (CA) repeats, as much in compound loci as in imperfects ones (see Table 2).

Successful cross-amplification was attained by testing loci AICA135 and AIGA138 in a sample of *A. uruguayana*. PCR products showed approximately the expected size, were purified and sequenced. It was confirmed the homology between sequences from both species through alignment using program ClustalW 1.83 (Chena et al., 2003). Locus AICA135 was characterized as (AC)₁₃...(AC)₅GC(AC)₃ in *A. longirostri*, showed the same repeats and interruptions when amplified from the *A. uruguayana* genome, but for the later species both (AC) repeats are longer (Figure 2); as it was for the locus AIGA138 which showed one more (GA/TC) repeat unit when amplified from *A. uruguayana* (Figure 3).

The consensus sequence attained for locus AICA135 (from *A. uruguayana* genome) was not of good quality. Even so, similarity values were of 80% and 95% for loci AICA135 and AIGA138, respectively; suggesting that microsatellite markers developed from one *Aegla* species successfully should be applied for other.

To assess the potential of developed microsatellite loci as molecular genetic markers, they will be tested for conformance to expected Hardy-Weinberg equilibrium frequencies using exact tests implemented in GENEPOP (Raymond and Rousset, 1995), as microsatellite markers represent an excellent molecular tool for population structure, biogeography and other instigated and unclear aspects of aeglids.

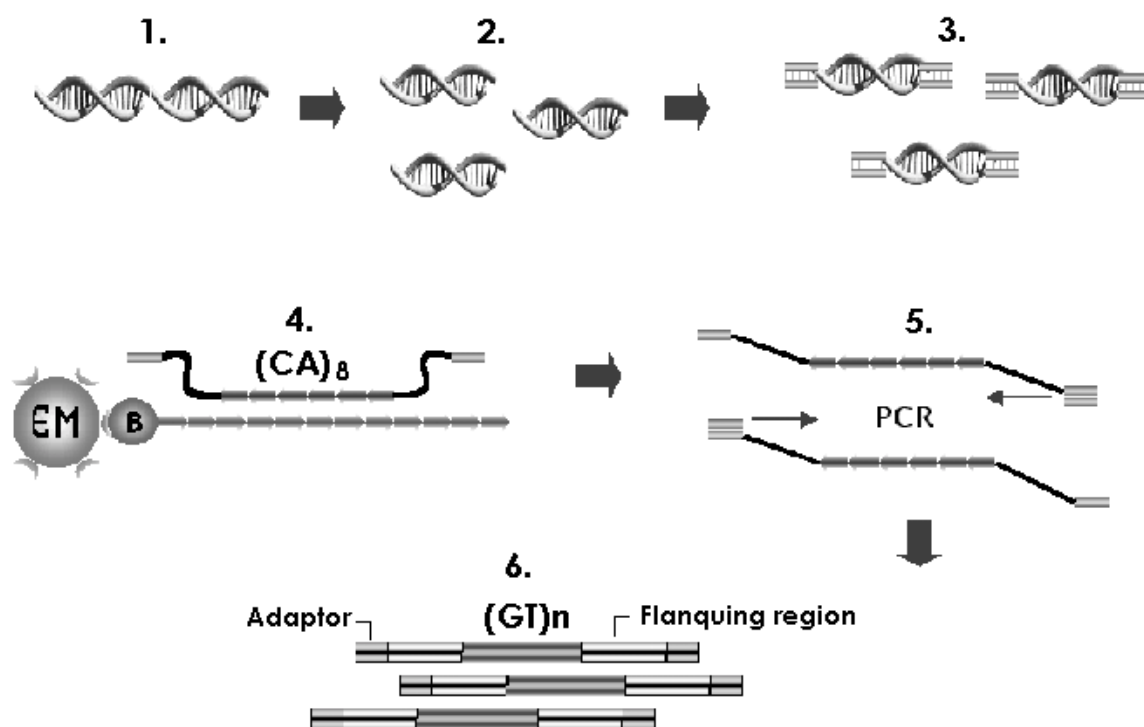


Figure 1 – Schematic representation of the selective hybridization protocol for microsatellite development, employed in these study: 1. Sample of *A. longirostri* genomic DNA; 2. Digestion with restriction enzyme; 3. Ligation of adaptor sequences; 4. Selective hybridization of DNA fragments with the probe $(CA)_8$ and magnetic separation; 5. Amplification of captured single strands by PCR; 6. Cloning and sequencing.


```

urug138      GATTATTAGCACCCTGCCCACACACCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 60
long138      GAT--TTAGCACCCTGCCCACACACCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 58
***          *****

urug138      CCCACACACCATTACCCCCACCTGTGCCCTCATGGATGTCTCAGTGAGGCTGTGATTATC 120
long138      CCCACACACCATTACCCCCACCTGTGCCCTCATGGATGTCTCAGTGAGGCTGTGATTATC 118
*****

urug138      TGGTCTACCCTCGCCT-TCTTTCACCTCCTCTCCCTCTCTTTTTTGCCACATCGTTCTCTCT 179
long138      TGGTCTACCCTCTCCTCTCTTTCACCTCCTCTCCCTCTCTTTTTTGCCACATCGTTCTCTCT 178
*****

urug138      CTCACTTTCTCTCTCCATATGAATCTTCTTATAACAATTCTTCATCTTATATCACCATTTC 239
long138      CTCACTTTCTCTCTCCATATGAATCTTCTTATAACAATTCTTCATCTTAT-TCACCATTTC 237
*****

urug138      GTTC----- 243
long138      GTTCACTCAGCTTTTCTATCTATGTCTGCCTTCCCTCTTTCTCTCTCATGTTTTATATTCT 297
****

urug138      -----
long138      TATTGCAATACTCCTTTCACCTCCTCAAAATTCACATTTTATGTTGCACTTGTTATTTTA 357

urug138      -----
long138      CCTTTGCGTTATTTCTCTTTCTTCATTTTGTCTCCTGCCTTTTCTCTTTTTTGATACATCCC 417

urug138      -----
long138      AGCATCTTCTGTAAGCAATTTTACTCTTTCTCTCTTTGTTTCTCTCTCTATCTATCT 477

urug138      -----
long138      ATCTATCTATCTATCTATCTATCTCTCTTTCTAACACTTATCCCTGGGCAAGTGAATCTC 537

urug138      -----
long138      TGGCGTCTGACTCCCTTGCGTCGCATCCCTGAGACCAACTTAATACCGGATAATGACCCC 597

urug138      -----
long138      TGCAGAGACACACAGAAATGGTTCG 620

```

Figure 3 - CLUSTAL W (1.83) multiple sequence alignment. urug138 – PCR product sequence of locus AIGA138 amplified from *Aegla uruguayana* sample; long138 – AIGA138 clone sequence of the *Aegla longirostri* microsatellite-enriched genomic library. Underlined sequences represent primer sites. The red sequence corresponds to the microsatellite.

Table 1 – First attempt to construct a microsatellite-enriched genomic library for *Aegla longirostri*. All isolated sequences lacked a flanking region as a result of internal priming of the biotinylated oligonucleotide probe (CA)₈. Real loci correspondent to these captured sequences may contain a longer repeat than those ones, which represent the PCR product cloned.

Classification	Locus	Repeat	Classification	Locus	Repeat
Perfect	AICA02	(CA) ₁₄	Perfect	AICA54	(CA) ₃₀
	AICA04	(CA) ₉		AICA57	(CA) ₆
	AICA10	(CA) ₁₁	Compound	AICA43	(TC) ₈ (AC) ₃₈
	AICA12	(AC) ₉		AICA27	(TC) ₉ AAA(AC) ₁₁
	AICA13	(CA) ₆		AICA29	(CA) ₁₁ AA(TC) ₁₁
	AICA16	(CA) ₁₆		AICA17	(AC) ₄ AA(AC) ₄
	AICA19	(GA) ₉		AICA24	(CA) ₄ AA(CA) ₂ G(AC) ₂ G (CA) ₄
	AICA25	(AC) ₁₂	Imperfect	AICA39	(AC) ₃₇ T(CA) ₇ ...(CT) ₉
	AICA31	(CA) ₁₄		AICA58	(AC) ₄ ...(AC) ₂ AA (AC) ₄
	AICA32	(CA) ₃₃	Complex	AICA09	(CA) ₃ AAGAA(AG) ₁₀ (GAAA) ₃
	AICA37	(CA) ₉	Without repeat	AICA01	-
	AICA46	(GA) ₇		AICA41	-
	AICA50	(CA) ₈		AICA45	-

Table 2 - Microsatellite-enriched genomic library attained for *Aegla longirostri*. Forty-two characterized loci, according to their classification, the captured and extra repeats and situation: UCS (Undefined Consensus sequence); IFR (Insufficient Flanking Region); SRR (Short Repetitive Run); OK (loci with a primer pair defined). The seventeen chimeric sequences cloned were not considered.

Classification	Locus	Repeat	Situation	Extra repeats
Perfect	AICA94	(CA) _{>40}	UCS	-
	AICA97	(CA) _{>30}	UCS	-
	AICA99	(AC) ₁₁ ...(CA) ₁₀	OK	-
	AICA117	(AC) ₃₆	IFR	-
	AICA122	(CA) _n	IFR	-
	AICA127	(CA) ₃₇	IFR	-
	AICA128	(CA) ₄₃	OK	-
	AICA129	(CA) ₂₉	IFR	-
	AICA131	(CA) ₂₁	OK	-
	AICA138 ^a	(GA) ₁₇	OK	(GA) ₂ (GATA) ₈ (GA) ₅
	AICA139	(CA) ₃₃	IFR	-
	AICA140	(CA) ₁₄	OK	-
	AICA142	(AC) ₁₁	UCS	(AC) _{>10} (ACGC) ₁₅ (AC) ₆ AG (AC) ₁₅
	AICA145	(CA) ₁₃	OK	-
	AICA147	(CA) ₄₄	IFR	-
	AICA152	(CA) _n	IFR	-
	AICA153	(CA) ₆ ... (CA) ₇	UCS	(TC) ₅ (AC) ₇ and (CA) ₆ CC(CA) ₄
	AICA158	(CA) _{>40}	UCS	(CACG) _{>9} (CA) ₁₂ (TACA) ₉
	AICA159	(CA) ₁₅	OK	-
	AICA162	(CA) ₃	SSR	-
Compound	AICA110	(TC) ₄ (AC) ₄ (CT) ₂ (CA) ₁₄	IFR	-
	AICA119	(CA) ₆ CT(CA) ₂₇	OK	-
	AICA134	(TC) ₈ TT(TC) ₃₃ (TG) _{>40}	IFR	-

Table 2, continuation

	AICA136	(CT) ₆ (CA) ₂₈	OK	-
	AICA137	(CA) ₇ (CG) ₂ (CA) _n	IFR	-
	AICA141	(CT) ₂ (CA) ₈	IFR	-
	AICA144	(AC) _{>30} (TGC) ₄ (AC) ₁₄	IFR	-
Imperfect	AICA120	(TC) ₃ (AC) ₆ (TC) ₂₃	OK	(TA) ₅ and (TC) ₅ AC(TC) ₂
	AICA130	(AC) _n GC(AC) ₃	UCS	-
	AICA92	(CA) ₄ CG(CA) ₂ AG(CA) ₄ C G (CA) ₃ GA(CA) _n	IFR	-
	AICA105	(AC) ₁₁ TC(AC) ₁₀ GC(AC) ₁₃	OK	-
	AICA108	(AC) ₁₆ CT(CA) ₂₆ CG(CA) ₁₉	OK	-
	AICA114	(AC) _{7...} (AC) ₂ TC(AC) ₁₅ CC(AC) ₅	SRR	(CT) ₆ and (AG) ₄
Complex	AICA135	(AC) _{13...} (CT) _{6...} (AC) ₅ CG(AC) ₃	OK	(AG) ₆ (AC) ₂ ; (CAA) ₈ ; (AG) ₆ and (CA) ₄ CG(CA) ₄
	AICA143	(TC) ₃ TG(TC) _{4...} (CACG) ₄ CA(CACT) ₂ CG (CA) ₃	UCS	-
	AICA149	(AC) ₆ (GC) ₄ (AC) ₂₆ (GC) ₅ (AC) ₁₄ (GC) ₄ (AC) ₂₅ (GC) ₄ (AC) ₁₁	IFR	-
	AICA154	(GC) ₂ (CA) ₄ CG(CA) _{2...} (CA) ₅ (GC) ₂ (CA) ₄	UCS	-
	AICA157	(CA) _{>50} CT(CA) ₄ CG (CA) ₃	IFR	-
	AICA167	(AC) _{>28} AG(AC) ₃ AG (AC) ₄ (AG)	UCS	-
		AICA101 ^a	-	-
-	AICA163 ^a	-	-	(ACC) ₅
	AICA126 ^a	-	UCS	-

^a Without (CA) repeatⁿ Undefined number of repeats due interrupted or low quality sequencing reading.

