GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

New microsatellite markers for the abalone *Haliotis midae* developed by 454 pyrosequencing and *in silico* analyses

R. Slabbert[1,2], J.-A. Hepple[1,2], C. Rhode[1], A.E. Bester-Van der Merwe[1] and R. Roodt-Wilding[1]

[1]Molecular Aquatic Research Group, Department of Genetics, Stellenbosch University, Matieland, South Africa
[2]Central Analytical Facilities, Stellenbosch University, Matieland, South Africa

Corresponding author: R. Slabbert
E-mail: rslabbrt@sun.ac.za

Genet. Mol. Res. 11 (3): 2769-2779 (2012)
Received September 2, 2011
Accepted July 10, 2012
Published August 24, 2012
DOI http://dx.doi.org/10.4238/2012.August.24.2

ABSTRACT. Farming of *Haliotis midae* is the most lucrative aquaculture venture in South Africa. The genome of this species needs to be studied to assist in selective breeding programs aimed at increasing overall yield, and molecular markers will be required to attain this goal. We identified and characterized 82 polymorphic microsatellite loci by using repeat-enriched genomic libraries and high-throughput pyrosequencing technology. The observed number of alleles ranged from 2 to 21, expected heterozygosity from 0.063 to 0.968, observed heterozygosity from 0.000 to 1.000, and polymorphic information content from 0.059 to 0.934. Three loci gave significant hits to other haliotid genes and/or microsatellite loci; hits to genes were always located in the 5'/3'-UTR or intron region. Many of these newly designed markers would be useful for parentage, population and linkage studies.

Key words: 454 Sequencing; Aquaculture; Bioinformatics; Molecular markers; *Haliotis midae*

INTRODUCTION

Five abalone species live in the waters of South Africa but only one, *Haliotis midae*, is cultivated and exported. The *H. midae* industry is the most lucrative aquaculture sector in South Africa, with 14 active hatcheries and grow-out facilities (DAFF, 2012). Commercial abalone farming of *H. midae* was initiated in 1990 (Cook and Britz, 1991), and 15 years later, a genetics research program was introduced as a collaborative effort among academic institutions, industry, and government (Slabbert et al., 2009b). The overall aim of the program is the genetic characterization and enhancement of *H. midae* within its natural and commercial settings. To achieve this aim, various molecular genetic markers are needed to facilitate different aspects of the program such as genetic diversity studies, pedigree reconstructions, linkage mapping, and quantitative trait loci (QTL) discovery.

Microsatellite markers are both extremely popular in modern molecular disciplines and widely used in abalone genetic studies such as population structure and genetic diversity studies, parentage studies, linkage mapping, and QTL mapping (Baranski et al., 2006; Gutiérrez-Gonzales et al., 2007; Li et al., 2007; Slabbert et al., 2009a; Shi et al., 2010). Microsatellite loci have already been isolated for various abalone species, including *Haliotis rubra* (Evans et al., 2000), *H. asinina* (Selvamani et al., 2000), and *H. discus hannai* (Li and Akihiro, 2007). A number of microsatellite markers have also been published as a result of the *H. midae* genetic research program (Bester et al., 2004; Slabbert et al., 2008, 2010).

The development of microsatellite markers is labor intensive and costly. Genomic libraries must be constructed and enriched for microsatellite repeats, and clones must be screened and then sequenced (Zane et al., 2002). Even more advanced protocols such as fast isolation by amplified fragment length polymorphism (AFLP) of sequences containing repeats (FIASCO; Zane et al., 2002) and the SNX-unilinker method (Hamilton et al., 1999) are also unsuitable for high-throughput marker development owing to labor-intensive cloning and screening steps. The advent of new-generation sequencing technologies such as sequencing by synthesis (Ronaghi et al., 1998) and ligation-mediated sequencing (Shendure et al., 2005) could therefore provide alternative methods for generating large data sets, minimizing effort, and decreasing costs.

Pyrosequencing is based on the real-time acquisition of DNA synthesis data via bioluminescence and is driven by four enzymes: Klenow DNA polymerase I, ATP sulfurylase, luciferase, and apyrase (Ronaghi et al., 1998; Ahmadian et al., 2006). This technology has been used in single-nucleotide polymorphism genotyping (Ahmadian et al., 2000) and expressed sequence tag sequencing (Galindo et al., 2010). Pyrosequencing has also been used to develop microsatellite markers for a number of fungal, insect, bird, reptile, and plant species (Abdelkrim et al., 2009; Allentoft et al., 2009; Santana et al., 2009; Castoe et al., 2010; Blanca et al., 2011). These studies have shown that pyrosequencing is an effective platform for the automation of certain analytical steps within a standard marker development protocol, making it a more time-efficient strategy.

In this study, pyrosequencing was used to generate data from a repeat-enriched genomic library. Primers for microsatellite loci were designed and further characterized. The data generated using pyrosequencing were also screened against National Center for Biotechnology Information (NCBI) databases to identify possible orthologs and gene associations.

MATERIAL AND METHODS

Sample collection and DNA extractions

Sixteen *H. midae* samples were collected from wild fish at Saldanha Bay (west coast of South Africa). Genomic DNA was isolated from muscle tissue using a cetyltrimethylammonium bromide protocol (Saghai-Maroof et al., 1984).

