High incidence of cryptic repeated elements in microsatellite flanking regions of galatheid genomes and its evolutionary implications

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

HIGH INCIDENCE OF CRYPTIC REPEATED ELEMENTS IN MICROSATELLITE FLANKING REGIONS OF GALATHEID GENOMES AND ITS PRACTICAL IMPLICATIONS FOR MOLECULAR MARKER DEVELOPMENT

Deborah A. Bailie, Hugh Fletcher, and Paulo A. Prodöhl

(DAB, HF, PAP, correspondence, p.prodohl@qub.ac.uk) School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast BT9 7BL, Northern Ireland, UK

ABSTRACT

During the development of PCR primer sets for microsatellite marker loci from enriched genomic libraries for three squat lobster species from Galatheidae (Decapoda: Anomura); *Munida rugosa* (Fabricius, 1775), *M. sarsi* (Huus, 1935), and *Galathea strigosa* (Linnaeus, 1761) (collectively known as squat lobsters), a number of unforeseen problems were encountered. These included PCR amplification failure, lack of amplification consistency, and the amplification of multiple fragments. Careful examination of microsatellite containing sequences revealed the existence of cryptic repeated elements on presumed unique flanking regions. BLAST analysis of these and other VNTR containing sequences ($N = 252$) indicates that these cryptic elements can be grouped into families based upon sequence similarities. The unique features characterising these families suggest that different molecular mechanisms are involved. Of particular relevance is the association of microsatellites with mobile elements. This is the first reported observation of this phenomenon in crustaceans, and it also helps to explain why microsatellite primer development in galatheids has been relatively unsuccessful to date. We suggest a number of steps that can be used to identify similar problems in microsatellite marker development for other species, and also alternative approaches for both marker development and for the study of molecular evolution of species characterised by complex genome organisation. More specifically, we argue that new generation sequencing methodologies, which capitalise on parallel and multiplexed sequencing may pave the way forward for future crustacean research.

KEY WORDS: Galatheidae, microsatellites, mobile elements, squat lobsters

DOI: 10.1651/09-3252.1

INTRODUCTION

Microsatellite loci are ubiquitous in eukaryotes. Their co-dominant and often highly polymorphic nature now renders them one of the most widely used molecular markers for population genetics studies (Wilder and Hollocher, 2001). These marker loci have been relatively easily developed for many species using a range of readily available protocols (Selkoe and Toonen, 2006; Zane et al., 2002; Zhang, 2004). However, this has not been the case for a number of other species where development has been problematic such as shrimps (Weetman et al., 2007 and references therein), prawns (Brooker et al., 2000), gastropods (McInerney et al., 2010), and insects (Anderson et al., 2007; Meglécz et al., 2004, 2007; Van't Hof et al., 2007).

Thus far, the difficulty with microsatellite development has been predominantly associated with their presumed low abundance in the genome of particular species from different taxonomic groups (Meglécz et al., 2007; Ross et al., 2003; Van't Hof et al., 2007). However, while enrichment techniques have been useful in increasing yields of microsatellite containing sequences, this approach has not always worked. This suggests that the relatively low abundance of microsatellites is not always the main impediment in developing markers. Recent work on lepidopterans has reported on the difficulties associated with the isolation and characterization of microsatellite markers for this group of insects (Bogdanovicz et al., 1997; Keyghobadi et al., 2002; Meglécz et al., 2004, 2007; Meglécz and Solignac, 1998; Palo et al., 1995; Van't Hof et al., 2007).

Whilst microsatellites are not uncommon in the genome of lepidopterans, it is particularly difficult to design reliable single locus primer sets and to establish adequate conditions for reliable and reproducible PCR amplification. Even in instances where microsatellites have been successfully developed for lepidopterans, in most cases, these were associated with significant departures from Hardy-Weinberg expectations (Meglécz et al., 2004). Previous studies have suggested that these problems are related to the complex nature of microsatellite flanking regions in this group of insects (Meglécz et al., 2004, 2007; Van't Hof et al., 2007).

By comparing microsatellite flanking regions in two butterfly species, Meglécz et al. (2004) reported on the incidence of a high degree of sequence similarity among flanking regions of different microsatellites, and in particular for those from the same species. Further comprehensive work (Meglécz et al., 2007; Van't Hof et al., 2007) confirmed this finding across many insect species. The authors argued that this phenomenon can be explained by a number of alternative or combinatorial hypotheses including amplification/multiplication events involving microsatellite containing regions, microsatellites embedded in minisatellite regions or recombination mediated events. The authors also suggested a possible association between microsatellites and mobile elements, a phenomenon which has been commonly observed in plants, e.g., species of *Eucalyptus* (Rabello et al., 2005), barley (Ramsey et al., 1999), and rice (Temnykh et al., 2001), and also reported in vertebrates such as humans (Nadir et al.,

1996), primates (Arcot et al., 1995), and pigs (Alexander et al., 1995).

During the development of microsatellite PCR primer sets from enriched genomic libraries for three species of galatheid squat lobster, *Munida rugosa* (Fabricius, 1775), *M. sarsi* Huus, 1935, and *Galathea strigosa* (Linnaeus, 1761), a number of unforeseen problems were encountered. While a large number of clones from these libraries were found to contain clear microsatellites and apparent unique flanking sequences for PCR primer design, the majority (93%) of the primer sets designed failed to work for no obvious reasons. Our hypothesis is that, similar to what has been observed in insects; microsatellite flanking regions of galatheid crustaceans are very complex. To test this, we compare both microsatellites and other Variable Number of Tandem Repeat (VNTR) containing sequences of three galatheid species *M. rugosa*, *M. sarsi* and *G. strigosa*.

MATERIALS AND METHODS

Microsatellite Library Development and Sequencing

The approach used for the development and isolation of microsatellite containing regions for *M. rugosa*, *M. sarsi*, and *G. strigosa* followed the protocol described by Kijas et al. (1994) involving microsatellite enrichment using biotinylated oligonucleotides with modifications as reported in Boston et al. (2009) and McInerney et al. (2008; 2009). A by-product of the microsatellite development protocol used, were sequences containing small minisatellite repeat motifs (i.e., repeat units > 10 bp). While these minisatellite containing sequences were not used for PCR primer design (i.e., resulting products are usually too large for reliable PCR amplification), they were considered for subsequent BLAST analyses (see below). In combination, microsatellite and minisatellite containing regions are therein referred to as VNTRs.