Table 3 - Summary of the second dinucleotide (CA) microsatellite isolation for the freshwater crab *Aegla longirostri*.

Total number of sequenced clones	61
Redundancy	2
Total number of clones with microsatellite inserts	56 (94.9%)
Microsatellites with more than 10 repeats	54 (91,5%)
Doubtful clones ^a	17 (28.8%)
Loci without sufficient flanking region	15 (25.4%)
Loci without a consensus sequence	9 (15,25%)
Designed primer pairs	13 (22,05%)
Successfully optimized PCR conditions ^b	2 (15,4%)
Polymorphic loci	-

^aChimeric PCR products, showing shuffled flanking regions (see Roratto et al., 2007).

^bNot all primer pairs were evaluated at the moment.

Table 4 - Description of 13 microsatellite primer pairs developed for freshwater crab *Aegla longirostri*, the repeat motif, primers sequence, PCR conditions (when determined) and stage of characterization. T_m, temperature melting; MgCl₂, Magnesium Chloride concentration.

locus	Repeat motif	Primer sequence (5'-3')	PCR conditions		Characterization
			T _m (°C)	MgCl ₂	
AICA99	(AC) ₁₀ ...(CA) ₈	F-TCCGTGAGGGTGTAAAGGTGA	59.58	-	Ordered primer pair
		R-TCCCTTCGTGTAAGTTTCTGTCTTG	59.86		
AICA105	(AC) ₁₁ T(CA) ₁₁ CG(CA) ₁₄	F-ATAGGCACACACTCACAAAGC	57.3	-	Not defined
		R-CGACGTTTCCGAGTTACCTA			
AICA108	(AC) ₁₆ ...(AC) ₂₆ GC(AC) ₁₉	F-GAAAGAGATGCTATGGCGTG	57.3	-	Testing amplification conditions
		R-GGACACCAGGGCAACATAAG	59.4		
AICA119	(CA) ₆ CT(CA) ₂₇	F-GACTCACATCGGGTTCACAT	57.3	-	Not defined
		R-TGGCAATAACAGCAGTCAAG	55.3		
AICA120	(TC) ₃ (AC) ₆ (TC) ₂₃	F-ATCACCCCTCCCTTCTTTTT	58.9	-	Ordered primer pair
		R-GACCAATTTGCTTTTCAGTGGTG	59.5		
AICA128	(CA) ₄₃	F-TTTACTGGTGATCATTCTGTGG	56.5	-	Not defined
		R-GTAGCACCAATTTCTTTGTTTC	54.7		
AICA131	(CA) ₂₁	F-CATGCAATTTCCGAAAGAAAA	60.06	-	Ordered primer pair
		R-CTATTTCTGCACCCGGTGTT	59.99		
AICA135	(AC) ₁₃	F-CATCTCGCTCACGTGCTTTA	60.0 ^a	2,0M	In progress
		R-TATGAGGGAATGGGGGTGTA			
AICA136	(CT) ₆ (CA) ₂₈	F-CGAACTGGTCTAGCAGAATG	59.88	-	Ordered primer pair
		R-GACCTCACGTGCTGACTG	59.44		
AIGA138	(GA) ₁₇	F-GATTTAGCACCACTGCCACA	TD	1,5M	In progress
		R-TGAGTGAACGAATGGGTGAA	63/58 ^{a,b}		

Table 4, continuation

AICA140	(CA) ¹⁴	F-CAGGCACCTTTAAAGCCAAG R-CCCACACAAAAGCTTCACTCA	57.3	-	Testing amplification conditions
AICA145	(CA) ¹³	F-CAAGGGACATACATGCCCTCA R-AGGGTGAGGTTGCCGTATTTG	57.3	-	Testing amplification conditions
AICA159	(CA) ¹⁵	F-CCTGTCTTGACCCAATGCTT R-TCAAACAGGAAGAGGGGGCTA	57.3	-	Testing amplification conditions

^a Annealing temperature

^b Touch Down 63°C to 58°C, -0,5°C per cycle

2.5 References

BITENCOURT, J. V. T. et al. Comparison of different methodologies for DNA extraction from *Aegla longirostri* (Crustacea, Decapoda, Anomura). **Brazilian Archives of Biology and Technology**. *In press*, 2006.

BOND-BUCKUP G.; BUCKUP L. A família Aeglidae (Crustacea, Decapoda, Anomura). **Arquivos de Zoologia**. v. 32, p. 159-347, 1994.

CHENA, R. et al. Multiple sequence alignment with the Clustal series of programs. **Nucleic Acids Research**. v. 31, n. 13, p. 3497-5000, 2003.

D'AMATO, M. E.; CORACH, D. Population genetic structure in the fresh water anomura *Aegla jujuyana* by RAPD analysis. **Journal of Crustacean Biology**. v. 17, n. 2, p. 269-274, 1997a.

D'AMATO, M. E.; CORACH, D. Highly repetitive DNA sequences unique to Aeglidae (Anomura). **Journal of Crustacean Biology**. v. 17, n.1, p. 184-191, 1997b.

FELDMANN, R. M. *Haumuriaegla glaessneri* n.gen. and sp. (Decapoda: Anomura: Aeglidae) from Haumurian (Late Cretaceous) rocks near Cheviot, New Zealand. **New Zealand Journal of Geology and Geophysics**. v. 27, p. 379-385, 1984.

FELDMANN, R. M. et al. Early Cretaceous Arthropods from the Tlayúa Formation at Tepexi de Rodríguez, Puebla, México. **Journal of Paleontology**. v. 72, n. 1, p. 79-90, 1998.

GOPURENKO, D.; HUGLES, J. M.; MA, J. Identification of polymorphic microsatellite loci in the mud crab *Scylla serrata* (Brachyura: Portunidae). **Molecular Ecology Notes**. v. 2, p. 481-483, 2002.

HÄNFLING, B.; WEETMAN, D. Characterization of microsatellite loci for the Chinese mitten crab, *Eriocheir sinensis*. **Molecular Ecology Notes**. v. 3, p. 15-17, 2003.

JENSEN, P. C.; BENTZEN, P. Isolation and inheritance of microsatellite loci in the Dungeness crab (Brachyura: Cancridae: *Cancer magister*). **Genome**. v. 47, n. 2, p. 325-331, 2004.

MAGGIONI, R.; ROGERS A. D. Microsatellite primers for three Western Atlantic *Farfantepenaeus* prawn species. **Molecular Ecology Notes**. v. 2, p. 51-53, 2002.

MEEHAN, D.; XU, Z.; ZUNIGA, G.; ALCIVAR-WARREN, A. High Frequency and Large Number of Polymorphic Microsatellites in Cultured Shrimp, *Penaeus (Litopenaeus) vannamei* [Crustacea:Decapoda]. **Marine Biotechnology**. v. 5, p. 311-330, 2003.

PÉREZ-LOSADA, M. et al. Phylogenetic relationships among the species of *Aegla* (Anomura: Aeglididae) fresh-water crabs from Chile. **Journal of Crustacean Biology**. v. 22, n. 2, p. 304-313, 2002a.

PÉREZ-LOSADA, M. et al. Phylogenetic position of the freshwater anomuran family Aeglididae. **Journal of Crustacean Biology**. v. 22, n. 3, p. 670-676, 2002b.

PONGSOMBOON S. et al. Characterization of Tri- and Tetranucleotide Microsatellites in the Black Tiger Prawn, *Penaeus monodon*. **Science Asia**. v. 26, p. 1-8, 2000.

RAYMOND, M.; ROUSSET, F. Genepop (version 1.2): population genetic software for exact tests and acumenecism. **Journal of Heredity**. v. 86, p. 248-249, 1995.

REFSETH, U. H.; FANGAN, B. M.; JAKOBSEN, K. S. Hybridization capture of microsatellites directly from genomic DNA. **Electrophoresis**. v. 18, p. 1519-1523, 1997.

ROBAINAS, A. et al. Microsatellite loci from the pink shrimp *Farfantenaesus notialis* (Crustacea, Decapoda). **Molecular Ecology Notes**. v. 2, n. 3, p. 344-345, 2002.

RORATTO, P. A. et al. PCR-mediated recombination in microsatellite markers development: mechanism and implications. (Submetido a *Molecular Ecology Notes*).

ROZEN, S.; SKALETSKY, H. Primer 3 on the WWW for general users and for biologist programmers. In: KRAVETZ, S.; MISENDER, S. (Org.). **Bioinformatics methods and protocols: methods in Molecular Biology**. Human Press, Totowa, NJ, 2000. p. 365-386.

RYCHLIN, W.; RHOADS, R. E. A computer program for choosing optimal oligonucleotides for filter hibridization, sequencing and in vitro amplifications of DNA. **Nucleic Acids Research**. v. 17, p. 8543-8551, 1989.

SANTOS, S. et al. 2007. Genetic Differentiation of *Aegla longirostri* populations isolated by a geographical barrier. (Submetido ao *Journal of Crustacean Biology*).

STADEN R.; BEAL, K.F.; BONFIELD, J.K. The Staden Package, 1998. **Methods in Molecular Biology**. v. 132, p.115-130, 1998.

STEVEN, C.R. et al. Genetic markers in blue crabs (*Callinectes sapidus*) I: Isolation and characterization of microsatellite markers. **Journal of Experimental Marine Biology and Ecology**. v. 319, p. 3-14, 2005.

TEPOLT, C.K. et al. Characterization of microsatellite loci in the European green crab (*Carcinus maenas*). **Molecular Ecology Notes**. v. 6, p. 343–345, 2006.

TOONEN, R. J. Genetic evidence of multiple paternity of broods in the intertidal crab *Petrolisthes cinctipes*. **Marine Ecology – Progressive Series**. v. 270, p. 259-263, 2004.

WANG, H.; LI, F.; XIANG, J. Polymorphic EST–SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. **Aquaculture**. v. 249, p. 107-114, 2005.