Genomic library construction

The FIASCO method (Zane et al., 2002) was used to construct a repeat-enriched genomic library. A total of 250 ng *H. midae* DNA was digested with *Mse*I (New England Biolabs) and ligated to *Mse*I adaptors. This sample was then selectively amplified and separately enriched with biotinylated $(AC)_{12}$, $(GATC)_6$, $(CAA)_8$, and $(GTGC)_6$ probes and recovered using streptavidin magnetic particles. These enriched particles were again selectively amplified using an *Mse*I-specific primer mix.

Pyrosequencing and primer design

A final amount (5 µg) of polymerase chain reaction (PCR) product was sequenced using the Roche 454 GS-FLX system at Inqaba Biotech (Pretoria, South Africa). Samples were prepared and analyzed according to the manufacturer protocol. Single reads were obtained and contiguous sequences were constructed using the Newbler version 1.1.03.24 software. All the contiguous sequences were trimmed of any adaptor sequences using the Find and Replace function of Microsoft Word. These sequences were then analyzed for length, GC content, and repeat motifs using the online software program BatchPrimer3 version 1 (You et al., 2008). The same software package was used to design primers for repeats containing contiguous sequences for which adequate flanking regions were available. To avoid primer redundancy, all contiguous sequences for which primers were designed were screened against a local Basic Local Alignment Search Tool (BLAST) database (created using BioEdit version 5.0.9; Hall, 1999), which contained all the microsatellite sequences generated thus far for *H. midae*.

Genotyping

All PCR applications were conducted in a Geneamp 2700 thermo cycler (Applied Biosystems; Johannesburg, South Africa) in 10-µL reaction volumes containing 20 ng DNA, 0.2 µM of each primer, 200 µM deoxyribonucleotide triphosphates, 0.1 U 2G Fast *Taq* polymerase (KAPA Biosystems; Cape Town, South Africa), 1X buffer B (KAPA Biosystems), and 2 mM $MgCl_2$. A fast touchdown PCR program was used: an initial activation and denaturing step at 95°C for 2 min, followed by 10 cycles of 1 s at 94°C and 5 s at 65°C and 30 cycles of 1 s at 94°C and 5 s at 55°C. A final elongation step was performed at 72°C for 10 s. PCR products were separated on a 3730xl DNA Analyzer (Applied Biosystems) and scored using GeneMapper version 4 (Applied Biosystems).

Statistical analyses and bioinformatics

Sixteen Saldanha Bay samples were used to characterize the microsatellite loci. The number of alleles, observed and expected heterozygosities, and polymorphic information content (PIC) were calculated using CERVUS version 3.0.3 (Kalinowski et al., 2007). Deviations from Hardy-Weinberg equilibrium (Weir and Cockerham, 1984) were calculated using Genepop version 4 (Rousset, 2008). Sequential Bonferroni's correction was performed for multiple tests.

The bioinformatics protocol described by Farber and Medrano (2003) was used to search for possible homologous loci in related species. In brief, repeat motifs were masked using RepeatMasker (www.repeatmasker.org/cgi-bin/WEBRepeatMasker) to omit significant hits owing to repeat motif similarities. Masked sequences were then subjected to BLASTN and BLASTX in the nr-nucleotide and nr-protein databases of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple alignments of query and subject sequences were used to determine the position of the microsatellite repeat within genes.

RESULTS

Pyrosequencing and primer design

A total of 11,271 single-sequence reads were generated covering 1.82 Mb and assembled into 1067 contiguous sequences. One hundred and forty-one dinucleotide, 22 trinucleotide, 264 tetranucleotide, 20 pentanucleotide, and 15 hexanucleotide repeats were detected within 297 contiguous sequences. Primer pairs were successfully designed for 185 repeat-containing contiguous sequences using BatchPrimer3. Of these primer pairs, 27 were discarded because some of the individual primers were situated within repeat tracts, which may have caused difficulties in downstream applications such as PCR amplification and size calling of alleles. Another three sequences were discarded owing to similarity to previously isolated loci for which primers already existed. A total of 155 primer pairs were chosen for further analysis.

Statistical analyses and bioinformatics

The screening of the wild population revealed that 82 of 155 loci were polymorphic (Table 1). Ten of these loci had more than three alleles, indicating a duplication event within the same locus or possibly between loci. The number of observed alleles for the nonduplicated alleles ranged from 2 to 21, expected heterozygosity ranged from 0.063 to 0.968, observed heterozygosity ranged from 0.000 to 1.000, and PIC ranged from 0.059 to 0.934. Seventeen of the 82 microsatellite loci did not conform to Hardy-Weinberg equilibrium after sequential Bonferroni's correction ($P < 0.05$).

Three loci gave statistically significant hits to sequences in the NCBI databases; however, all three had multiple significant hits to other haliotid genes or microsatellite loci. Hits to genes were always located in the 5'/3'-untranslated or intron regions (Table 2). Only one microsatellite locus, *Hmid*PS1.588C, was strictly conserved between *H. midae* and *H. d. hannai*, taking into consideration the reverse complement.

Table 1. Eighty-two polymorphic microsatellite loci isolated using pyrosequencing for *Haliotis midae*.

