Conservation Genet Resour (2012) 4:257–260
DOI 10.1007/s12686-011-9519-z

TECHNICAL NOTE

The development of 20 microsatellite loci for the Australian marine mollusk, *Donax deltoides*, through next generation DNA sequencing

Adam D. Miller · Vincent L. Versace · Ty Matthews · Kate C. Bowie

Received: 4 August 2011 / Accepted: 28 August 2011 / Published online: 10 September 2011

Abstract Next Generation DNA sequencing was used to develop a suite of microsatellite markers for the marine mollusk, *Donax deltoides*. A total of 20 polymorphic loci were identified and 12 characterized using 30 individuals from a single population (Venus Bay) in south eastern Australia. We observed moderate to high genetic variation across most loci (mean number of alleles per locus = 7.3; mean heterozygosity = 0.633) with only a single locus (Ddel32) displaying significant deviation from Hardy–Weinberg equilibrium. Marker independence was confirmed with tests for linkage disequilibrium, however two loci were found to be influenced by null alleles. The 10 viable markers characterized in the present study provide a valuable resource for future population genetic assessments and fisheries management of *D. deltoides* in Australia.

Keywords Microsatellite development · Next generation DNA sequencing · *Donax deltoides* · Marine mollusk

A. D. Miller (✉)
Department of Zoology, The University of Melbourne, Parkville, VIC 3010, Australia
e-mail: admiller@unimelb.edu.au

A. D. Miller
Cesar, 293 Royal Parade, Parkville 3052, Australia

A. D. Miller · T. Matthews · K. C. Bowie
School of Life and Environmental Sciences, Deakin University, Warrnambool, VIC 3280, Australia

V. L. Versace
School of Information Systems, Deakin University, Warrnambool, VIC 3280, Australia

Introduction

Donax deltoides (Vereroida: Donacidae), more commonly known as the 'pipi', is a benthic marine bivalve commonly found in the swash zone on high energy beaches in south eastern Australia. The species is currently harvested by recreational fishers across most of its range with commercial fisheries also operating in regions of South Australia, Victoria and New South Wales (Ferguson and Mayfield 2006; Murray-Jones 1998). In recent years localized population declines have been observed in the respective states where fisheries currently operate reflecting the species' vulnerability to over-exploitation (Ferguson and Mayfield 2006; Lewis and Scarpaci 2010; Murray-Jones 1998; Murray-Jones and Johnson 2003). In order to implement sustainable management strategies for the fishery an improved understanding of the species biology, ecology and genetics is needed. In particular, a comprehensive assessment of population genetic structure will provide managers with an effective spatial framework for managing the fishery and baseline data for future population monitoring. Prior to this study, effective genetic markers for population genetic assessments of *D. deltoides* were not available.

The 454 next generation sequencing platform was used to identify microsatellite markers for *D. deltoides*. Approximately 10 μg of genomic DNA was extracted from muscle tissue from a single *D. deltoides* specimen using a QIAGEN DNA Easy kit (Qiagen). DNA was subsequently processed by the Australian Genome Research Facility where it was nebulized, ligated with 454 sequencing primers and tagged with a unique oligo sequence allowing sequences to be separated from pooled species DNA sequences using post-run bioinformatic tools. The DNA sample was analyzed using high throughput DNA sequencing on 1/16 of a 70 × 75 mm PicoTiterPlate using

Table 1 Primers sequences and characteristics of 20 microsatellite loci isolated from *Donax deltoides*