PCR-mediated recombination in microsatellite markers development: mechanism and implications¹.

¹ Artigo formatado conforme as normas da Revista Molecular Genetics and Genomics.

PCR-mediated recombination in microsatellite markers development: mechanism and implications.

Paula A. Roratto, Darine Buchmann, Sandro Santos and Marlise L. Bartholomei-Santos.

Laboratório de Diversidade Genética, Curso de Pós-graduação em Biodiversidade Animal, Prédio 16, sala 3129, Universidade Federal de Santa Maria, Av. Roraima 1000, Camobi, CEP 97105-900, Santa Maria, RS, Brazil.

Correspondence author:

Marlise Ladvocat Bartholomei-Santos (marlise@smail.ufsm.br)

Universidade Federal de Santa Maria, Curso de Pós-graduação em Biodiversidade Animal, CEP 97.105-900 Santa Maria, RS, Brazil. Tel: +55 3220 8339. Fax: +55 3220 8628.

Note: Nucleotide sequence data reported are available in the GenBank database under the accession numbers EF025136-EF025166.

Abstract Protocols of microsatellite enrichment library are widely applied for several species in order to supply the most informative molecular markers for population and inbreeding studies, among other assays. A drawback is the ratio of designed primer pairs that do not amplify the expected fragment, even after exhaustive optimization attempts. A possible cause of unsuccessful microsatellite primers can be that such loci are artifacts resulting from chimeric PCR products instead of real genomic sequences. The microsatellite-enriched library constructed for *Aegla longirostri* (Crustacea, Decapoda, Anomura) showed that 29% of sequenced clones are chimeric products, as these sequences shared one of the flanking regions around the same repeat motif, but not the other. PCR-mediated recombination is a well-known event described for several procedures at which related sequences are used as template. We have associated this phenomenon with microsatellite markers development. This study expounds the high ratio of recombinant sequences generated in *A. longirostri* microsatellite enriched library and discuss the mechanism and implications of PCR chimeric products formation during microsatellite isolation.

Key words Microsatellite isolation; Chimeric PCR product; Recombination; SSR; Unfruitful amplification

Introduction

Microsatellites or Simple Sequence Repeats (SSR) are genetic markers widely used in individual identification and population level analysis due their high power of genetic resolution (Chambers and Macavoy 2000). Several techniques for microsatellite isolation have been developed, specially the magnetic bead protocol, which basically relays in the hybridization of digested DNA with biotinylated oligonucleotide probes and subsequent separation using streptavidin-coated magnetic particles.

Despite the modern technology, results are not always satisfactory. The percentage of designed primer pairs that are successfully optimized is variable and generally low (Table 1). However, authors usually focus on amplifiable loci and do not discuss the possible factors behind loci that did not work, as publications about microsatellite isolation generally are limited to notes. Loci that failed to amplify might be attributed to artifacts of the PCR-based isolation process.

(Insert Table 1 after the paragraph above)

Chimeric PCR products can arise when related sequences are present in a reaction as templates. An incompletely extended primer in the elongation phase of the PCR cycle

generates a shorter nascent strand that, in a subsequent cycle, can prime off a heterologous target sequence and be completely extended. The outcoming chimeric sequence has its 5' end corresponding to the first template and its 3' end to the heterologous template.

Formation of hybrid sequences by PCR has been related for attempts to amplify genes belonging to multiple families (Bhavsar et al. 1994), characterization of alleles from a heterozygous subject (Bradley and Hillis 1997), forensic applications in which the template is generally ancient DNA (Pääbo et al. 1990), reverse transcription (Brakenhoff et al. 1991), in PCR-derived clones from polyploid genomes (Cronn et al. 2002), and for environmental DNA samples (von Wintzingerode et al. 1997). Even false sequences resulting from chimeric PCR have been deposited in public databases (Hugenholz and Huber 2003; Ashelford et al. 2005). On the other hand, chimeric PCR product is a powerful tool widely used to create recombinant molecules in biotechnology assays (Coljee et al. 2000).

In our attempt to isolate microsatellite sequences from *Aegla longirostri* (a South American freshwater crab), besides 21,43% of clones without sufficient flanking sequences for primer design and 3,45% of redundancy, it took our attention the occurrence of approximately 29% of doubtful clones. They are a kind of “shuffled” sequences that vary combinations of identical flanking regions around the (CA)_n repeat.

Suspicion of chimeric products led us to an exhaustive search for such phenomenon related with microsatellite development works. Despite the evident propensity of SSR to produce PCR-based recombination due the presence of the same motif in all templates, few works related the possibility of such event occurrence.

The authors of the isolation methodology we have been using (Refseth et al. 1997) mentioned a theoretical possibility of chimeric clones occurrence due to formation of truncated PCR products, which would act as an internal primer in a different microsatellite locus, but they did not seek for chimeras aligning their sequences. Moreover, the authors hypothesize that this phenomenon would not be a major problem, since they obtained successful amplification of some of their clones.

The first report of amplification artifacts related with microsatellite isolation found 11 chimeric sequences out of 19 sequenced clones (Koblízková et al. 1998). The authors suggested the presence of free oligonucleotides (GA)₁₃, used for the capture of microsatellites process, in the recovered DNA from the enrichment step as the cause of truncated fragments formation which would be priming a heterologous locus in subsequent PCR cycles.

Another study mentioned a chimera within their findings, but did not specify it as a chimeric PCR product (Poteaux et al. 1999) and more recently, one microsatellite isolation work mentioned the possibility of occurrence of chimeric clones (Hughes et al. 2002).

Excepting the study of Koblízková et al. (1998), none of the works which mentioned the possibility of chimeras in microsatellite development proposed a mechanism to explain this phenomenon or associated its occurrence with unsuccessfully amplifiable microsatellite loci.

The present study relates a surprising result from microsatellite development procedure. Here we propose a mechanism for chimeric microsatellite loci which differs from that proposed by Koblízková et al. (1998). We also discuss about ways to detect and to avoid chimeric clones, since they can be responsible for many microsatellite primers described in the literature for several organisms, which failed to amplify the expected products, despite the attempts to optimize PCR conditions.

Materials and Methods

Microsatellites were developed following the method described in Refseth et al. (1997), with some modifications. Genomic DNA from *A. longirostri* was digested with TaqI and fragments were ligated to adapters. A 5' biotinylated probe (CA)₈ was hybridized to the fragmented DNA. Hybrids were captured by streptavidin-coated magnetic beads (Dynabeads® M-280 Streptavidin, Dynal). Recovered DNA was amplified by PCR, purified and cloned using TA Cloning® kit (Invitrogen).

Positive clones, checked by the presence of insert, were sequenced. Before primer design, repetition was taken off and all 3' and 5' flanking regions around (CA)_n repeat were aligned as separated queries using ClustalW (Chenna et al. 2003).

Results

The first microsatellite-enriched library that we developed yielded 62 positive colonies. Three of 32 sequenced clones did not contain repeats. All the others presented only one flanking region to the microsatellite (Figure 1). These sequences are originated by the internal priming of the biotinylated oligonucleotide probe leaking from the magnetic beads. Absence of the adapter sequence at one microsatellite end of the insert confirms the repeat as the primer site.

(Insert Figure 1 after the paragraph above)

In order to avoid carrying oligonucleotide probes with the recovered DNA, we adopted the following strategies: reduction of probe concentration from 0,4 μ M to 0,3 μ M; reduction of the elution temperature from 90°C to 80°C (high temperatures could break the strong ligation between the biotin in the probe and the streptavidin bound to the magnetic beads); and an additional elution step, on which we add magnetic beads again to interact with any biotinylated probe that could still be in solution.

Our second attempt yielded 61 positive clones, among which only two resulted from free oligonucleotide extension. Different clones that showed high identity for both, upstream and downstream regions around the (CA) $_n$ repeat, were considered redundant (two cases). However, the outcome revealed that some inserts shared one of the flanking regions, but not the other. These 17 doubtful sequences corresponded to 28,8% of our clones and formed six subsets (A to F), as represented in Table 2, with high alignment scores. Only the 5' end of clone *AICA112* showed not too high values of identity with the related sequences 5' *AICA121*, 5' *AICA124* and 5' *AICA166*.

(Insert Table 2 after the paragraph above)

BLAST analyses (Altschul et al. 1990) were performed to verify the possibility of these sequences being part of repetitive genes, but no matches were found. Sequences were also submitted to NEBcutter V 2.0 (Vincze et al. 2003) to check the presence of sites for TaqI endonuclease that was used for the DNA digestion before isolation, which could represent a point of ligation between two different loci that had been cloned together. No sites for TaqI were found in these shuffled sequences.

Discussion

The mechanism implicated in chimeric microsatellite loci

Our favored explanation is that these results suggest *in vitro* recombination events during the PCR preceding cloning. A captured microsatellite locus not completely extended (extended until the (CA) $_n$ repeat) in one cycle annealed its 3' end with a (TG) $_n$ repeat of other microsatellite locus in the subsequent cycle, working as a priming site to follow extension. The generated nascent strand is a chimera formed by the 5' flanking region of one locus, a (CA) $_n$ repeat that was the crossover point and a 3' flanking region of another locus (Figure 2). Such chimera does not represent a contiguous sequence present in *A. longirostri*. Hence, primers designed for these PCR artifacts will certainly not amplify.

(Insert Figure 2 after the paragraph above)

Koblízková et al. (1998) defended that chimeric PCR products result from elongation of free oligonucleotide probes, which firstly would generate amplification fragments without one flanking region, and in a subsequent cycle would generate the chimeric products. They also defended that the problem could be eliminated with 3' modified oligonucleotides. This supposition could be applied for our 29% chimeric sequences attained, as in our second library we adopted strategies to avoid contamination of the recovered DNA with the biotinylated (CA)₈ probe and the outcome revealed significant reduction of clones lacking a flanking end (only two of 59 clones), but clone AICA161 shared its 5' flanking region with three other clones (Table 2) and had its 3' flanking region left out by the biotinylated oligonucleotide (CA)₈ extension from the (GT)_n repeat. This fact could merely be a coincidence, since redundancy was observed, but also could indicate that this sequence in particular generated the related chimeric clones, as proposed by Koblízková et al. (1998).

Although we agree with their hypothesis, we do not realize the presence of free oligonucleotides as the only explanation for the recombinant sequences attained from *A. longirostri* microsatellite-enriched library, by the complexity of this mechanism and by the reasons described bellow.

Once all captured single strand sequences have a (GT)_n, the Koblízková et al. (1998) proposal requires : (a) for a hypothetical locus A, the probe works as primer generating a strand with only one flanking region and its 5' end containing a (CA)_n repeat. The adaptor should extend this new strand in a subsequent PCR cycle to finally generate a sequence with only one flanking region and its 3' end containing a (GT)_n repeat; (b) another hypothetical locus B should be completely extended for both adaptors as primers, generating a new strand with both flanking regions and a (CA)_n repeat within. This new strand will anneal with that sequence ending in 3' (GT)_n (described on item a) to finally generate a chimeric sequence. Moreover:

1. Positive clones without a flanking region, caused by extended free oligonucleotides, generally correspond to the totality of yielded clones containing repeats, as it was found in our first attempt, as well as for a library constructed for a flatworm (M. L. Bartholomei-Santos, personal communication). Koblízková et al. (1998) did not report the occurrence of clones with only one flanking region (Fig 1) within their 19 sequences, just the 11 chimeric ones;

2. According to the authors, the modified procedure successfully blocked the formation of PCR artifacts. However, they sequenced 19 of 359 positive clones with strong hybridization signal to (GA)_n probe, which could represent an inadequate sample size. We

were able to identify chimeric sequences only after sequencing and aligning around fifty positive clones;

3. The mechanism proposed by Koblízková et al. (1998) demands that both the adapter oligonucleotide and the free repetitive oligonucleotide probe act as primers in the PCR. However, our first library indicated that when present in the reaction, the repetitive oligonucleotide will compete with the primer and prevent the action of the adapter in the amplification of one extremity of the microsatellite sequence, probably due to its more rapid annealing (Figure 1). This would lead to a preferential amplification of microsatellites sequences containing only one flanking region.