Sequence Analysis

While Meglécz et al. (2007) have described a computer program (Microfamily) for detecting flanking regions similarities among different microsatellite loci, we elected to use the approach described by Meglécz et al. (2004) as we believe that it allows for a more thorough examination of the results of VNTR containing regions with inconspicuous short or semi-repetitive flanking regions. The approach of Meglécz et al. (2004) basically involves an all-against-all pairwise comparison of all VNTR containing sequences; including those with short or with only one flanking region. This was conducted for the three galatheid species through a BLASTN search (Altschul et al., 1990).

To test for the presence of transposable elements, all VNTR containing sequences identified within the data set were blasted against an online reference collection of transposable elements using the Censor software tool (Kohany et al., 2006). A comparison was also carried out with all available galatheid mobile elements currently reported on GenBank, using default parameters.

RESULTS

Development and Analysis of VNTR Containing Sequences

Over the three species, a total of 10,950 recombinant clones were screened for microsatellites. From these, 393 clones (4%) that were identified as positive by hybridisation were sequenced in both directions. Removal of false positives (14%) and poor sequences (22%) resulted in a total of 252 VNTR containing sequences (see Table 1 for details). Of these, 167 (66%) were not considered for primer design for the following reasons: 56 contained minisatellite DNA, 104 lacked suitable flanking regions; and seven contained

Table 1.   Classification of 252 clones screened for suitability to produce primer-pairs for PCR amplification (of microsatellites only).

Species	*Galathea strigosa*	*Munida rugosa*	*Munida sarsi*	All
	Satellite and flanking sequences suitability for primer design			
Clones	59	90	103	252
Potentially useable	16	32	37	85
Flanks not suitable	27	31	46	104
Minisatellite	12	25	19	56
	Sequence similarity in flanking regions			
Apparently unique	28	53	52	133
Identical	11	1	1	13
Shared sequence > 40 bp	10	20	25	55
Shared sequence < 40 bp	10	16	25	51

interspaced complex microsatellites with a flanking region similar to the SSR motif.

The remaining 85 microsatellite containing sequences had potentially suitably sized flanking regions for primer design. PCR primer sets were designed using PrimerSelect software (DNAstar Inc) for 60 (70%) of these. Despite extensive optimisation attempts (e.g., varying annealing temperature, $MgCl_2$ and DNA concentration), including redesigned primers, 85% of all primer sets resulted in the amplification of products of incorrect size and/or multiple fragments. The remaining primers sets (15%) failed to amplify any products. Careful examination of the latter, indicate that they were associated with microsatellite containing sequences with limited (i.e., poor) flanking regions, thus it is not entirely surprising that they were not successful.

Subsequent BLAST analysis identified a surprisingly high number of cryptic repeated elements consisting of short DNA stretches of ≥ 40 bp in length in the 'unique' flanking regions of these microsatellites. To test whether these cryptic repeated elements were responsible for the PCR amplification problems, a number of new primer sets ($N = 10$) were redesigned avoiding them, and used in further PCR amplification attempts. These, however, also failed to produce reliable and/or consistent PCR amplification.

Further detailed examination of results from the BLAST analysis, revealed many additional cryptic repeated elements (shorter than 40 bp) in the microsatellite flanking regions of the different clones. Taking these in consideration, only 32 of the initial 85 suitable microsatellite containing sequences were clearly 'unique' within our data set (six *G. strigosa*, 13 *M. rugosa* and 13 *M. sarsi*). Based on this information, additional primers sets were designed for these sequences.

In this third attempt, an additional 16 primer sets were designed (total no. of primers = 86), of which six sets proved successful in the amplification of products corresponding to single microsatellite loci (i.e., one or two allelic fragments per individual). In other instances, however, the same amplification anomalies described earlier were observed, i.e., amplification of products of incorrect size and/or multiple fragments. These results clearly confirm the high degree of complexity in otherwise 'unique' DNA regions flanking microsatellites in these species.

Table 2. Number of cryptic repeated elements appearing (above 40 bp in length) in multiple clones within and between galatheid species.

Species (total no. of clones)	*Galathea strigosa* (59)	*Munida rugosa* (103)	*Munida sarsi* (90)
Galathea strigosa	9	0	1
Munida rugosa		2	41
Munida sarsi			2

Families of Cryptic Repeats in VNTR Flanking Regions

In addition to the identification of cryptic repeated elements in the microsatellite containing sequences, the BLAST analysis incorporating all 252 VNTR containing sequences revealed a number of interesting features. Partial results of this analysis are summarised in Table 2. Species

specific cryptic elements ranged from two to nine. Cryptic elements shared among flanking regions of VNTRs from different species were primarily observed in comparisons involving species of *Munida* (41) with only a single instance of an inter-genus cryptic element.