Locus	Primer sequence (5′–3′)	Repeat motif	*n*	*r*	Size range (bp)	H_O/H_E	HWE	F_{IS}	Frequency of null alleles	Genbank accession number
MPlexA										
Ddel7	TCCGACGTAAAACAAATGTCA AATAAGAAATACATTTTATGCAAGGAT	$(GT)^{15}$	30	11	107–145	0.857/0.830	0.924	−0.034	0.000	JN391488
Ddel10	CCCCATTAGATCAAGCTCCA CACTTTGGCTTCCAACAGGT	$(CT)^{8}$	30	4	81–87	0.586/0.591	0.820	0.008	0.005	JN391489
Ddel12	AGCTGTAGTCGATTTGAGCTGT TTATCCCAAAGCCGATACGA	$(CTC)^{13}$	30	8	177–197	0.615/0.621	0.630	0.009	0.000	JN391490
Ddel16	ATGTCGGCAAGATTCAGTCC AAATTGCAACTTGTTTGATAATGA	$(CAT)^{8}$	30	9	102–136	0.724/0.751	0.802	0.036	0.014	JN391491
Ddel24	GCACAGTAAAACTTCACTTAAACG AGGCGGTGAAGAAACAGGTA	$(AC)^{11}$	30	14	149–176	0.821/0.860	0.311	0.046	0.013	JN391492
Ddel28	TGTTAGTTGTGTAATTGTAAGCATTTT CCTTGCATTTCTCGTCACAA	$(TCA)^{12}$	30	9	211–235	0.750/0.752	0.487	0.003	0.000	JN391493
Ddel32	AATTGTTATGAAAGGTCTAAATGATTG CCTGAACCAATCATCACAAGTC	$(TA)^{12}$	30	8	92–106	0.370/0.691	<0.001*	0.469*	0.199	JN391494
MPlexB										
Ddel4	GGTTTTCTGCAGTCTGCTCC TTCCTGCGGAAAGGTTAGTG	$(CT)^{8}$	30	2	192–194	0.207/0.189	1.000	−0.098	0.000	JN391495
Ddel6	TTCTCAGAAATCGAAATGACG TTTCTACAAAGGAAGTAGAGCCA	$(GT)^{9}$	30	3	113–125	0.222/0.469	0.002*	0.531*	0.183	JN391496
Ddel14	CCATGTAAGTAGAACGGGGC CCATTAAATTCCAATGTCCCA	$(TAA)^{9}$	30	3	143–151	0.276/0.296	0.621	0.069	0.020	JN391497
Ddel15	AATCCAAATTTAGCATTCTGCC CCAAAAGCACATCTTTGGAGA	$(AG)^{8}$	30	9	85–101	0.586/0.775	0.017	0.247	0.102	JN391498
Ddel20C	ATTCGACAGAGCACATGCC TAATTCCAGCAGTCGCAAGA	$(ATT)^{11}$	30	8	224–248	0.783/0.771	0.240	−0.015	0.000	JN391499
Additional loci										
Ddel5	TGTGTTGTCGTAACGGATGC CAAGGTGTGAAAATGTTGTTGC	$(ATC)^{12}$	8	2	198–216	–	–	–	–	JN391500
Ddel9	GCATCTGATACTCCACAACAACT GATTCTTGGGTTCCCACAAC	$(CT)^{11}$	8	6	76–86	–	–	–	–	JN391501
Ddel23	CCTGCTTTGATTGTCGTCAG TAAACAATGTGCCCTGTTCG	$(ATC)^{16}$	8	3	110–123	–	–	–	–	JN391502

Table 1 continued

Locus	Primer sequence (5′–3′)	Repeat motif	*n*	*r*	Size range (bp)	H_O/H_E	HWE	F_{IS}	Frequency of null alleles	Genbank accession number
Ddel26	GGATTGCACATAACACGTCG GCTTGTCCAGTATGGGTCTTTT	$(TG)^{11}$	8	5	151–171	–	–	–	–	JN391503
Ddel27	AGGTTGAATGCGGATCACTC TTTGTATATTGTCTTTGTTTTGTTTGT	$(AC)^{16}$	8	3	95–101	–	–	–	–	JN391504
Ddel35	CGACAAAAGCAGTATTCCAAA AAATTCCAGTTAGAGACGGCA	$(CA)^{16}$	8	5	93–109	–	–	–	–	JN391505
Ddel36	GAGATAGAGGCCCAATACTGATG TCCAATACAGATGATAGAATCATGAAC	$(GA)^{16}$	8	5	124–132	–	–	–	–	JN391506
Ddel39	CAAGCACCACCCTAGTATAAGGA AGAGGTAGGTCAAGTCCGGG	$(ATG)^{13}$	8	5	143–158	–	–	–	–	JN391507

Number of alleles (*r*), observed (H_O) and expected (H_E) heterozygosities, Hardy–Weinberg equilibrium *P* values (HWE) and inbreeding (F_{IS}). * Significant *P* value after correction for multiple comparisons

the Roche GS FLX (454) system (Margulies et al. 2005). A total of 79,905 reads were obtained from the analysis, from which 721 unique sequence contigs possessing microsatellite motifs were identified using the software GDD (Meglécz et al. 2010). Primer3 (Rozen and Skaletsky 2000) was used to design optimal primer sets for each unique contig where possible, with a total of 90 contigs found to possess optimal priming sites. A selection of 40 contigs was used for subsequent analysis, 25 of which consisted of di-nucleotide repeats and 15 consisting of tri-nucleotide repeats.

Loci were screened for polymorphism using template DNA from eight individuals, representing 3 wild populations from south eastern Australia. These included two sites from Victoria (Nelson, −38.067 141.014, and Venus Bay, −38.706 145.811) and one site from South Australia (Coorong, −35.910 139.395). Loci were pooled into 10 groups of four, labelled with unique fluorophores (FAM, NED, VIC, PET) and co-amplified by multiplex PCR using a Qiagen multiplex kit (Qiagen) and an Eppendorf Mastercycler S gradient PCR machine following the protocol described by Blackett et al. (pers comm). Genotyping was subsequently performed using an Applied Biosystems 3730 capillary analyzer and product lengths were scored manually and assessed for polymorphisms using GeneMapper version 4.0 (Applied Biosystems). From 40 loci a total of 20 were found to be polymorphic (Table 1), 10 were monomorphic and nine failed to amplify.