4. Finally, and mainly, some of our chimeric clones showed smaller similarity scores, since they presented a few differences at their flanking end shared with other clone, although they were clearly related. These hybrids could be generated if the incomplete extension during amplification passes beyond or stops before the microsatellite, and the 3' portion of this sequence act as a primer in a subsequent cycle (Figure 3a). Clone *AICA111* shares its 5' flanking region with sequences 3'*AICA95*, 5'*AICA116* and 5'*AICA161* until position 70 (...TG), where begin the repeat for that clone, but not for the others (Figure 3b). Following the alignment in Figure 3b, the same was observed for sequence *AICA161* at position 85. Probably an incompletely extended strand until these points (70 and 85), finishing in TG, primed off another microsatellite locus and generated a chimeric sequence. This mechanism seems very likely to have occurred in our microsatellite enrichment step and could not be attributed to products generated by contaminant repetitive probe, because this one can only generate identical flanking regions.

(Insert Figure 3 after the paragraph above)

As described in the introduction, formation of hybrid sequences by a similar mechanism to that we are proposing for microsatellite development have been demonstrated for several types of closely related sequences, and hence, microsatellites are also good candidates to this PCR artifact.

Besides prematurely finished products, chimeric PCR could be occurring by a template-switching process in which polymerase or the nascent strand switches from the original to a secondary template during DNA synthesis. Template-switching is a well-known recombination mechanism for virus (Hu and Temin 1990). It was demonstrated that this process can also occur during *in vitro* DNA synthesis with two experimental templates of 228 bp that contain a (CA)₂₅ repeat (Odelberg et al. 1995). However, that work intensified the

template-switching occurrence by carrying out a 10 μ L reaction with 125 ng of DNA template, in only one cycle-amplification procedure. Such situation does not apply to our microsatellite isolation protocol as a whole. We do not rule out the template-switching as the cause of chimeric sequences yielded for the *A. longirostri* microsatellite enriched library, but the chimera formation by prematurely nascent strands seems to be the most suitable explanation.

How to detect chimeric sequences?

Some methods have been proposed to detect chimeric sequences (see reference Gonzales et al. 2005 for a review). Those tools are based in comparison of the “supposed” chimera portions with a database of closely related sequences. Ribosomal RNA sequences analysis can count on the Ribosomal Database Project (RDP-II), an online database that, among other tools, offers a Chimera Check service (Maidak et al. 2001).

That procedure is worth for sequences that share substantial similarity, like genes. In microsatellite enriched library we do not work with closely related sequences. It will only be possible to generate a database with the own loci obtained with the protocol used, or maybe to use microsatellite sequences of some close species.

For microsatellite isolation from *Silene tartarica*, a database was created from the obtained sequences and BLAST was used to search it for matching clones, which was useful to find redundant clones. The remaining clones did not present identity, suggesting the absence of PCR artifacts. Most loci yielded PCR products in the expected size range in that case (Tero et al. 2006).

What is really worth pointing out is the importance of aligning microsatellite flanking ends, after taking off the repetitive region, trying to find shuffled sequences originated from chimeric PCR. It should be taken into account the low possibility of obtaining related recombinant sequences when only a subset of cloned PCR products are sequenced, or if a small number of clones is obtained from a isolation procedure. We realized that we fortunately – or unfortunately, for the microsatellite-enriched library – cloned a great variety of chimeras within all positive colonies obtained, which could be aligned and detected. Hence, the more clones are sequenced, the more probable is the detection of chimeras. If a flanking region is not present in shuffled clones, we can trust it as real, and it can be considered for primer design.

Minimizing the PCR recombination

Judo et al. (1998) performed different approaches with two related experimental templates in order to stimulate and suppress PCR-mediated recombination, based in factors that might affect the degree of incomplete elongation: the processivity of the enzyme and the length of elongation time. The importance of the number of cycles was also evaluated. The authors demonstrated that short extension time, over-PCR and lower processivity DNA polymerase contribute to recombination. It was clear that recombination events occur late in PCR reaction, as previously suggested (Meyerhans et al. 1990). According to both studies, the possibility of premature extension products to compete successfully with the normal PCR primers for target sequences increase with each subsequent round of amplification, as the normal PCR primers to target DNA progressively decrease in concentration.

These findings, despite not directly related to SSR, are important to take into consideration in microsatellite-enriched library. Our procedure for *A. longirostri* microsatellite development was performed with 30 PCR cycles of 2 minutes elongation step. As an attempt to minimize the recombination ratio, we will adopt and recommend a reaction with fewer cycles, larger elongation time, replacement of Taq DNA polymerase by a higher processivity polymerase, besides the utilization of a 3' modified oligonucleotide to eliminate the possibility of probes working as primers, as recommended by Koblízková et al. (1998).

Frequency of recombinant formation

Some studies described the expected recombination frequency range for small amplification products between 1% and 12,5% (Odelberg et al. 1995; Judo et al. 1998; Meyerhans et al. 1990). However, it is important to note that they used experimental templates of high homology, whilst a PCR for microsatellite enrichment works with different loci as template.

Only one of the studies cited above (Odelberg et al. 1995) included template with microsatellite for the evaluation of recombination frequency *in vitro*. Nevertheless, only one template was used (the lipoprotein lipase gene) and the resultant ratios ~2% to ~17% were attributed exclusively to template-switching.

Differently from the studies mentioned above, the isolated microsatellites have completely distinct flanking regions. Even so, microsatellites library construction showed the highest recombination ratios: 28,8% to *A. longirostri* and 57,9% to *Vicia faba* (Koblízková et al. 1998).

We suspected that the ratio of ~29% of recombination products in our clones could be due to the presence of long repeats. The fact of some sequences are involved in several shuffled clones (Groups B, C and F, Table 2) seems to indicate that the microsatellites in these sequences are more propense to recombination during PCR. We can speculate that the longer the repetition, the more susceptible to recombination. It is not possible to access whether this is true because the real loci were not identified, and the number of repetitions in a chimeric sequence probably does not correspond to the real one.

Concluding remarks

The main goal of this work was to discuss the occurrence of the PCR-recombinant products associated with microsatellite isolation, based on results attained from the microsatellite-enriched library developed for *A. longirostri* (data not published yet). Such recombination event would have been overlooked in several SSR development studies. Since efforts, time and money are often employed in primer design and optimization of PCR conditions for microsatellite loci, we strongly suggest the alignment of repeat's flanking regions to detect these artifacts.

It was not possible to define which, among shuffled clones, represent the real locus. So, none of these doubtful sequences should be considered for primer design. Refseth et al. (1997) recommended to perform a linear PCR on the genomic DNA as described by Espelund and Jakobsen (1992), when obtaining only one flanking region due to either generation of chimeric structures or by the oligonucleotide probe elongation

Here, we presented our data and conclusion as a possible, but not the only, explanation for the unfruitful loci for which the designed primer pairs fail to amplify microsatellite markers in several studies.

Future studies of PCR-mediated recombination with different microsatellite loci as templates would be of great value to estimate ratios of recombination and evaluate factors that affect the formation of abortive extension products and template-switching for repetitive sequences.

Acknowledgments We gratefully acknowledge Dr. Élgion Loreto for the sequencing service and the Brazilian Agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *J Mol Biol* 215: 403-410
- An HS, Han SJ (2006) Isolation and characterization of microsatellite DNA markers in the Pacific abalone, *Haliotis discus hannai*. *Mol Ecol Notes* 6:11-13
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At Least 1 in 20 16S rRNA Sequence Records Currently Held in Public Repositories Is Estimated To Contain Substantial Anomalies. *Appl Environ Microbiol* 71:7724-7736
- Bhavsar D, Zheng HD, Drysdale J (1994) Chimerism in PCR products from a multigene family. *Biochem Biophys Res Commun* 205:944-947
- Blanquer A, Uriz MJ, Pascual M (2005) Polymorphic microsatellite loci isolated from the marine sponge *Scopalina lophyropoda* (Demospongiae: Halichondrida). *Mol Ecol Notes* 5:466-468
- Bradley RD, Hillis DA (1997) Recombinant DNA Sequences Generated by PCR Amplification. *Mol Biol Evol* 14:592-593
- Brakenhoff RH, Schoenmakers JG, Lubsen NH (1991) Chimeric cDNA clones: a novel PCR artifact. *Nucleic Acids Res* 19:1949
- Cesari M, Mularoni L, Scanabissi F, Mantovani B (2004) Characterization of dinucleotide microsatellite loci in the living fossil tadpole shrimp *Triops cancriformis* (Crustacea Branchiopoda Notostrace). *Mol Ecol Notes* 4:733-735
- Chambers GK, Macavoy ES (2000) Microsatellites: consensus and controversy. *Comp Biochem Physiol, B* 126:455-476
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497-500
- Coljee VM, Murray HL, Donahue WF, Jarrell KA (2000) Seamless gene engineering using RNA- and DNA- overhang cloning. *Nat Biotechnol* 18:788-791
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor Appl Genet* 104:482-489
- Dutech CL, Amsellem L, Billotte N, Jarne P (2000) Characterization of (GA)_n microsatellite loci using an enrichment protocol in the neotropical tree species *Vouacapoua americana*. *Mol Ecol* 9:1433-1435

- Espelund M, Jakobsen KS (1992) Cloning and direct sequencing of plant promoters using primer-adapters mediated PCR on DNA coupled to a magnetic solid phase. *Biotechniques* 13: 74-81
- Fisher D, Bachmann K (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *Biotechniques* 24:796-802
- Gardner MG, Cooper SJB, Bull CM, Grant WN (1999) Brief communication. Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *J Hered* 90:301-304
- Gaublomme E, Dhuyvetter H, Verdyck P, Mondor-Genson G, Rasplus JY, Desender K (2003) Isolation and characterization of microsatellite loci in the ground beetle *Carabus problematicus* (Coleoptera, Carabidae). *Mol Ecol Notes* 3:341-343
- Gonzalez JM, Zimmerman J, Saiz-Jimenez C (2005) Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics* 21:333-337
- Hadonou AM, Walden R, Darby P (2004) Isolation and characterization of polymorphic microsatellites for assessment of genetic variation of hops (*Humulus lupulus* L.). *Mol Ecol Notes* 4:280-282
- Hamilton MB, Pincus EL, DiFiore A, Fleischer RC (1999) A universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 7:500-507
- Herrera EA, Chemello, ME, Lacey EA, Salas V, Sousa BF (2004) Characterization of microsatellite markers from capybaras, *Hydrochoerus hydrochaeris* (Rodentia: Hydrochoeridae). *Mol Ecol Notes* 4:541-543
- Hu WS, Temin HM (1990) Retroviral recombination and reverse transcription. *Science* 250:1227-1233
- Hugenholtz P, Huber T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int. J. Syst Evol Microbiol* 53:289-293
- Hughes M, Russel J, Hollingsworth PM (2002) Polymorphic microsatellite markers for the Socotran endemic herb *Begonia socotrana*. *Mol Ecol Notes* 2:159-160
- Judo MSB, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res* 26:1819-1825
- Koblízková A, Dolezel J, Macas J (1998) Subtraction with 3' modified oligonucleotides eliminates amplification artifacts in DNA libraries enriched for microsatellites. *Biotechniques* 25: 32-38