All cryptic repeated elements identified in the flanking regions could be reliably grouped according to sequence similarity into distinct DNA families (see Fig. 1 for an illustration). Within these families, matches among cryptic repeated elements often involved ~ 100 bp regions with ~ 96% sequence similarity. Among these families, distinct patterns were noticed. For instance, the microsatellite motif within Families 1 (motif-CAG), 4 (motif-GTC) and 5 (motif-GTC) were the same among constituent sequences. Families 2 and 3, however, while sharing cryptic repeated
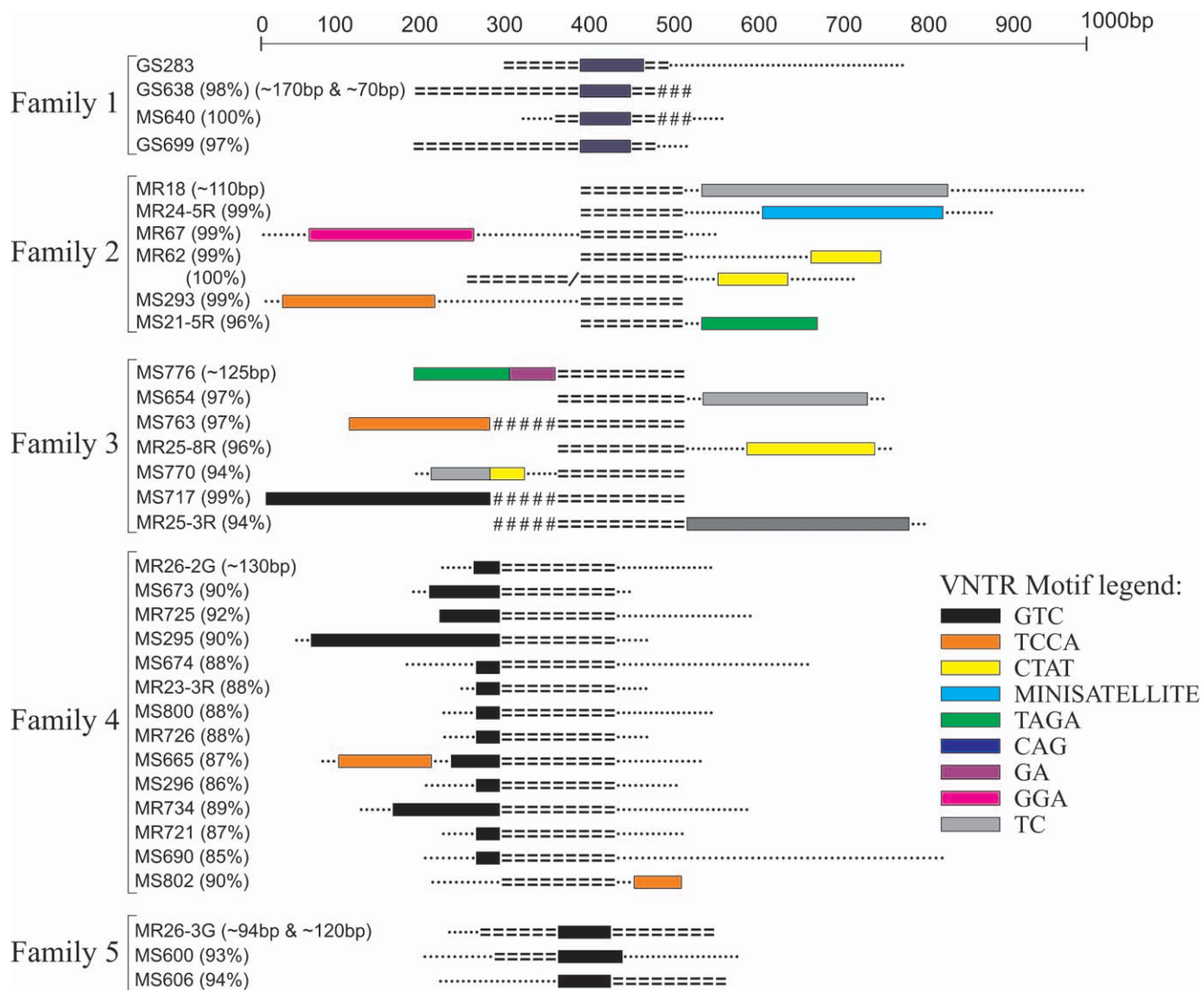

Fig. 1. Diagram illustrating five DNA families of related cryptic repeated elements within flanking regions of distinct microsatellite containing regions identified from *Munida rugosa* (MR), *M. sarsi* (MS) and *Galathea strigosa* (GS). Percentage values (in parenthesis) correspond to the results of BLAST sequence comparison of the flanking regions of each member of a particular group against a reference within the group (the first sequence in each group), and do not take into account the microsatellite repeat motif. The length of the sequences/regions is roughly proportional to the scale on the top of the diagram. The motif legend details the motif repeat identified by colour in each clone. The tram symbol (=) represent DNA regions that are shared (with varying degree of similarity) among sequences belonging to a given group, this element is not repetitive in nature within itself but appears to be a stretch of unique DNA. Multiple dots (…) represent putative unique DNA sequence within flanking regions and (####) represent regions of similarity within a group not shared by all members.
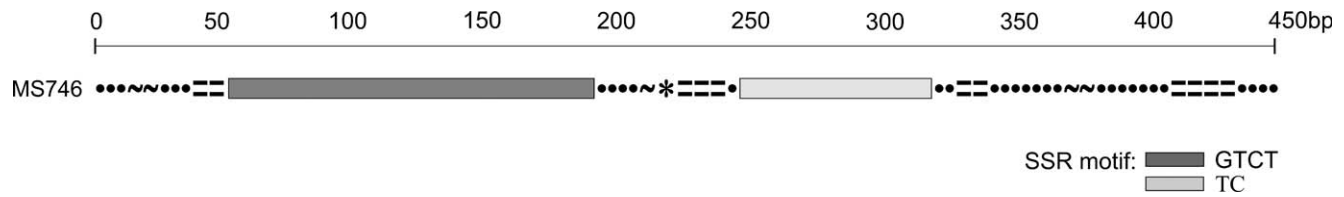
Fig. 2.   Diagrammatic illustration of sequence (repeat motif and flanking region) MS746 for which four microsatellite primers were designed to attempt PCR amplification. The tram lines (===) indicate the regions which were shared (BLAST search) with similar sequences in the flanking regions of other distinct microsatellite sequences within the squat lobster database. The symbol (~~) underlines the sequence where PCR primers were designed upon. A region between the two SSR repeats (motif identified in legend) illustrated overlaps where two PCR primers were designed on a repetitive region (*). Multiple dots (…) represent the parts of the sequence that do not match any other sequence in the squat lobster database therefore are assumed to be 'unique'.

elements in the flanking regions of each microsatellite containing sequence, did not share the same microsatellite motifs (see Fig. 1). Families 1 (comprised of 4 clones) and 5 (comprised of 3 clones) display similarities in the cryptic repeated elements in the flanking region on both sides of the microsatellite, while the three other families (2, 3 and 4) share cryptic repeated elements on a single side only. Interestingly, an instance of a duplicated cryptic repeated element occurring twice within a single flanking region was observed in Family 2 (clone MR62). In Family 3, all clones share one cryptic repeated element, but half also share a second unrelated cryptic repeated element. Family 4, the largest in our data set, is comprised of 14 clones of *Munida* from both species.