Locus name	Repeat tract	Primer sequence (5'-3')	PCR programme	N_A	H_O	H_E	HWE (P)	PIC	Accession No.
*Hmid*PS1.38T	$(CCAA)_n$	F: TCACATCCTCCTATCATCCAC R: GTTGGAAGATGTAATCGTTGG	TD-PCR	6	0.500	0.518	0.372	0.478	GU256656
*Hmid*PS1.42C	$(ATCC)_n...(ATCC)_n$	F: CATTTTCCCATGTATCCAAAC R: GGATTTGGATGGAAAAATCTC	TD-PCR	3	0.714	0.532	0.394	0.450	GU256657
*Hmid*PS1.95H	$(CACACG)_n$	F: ATTCAGACTGGTACGATTTCC R: ACCTATGTGTATGTGAGTGTGTG	TD-PCR	2	0.125	0.121	1.000	0.110	GU256658
*Hmid*PS1.124D	$(AC)_n$	F: ATTTTATTGGGGGAAAGAATG R: ATGGGTGAGTGACTCTGTGTT	TD-PCR	2	0.375	0.315	1.000	0.258	GU256659
*Hmid*PS1.138D	$(CA)_n$	F: AAACACATACAGGCACTCACA R: AGACAGGGTGTAACATTCTATTCA	TD-PCR	8	0.375	0.760	0.000***	0.706	GU256660
*Hmid*PS1.147M	$(TG)_n(GA)_n$	F: ATGCGTTGCGTACGTGT R: TTTTCCTCTGTCTCTCATCTCTC	TD-PCR	3	0.063	0.179	0.035	0.166	GU256661
*Hmid*PS1.150C	$(CA)_n...(CA)_n...(CACT)_n$ $(CA)_n...(CA)_n$	F: AAACGCTCATGCTCACATACT R: AGCCTAACAAACACTTTGCTG	TD-PCR	21	0.750	0.968	0.000***	0.934	GU256662
*Hmid*PS1.155D	$(CA)_n$	F: GGACCAACAGACAATTGAAAC R: AGGATCTGTCACCTACAGACG	TD-PCR	2	0.063	0.063	-	0.059	GU256663
*Hmid*PS1.156D	$(GT)_n$	F: ACGTAAGGCAGATTGATTTTG R: CATACACACATACACACATACG	TD-PCR	6	0.111	0.739	0.000***	0.669	GU256664
*Hmid*PS1.160T	$(CACT)_n$	F: AAAGTTTTGGTCACAATACCG R: CTTTGTGTGAGTGCGTGAGT	TD-PCR	4	DUP				GU256665
*Hmid*PS1.171C	$(CACCA)_n...(CACAC)_n$	F: TATTCAACGACTGACCATTCC R: TTGGTGTGTGGGTGTGC	TD-PCR	2	0.125	0.121	1.000	0.110	GU256666
*Hmid*PS1.179R	$(CAA)_n$	F: CAGCAAACAAGTATCAACAGC R: AATCTTGTGTTCTGTTTTGG	TD-PCR	6	DUP				GU256667
*Hmid*PS1.188C	$(GTGC)_n...(GT)_n$	F: GGACTGACCATTTGATAATGTG R: CTGCAAAGAATACTATTTGAGGAA	TD-PCR	4	0.188	0.236	0.192	0.220	GU256668
*Hmid*PS1.193C	$(ACTC)_n...(TCAC)_n...$ $(CACT)_n...(CACT)_n...(TCAC)_n$	F: CTCAATATATCCCCACGAGAA R: TTGAGTCGGATATTGTCACAG	TD-PCR	9	0.500	0.845	0.011	0.800	GU256669
*Hmid*PS1.195T	$(ACGC)_n$	F: GCTTTGAACCCGTTATTTGTT R: GGATGTTGACCGATTATTCAT	TD-PCR	2	0.313	0.272	1.000	0.229	GU256670
*Hmid*PS1.197C	$(TG)_n...(TG)_n$	F: GTCTGTCAGGTGTGTGTGAGA R: CGAACACAAGATCTCTATCTCTCT	TD-PCR	2	0.000	0.121	0.032	0.110	GU256671
*Hmid*PS1.206C	$(GACT)_n...(GTGA)_n...(AGTG)_n$	F: ATCCAGGCTGATTGTGAGA R: TGTGACAGATGAGTTTGACAAC	TD-PCR	6	0.375	0.601	0.017	0.517	GU256672
*Hmid*PS1.207C	$(AGTG)_n...(GTGA)_n$	F: TTGATTGATCGATTGATTGAG R: GTCACCATGTGTCACAGAGAA	TD-PCR	2	0.250	0.226	1.000	0.195	GU256673
*Hmid*PS1.208D	$(AC)_n$	F: CTATTTCGCACACGCTGAT R: GTGTGTGCGTGCTTGC	TD-PCR	4	0.143	0.550	0.001***	0.483	GU256674
*Hmid*PS1.222T	$(AGTG)_n$	F: GAAATGCGATCTCTGATTGAG R: AGCGTTACACTCATTCACTCAC	TD-PCR	3	0.125	0.123	1.000	0.116	GU256675
*Hmid*PS1.227T	$(ATGT)_n$	F: TAACACGTACACTCCAGTCC R: ACATGCTTGACTGTGACTCTCT	TD-PCR	4	0.250	0.603	0.002	0.496	GU256676

Continued on next page

Table 1. Continued.