- Maidak BL, Cole JR, Lilburn TG, Parker CT, Saxman PR, Farris RJ, Garrity GM, Olsen GJ, Schmidt TM, Tiedje JM (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* 29:173-174
- Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Res* 18:1687-1691
- Mottura MC, Finkeldey R, Verga AR, Gailing O (2005) Development and characterization of microsatellite markers for *Prosopis chilensis* and *Prosopis flexuosa* and cross-species amplification. *Mol Ecol Notes* 5:487-489
- Nutt KJ (2003) Characterization of microsatellite markers for the North African gundi, *Ctenodactylus gundi* (Rodentia: Ctenodactylidae). *Mol Ecol Notes* 4:511-513
- Odelberg SJ, Weiss RB, Hata A, White R (1995) Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res* 23:2049-2057
- Pääbo S, Irwin DM, Wilson A (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265:4718-4721
- Poteaux C, Bonhomme F, Berrebi P (1999) Microsatellite polymorphism and genetic impact of restocking in Mediterranean brown trout (*Salmo trutta* L.). *Heredity* 82:645-653
- Refseth UH, Fangan BM, Jakobsen KS (1997) Hybridization capture of microsatellites directly from genomic DNA. *Electrophoresis* 18:1519-1523
- Sargent DJ, Hadonou AM, Simpson DW (2003) Development and characterization of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry. *Mol Ecol Notes* 3:550-552
- Schwartz TS, Jenkins F, Beheregaray LB (2005) Microsatellite DNA markers developed for the Australian bass (*Macquaria novemaculeata*) and their cross-amplification in estuary perch (*Macquaria colonorum*). *Mol Ecol Notes* 5:519-520
- Tero N, Neumeier H, Gudavalli R, Schötterer C (2006) *Silene tartarica* microsatellites are frequently located in repetitive DNA. *J Evol Biol* 19: 1612-1619
- Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31:3688-3691
- von Wintzingerode F, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21:213-229.
- Zane L, Baegelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1-16

Zhou Y, Gu H, Dorn S (2005) Polymorphic microsatellite loci in the parasitic wasp *Cotesia glomerata* (Hymenoptera: Braconidae). Mol Ecol Notes 5:475-477

Figure Legends:

Figure 1: Schematic representation of the recombinant sequences attained from the first *Aegla longirostri* microsatellite enriched library, in which biotinylated oligonucleotide probes worked as an internal primer to the captured locus. The PCR products yielded by the isolation procedure consist of the adapter sequence (checkered box) at one end, followed by a flanking region (white box) and the repeat motif (light hatched region). Vector sequence is represented by black boxes.

Figure 2. Schematic representation of three successive PCR cycles related to chimeric products formation. Illustrations represent two isolated microsatellite loci with the same repeat motif (light hatched region), but different flanking regions: white boxes - flanking regions of locus A; black boxes - flanking regions of locus B. Adapters that work as primers are shown as checkered boxes. In the n cycle it is produced an Incompletely Extended Strand (IES) from locus A, which ends in the repeat and serves as a primer for locus B in the $n + 1$ cycle. It is generated a chimeric strand that becomes double stranded in the next cycle ($n + 2$), and which can be amplified in the following cycles (not represented).

Figure 3. Chimeric sequences containing microsatellite, on which the Incompletely Extended Strand (IES) did not finish in the repeat motif. a. Schematic representation of three successive PCR cycles related to chimeric products formation. In the cycle n , the IES stops at a point of the upstream flanking region (locus A - white boxes) containing a sequence similar to the repeat, which primes off the repetition of a different microsatellite locus (locus B - black boxes) in the $n + 1$ cycle. It is generated a chimeric strand with a partial upstream flanking region of locus A, the repeat and the downstream flanking region of locus B, that becomes double stranded in the next cycle ($n + 2$). b. Alignment of the isolated sequences *AICA95*, *AICA111*, *AICA116* and *AICA161*, showing the high similarity of them to the upstream flanking region until the position 70 and 85, where begins the repeat for sequences *AICA111* and *AICA161*, respectively. The (TG) n repeat begins at position 298 for locus *AICA95* and *AICA116*. Both sequences present a (TG) at the positions 70 and 85.

Table 1. Variable efficiency of microsatellite-enrichment protocols, based on the percentage of designed primer pairs for different loci that successfully amplify the target region, according with different methodologies. Publications were randomly chosen to exemplify.

Study (Reference)	Number of designed primer pairs	Locus successfully optimized	Efficiency (%)	Methodology ^a (Reference)
An & Han 2006	150	50	33,33	Hamilton <i>et al.</i> 1999; Gardner <i>et al.</i> 1999
Schwartz <i>et al.</i> 2005	31	6	19,35	Fisher & Bachmann 1998
Zhou <i>et al.</i> 2005	14	7	50,0	Glenn (web) ^b
Blanquer <i>et al.</i> 2005	21	7	33,33	FIASCO (Zane <i>et al.</i> 2002)
Mottura <i>et al.</i> 2005	18	9	50,0	Fisher & Bachmann 1998
Cesari <i>et al.</i> 2004	31	5	16,13	FIASCO (Zane <i>et al.</i> 2002)
Herrera <i>et al.</i> 2004	28	14	50,0	Hamilton <i>et al.</i> 1999; Nutt 2003
Hadonou <i>et al.</i> 2004	26	10	38,46	Sargent <i>et al.</i> 2003
Gaublomme <i>et al.</i> 2003	22	6	27,27	Dutech <i>et al.</i> 2000

^aMethodologies including a PCR step before cloning.

^bProtocol modified from Glenn *et al.*, available at (http://www.uga.edu/srel/DNA_Lab/protocols.htm)

Table 2. Identity values (Clustal W) for microsatellite flanking regions shared among different clones attained from *Aegla longirostri* microsatellite-enriched library. The 5' and 3' flanking sequences of each microsatellite were separately analyzed, without the (CA) repeat. For seven groups of clones (A to F), one flanking end presented high homology values, which was not verified for the other end of the same clones. Scores for the non-shared flanking sequences of these shuffled clones are also showed, denoting low identity among them. 5' – upstream flanking region; 3' – downstream flanking region. Sequence length (bp) in parenthesis.

Group	Flanking regions shared among chimeric clones ^a	Score	Flanking regions not shared among chimeric clones	Score
A	5' 118 (179) ^b - 3' 146 (206)	93	3' 118 (314) - 5' 146 (392)	3
B	5' 106 (106) - 3' 129 (108)	95	3' 106 (216) - 5' 129 (116)	12
	5' 106 (106) - 5' 165 (115)	89	3' 106 (216) - 3' 165 (232)	5
	5' 106 (106) - 3' 174 (105)	95	3' 106 (216) - 5' 174 (171)	7
	3' 129 (108) - 5' 165 (115)	92	5' 129 (116) - 3' 165 (232)	8
	3' 129 (108) - 3' 174(105)	98	5' 129 (116) - 5' 174(171)	7
	5' 165 (115) - 3' 174(105)	94	3' 165 (232) - 5' 174(171)	5
C	5' 112 (61) - 5' 121 (60)	51	3' 112 (277) - 3' 121 (205)	4
	5' 112 (61) - 5' 124 (56)	67	3' 112 (277) - 3' 124 (181)	6
	5' 112 (61) - 5' 166 (56)	64	3' 112 (277) - 3' 166 (455)	11
	5' 121 (60) - 5' 124 (56)	92	3' 121 (205) - 3' 124 (181)	8
	5' 121 (60) - 5' 166 (56)	96	3' 121 (205) - 3' 166 (455)	9
	5' 124 (56) - 5' 166 (56)	96	3' 124 (181) - 3' 166 (455)	15
D	3' 96 (173) - 3' 113 (172)	98	5' 96 (330) - 5' 113 (269)	10
E	5' 96 (330) - 5' 133 (213)	92	3' 96 (173) - 3' 133 (89)	15
F	3' 95 (297) - 5' 111 (67)	98	5' 95 (298) - 3' 111 (128)	8
	3' 95 (297) - 5' 116 (297)	99	5' 95 (298) - 3' 116 (62)	19
	3' 95 (297) - 5' 161 (78)	85	5' 95 (298) - 3' 161 ^c	-
	5' 111 (67) - 5' 116 (297)	98	3' 111 (128) - 3' 116 (62)	16
	5' 111 (67) - 5' 161 (78)	98	3' 111 (128) - 3' 161 ^c	-
	5' 116 (297) - 5' 161 (78)	85	3' 116 (62) - 3' 161 ^c	-

^aClones are designed/nominated A/C Axx. Here they are represented only by the clone number (96, 106, 118, etc.)

^bLength of flanking ends of chimeric clones differs because only regions with good quality of sequencing lectures were considered in the analysis, but similarity scores refers to aligned regions of the same size.

^cSequence without 3' flanking region.

Figure 1:



Figure 2:

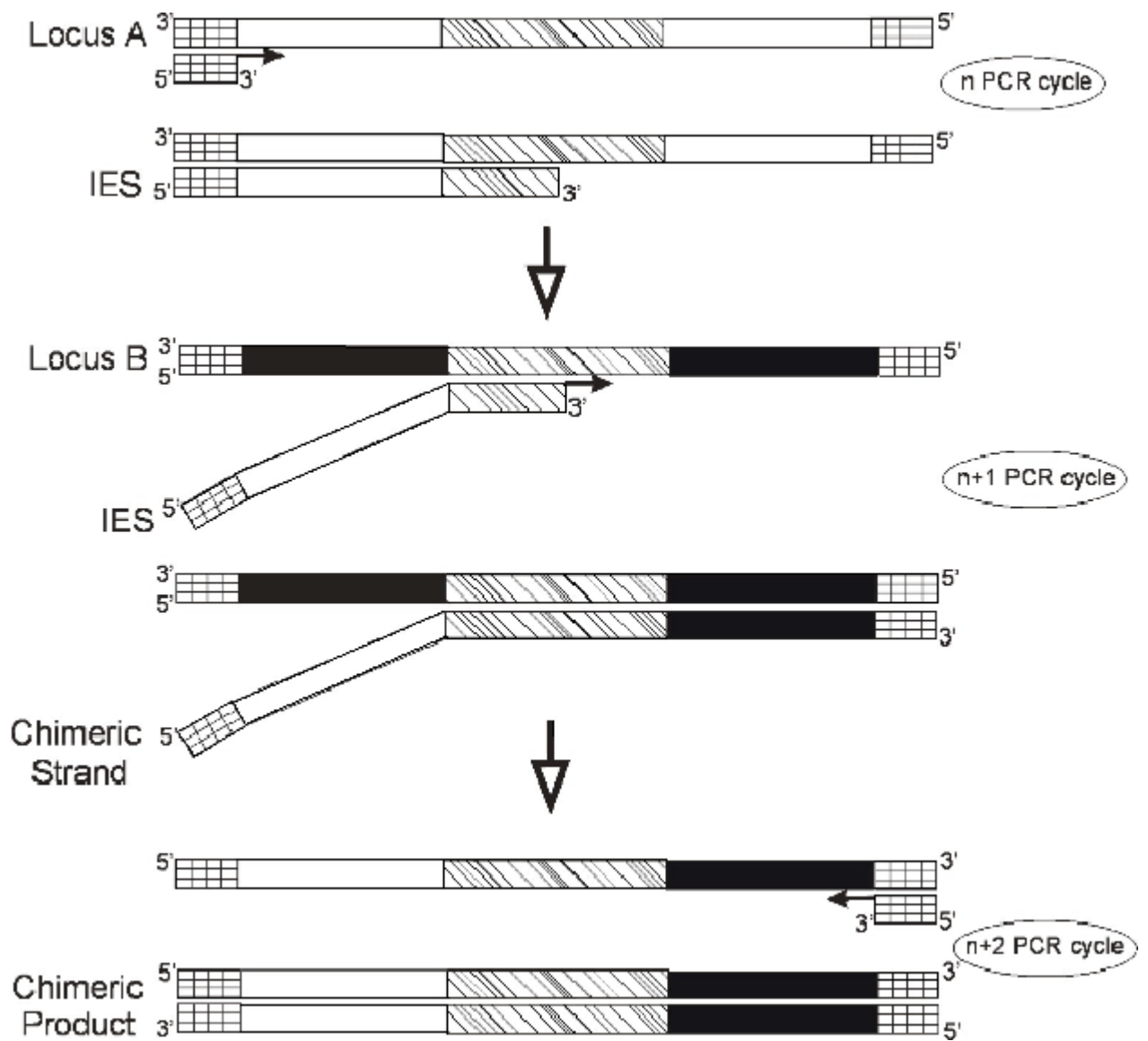
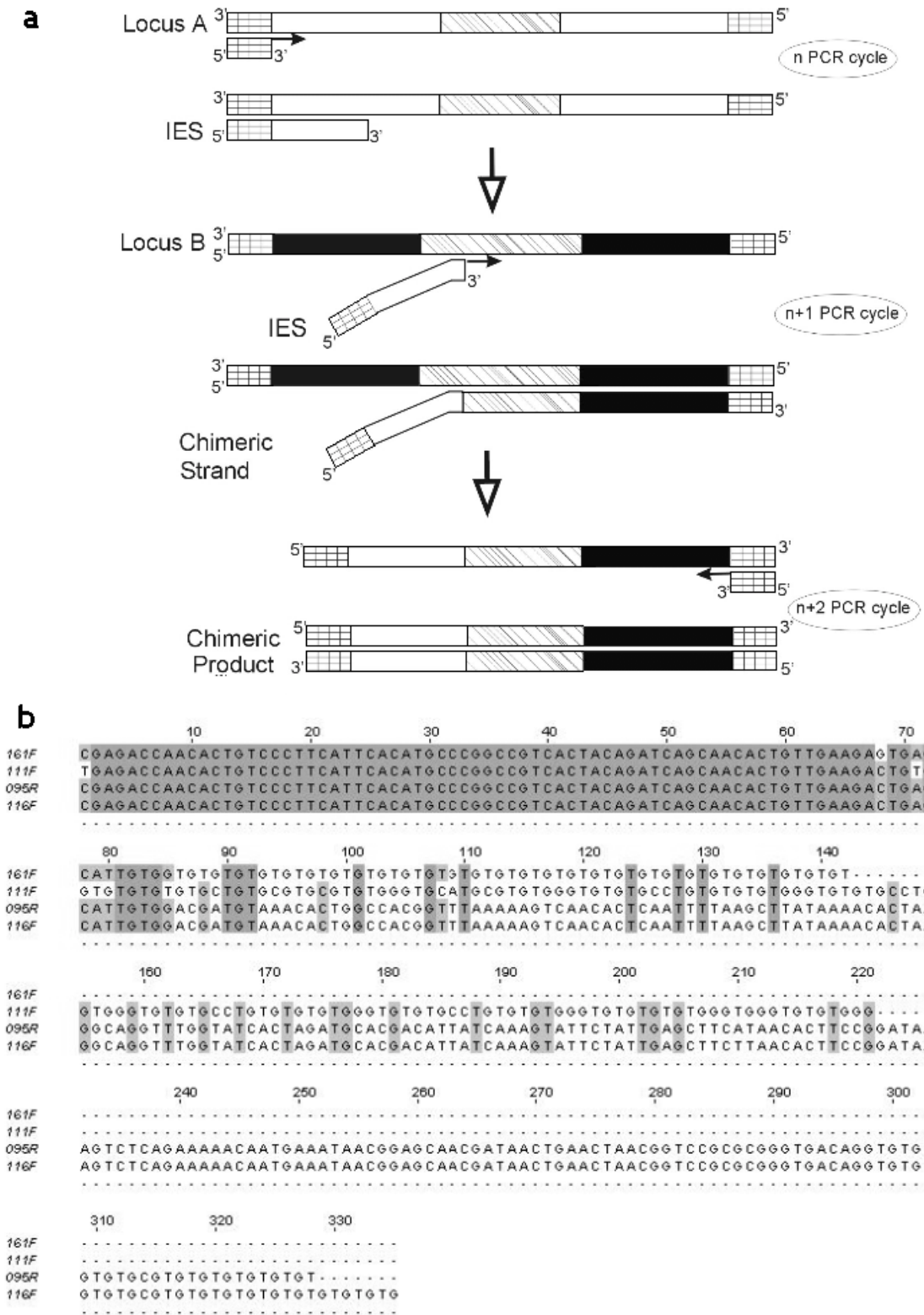


Figure 3:



4 DISCUSSÃO

De uma forma geral, o desenvolvimento de marcadores microssatélites se baseia em uma metodologia moderna que, apesar de utilizar técnicas simples de Biologia Molecular, é bastante trabalhosa na medida em que muitos são os clones positivos obtidos, mas ao final do processo, poucos são os locos caracterizados que poderão ser empregados produtivamente em estudos populacionais, entre outras aplicações (Zane et al., 2002).

A metodologia utilizada neste estudo mostrou-se bastante eficiente em termos de quantidade de seqüências-alvo obtidas. A primeira biblioteca genômica enriquecida com microssatélites para *A. longirostri* forneceu 62 colônias recombinantes, sendo que das 32 seqüenciadas 90,6% apresentaram insertos contendo elementos repetitivos, embora 100% destes insertos tenham sido fruto da extensão do oligo biotilado (CA)₈ e, portanto, não apresentam uma das regiões flanqueadoras.

A segunda tentativa também se mostrou bastante rentável, com 68 colônias recombinantes obtidas e índices de apenas 5,1%, de insertos sem repetição e 3,4% com menos de 10 unidades repetitivas. Estes índices são baixos em comparação ao obtido por Refseth et al. (1997), ao descrever a técnica. Naquele estudo, 40% das seqüências isoladas apresentaram número reduzido de unidades repetitivas e 23% não continham microssatélites.

O presente trabalho descreve um número significativo de locos de microssatélites isolados com numerosas unidades repetitivas: 7 locos (~12%) apresentaram entre 20 e 30 repetições (CA), 5 locos (8,5%) entre 30 e 40 repetições e 6 locos (10%) apresentaram mais de 40 repetições (CA). Destas seqüências contendo elementos repetitivos longos, provavelmente caracterizariam-se locos de microssatélites polimórficos, no entanto apenas os locos AICA108 [(AC)₁₆CT(AC)₂₆CG(AC)₁₉], AICA119 [(CA)₆CT(CA)₂₇], AICA128 [(GT)₄₃] e AICA136 [(CT)₆(CA)₂₈] apresentaram região flanqueadora com comprimento suficiente e ideal para a projeção de pares de *primers*. Os demais locos para os quais foram projetados pares de *primers* neste trabalho para repetições (CA) apresentam microssatélites com menos de 20 unidades repetitivas. Entretanto, destaca-se o loco

de microssatélite composto AICA120 que, apesar da pequena repetição (CA)₆, apresenta também a repetição (CT)₂₃.

Cerca de 70% das seqüências isoladas não puderam ser empregadas na elaboração de pares de *primers*: 15 (25,4%) não permitiram o desenvolvimento de pares de *primers* por falta de região flanqueadora, 17 (28,8%) corresponderam a seqüências quiméricas resultantes de artefatos da PCR e 9 (15,25%) não apresentaram uma seqüência consenso para as leituras de seqüenciamento obtidas.

Um passo recomendado pela metodologia que não foi aplicado em nossos procedimentos é a seleção dos fragmentos de DNA aplicados em gel de agarose, antes da captura (Refseth et al., 1997). Em geral os protocolos de isolamento de microssatélites recomendam a utilização de fragmentos entre 300 e 1000 pb, pelo fato de que os insertos não podem ser muito pequenos a ponto de não haver região flanqueadora suficiente para obtenção de *primers*; nem muito grandes, o que dificultaria o seqüenciamento. Entre os nove clones para os quais não se obteve uma leitura consenso, apenas dois deles apresentaram leituras *forward* e *reverse* com mais de 500pb, as quais não se sobrepuseram, o que justificaria a seleção dos insertos antes da clonagem. Entretanto, os demais insertos apresentaram leituras confiáveis menores e, em geral, o decaimento da leitura ocorreu justamente quando começa a repetição.

Embora importante, esse passo do procedimento não se mostrou tão vital quanto a captura por afinidade e a PCR, as quais apresentaram inconveniências que marcaram os resultados de isolamento e serviram para que modificações fossem agregadas à metodologia. A obtenção de 100% dos clones (contendo repetição) sem uma extremidade flanqueadora na 1ª biblioteca construída, expõe a captura como um ponto crítico do procedimento (veja Figura 1 do artigo 1), no qual deve-se evitar ao máximo a liberação do oligo biotilado durante a eluição dos fragmentos de interesse.

A presença do oligo biotilado (CA)₈, atuando como *primer* na PCR que precede a clonagem, é evidenciada pelo fato de os insertos do primeiro isolamento apresentarem a seqüência do adaptador apenas na extremidade oposta à que inicia com a repetição. Além disso, dos 32 clones seqüenciados da primeira biblioteca genômica, os três que não apresentaram elementos repetitivos possuíam a seqüência do adaptador em ambas as extremidades, demonstrando que para insertos sem microssatélites não houve interação com o oligo biotilado atuando

como *primer*. Embora imprópria para o desenvolvimento de pares de *primers*, esta biblioteca genômica enriquecida com microssatélites apresentou repetições (CA) que variaram de seis a 38 unidades e seqüências compostas interessantes (veja Tabela 1 do artigo 1).

A outra inconveniência reside nos artefatos correspondentes a quase 29% dos clones positivos obtidos na segunda biblioteca enriquecida, os quais exigiram todo um trabalho de bancada e análise computacional, sendo que jamais atenderão a expectativa maior do projeto de desenvolver ferramentas moleculares funcionais (ver artigo 2).

Tanto a perda de uma extremidade flanqueadora causada pelo anelamento interno do oligo biotilado, quanto à formação de produtos quiméricos durante a PCR, foram preditas por Refseth et al. (1997) que descreveram a técnica. Para ambos os casos, os autores apresentam uma solução para recuperar os locos, baseados em Espelund e Jakobsen (1992 apud Refseth et al., 1997), que propõem a técnica do PCR linear com o DNA inicial empregado na hibridização, usando um *primer* biotilado para a única região flanqueadora conhecida. O produto da PCR pode ser extraído com a utilização de esferas magnéticas e a fita complementar é posteriormente gerada com o adaptador como *primer*.