A varying number of shorter additional distinct cryptic repeated elements (~ 12-16 bp in average) were also found in the flanking regions of 'suitable' microsatellite containing sequences (see Fig. 2). Their number, i.e., different 12-16 bp elements, ranged from one to five within microsatellite flanking regions of a particular clone, but no more than once per clone and the same 12-16 bp repeat can occur up to 15 times across different clones. Among different VNTR containing sequences, however, these shorter cryptic elements can also be grouped into families. One particular 16 bp element was characterised as a direct or inverted repeat in 11% of all sequences analysed and in 21% of sequences already grouped into DNA families. This cryptic repeated element is identified in four of the seven clones of Family 3 (data not shown). Close examination of microsatellite containing sequences for which primers failed, suggests that these shorter cryptic repeated elements are also likely to be responsible, i.e., primers annealing to these inconspicuous short repeated regions (Fig. 2) results in the PCR amplification of either multiple fragments or single fragments of incorrect size.

Transposable Elements

Comprehensive scanning of all 252 squat lobster VNTR containing sequences against the Censor software tool using default settings resulted in 111 matches with known repetitive/mobile elements. These included transposons (35%), LTR retrotransposons (5%), non-LTR retrotransposons (19%) and endogenous retroviruses (8%). No noticeable associations were observed between galatheid species or genus and particular types of transposable elements.

Of the 27 short direct or inverted repeats identified (~ 16 bp), 56% of sequences containing these correlated to a known mobile element through BLAST analysis. The 28 highest similarity scores resulting from the BLAST analysis (involving all sequences) are shown in Table 3. These cover all of the main eukaryote groups, i.e., 27% matched mammals, 17% plants, 20% insects and 13% birds. No significant matches were found with the limited number of the reported crustacean and/or galatheid (Terrat et al., 2008; Piednoel and Bonnivard, 2009) mobile elements.

DISCUSSION

While microsatellites have been successfully used in crustacean population genetic studies (Beacham et al., 2008; Ball et al., 2003; Ferguson et al., 2002; Kenchington et al., 2009), an increasing number of investigations report on the difficulties in their isolation and optimisation (Brooker et al., 2000; Moore et al., 1999; Robainas et al., 2002; Weetman et al., 2007). Excluding the present investigation, only one additional study has reported on the occurrence of VNTRs in galatheid species (Cabezas et al., 2009), and the authors hint at the difficulties associated with successful microsatellite optimisation and application in this group.

The detailed breakdown and characterisation of sequences reported here suggest that VNTRs are abundant for the three galatheid species despite only a small number of reliable PCR primer sets being developed. Bias due to technical artefacts resulting from the enrichment protocol used for microsatellite development can be dismissed as a possible reason for the lack of success of PCR primer sets. The same enrichment protocol has been successfully used within our research group for the development of microsatellite primer sets for mammals (Boston et al., 2009), gastropods (McInerney et al., 2008, 2009), fish, and other crustacean species at the time of the galatheid microsatellite development. It is unlikely that technical bias was restricted only to the three independent galatheid enriched libraries, thus, the observations reported in this study reflect true genome complexities in these species. Furthermore, at least in insects, Meglecz et al. (2007) have demonstrated that similarities among microsatellite flanking regions is independent both from methodological approaches (whole genomes comparisons, whole genome shotgun sequences comparisons and partial genomic

Table 3.   Positive matches of squat lobster sequences with known mobile elements detected through BLAST searches against the Censor GIRI database. Identified mobile elements are listed including type and associated species/group. The 'Similarity' statistics is calculated by Censor, and converted to percentage for simplification. BLAST score is simply the alignment score provided through the BLAST search.

Galatheid sequence ID	Size of match (bp)	Matching mobile element	Size of mobile element (bp)	Element type	Element associated species	Species group	Accession no.	Similarity (%)	Blast alignment score
MS649	214	SQR2_MM	632	Satellite	*Mus musculus*	Mammal	GU384896	77.6	626
MS647	177	SQR2_MM	632	Satellite	*Mus musculus*	Mammal	GU384897	80.13	559
GS695	114	RMER17D_MM	923	Non-LTR Retrotransposon	*Mus musculus*	Mammal	GU384898	78.95	391
GS638	113	GOLEM	3029	Mariner DNA transposon	*Eutheria*	Mammal	GU384899	72.17	372
MS775	75	RMER17A	770	Endogenous Retrovirus	*Mus musculus*	Mammal	GU384900	80.52	305
GS110	84	LTR18F_ML	761	Endogenous Retrovirus	*Myotis lucifugus*	Mammal	GU384901	78.67	280
GS796	82	Tigger2f	3455	Mariner DNA transposon	*Carnivora*	Mammal	GU384902	77.5	279
MS604	52	BGLII_Rat_LTR	420	Endogenous Retrovirus	*Rattus*	Mammal	GU384903	79.25	237
MS298	229	TREP60	334	microsatellite DNA	*Hordeum vulgare*	Plant	GU384904	75.45	671
MR58	216	Harbinger2N_TP	1430	nonautonomous DNA transposon	*Thalassiosira pseudonana*	Plant	GU384905	78.46	282
GS752	36	ATLANTYS_LC_I	12848	LTR retrotransposon	*Lotus japonicus*	Plant	GU384906	89.19	279
MS758	57	EnSpm-12_ZM	12258	DNA transposon	*Oryza sativa*	Plant	GU384907	78.95	260
MS758	66	MERMITE	1401	non-autonomous DNA transposon	*Oryza sativa*	Plant	GU384907	76.92	206
MR22-3G	80	MINIME_DN	1218	mini-me retrotransposon	*Drosophila*	Insect	GU384908	85.53	432
MS292	63	Copia17-NVi_I	7603	LTR retrotransposon	*Nasonia vitripennis*	Insect	GU384909	82.81	322
MR23-3R	33	MINIME_DN	1218	mini-me retrotransposon	*Drosophila*	Insect	GU384910	91.18	279
MR65	49	NAVIRTE1	5663	Non-LTR Retrotransposon	*Nasonia vitripennis*	Insect	GU384911	80.39	232
MR721	40	ISFUN1	928	direct repeat	*Drosophila funebris*	Insect	GU384912	80.49	227
MR65	43	G2_DM	3102	non_LTR retrotransposon	*Drosophila melanogaster*	Insect	GU384911	77.27	222
MR677	48	PlatCR1	3589	Non-LTR Retrotransposon	*Ornithorhynchus*	Bird	GU384913	77.55	254
GS702	50	CR1-X3_Pass	4471	Non-LTR Retrotransposon	*Passeriformes*	Bird	GU384914	82.69	245
MR27-2G	45	TguLTRL2a7	1387	Endogenous Retrovirus	*Estrildidae*	Bird	GU384915	83.72	231
MS745	36	CVA	465	DNA transposon	*Crassostrea virginica*	Mollusc	GU384916	86.84	214
MS80	46	Penelope-4_NV	3143	non-LTR retrotransposon	*Nematostella vectensis*	Cnidarian	GU384917	83.33	209
GS635	51	hAT-60_HM	3293	DNA transposon	*Hydra magnipapillata*	Cnidarian	GU384918	79.63	234
GS636	41	REX3	2223	non-LTR retrotransposon	*Xiphophorus maculatus*	Fish	GU384919	85.71	256
GS699	32	Copia2-I_CR	5509	Retrotransposon	*Chlamydomonas reinhardtii*	Algae	GU384920	93.94	239