Locus name	Repeat tract	Primer sequence (5'-3')	PCR programme	N_A	H_O	H_E	HWE (P)	PIC	Accession No.
*Hmid*PS1.228T	$(ACTC)_n$	F: TAACAAGTCAGCCACTCAACA R: GTGTGAGTGAGGTGAGGAGAC	TD-PCR	4	0.500	0.575	0.697	0.459	GU256677
*Hmid*PS1.247M	$(GA)_n(GAAT)_n$	F: ATCAGAGTGTGTGCATCAGTG R: CTGACATACACACAGACATCCA	TD-PCR	4	0.875	0.665	0.345	0.573	GU256678
*Hmid*PS1.305T	$(GCAC)_n$	F: CTCGAGTTTCAACCATTGAGT R: GGGTGGGTGTTACGAGTG	TD-PCR	5	0.714	0.762	0.307	0.688	GU256679
*Hmid*PS1.332D	$(AC)_n$	F: TGAAACACTCACACATCGACT R: TGGTTCATGCATAAATGTTGT	TD-PCR	18	0.563	0.950	0.000***	0.915	GU256680
*Hmid*PS1.353T	$(CACG)_n$	F: CGGAATAGAAGACGAGCAAT R: ATTTGACAGAAGGTGGTGGT	TD-PCR	3	0.188	0.365	0.048	0.309	GU256681
*Hmid*PS1.355R	$(TCA)_n$	F: GCCCATGTGCGAAGTT R: ATGTTTTGAGGAGGGAGTTTC	TD-PCR	7	0.375	0.716	0.005	0.663	GU256682
*Hmid*PS1.370C	$(CAACC)_n...(CACT)_n$	F: ACAACCAAACTCAACCCAAC R: TCATGTGACACGAGTGTGTG	TD-PCR	4	0.600	0.561	1.000	0.454	GU256683
*Hmid*PS1.374T	$(GAGT)_n$	F: TGACAAGTTTGGATTTGTTTTC R: TAGCTGGAATATTGCTGAGTG	TD-PCR	4	0.375	0.333	1.000	0.299	GU256684
*Hmid*PS1.375C	$(GTGA)_n...(GTGA)_n$	F: GGAGTGAACGAGTGAAGAAGT R: ACAACTCACCTGTCTTGTTTGT	TD-PCR	2	0.182	0.485	0.059	0.356	GU256685
*Hmid*PS1.379T	$(GTGC)_n$	F: TACTGTCTCTCTCGACGGTTC R: GCAAACACAATAAAACACCAA	TD-PCR	8	0.625	0.696	0.145	0.648	GU256686
*Hmid*PS1.382D	$(TG)_n$	F: TGGAATAACTGTCTATTTCGTCA R: TCAGACAGAAAGACACACACG	TD-PCR	7	0.385	0.812	0.002	0.749	GU256687
*Hmid*PS1.398P	$(AGGTG)_n$	F: ACAAGCCTCTAAAATGCCTCT R: TGCACGGTAAACTCAATCT	TD-PCR	4	0.125	0.579	0.000***	0.482	GU256688
*Hmid*PS1.405T	$(CAAC)_n$	F: CCTGCCACTCACTCAACTATT R: GGATAAGTGATTGGATGGGATA	TD-PCR	7	DUP				GU256689
*Hmid*PS1.433H	$(CACACG)_n$	F: AGTCCTGACCAGAACAAACAG R: GAGGTGAAAAGGATTTTGATG	TD-PCR	5	DUP				GU256690
*Hmid*PS1.457T	$(GAGT)_n$	F: TGGATGAGTGAATGAGAATGA R: ACACGATGACAAACAATGTGA	TD-PCR	8	0.308	0.665	0.007	0.620	GU256691
*Hmid*PS1.469R	$(ATC)_n$	F: TTGGTCAGCCATGTAGTCATA R: ATGATGGTGGGGATGATG	TD-PCR	2	1.000	0.516	0.000***	0.375	GU256692
*Hmid*PS1.484C	$(GAGT)_n...(GTGA)_n...(GTGA)_n$	F: ATATCTGACGTCTACCCCACA R: CCCATCCTGTGAAGAACATAC	TD-PCR	4	0.250	0.288	0.283	0.267	GU256693
*Hmid*PS1.487T	$(TGAG)_n$	F: ACGTACACGGACTGACAATTT R: ACTTGAAGAAGCGTAAAACCA	TD-PCR	6	0.357	0.638	0.001***	0.575	GU256694
*Hmid*PS1.521T	$(GAGT)_n$	F: ATCTGTGTGCACTCAATCTGT R: ACCCTACCACCCATTTAACTT	TD-PCR	4	0.333	0.303	1.000	0.276	GU256695
*Hmid*PS1.549D	$(TG)_n$	F: GTTGTGTGGGAGCGTATGTAT R: TACACCCCACATATACACCAA	TD-PCR	10	0.813	0.885	0.225	0.842	GU256696
*Hmid*PS1.551C	$(TATG)_n...(TGTA)_n$	F: GCTCCACCAAAATTGATATGT R: ACATACACATAGGTACACACACA	TD-PCR	4	0.688	0.595	0.905	0.506	GU256697

Continued on next page

Table 1. Continued.