Com o emprego desta técnica, há muito a ser explorado tanto nas seqüências sem extremidade flanqueadora da 1ª captura quanto para os casos de extremidades flanqueadoras curtas e clones quiméricos da 2ª captura. Além disso, mais reações de sequenciamento podem ser investidas para os locos sem consenso definido, com o desenho de *primers* internos.

Treze seqüências obtidas na 2ª captura apresentaram regiões flanqueadoras suficientes para o desenvolvimento de *primers* em torno de repetições com mais de 10 unidades. Condições de amplificação foram exaustivamente testadas com os pares de *primers* designados para os locos AICA105, AICA119, e AICA128, sem sucesso. Quando a PCR apresentava algum produto, o mesmo não era do tamanho esperado e nem reproduzível.

Pares de *primers*, destinados a locos de microssatélites, que não amplificam são freqüentemente relatados em trabalhos de isolamentos de microssatélites (veja revisão no artigo 2), sem uma justificativa plausível. A atribuição destes locos a seqüências quiméricas foi abordada no artigo 2, onde se argumentou também que uma seqüência quimérica pode não ser detectada pelo alinhamento dos clones

quando, entre as seqüências alinhadas, não há nenhuma outra relacionada a ela. Considerando a taxa de quase 29% de clones quiméricos detectados neste estudo, uma das justificativas para os pares de *primers* designados aos locos AICA105, AICA119 e AICA128 não amplificaram, poderia ser o fato de estas seqüências se tratarem de clones quiméricos não detectados.

A escolha do programa para a elaboração dos pares de *primers* também foi avaliada. Estes primeiros locos (AICA105, AICA119 e AICA128) foram submetidos ao programa OLIGO (Rychlin and Rhoads, 1989), sendo que nenhum amplificou. A segunda remessa de locos que obtiveram seqüência-consenso bem definida foi submetida ao programa PRIMER3 (Rozen and Skaletsky, 2000) numa tentativa de comparação e, até o momento, pelo menos dois locos já produziram produto de amplificação esperado. Não é possível afirmar, com base nestes resultados, que um programa apresenta pares de *primers* mais confiáveis em detrimento do outro. Entretanto é interessante observar que as 13 seqüências foram submetidas a ambos os programas e somente para uma delas os pares de *primers* ideais foram concordantes.

Dos últimos pares de *primers* elaborados, as condições ideais de amplificação foram estabelecidas para os locos AICA135 e ALGA138 e análises recentes em gel de poliacrilamida têm demonstrado polimorfismo, embora ainda não haja avaliação estatística para os mesmos. Os demais estão sob processo de otimização das condições de amplificação. Alguns destes já apresentaram produto de amplificação do tamanho esperado. Entretanto, sob as mesmas condições, ainda não foi possível reproduzir os resultados e/ou estender aos demais indivíduos.

Embora as bibliotecas genômicas tenham sido geradas exclusivamente com a sonda (CA)₈, 13 repetições (GA/CT) foram obtidas ao acaso, além de outras três repetições tri e quatro tetranucleotídicas (veja Tabela 2 do artigo 1). Neste estudo, o loco A/GA138 contendo a repetição (GA)₁₇, poderá somar aos demais locos (CA)_n caracterizados em estudos futuros.

Um estudo bem fundamentado, destinado a analisar a estruturação genética de populações de *Aegla*, é fundamental para a compreensão da história evolutiva do grupo, sua situação atual e para o planejamento de ações que garantam sua sobrevivência. Este trabalho descreve o desenvolvimento de pelo menos dois marcadores microsatélites (até o momento) para *A. longirostri*, endêmica do estado do Rio Grande do sul.

Além de avaliar o grau de divergência entre populações de *Aegla longirostri* sob efeito de uma barreira geográfica montanhosa (Santos et al., 2007), a análise por microssatélites pode revelar também se há divergência entre populações que compartilham a mesma bacia. O fato de pertencerem à mesma rede hidrográfica, o que teoricamente permitiria o fluxo gênico entre estas populações, pode estar sendo superado pela deterioração dos ambientes límnicos, diminuindo a qualidade das águas e restringindo cada vez mais a ocorrência de espécies às nascentes.

Demais espécies do gênero *Aegla*, além da *A. longirostri* para a qual os locos de microssatélites foram desenvolvidos, também poderão ser alvo de estudos deste gênero se estes marcadores se mostrarem polimórficos para as mesmas, como já foi demonstrado para *Aegla uruguayana*.

5 CONCLUSÃO

O desenvolvimento deste estudo permitiu concluir que:

- A metodologia utilizada levou ao isolamento de um grande número de seqüências repetitivas. Embora muitas não possam ser empregadas como marcadores moleculares por não permitirem o desenvolvimento de pares de *primers*, a biblioteca genômica construída neste trabalho forneceu 13 seqüências aptas para este fim. Após serem devidamente caracterizadas e, associadas aos demais locos que aguardam novas reações de seqüenciamento para posterior avaliação, os marcadores microssatélites desenvolvidos neste estudo são promissores para análises em nível de estrutura populacional, biogeografia entre outros de *A. longirostri*, bem como para demais espécies do gênero para as quais estes locos amplificarem e se mostrarem polimórficos;

- A identificação de cerca de 29% dos clones isolados oriundos de produtos quiméricos da PCR, apesar de diminuir o número de locos apropriados à utilização como marcadores moleculares, permitiu um aprofundamento a respeito dos artefatos que podem ser gerados durante o processo de isolamento de microssatélites, bem como a elaboração de um segundo trabalho, cujo principal objetivo foi a divulgação deste evento e das modificações agregadas à técnica para preveni-lo.

6 PERSPECTIVAS

- Investir em reações de seqüenciamento aos insertos que não possuem uma seqüência consenso definida e aos que ainda não foram seqüenciados, para designar pares de *primers* adicionais;
- Otimizar condições de amplificação aos pares de *primers* obtidos;
- Avaliar o grau de polimorfismo dos locos isolados em uma amostra de 15 indivíduos de *A. longirostri*;
- Utilizar os locos isolados em estudos de genética de populações de *A. longirostri*;
- Testar a amplificação destes locos em outras espécies do gênero;
- Avaliar a eficiência das estratégias sugeridas no Artigo 2 para evitar a formação de produtos de PCR quiméricos na construção de futuras bibliotecas genômicas enriquecidas com microssatélites.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ABELL, R. Conservation biology for the biodiversity crisis: a freshwater follow-up. **Conservation Biology**. v. 16, n. 5, p. 1435-1437, 2002.

AHYONG, S. T.; O'MEALLY, D. Phylogeny of the decapoda reptantia: resolution using three molecular loci and morphology. **The Raffles Bulletin of Zoology**. v. 52, n. 2, p. 673-693, 2004.

ARENAS, J. La Cordillera de la Costa como refugio de la fauna dulcícola preglacial. **Archives de Biología y Medicina Experimentales**. v. 10, p. 40. 1974.

BAHAMONDE, N. et al. Categorías de conservación de decápodos nativos de aguas continentales de Chile. **Boletín del Museo Nacional de Historia Natural**. v. 47, p. 91-100, 1998.

BENZIE, J. A. H. Population structure in penaeid prawns. **Aquaculture Research**. v. 31, n. 1, p. 95-119, 2000.

BILODEAU, A. L.; FELDER, D. L.; NEIGEL, J. E. Multiple paternity in the thalassinidean ghost shrimp, *Callinectes islagrande* (Crustacea: Decapoda: Callinassidae). **Marine Biology**. v. 146, n. 2, p. 381-385, 2005.

BOND-BUCKUP G.; BUCKUP L. A família Aeglidae (Crustacea, Decapoda, Anomura). **Arquivos de Zoologia**. v. 32, p. 159-347, 1994.

BOND-BUCKUP G. Família Aeglidae. In: MELLO, G. A. S. **Manual de identificação dos Crustacea Decapoda de água doce do Brasil**. Loyola, São Paulo, 2003. p. 21-116.

BOND-BUCKUP, G.; BUCKUP, L.; ARAUJO, P. B. Crustáceos. In: FONTANA, C. S., BENCKE, G. A.; REIS, R. E. (Org.). **Livro Vermelho da fauna ameaçada de extinção no Rio Grande do Sul**. EDIPUCRS, Porto Alegre, 2003. p. 632.

BUENO, A. A. P.; BOND-BUCKUP, G. 2. Natural diet of *Aegla platensis* Schmitt and *Aegla ligulata* Bond-Buckup & Buckup (Crustacea, Decapoda, Anomura) from Brazil. **Acta Limnológica Brasileira**. v. 16, p. 155-127, 2004.

CASTRO-SOUZA, T.; BOND-BUCKUP, G. O nicho trófico de duas espécies simpátricas de *Aegla* Leach (Crustácea, Aeglidae) no tributário da bacia hidrográfica do Rio Pelotas, Rio Grande do Sul, Brasil. **Revista Brasileira de Zoologia**. v. 21, n. 4, p. 805-813, 2004.

CHAMBERS, G. K.; MACAVOY, E. S. Microsatellites: consensus and controversy. **Comparative Biochemistry and Physiology Part B**. v. 126, p. 455-476, 2000.

CHAND, V.; DE BRUYN, M.; MATHER P. B. Microsatellite loci in the eastern form of giant freshwater prawn (*Macrobrachium rosenbergii*). **Molecular Ecology Notes**. v. 5, p. 308-310, 2005.

CHAROENTAWEE, K.; POOMPIANG, S.; NA-NAKORN, U. Isolation and characterization of microsatellites in giant freshwater prawn *Macrobrachium rosenbergii*. **Molecular Ecology Notes**. v. 6, n. 3, p. 823-825, 2006.

D'AMATO, M. E.; CORACH, D. Population genetic structure in the fresh water anomura *Aegla jujuyana* by RAPD analysis. **Journal of Crustacean Biology**. v. 17, n. 2, p. 269-274, 1997.

EISEN, J. A. Mechanistic basis for microsatellite instability. In: GOLDSTEIN, D.B.; SCHÖLTTERER, C. (Org.). **Microsatellites: evolution and applications**. Oxford University Press, Oxford, 1999. p. 34-48.

ESTOUP, A. et al. Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). **Molecular Ecology**. v. 7, p. 339-353, 1998.

FELDMANN, R. M. *Haumuriaegla glaessneri* n.gen. and sp. (Decapoda: Anomura: Aeglidae) from Haumurian (Late Cretaceous) rocks near Cheviot, New Zealand. **New Zealand Journal of Geology and Geophysic**. v. 27, p. 379-385, 1984.

FELDMANN, R. M. Paleontology of two decapod taxa in Southern Hemisphere: Global conclusions with sparse data. In: GORE, R. H.; HECK, K. L. (Org.). **Crustacean issues**, volume 4. Crustacean biogeography. A. A. Balkema, Rotterdam, The Netherlands, 1986. p. 5-19.

FELDMANN, R. M. et al. Early Cretaceous Arthropods from the Tlayúa Formation at Tepexi de Rodríguez, Puebla, México. **Journal of Paleontology**. v. 72, n. 1, p. 79-90, 1998.