libraries for microsatellites) and species investigated. While such a comparison has not been done in here, it is reasonable to assume this will also apply for galatheids.

Alternative primer development characterised as REpetitive Flanking sequences (ReFS) was introduced by Anderson et al. (2007) and undertaken in the current study. This single primer approach confirmed that the cryptic repeated elements identified here are not artefacts but elements that are common throughout the galatheid species genomes. Thus, the ReFS approach produced multi-banded fingerprints, which were polymorphic both between and among species (for further details see Bailie, 2008). Therefore, similar to the findings of Meglecz, et al. (2004; 2007) it would appear that, at least some crustacean species have increased levels of complexities in their flanking regions when in comparison to others.

The high degree of sequence similarity of cryptic repeated elements among flanking regions of different microsatellite loci has previously been reported in insects (Anderson et al., 2007; Meglecz et al., 2004, 2007; Sunnucks and Wilson, 1999; Van't Hof et al., 2007; Wu et al., 1999). It is often the case that negative results are seldom published and despite crustacean investigations having reported difficulties in obtaining microsatellite markers (references aforementioned), this is the first to elaborate on the respective causes. Results of the current investigation are the first to relate the lack of success in microsatellite development for any crustacean species to the similarities in cryptic repeated elements identified. This pattern was observed across all three galatheid species investigated, and it is likely that such similarities are more widespread in other unrelated taxonomic groups facing difficulties (detailed above) but have yet to be investigated and reported.

Galatheid Microsatellite DNA Families

The results of this investigation combined with previously published work on lepidopterans raises many questions concerning the origin and successful use of microsatellites as genetic markers. The origin of microsatellites has been the subject of much debate in the literature (Goldstein and Schlötterer, 1999). The common assumption is that a VNTR arises by a fortuitous generation of a small tandem duplication, which then undergoes length changes by replication slippage (Schlötterer and Tautz, 1992). In this case, the association of microsatellites with interspersed repeats would be random and depend on their frequency in the genome. While it is impossible to determine this without analysis of a vast amount of genome sequences, Family 2 and 3, with shared sequences at various distances from different VNTR regions, support this model. A high incidence of these cryptic repeats would be expected in genomes that have a large quantity of repeated sequences and that are relatively young, (less than 20-50 Myrs old) depending on the original copy number.

Another possible explanation of the cryptic repeats is that they are derived from tandem repeats that have

mutated randomly so the fragments of the original sequence are separated by regions mutated to 'uniqueness'. Twenty of the sequenced clones contained evidence of tandem repeats of units ranging from $\sim$ 12-60 bp, some almost perfect, some containing longer units comprised of smaller components and some with recurrent motifs of 4-6 bp separated by widely diverged sequences. To put this into perspective, primer sized units of $\sim$ 25 bp at a mutation rate of 10^{-8} per base pair per year would have a half-life of 2.7Myr and 10% would survive without mutation for 9 Myr. We found no evidence for recombination and/or conversion mediated events as plausible explanations for the apparent association of cryptic repeated elements with VNTR containing regions as suggested by Van't Hof et al. (2007).

Mobile Element Association to Galatheid DNA Families

The pattern observed in Families 1, 4 and 5, which indicates an association of VNTRs with similar cryptic repeated elements in their flanking regions, suggests that the whole region, microsatellite and flank, pre-existed and were duplicated together. The abrupt loss of homology with distance implies that the transposition or insertion of a short region, or that this association has taken place in conjunction with a mobile element, which has since been excised. Consequently, a number of investigations have reported an association between transposable elements and VNTRs, e.g., species of *Eucalyptus* (Rabello et al., 2005), barley, (Ramsey et al., 1999), rice (Temnykh et al., 2001), and Diptera (Wilder and Hollocher, 2001).

Under this model, the original region containing a proto-VNTR would require a small mutation to create the initial repeat, a proportion of which would eventually develop into a VNTR at that position. The flanking sequences will only be homologous if they are co-transposed with the microsatellite suggesting that the transposable element may contain the proto-VNTR (Wilder and Hollocher, 2001) and the multiplication/integration event is a hitchhiking process of both the microsatellite and its flanking regions. Therefore, it is possible that the short repeats are generated occasionally as part of the integration or excision of transposable elements. The integrating element might be reverse transcribed from cellular RNA, as are human *Alu* elements (a retrotransposon), and might coincidentally contain or generate a short microrepeat, or might be a Class II transposon, which move by a "cut and paste" process. The variable nature of transposon excision events makes this model compatible with all the repeat families described here.