Locus name	Repeat tract	Primer sequence (5'-3')	PCR programme	N_A	H_O	H_E	HWE (P)	PIC	Accession No.
*Hmid*PS1.559M	$(CA)_n(G)(TCAC)_n$	F: TAAGGATCATACACACTCGT R: CTTCAGTCGAAATGTGTATAACG	TD-PCR	2	0.000	0.121	0.032	0.110	GU256698
*Hmid*PS1.561C	$(TTGC)_n...(TGTT)_n$	F: AGGGTAGAACAATTTCCTGCT R: ATGGTCACTTGGGTTTGTCTA	TD-PCR	3	0.063	0.284	0.003	0.257	GU256699
*Hmid*PS1.588C	$(CACT)_n...(TCAC)_n...(CACT)_n$	F: GGAATATTGCTAAATGGTGGA R: TGAGTGAGTAATTGGGTAGGTG	TD-PCR	5	0.250	0.716	0.000***	0.641	GU256700
*Hmid*PS1.629C	$(TTGT)_n...(TGGG)_n...(TAGG)_n$	F: TGCATTTTGTAGTGTGTTTGC R: CCCTCCCTACCTACCTACCTA	TD-PCR	6	DUP				GU256701
*Hmid*PS1.635D	$(AC)_n$	F: CCAACAGTTTTCTGAATGTGA R: TGAAGTAGATGATGGGTGCAT	TD-PCR	2	0.063	0.063	-	0.059	GU256702
*Hmid*PS1.638T	$(GTGA)_n$	F: CAAGATCTAAAAATGGCCTCA R: CTTCACACTCACAAACTCAGC	TD-PCR	3	0.375	0.331	1.000	0.294	GU256703
*Hmid*PS1.692T	$(ATAC)_n$	F: TAAGACTGAGGGCGCTTTT R: TGGTGTGGATATGTGAAAATG	TD-PCR	2	0.063	0.063	-	0.059	GU256704
*Hmid*PS1.711T	$(GTGA)_n$	F: TAAACTGCTGTCACCAAGGA R: TACCCACACAACAGACTACCC	TD-PCR	5	DUP				GU256705
*Hmid*PS1.728D	$(CA)_n$	F: ACTTCACACATGAATGCACAC R: GTTGTTTTACAGACCGTCGAG	TD-PCR	2	0.125	0.121	1.000	0.110	GU256706
*Hmid*PS1.768T	$(ACTC)_n$	F: TAAAGCGGCGTAAAACTGA R: AATAGCCTGTCAGGTCATCG	TD-PCR	2	0.000	0.444	0.000***	0.337	GU256707
*Hmid*PS1.805T	$(CACG)_n$	F: AGAGGTTTGACATGACTTCCA R: ATGCGTGTTTGTATATGTG	TD-PCR	4	0.200	0.598	0.002	0.511	GU256708
*Hmid*PS1.807T	$(GAGT)_n$	F: TGTTTGAATAACCACCCTCTT R: CATTAGCTAACCCACAAATCC	TD-PCR	4	0.438	0.760	0.006	0.688	GU256709
*Hmid*PS1.811C	$(TTGT)_n...(TG)_n$	F: ATTGAAATAAATGCGCTTCAG R: CAAACACATAGATAGACGCACTT	TD-PCR	11	0.462	0.837	0.004	0.789	GU256710
*Hmid*PS1.818C	$(ATGG)_n...(TGGA)_n...(AC)_n$	F: AATGTAGGGTTGCTTCAAATG R: GAGTGTGTGGGTGTCTCTTTC	TD-PCR	9	0.625	0.738	0.165	0.696	GU256711
*Hmid*PS1.831M	$(CACC)_n(CACT)_n(CACC)_n(CACT)_n(CAC)(CACT)_n$	F: CTCACTCACTCCCTCATTCAC R: CCTGTACTGGTTAACAATATTTGAG	TD-PCR	17	DUP				GU256712
*Hmid*PS1.840D	$(AC)_n$	F: CATACAGAACACTGCGGAAC R: AACCACTTAGTGTGTGCGATT	TD-PCR	2	0.063	0.417	0.002	0.323	GU256713
*Hmid*PS1.844M	$(GAGT)_n(GTGA)_n$	F: ACATATGCGCCTTTGTGTTAT R: CAGGTAACTCACTCACTCACG	TD-PCR	7	DUP				GU256714
*Hmid*PS1.859T	$(CTCA)_n$	F: AAGACCGTACACTCTCACTCG R: TGGTGAGATATACAGGGTGAAA	TD-PCR	6	0.250	0.435	0.002	0.404	GU256715
*Hmid*PS1.860D	$(GT)_n$	F: AGTAGGTGGACCTCTCTCCAT R: ACAGAATCTACACGCACACAC	TD-PCR	3	0.438	0.365	1.000	0.309	GU256716
*Hmid*PS1.868T	$(TGAG)_n$	F: TGTAGGGATGAGAACGAAAAG R: GGCGTAAAACCATACTCACTC	TD-PCR	3	0.875	0.534	0.007	0.412	GU256717
*Hmid*PS1.870C	$(CACACG)_n...(AC)_n$	F: ACAACAACACACACGCACA R: GTGCCAAAACATATTTCAAAAC	TD-PCR	15	0.875	0.938	0.493	0.901	GU256718

Continued on next page

Table 1. Continued.