FESSEHAYE, Y. et al. Mating systems and male reproductive success in Nile tilapia (*Oreochromis niloticus*) in breeding hapas: A microsatellite analysis. **Aquaculture**. v. 256, p.148-158, 2006.

GEIST, J.; KUEHN, R. Genetic diversity and differentiation of central European freshwater pearl mussel (*Margaritifera margaritifera* L.) populations: implications for conservation and management. **Molecular Ecology**. v. 14, p. 425-439, 2005.

GOLDSTEIN, D. B.; SCHÖLTTERER, C. **Microsatellites: evolution and applications**. Oxford University Press, Oxford, 1999. 352 p.

GOYARD, E. et al. Residual genetic variability in domesticated populations of the Pacific blue shrimp (*Litopenaeus stylirostris*) of New Caledonia, French Polynesia and Hawaii and some management recommendations. **Aquatic Living Resources**. v. 16, p. 501-508, 2003.

GOW, J.L. et al. Breeding system and demography shape population genetic structure across ecological and climatic zones in the African freshwater snail, *Bulinus forskalii* (Gastropoda, pulmonata), intermediate host for schistosomes. **Molecular Ecology**. v. 13, p. 3561-3573, 2004.

IMGRUND, J.; GROTH, D.; WETHERALL, J. Genetic analysis of the freshwater crayfish *Cherax tenuimanus*. **Electrophoresis**. v. 18, n. 9, p. 1660-1665, 1997.

JARA, C.; PÉREZ-LOSADA, M.; CRANDALL, K. A. *Aegla occulentalis*, new species of fresh-water anomuran crab (Crustacea: Decapoda) from the Nahuelbuta Coastal Range, Chile. **Proceedings of the Biological Society of Washington**. v. 116, p. 933-942, 2003.

JIN, L.; ZHONG, Y.; CHAKRABORTY, R. The exact number of possible microsatellite motifs. **American Journal of Human Genetics**. v. 55, p. 582-583, 1994.

KASHI, Y; SOLLER, M. Functional roles of microsatellites and minisatellites. In: Goldstein, D.B.; Schöltterer, C. (Org.). **Microsatellites: evolution and applications**. Oxford University Press, Oxford, 1999. p 10-23.

LUCENTINI, L. et al. Microsatellite polymorphism in Italian populations of northern pike (*Esox lucius* L.). **Fisheries Research**. v. 80, p. 251-262, 2006.

LUNDBERG, J. G. et al. The stage for Neotropical fish diversification: a history of tropical South America rivers. In: MALABARBA, L. R. et al. (Org.). **Phylogeny and classification of Neotropical Fishes**. EDIPUCRS. Porto Alegre, 1998. p. 13-48.

MACPHERSON, E.; JONES, W.; SEGONZAC, M. A new squat lobster family of Galatheaidea (Crustacea, Decapoda, Anomura) from the hydrothermal vents of the Pacific-Antartic Ridge. **Zoosystema**. v. 27, n. 4, p. 709-723, 2005.

MAGNI, S. T.; PY-DANIEL, V. *Aegla platensis* Schmitt, 1942 (Decapoda, Anomura) um predador de imaturos de Simuliidae (Diptera: Culicomorpha). **Revista de Saúde Pública**. v. 23, n. 3, p. 258-259, 1989.

MARTIN, J. W.; ABELE, L. G. Phylogenetic relationships of the genus *Aegla* (Decapoda: Anomura: Aeglidae), with comments on anomuran phylogeny. **Journal of Crustacean Biology**. v. 6, p. 576-616, 1986.

MARTIN, J. W.; ABELE, L. G. External morphology of the genus *Aegla* (Decapoda: Anomura: Aeglidae). **Smithsonian Contributions to Zoology**. v. 453, p. 46, 1988.

MENIER, C. et al. Lack of molluscan host diversity and the transmission of an emerging parasitic disease in Bolivia. **Molecular Ecology**. v. 10, p. 1333-1340, 2001.

MORRISON, C. L. et al. Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. **Proceedings of the Royal Society of London B**. v. 269, p. 345-350, 2002.

MORRONE, J. J. On the geographical distribution of the species of *Aegla* Leach: identification of areas of endemism (Decapoda, Aeglidae). **Biogeographica**. v. 72, n. 3, p. 113-119. 1996.

MORRONE, J. J.; LOPRETTO, E. C. Distributional patterns of freshwater Decapoda (Crustacea: Malacostraca) in southern South America: a panbiogeographical approach. **Journal of Biogeography**. v. 21, p. 97-109, 1994.

ORTMANN, A. E. The geographical distribution of the freshwater decapods and its bearing upon ancient geography. **Proceedings of the Philadelphia Society**. v. 41, p. 267-400, 1902.

OSTELLARI, L. et al. Novel microsatellite loci isolated from northern krill, *Meganyctiphanes norvegica* (Crustacea: Euphausiacea). **Marine Ecology**. v. 9, n. 3, p. 377, 2000.

PARKER, K. M.; SHEFFER, R. J.; HEDRICK, P. W. Molecular variation and Evolutionary Significant Units in the endangered *Gila topminnow*. **Conservation Biology**. v. 13, n. 1, p. 108-116, 1999.

PÉREZ-LOSADA, M. et al. Phylogenetic relationships among the species of *Aegla* (Anomura: Aeglidae) fresh-water crabs from Chile. **Journal of Crustacean Biology**. v. 22, n. 2, p. 304-313, 2002a.

PÉREZ-LOSADA, M. et al. Phylogenetic position of the freshwater anomuran family Aeglidae. **Journal of Crustacean Biology**. v. 22, n. 3, p. 670-676, 2002b.

PÉREZ-LOSADA, M. et al. Conservation phylogenetics of Chilean freshwater crabs *Aegla* (Anomura: Aeglidae): assigning priorities for aquatic habitat protection. **Biological Conservation**. v. 105, p. 345-353, 2002c.

PÉREZ-LOSADA, M. et al. Molecular systematics and biogeography of the southern South American fresh-water "crabs" *Aegla* (Decapoda: Anomura: Aeglidae) using multiple heuristic tree search approaches. **Systematic Biology**. v. 53, n. 5, p. 767-780, 2004.

PORTER, M. L.; PÉREZ-LOSADA, M.; CRANDALL, K. A. Model-based multi-locus estimation of decapod phylogeny and divergence times. **Molecular Phylogenetics and Evolution**. v. 37, p. 355-369, 2005.

REFSETH, U. H.; FANGAN, B. M.; JAKOBSEN, K. S. Hybridization capture of microsatellites directly from genomic DNA. **Electrophoresis**. v. 18, p. 1519-1523, 1997.

RINGUELET, R. A. Consideraciones sobre las relaciones filogenéticas entre las especies del género *Aegla* Leach (Decapoda: Anomura). **Notas del Museo de La Plata, Zoología**. v. 14, p. 111-118, 1949.

ROBAINAS, A. et al. Microsatellite loci from the pink shrimp *Farfantenaues notialis* (Crustacea, Decapoda). **Molecular Ecology Notes**. v. 2, n. 3, p. 344-345, 2002.

ROZEN, S.; SKALETSKY, H. Primer 3 on the WWW for general users and for biologist programmers. In: KRAVETZ, S.; MISENDER, S. (Org.). **Bioinformatics methods and protocols: methods in Molecular Biology**. Human Press, Totowa, NJ, 2000. p. 365-386.

RYCHLIN, W.; RHOADS, R. E. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplifications of DNA. **Nucleic Acids Research**. v. 17, p. 8543-8551, 1989.

SANTOS, S. et al. 2007. Genetic Differentiation of *Aegla longirostri* populations isolated by a geographical barrier. (Submetido ao *Journal of Crustacean Biology*).

SCHMITT, W. L. The species of *Aegla*, endemic south American fresh-water crustaceans. **Proceedings of the United States National Museum**. v. 91, n. 31-32, p. 431-524, 1942.

TOONEN, R. J. Genetic evidence of multiple paternity of broods in the intertidal crab *Petrolisthes cinctipes*. **Marine Ecology – Progressive Series**. v. 270, p. 259-263, 2004.

TUDGE, C. C.; SCHELTINGA, D. M. Spermatozoal morphology of the freshwater anomuran *Aegla longirostri* Bond-Buckup & Buckup, 1994 (Crustacea: Decapoda: Aegliidae) from South America. **Proceedings of the Biological Society of Washington**. v. 115, n. 1, p. 118-128, 2002.

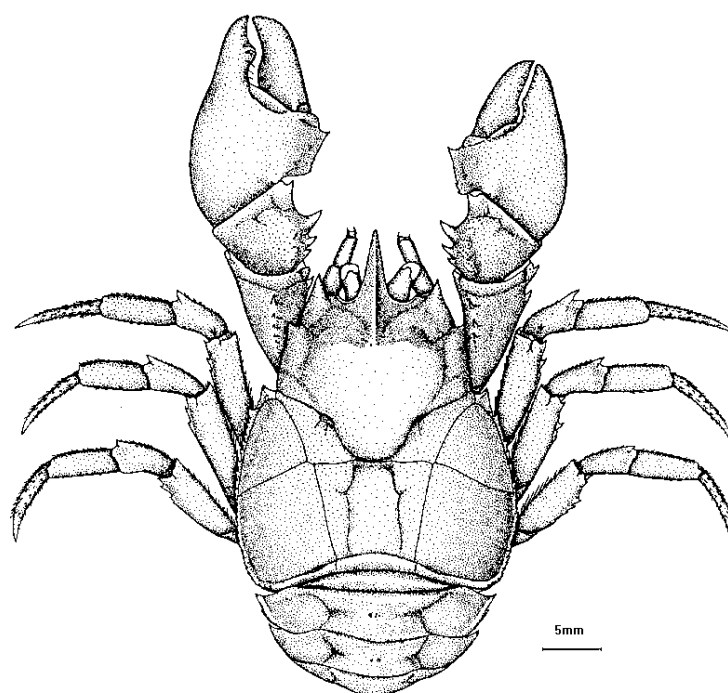
XU, Z. et al. Genetic diversity of wild and cultured Black Tigre Shrimp *Penaeus monodon* in the Philippines using microsatellites. **Aquaculture**. v. 199, p. 13-40, 2001.

YAN, J. et al. RAPD and microsatellite análisis of diploid gynogens from allotetraploid hybrids of red crucian carp (*Carassius auratus*) X common carp (*Cyprinus carpio*). **Aquaculture**. v. 243, p. 49-60, 2005.

ZANE, L.; BARGELLONI, L.; PATARNELLO, T. Strategies for microsatellite isolation: a review. **Molecular Ecology**. v. 11, p. 1-16, 2002.

ANEXO A – Descrição da Espécie*Aegla longirostri*

Margens laterais da área branquial anterior da carapaça pouco expandidas com tubérculos escamiformes; espinho ântero-lateral da carapaça ultrapassando a metade da córnea; rostro muito longo, estiliforme, carenado em todo o seu comprimento; ângulo anterior da margem ventral do epímero dois projetado por espinho robusto; margem externa proximal do dedo móvel do quelípodo sem lobo; crista palmar retangular, escavada; margem interna da face ventral do ísquio do quelípodo com dois espinhos robustos e entre eles tubérculos; margem ventral do mero do segundo pereiópodo com espinho ântero-medial seguido de tubérculo escamiformes (Bond-Buckup e Buckup, 1994).



Vista dorsal de um indivíduo de *Aegla longirostri*, extraído de Bond-Buckup e Buckup, 1994.