A BLAST search of squat lobster sequences against a database of mobile elements yielded a surprisingly high number of positive hits to other taxa, a trend not observed in other studies (Meglécz et al., 2004; Van't Hof et al., 2007). A comprehensive investigation by Meglécz et al. (2007) included 1236 sequences from 23 insect species and reported, in the case of *Drosophila melanogaster*, that 6.4% of microsatellite sequences showed similarities to a repetitive element bank for *Drosophila*. As a result, Meglécz et al. (2007) suggested that repetitive element banks are generally species specific for each genus or target species. It is therefore remarkable that the percentage of positive hits in each galatheid species against a general mobile element bank were 12.3% for *G. strigosa*, 15.5% for *M. rugosa* and 16.3% for *M. sarsi*. This would lead one to hypothesise that the associations observed in the present study are an underestimation, due to the scarcity of information on the mobile elements in galatheid species. This factor is also reinforced by the lack of success in similar comparisons with the only known transposable elements reported in galatheids (Terrat et al., 2008; Piednoel and Bonnivard, 2009). The identification of cryptic repeated elements without homology to the mobile element database suggests that there are galatheid transposable elements yet to be characterised.

The short direct or inverted repeats identified in the current study are characteristic of class II transposons, which utilize a transposase enzyme that binds to direct/inverted repeats flanking the transposable element during the process of transposition. These repeats occur in one fifth of the DNA families clones identified. More than half of these clones containing the direct short/inverted repeat relate to a transposable element, e.g., from families *Copia*, *hAT*, *Polinton* and *mini-me*, and a further 19%, relate solely to galatheid microsatellite DNA families. These findings seem to provide evidence for a novel galatheid mobile element, which is/has been recently active. Unfortunately the short nature of the sequences of the current dataset only permitted the identification of a single short direct/inverted repeat per sequence indicating that the mobile element is much larger than this study can characterise.

Conclusions

The discovery of inconspicuous cryptic repeats (both above and below 40 bp) adjacent to microsatellites explains the difficulties encountered in microsatellite marker development because the flanking regions are not, in fact, "unique". The frequent presence of sequences homologous to transposable elements suggests a role for transposition in the production or duplication of some of the galatheid microsatellite DNA families observed. In other cases the association appears to be random, and is consistent with microsatellites simply occurring in a genome with a large number of dispersed repeated sequences. There is also the possibility that some of the microsatellites are simply embedded in very old blocks of tandem repeated sequences where mutations have exponentially obscured many of the repeats, but left some almost intact. There is no evidence for recombination or unequal crossing-over, but in a short sequence this would be indistinguishable from one end of a transposition. Apart from this, these results concur generally with the theories of Meglecz et al. (2004; 2007) and Van't Hof et al. (2007), and also show that this phenomenon is not confined to insects. These conclusions suggest the hypothesis that difficulty in developing microsatellite markers for PCR is increased in genomes with large numbers of dispersed repeats or degenerate tandem repeats.

Future Considerations

From the outset of any investigation aiming at microsatellite development, it is difficult to predict problems as the

ones reported in here. In order to avoid wasteful expenditure, and both time and labour resources on the development of primers that are unlikely to work, at least for crustaceans, we suggest a number of steps that could be followed to maximize chances of success. For instance, $C_o t$ analysis for repeat content in a genome could be used to predict difficulties in primer design and it would be interesting to determine, in future investigations, if such a correlation does exist. This would be a more reliable indicator than genome size as the bulk of large genomes may be old degenerate repeats. A small number of transposable elements carrying the common microsatellites would disrupt primer development without being apparent in the genome.

A simple technical improvement aiming to identify reproducible microsatellites would be to clone larger fragments, i.e,. by using infrequent cutters restriction enzymes, which would permit the development of longer specific primers. This was attempted for a few sequences in the current database but to no avail. This suggests that the extent of complexities observed in disturbed regions of the genome require further study and are beyond the scope of the current investigation. In any event, local BLAST analysis, in combination with similar BLAST searches against transposable element databases, is recommended to compare all sequenced clones in order to attempt to uncover cryptic repeats or widespread families of transposable elements that might otherwise go unnoticed. If the underlying cause for unusual results is associated with tandem repeats, given that these usually occur in blocks of heterochromatin, it is plausible that useful microsatellites may be found near coding sequences. Thus, screening cDNA libraries for microsatellites could potentially yield a higher frequency of useable clones, as has been applied to *Aedes japonicus* (Widdel et al., 2005).

Alternatively or complementary to future microsatellite marker development, the mobile element-microsatellite association can be capitalised upon by molecular marker systems such as those applied to plant genomes, e.g., Sequence Specific Amplified Polymorphism (SSAP) (Kalendar et al., 2006), Inter-Retrotransposon Amplified Polymorphism (IRAP) (Schulman et al., 2004) and REtrotransposon-Microsatellite Amplified Polymorphism (RE-MAP) (Kalendar et al., 2006) or those more recently applied to lepidopteran species, e.g., ReFS (Anderson et al., 2007). The discovery of mariner transposons in the current study and in a wide spectrum of marine crustacean and mollusc genomes (Casse et al., 2006 and references therein: Bui et al., 2008; Halaimia-Toumi et al., 2004) raises the possibility of applying such marker systems, i.e., IRAP and/or REMAP to future galatheid studies.

Interestingly, the incidence of cross species similarities observed in the current study is extensive in comparison to the reports by Meglécz et al. (2004), in which few cross species similarities were observed. Alternatively, Van't Hof et al. (2007) identified four elements, referred to as *Lepidoptera*-specific core sequences (LSCS) generally associated with a microsatellite and/or retrotransposons that were shared across varying numbers of lepidopteran species. The higher incidence of shared similarities between closely related *Munida* species (Group 4, Fig. 2)

raises interesting questions in terms of the genome associations of microsatellites and mobile elements pre and post species divergence from one another. There would appear to have been extensive recent genomic reshuffling and the species divergence is most probably recent ($\sim$ 3-11 million years ago, Bailie, 2008). An alternative phylogenetic approach could be considered, given the existence of such retrotransposed elements and their flanking regions, this could be used to clarify the hazy evolutionary history of galatheids, as has recently been conducted in mammals (Moller-Krull et al., 2007). Despite laborious time and effort invested in microsatellite development, three polymorphic and reproducible microsatellite primers were successfully obtained; these will permit testing of hypotheses relating to paternity and population genetics in the galatheid species (Bailie and Prodöhl, submitted).