Locus name	Repeat tract	Primer sequence (5'-3')	PCR programme	N_A	H_O	H_E	HWE (P)	PIC	Accession No.
*Hmid*PS1.873T	$(TCAC)_n$	F: AACAGTGTCAATAAAGCTGGAA R: TGGACCAGAGTGATAATAGTGTG	TD-PCR	3	0.467	0.393	1.000	0.342	GU256719
*Hmid*PS1.874C	$(CACG)_n...(AC)_n$	F: AACGAAGGACAGTAAAAACAACT R: CAGCTAGACTGAGTGTGACCA	TD-PCR	14	0.625	0.927	0.000***	0.890	GU256720
*Hmid*PS1.890M	$(CACT)_n(CT)_n$	F: TTCTCATTTCACACACACAGG R: CCTTTTCACTTCATAGCGTGT	TD-PCR	14	DUP				GU256721
*Hmid*PS1.906T	$(CCAC)_n$	F: CACTCACTCACTAACCCACCT R: ATAAACCCTGAAACCCTGAAA	TD-PCR	3	0.231	0.218	1.000	0.198	GU256722
*Hmid*PS1.952D	$(TG)_n$	F: TGAGTCCTGAGTAACTGCAAA R: TGGATTGAACAACTTTCAACA	TD-PCR	8	0.357	0.690	0.000***	0.638	GU256723
*Hmid*PS1.961T	$(GTAG)_n$	F: AAACTAGAAAGGAGGCACGTT R: ATACTACACGCGCACACATAC	TD-PCR	2	0.125	0.444	0.007	0.337	GU256724
*Hmid*PS1.967M	$(TGTC)_n(TG)_n$	F: ATATGCACACCGAGTGAAATC R: CTAACATGACCAGCGATTGTT	TD-PCR	6	0.563	0.798	0.021	0.738	GU256725
*Hmid*PS1.972T	$(TCAC)_n$	F: CCCACTCACTCACATATCCAC R: GCATGGAAAAACAAAATGTCT	TD-PCR	5	0.688	0.593	0.446	0.492	GU256726
*Hmid*PS1.981T	$(CTCA)_n$	F: CTGGAATATTGCTAAAAGTGG R: TTAGAGACTGAAGACGGGATG	TD-PCR	9	0.600	0.869	0.057	0.821	GU256727
*Hmid*PS1.982M	$(TGTA)_n(TG)_n$	F: TCCTGAGTAATCGTACTCTGTGT R: CCAACAATGTAACCACAGGAT	TD-PCR	3	0.214	0.415	0.019	0.359	GU256728
*Hmid*PS1.1007C	$(ACTC)_n...(TCAA)_n$	F: ATATTGCCGATGTGACGTTAT R: TGATTGATTGTAGTGATTGAGTTG	TD-PCR	5	0.800	0.651	0.001***	0.566	GU256729
*Hmid*PS1.1009H	$(GTGGGT)_n$	F: TGTAAGAAGTGGACCAGCAGT R: TGCTCCACAAAACTGAGTACA	TD-PCR	9	DUP				GU256730
*Hmid*PS1.1012R	$(CAT)_n$	F: CCCACACACATGAGAAATGT R: CAAGTATGTGTGGGTGGTGTT	TD-PCR	9	0.733	0.885	0.166	0.839	GU256731
*Hmid*PS1.1018T	$(TTGT)_n$	F: CTGCGTCCTTGTGTGTGT R: TGACGAGAGGGTCAGATTAGA	TD-PCR	6	0.500	0.694	0.034	0.630	GU256732
*Hmid*PS1.1026M	$(GTGA)_n(GTGC)_n$	F: GTGTGCGTGAGTGAGTGAGT R: CACCAAATTACATCCCATACC	TD-PCR	6	0.250	0.679	0.000***	0.610	GU256733
*Hmid*PS1.1038T	$(GTGA)_n$	F: TATGTGCATGTGGGGTTATG R: AACAAACACATGGATACACCA	TD-PCR	5	0.688	0.653	1.000	0.562	GU256734
*Hmid*PS1.1058C	$(TGAG)_n...(AGTG)_n...(AGTG)_n$	F: GTAATTGGATCAAAGATGC R: AAATGACAGCTCTCAGATTGC	TD-PCR	11	0.600	0.903	0.000***	0.860	GU256735
*Hmid*PS1.1063C	$(TC)_n...(CGTG)_n$	F: AAAGGTTTGTGGAATGTGTGT R: TACCACACACCCTCAAGTATG	TD-PCR	11	0.533	0.903	0.001***	0.860	GU256736
*Hmid*PS1.1066M	$(GT)_n(TG)_n(TGTT)_n$	F: AATCCAACAAAGGAAATACCC R: CACACCAAACAAACAAACAAA	TD-PCR	6	0.188	0.516	0.000***	0.474	GU256737

N_A = number of observed alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; ***P < 0.05 = significant departure from Hardy-Weinberg equilibrium (HWE) after sequential Bonferroni's correction; PIC = polymorphic information content; TD-PCR = touch-down PCR; DUP = duplication of locus.