A subset of 12 polymorphic loci were selected, pooled into two groups for multiplexing based on observed locus specific allele size ranges and further characterized using 30 individuals from the Venus Bay population. Microsatellite profiles were again examined using GeneMapper version 4.0 and alleles were scored manually. The *Excel Microsatellite Toolkit* (Park 2001) was then used to estimate expected (H_E) and observed (H_O) heterozygosities and number of alleles (N_A), while examination of Hardy–Weinberg proportions (HWE), the inbreeding coefficient (F_{IS}) and linkage disequillibrium between all pairs of loci were conducted using GENEPOP version 4 (Raymond and Rousset 1995). Significance values were adjusted to allow for multiple statistical tests using Bonferroni corrections where necessary (Rice 1989). Finally to check for null alleles and scoring errors all loci were assessed using MICRO-CHECKER (Van Oosterhout et al. 2004). The frequency of null alleles per locus was obtained using the 'Brookfield 1' formula as null homozygotes were not observed (Brookfield 1996).

Majority of loci were characterized by moderate to high genetic variation, with an average of 7.3 alleles per locus (range = 2–14 alleles) and heterozygosity estimates ranging between 0.189 and 0.860 (mean = 0.633). Marker independence was confirmed as linkage disequilibrium

analyses indicated no significant linkage between loci. Only two loci, Ddel6 and Ddel32, were found to deviate significantly from Hardy–Weinberg Equilibrium ($P = 0.002$ and $P < 0.001$, respectively; Table 1) and display significant F_{IS} estimates ($F_{IS} = 0.2989$, $P = 0.0127$ and $F_{IS} = 0.2989$, $P = 0.0127$, respectively; Table 1), suggesting heterozygote deficiencies. MICRO-CHECKER analyses identified evidence of null alleles at three loci, Ddel6, Ddel15 and Ddel32 (Table 1), while evidence of scoring issues or large allele dropouts was not detected. These findings suggest that three of the 12 characterized loci (Ddel6, Ddel15, Ddel32) might be problematic for future population genetic analyses.

The performance of markers Ddel6, Ddel15, and Ddel32 was further assessed using 32 individuals from a separate population (Coorong, South Australia). HWE and F_{IS} estimates for marker Ddel32 were again found to be significantly different from zero ($P < 0.001$; $F_{IS} = 0.397$, $P < 0.001$), and evidence of null alleles was again recorded (frequency = 0.136). Although HWE and F_{IS} estimates for Ddel6 on this occasion were found to be non-significant following corrections for multiple comparisons ($P = 0.069$; $F_{IS} = 0.299$, $P = 0.013$), null alleles were again detected (frequency = 0.104) further compromising the reliability of this marker. We recommend that Ddel6 and Ddel32 be excluded from future population genetic studies as these markers are likely to introduce ambiguity. Conversely Ddel15 was found to conform to Hardy–Weinberg expectations ($P < 0.340$), produced non-significant F_{IS} estimates and null alleles were not detected. Although we cannot explain the discrepancy between the Ddel15 estimates derived from the Venus Bay and Coorong populations, this might be associated with stochastic population processes, however this is most likely associated with differences in null allele frequencies between populations. Therefore we recommend that marker Ddel15 be included but treated with caution in future population genetic studies.

The 10 viable and 9 additional uncharacterized microsatellite markers described in this study provide a valuable resource for future population genetic assessments of *D. deltoides* in Australia. Estimates of gene flow and population genetic structure using these markers will provide managers with an effective spatial framework for managing the *D. deltoides* fishery and valuable baseline data for future temporal population monitoring that will assist managers to mitigate the potential for over-exploitation. More importantly this research will promote the future conservation of *D. deltoides* populations in south eastern Australia.

Acknowledgments The authors would like to acknowledge Deakin University for financial support through the Environmental Sustainability Faculty Research Cluster (ESFRC). This project was also assisted by the Victorian Government through revenue from Recreational Fishing Licenses.

References

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5:453–455

Ferguson G, Mayfield S (2006) The South Australian Goolwa cockle (*Donax deltoides*) fishery: fishery Assessment Report for PIRSA Fisheries. SARDI (Aquatic Sciences) Adelaide. Report No. RD06/0005-1, p 30

Lewis Z, Scarpaci C (2010) Recreational harvest: implications on management and sustainability of the pipi (*Donax deltoides*) at Venus Bay, Victoria. Final report to fisheries Victoria. Melbourne, Australia

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26(3):403–404

Murray-Jones S (1998) Harvesting of the pipi, *Donax deltoides* in NSW. FRDC final report (95/152). Wollongong, University of Wollongong

Murray-Jones S, Johnson J (2003) Goolwa cockle (*Donax deltoides*). SARDI Aquatic Sciences, Adelaide, pp 1–54

Park SDE (2001) Trypanotolerance in West African Cattle and the populationgenetic effects of selection. PhD. University of Dublin

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43(1):223–225

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365–386

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4(3):535–538