While commonly available protocols for microsatellite development are likely to continue to be used for producing species specific markers, recent advances in sequencing methodology can potentially provide a better alternative to avoid many of the problems detailed in the present study. New generation sequencing (NGS) platforms based on very high throughput parallel and multiplexing genotyping, e.g., Roche 454, Illumina GA, and the ABI SOLiD technologies, can analyse the equivalent of a cloned library in a day. This easily allows for the provision of sufficient data to cross align the microsatellite flanking regions with enough of the genome to get an idea of their uniqueness. This approach also permits the design of longer PCR primers and the identification of useful microsatellites, which are near coding regions. Furthermore, such an approach would completely eliminate any possibility of sampling bias resulting from enrichment processes used to produce microsatellite libraries. This strategy should dramatically improve the ability to obtain reliable genetic markers for population studies and also to facilitate the identification of other unique sequence polymorphisms. While NGS technology is still expensive, we anticipate this will rapidly change with more commercial companies offering affordable access to the new technology. Given the problems reported in the present study, we argue that NGS may be the way forward for crustacean genetic research.

REFERENCES

Alexander, L. J., G. A. Rohrer, R. T. Stone, and C. W. Beattie. 1995. Porcine SINE associated microsatellite markers: evidence for the new artiodactyls SINEs. Mammalian Genome 6: 464-468.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

Anderson, S. J., P. Gould, and J. R. Freeland. 2007. Repetitive flanking sequences (ReFS): novel molecular markers from microsatellite families. Molecular Ecology Notes 7: 374-376.

Arcot, S. S., Z. Wang, J. L. Weber, P. L. Deninger, and M. A. Batzer. 1995. *Alu* repeats: A source for the genesis of primate microsatellites. Genomics 29: 136-144.

Bailie, D. A. 2008. Phylogeny, population genetics and mating strategies of the squat lobster species *Munida rugosa* and *Munida sarsi* (Crustacea, Decapoda, Galatheidae). Ph.D. Thesis, Queens University, Belfast.

Ball, A. O., and R. W. Chapman. 2003. Population genetic analysis of white shrimp, *Litopenaeus setiferus*, using microsatellite genetic markers. Molecular Ecology 12: 2319-2330.

Beacham, T. D., J. Supernault, and K. M. Miller. 2008. Population structure of Dungeness crab (*Cancer magister*) in British Colombia. Journal of Shellfish Research 27: 901-906.

Bogdanovicz, S. M., V. S. Mastro, D. C. Prasher, and R. G. Harrison. 1997. Microsatellite variation among Asian and North American gypsy moths (Lepidoptera:Lymantriidae). Annals of Entomological Society of America 90: 768-775.

Boston, E. S. M., I. Montgomery, and P. A. Prodöhl. 2009. Development and characterisation of 11 polymorphic compound tetranucleotide microsatellite loci for the Leisler's bat, *Nyctalus leisleri* (Vespertilionidae, Chiroptera). Conservation Genetics 10: 1501-1504.

Brooker, A. L., J. A. H. Benzie, D. Blair, and J-J. Versini. 2000. Population structure of the giant tiger prawn *Penaeus monodon* in Australian waters, determined using the microsatellite markers. Marine Biology 136:149-157.

Bui, Q-T., N. Casse, V. Leignel, V. Nicolas, and B. Chenais. 2008. Widespread occurrence of *mariner* transposons in coastal crabs. Molecular Phylogenetics and Evolution 47: 1181-1189.

Cabezas, P., P. Bloor, I. Acevedo, C. Toledo, M. Calvo, E. Macpherson, and A. Machordom. 2009. Development and characterisation of microsatellite markers for the endangered anchialine squat lobster *Munidopsis polymorhpa*. Conservation Genetics 10: 673-676.

Casse, N., Q. T. Bui, V. Nicolas, S. Renault, Y. Bigot, and M. Laulier. 2006. Species sympatry and horizontal transfer of *Mariner* transposons in marine crustacean genomes. Molecular Phylogenetics and Evolution 40: 609-619.

Fabricius, J. C. 1775 Systema Entomologiae, sistens Insectorum Classes, Ordines, Genera, Species, adjectis Sysnonymis, Locis, Descriptionibus, Observationibus, pp. 832. Kortii, Flensburgi et Lipsiae.

Ferguson, A. (coord.). 2002. Genetic diversity in the European lobster (*Homarus gammarus*): population structure and impacts of stock enhancement. Available online at: http://www.qub.ac.uk/bb-old/prodohl/GEL/gel.html.

Goldstein, D. B., and C. Schlotterer. 1999. Microsatellites: Evolution and Applications. Oxford University Press, Oxford.

Halaimia-Toumi, N., N. Casse, M. Demattei, S. Renault, E. Pradier, Y. Bigot, and M. Laulier 2004. The GC-rich transposon Bytmar1 from the deep-sea hydrothermal crab, *Bythograea thermydron*, may encode three transposase isoforms from a single ORF. Journal of Molecular Evolution 59: 747-760.

Huus, J. 1935. Zur morphologisch-systematischen und biologischen Kenninis der nordischen Munida-Larven (Crustacea Decapoda). 8 (1934), 1-29, figs 1-6, pls 1-4. Bergen Museum Aarbok.

Kalendar, R., and H. A. Schulman. 2006. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nature Protocols 1: 2478-2484.

Kenchington, E. L., G. C. Harding, M. W. Jones, and P. A. Prodöhl. 2009. Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*. Molecular Ecology 18: 1654-1667.

Keyghobadi, N., J. Roland, and C. Strobeck. 2002. Isolation of novel microsatellite loci in the Rocky Mountain apollo butterfly, *Parnassius smintheus*. Hereditas 136: 247-250.

Kijas, J. M. H., J. C. S. Fowler, C. A. Garbett, and M. R. Thomas. 1994. Enrichment of microsatellites from citrus genome using biotynilated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16: 657-662.

Kohany, O., A. J. Gentles, L. Hankus, and J. Jurkas. 2006. Annotation, submission and screening of repetitive elements in Repbase: Repbase-Submitter and Censor. BMC Bioinformatics 25: 474.

Linnaeus, C. 1761. Fauna Suecica sistens Animalia Sueciae Regni: Mammalia, Aves, Amphibia, Pisces, Insecta, Vermes. Distributa per Classes, Ordines, Genera, Species, cum Differentiis Specierum, Synonymis Auctorum, Nominibus Incolarum, pp. 578. Locis Natalium, Descriptionibus Insectorum.