Table 2. A summary of loci with significant BLAST hits.

Locus name	Microsatellite locus hits			Gene hits		
	Locus name (Accession No.)	E-value	Identities-value	Gene name (Accession No.)	E-value	Identities-value
*Hmid*PS1.374T	*H. d. hannai* microsatellite $(AAAC)_n$ (GU995824)	5e-09	90%	*H. discus* lysin (FJ940391)	2e-08	87%
				H. rufescens lysin (AF076822)	2e-08	87%
				H. corrugata lysin (FJ940473.1)	3e-07	88%
*Hmid*PS1.588C	*H. d. hannai* microsatellite $(GAGT)_n$ (AB177913)	8e-09	78%	*H. rubra* ATPase alpha-subunit (AY043205)	2e-07	78%
	H. kamtschatkana microsatellite $(GT)_n$ (AY013579)	3e-05	79%	*H. d. discus* peroxiredoxin (EF103356)	3e-05	80%
	H. sieboldii microsatellite $(CT)_n$ (JF693957)	1e-10	79%			
*Hmid*PS1.1007C	*H. rubra* microsatellite $(CA)_n$ (AF194955)	7e-08	95%	*H. tuberculata* hemocyanin (AJ252741)	7e-05	90%
	H. d. hannai microsatellite $(CT)_n$ (AB177931)	7e-8	95%			

DISCUSSION

A total of 82 microsatellite markers were developed for *H. midae* using the FIASCO method and the 454 pyrosequencing. The usefulness of these newly designed microsatellite markers for future applications such as population structure analysis, parentage assignments, and linkage mapping was assessed by calculating various parameters. Deviations from Hardy-Weinberg equilibrium (see Table 1) were mostly the result of heterozygote deficiency caused by the presence of null alleles (O'Connell and Wright, 1997), allele dropout, or scoring errors (Jones and Ardren, 2003). Two observations (*Hmid*PS1.469R and *Hmid*PS1.1007C) were explained by heterozygote excess. Although loci such as *Hmid*PS1.469R and *Hmid*PS1.1007C could be interesting candidates for studying selection processes in the life history of *H. midae*, they should be used with caution. The high PIC values obtained for many of the loci make them good candidates for parentage assignments and linkage mapping owing to a strong likelihood of being informative in both parent and offspring.

Amplification of more than the expected two alleles has previously been observed in *H. midae* (Slabbert et al., 2010) and *H. rubra* (Evans et al., 2001). The exact mechanisms underlying this occurrence are still unclear but could be explained by genome duplication, polyploidy, aneuploidy, or conserved microsatellite repeat tracts and flanking regions found in mobile elements (Hubert et al., 2000). A recent study by Rhode and Roodt-Wilding (2011) found that 21% of all known *H. midae* microsatellite loci are associated with characterized transposable elements, which play a role in locus duplication. This high association may therefore be the most likely explanation for duplications in abalone microsatellites.

Three loci were found to have orthologs in other haliotids. The lack of strict repeat motif conservation is expected taking into account the life cycle hypothesis of microsatellite evolution (Ellegren, 2004). These loci also had significant hits to haliotid genes. BLAST alignment analysis showed that the repeat motif was located in the untranslated region or introns. Furthermore, because individual loci are present across haliotid species and not necessarily in the same genes, these loci might form parts of gene regulatory elements (Li et al., 2004). Two of these markers (*Hmid*PS1.588C and *Hmid*PS1.1007C) deviated from Hardy-Weinberg

expectations, which may indicate functional constraints and thus selective pressures. These loci are prime candidates for synteny mapping, QTL, and functional analysis.

The data generated using FIASCO and pyrosequencing 454 were accurate and adequate for the development and characterization of 82 polymorphic microsatellite markers. Pyrosequencing provides sequence information on all available DNA fragments present within an enriched library, in contrast to traditional cloning in which technical, time, and budget constraints cause the loss of significant information. The characterization of the newly designed markers showed that many of them would be useful for parentage and population studies. The additional markers will also contribute to the construction of a detailed linkage map.

ACKNOWLEDGMENTS

Research supported by the National Research Foundation of South Africa (The Innovation Fund) and a consortium of the following abalone hatcheries: Abagold (Pty) Ltd., Aquafarm (Pty) Ltd., HIK Abalone Farm (Pty) Ltd., Irvin and Johnson Abalone (I&J) Ltd., and Roman Bay Sea Farm (Pty) Ltd. We thank Dr. Peizheng Wang for technical assistance, Arnold Vlok for assistance during sample collection, and Stellenbosch University for the use of facilities.

REFERENCES

Abdelkrim J, Robertson B, Stanton JA and Gemmell N (2009). Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46: 185-192.

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, et al. (2000). Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* 280: 103-110.

Ahmadian A, Ehn M and Hober S (2006). Pyrosequencing: history, biochemistry and future. *Clin. Chim. Acta* 363: 83-94.

Allentoft M, Schuster SC, Holdaway R, Hale M, et al. (2009). Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* 46: 195-200.

Baranski M, Rourke M, Loughnan S, Austin C, et al. (2006). Isolation and characterization of 125 microsatellite DNA markers in the blacklip abalone, *Haliotis rubra*. *Mol. Ecol. Notes* 6: 740-746.