McInerney, C. E., A. L. Allcock, M. P. Johnston, and P. A. Prodöhl. 2008. Characterisation of polymorphic microsatellites for the periwinkle gastropod, *Littorina littorea* (Linnaeus, 1758) and their cross-species amplification in four congeners. Conservation Genetics 10: 1417-1420.

———, ———, ———, and ———. 2009. Characterization of polymorphic microsatellite loci for the rough periwinkle gastropod *Littorina saxatilis* (Olivi, 1972) and their conservation in four congeners. Conservation Genetics 10: 1989-1992.

———, ———, ———, D. A. Bailie, and P. A. Prodöhl. (in press.) Comparative genomic analysis reveals species-dependent genomic complexities that explain difficulties with microsatellite marker development in molluscs. Heredity.

Meglécz, E., and M. Solignac. 1998. Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). Hereditas 128: 179-180.

———, F. Petanian, E. Danchin, A. Coeur D'acier, J. Rasplus, and E. Faure. 2004. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. Molecular Ecology 13: 1693-1700.

———, S. J. Anderson, D. Bourguett, R. Butcher, A. Caldas, A. C. Cassel-Lundhagen, A. D'Acier, D. A. Dawson, N. Faure, C. Fauvelot, P. Franck, G. Harper, N. Keyghobadi, C. Kluetsch, M. Muthulakshmi, J. Nagaraju, A. Patt, F. Péténian, J-F. Silvain, and H. R. Wilcock. 2007. Microsatellite flanking region similarities among different loci within insect species. Insect Molecular Biology 16: 175-185.

Moller-Krull, M., F. Delsuc, G. Churakov, C. Marker, M. Superina, J. Brosius, E. J. P. Douzery, and J. Schmitz. 2007. Retroposed elements and their flanking regions resolve the evolutionary history of Xenarthran mammals (Armadillos, Anteaters, and Sloths). Molecular Biology and Evolution 24: 2573-2582.

Moore, S. S., V. Whan, G. P. Davis, K. Byrne, D. J. Z. Hetzel, and N. Preston. 1999. The development and application of genetic markers for the Kumura prawn *Penaeus japonicus*. Aquaculture 173: 19-32.

Nadir, E., H. Margali, T. Gallily, and S. A. Ben-Sasson. 1996. Microsatellite spreading through the human genome: evolutionary mechanisms and structural implications. Proceedings of the National Academy of Sciences USA 93: 6470-6475.

Palo, J., S-L. Varvio, I. Hanski, and R. Väinölä. 1995. Developing microsatellite markers for insect population structure: complex variation in a checker spot butterfly. Hereditas 123: 295-300.

Piednoel, M., and E. Bonnivard. 2009. *DIRS1*-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms. Evolutionary Biology 9: 86.

Rabello, E., A. Nunes de Souza, D. Saito, and S. Mui Tsai. 2005. *In silico* characterization of microsatellites in *Eucalyptus* spp: Abundance, length variation and transposon associations. Genetics and Molecular Biology 28: 582-588.

Ramsey, L., M. Macauley, L. Cardie, M. Morgante, S. Ivanissech, E. Maestri, W. Powell, and R. Waugh. 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. The Plant Journal 17: 415-425.

Robainas, A., M. Monnerot, M. Solignac, N. Dennebouy, G. Espinosa, and E. G. Machado. 2002. Microsatellite loci from the pink shrimp *Farfantepenaeis notialis* (Crustacea, Decapoda). Molecular Ecology Notes 2: 344-345.

Ross, C. L., K. A. Dyer, T. Erez, S. J. Miller, J. Jaenike, and T. A. Markow. 2003. Rapid divergence of microsatellite abundance among species of Drosophila. Molecular Biology and Evolution 20:1143-1157.

Schlötterer, C., and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acid Research 20: 211-215.

Schulman, A. H., A. J. Flavell, and T. H. Ellis. 2004. The application of LTR retrotransposons as molecular markers in plants. Methods of Molecular Biology 260: 145-173.

Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology letters 9: 615-629.

Sunnucks, P., and A. C. C. Wilson. 1999. Microsatellite markers for the onychophoran *Euperipatoides rowelli*. Molecular Ecology 8: 899-900.

Temnykh, S., G. DeClerk, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11: 1441-1452.

Terrat, Y., E. Bonnivard, and D. Higuet. 2008. GalEa retrotransposons from galatheid squat lobsters (Decapoda, Anomura) define a new clade of Ty1/copia-like elements restricted to aquatic species. Molecular Genetics and Genomics 279: 63-73.

Van't Hof, A. E., P. M. Brakefield, I. J. Saccheri, and B. J. Zwaan. 2007. Evolutionary dynamics of multilocus arrangements in the genome of the butterfly *Bicyclus anyana*, with implications for Lepidoptera. Heredity 98: 320-328.

Weetman, D., A. Ruggiero, S. Mariani, P. W. Shaw, A. R. Lawler, and L. Hauser. 2007. Hierarchical population genetic structure in the commercially exploited shrimp *Crangon crangon* identified by AFLP analysis. Marine Biology 151: 565-575.

Widdel, A. K., L. J. McCuiston, W. J. Crans, L. D. Kramer, and D. M. Fonesca. 2005. Finding needles in the haystack: Single copy microsatellite loci for *Aedes japonicus* (Diptera: Culicidae). American Journal of Tropical Medicine and Hygiene 73: 733-748.

Wilder, J., and H. Hollocher. 2001. Mobile elements and the genesis of microsatellites in dipterans. Molecular Biology and Evolution 18: 384-392.

Wu, L., L. Kaufman, and P. A. Fuerst. 1999. Isolation of microsatellite markers in *Astatoreochromas alluaudi* and their cross-species amplifications on other African cichlids. Molecular Ecology 8: 895-897.

Zane, L., L. Bargelloni, and T. Patarnello. 2002. Strategies for microsatellite isolation: a review. Molecular Ecology 11: 1-16.

Zhang, D. 2004 Lepidopteran microsatellite DNA: redundant but promising. Trends in Ecology and Evolution 19: 507-509.