Bester AE, Slabbert R and D'Amato ME (2004). Isolation and characterization of microsatellite markers in the South African abalone (*Haliotis midae*). *Mol. Ecol. Notes* 4: 618-619.

Blanca J, Canizares J, Roig C, Ziarsolo P, et al. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* 12: 104.

Castoe TA, Poole AW, Gu W, Jason de Koning AP, et al. (2010). Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Mol. Ecol. Resour.* 10: 341-347.

Cook P and Britz PJ (1991). Perlemoen farming in South Africa. Mariculture Association of Southern Africa, Fishing Industry Research Institute, Cape Town.

DAFF (Department of Agriculture, Forestry and Fisheries) (2012). Aquaculture Annual Report 2011, South Africa.

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5: 435-445.

Evans B, White RW and Elliott NG (2000). Characterization of microsatellite loci in the Australian Blacklip abalone (*Haliotis rubra*, Leach). *Mol. Ecol.* 9: 1183-1184.

Evans B, Conod N and Elliott NG (2001). Evaluation of microsatellite primer conservation in abalone. *J. Shellfish Res.* 20: 1065-1070.

Farber CR and Medrano JF (2003). Putative *in silico* mapping of DNA sequences to livestock genome maps using SSLP flanking sequences. *Anim. Genet.* 34: 11-18.

Galindo J, Grahame JW and Butlin RK (2010). An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *J. Evol. Biol.* 23: 2004-2016.

Gutiérrez-Gonzales JL, Cruz P, Del Rio-Portilla MA and Perez-Enriquez R (2007). Genetic structure of green abalone *Haliotis fulgens* population off Baja California, Mexico. *J. Shellfish Res.* 26: 839-846.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/

NT. *Nucleic Acids Symp. Ser.* 41: 95-98.
Hamilton MB, Pincus EL, Di FA and Fleischer RC (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 27: 500-507.
Hubert S, English LJ, Landau BJ, Guo X, et al. (2000). Microsatellite analysis of trisomic families in the Pacific oyster, *Crassostrea gigas*, Thunberg. *J. Shellfish Res.* 19: 615.
Jones AG and Ardren WR (2003). Methods of parentage analysis in natural populations. *Mol. Ecol.* 12: 2511-2523.
Kalinowski ST, Taper ML and Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16: 1099-1106.
Li Q and Akihiro K (2007). Sequence characterization of microsatellite DNA sequences in Pacific abalone (*Haliotis discus hannai*). *J. Ocean Univ. Chin.* 6: 47-52.
Li Q, Shu J, Yu R and Tian C (2007). Genetic variability of cultured populations of the Pacific abalone (*Haliotis discus hannai* Ino) in China based on microsatellites. *Aquac. Res.* 38: 981-990.
Li YC, Korol AB, Fahima T and Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21: 991-1007.
O'Connell M and Wright JM (1997). Microsatellite DNA in fishes. *Rev. Fish. Biol.* 7: 331-363.
Rhode C and Roodt-Wilding R (2011). Bioinformatic survey of *Haliotis midae* microsatellites reveals a non-random distribution of repeat motifs. *Biol. Bull.* 221: 147-154.
Ronaghi M, Uhlen M and Nyren P (1998). A sequencing method based on real-time pyrophosphate. *Science* 281: 363-365.
Rousset F (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103-106.
Saghai-Maroof MA, Soliman KM, Jorgensen RA and Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 81: 8014-8018.
Santana QC, Coetzee MPA, Steenkamp ET, Mlonyeni OX, et al. (2009). Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques* 46: 217-223.
Selvamani MJP, Degnan SM, Paetkau D and Degnan BM (2000). Highly polymorphic microsatellite loci in the Heron Reef population of the tropical abalone *Haliotis asinina*. *Mol. Ecol.* 9: 1184-1186.
Shendure J, Porreca GJ, Reppas NB, Lin X, et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309: 1728-1732.
Shi Y, Guo X, Gu Z, Wang A, et al. (2010). Preliminary genetic linkage map of the abalone *Haliotis diversicolor* Reeve. *Chin. J. Oceanol. Limn.* 28: 549-557.
Slabbert R, Ruivo NR, Van den Berg NC, Lizamore DL, et al. (2008). Isolation and characterization of 63 microsatellite loci for the abalone, *Haliotis midae*. *J. World Aquacult. Soc.* 39: 429-435.
Slabbert R, Bester AE and D'Amato ME (2009a). Analyses of genetic diversity and parentage within a South African hatchery of the abalone *Haliotis midae* Linnaeus using microsatellite markers. *J. Shellfish Res.* 28: 369-375.
Slabbert R, Roux A, Van der Merwe M, Watson L, et al. (2009b). Genetic Enhancement in the South African Abalone (*Haliotis midae*): Current Status. In: Book of Abstracts of the 10th International Symposium on Genetics in Aquaculture, Bangkok, 22-26.
Slabbert R, Hepple J, Venter A, Nel S, et al. (2010). Isolation and segregation of 44 microsatellite loci in the South African abalone *Haliotis midae* L. *Anim. Genet.* 41: 332-333.
Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
You FM, Huo N, Gu YQ, Luo MC, et al. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9: 253.
Zane L, Bargelloni L and Patarnello T (2002). Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11: 1-16.