Conservation Genet Resour
DOI 10.1007/s12686-013-9968-7

TECHNICAL NOTE

SNP discovery in the marine gastropod *Concholepas concholepas* by high-throughput transcriptome sequencing

Cristian Gallardo-Escárate · Gustavo Núñez-Acuña · Valentina Valenzuela-Muñoz

Received: 14 May 2013 / Accepted: 30 May 2013

Abstract Novel simple nucleotide polymorphisms (SNP) were identified using 454 transcriptome pyrosequencing in the snail *Concholepas concholepas.* From 2,991 contigs annotated with SNPs from De novo assembly, 7,069 putative polymorphisms were selected according their quality and functional annotation (gene ontology), of these, 18 SNPs were validated by HRM analysis using 50 individuals from two wild populations along the Chilean coast (29°S–71°W and 39°S–73°W). Herein, 9 loci evidenced Hardy–Weinberg equilibrium between both populations. F_{st} value between populations was estimated in 0.004, and F_{is} were calculated in 0.286 and 0.291 for North and South populations respectively. Here we carried out the first SNP mining in *C. concholepas* using pyrosequencing technology that promise to be useful for future conservation strategies of this overexploited marine snail species.

Keywords Pyrosequencing 454 · Expressed sequence tag (EST) · Single nucleotide polymorphisms · *Concholepas concholepas*

Concholepas concholepas, locally called as "loco", is a benthic species endemic to the Southeastern Pacific coast. The loco is an important component of intertidal and shallow sub-tidal communities and it is also one of the main invertebrates targeted by small-scale fisheries in Chile (Castilla 1999). This gastropod constitutes a flag species for marine management and conservation in Chile (Fernández and Castilla 2005). However, it has not been evaluated in terms of conservation category even though some local populations have disappeared. The loco has an extensive distribution ranging from tropical (Lobos Afuera Island, 6°S) to sub-antarctic habitats (Cape Horn, 56°S) (Cárdenas et al. 2008) Then, there is an urgent necessity to identified local adaptation across this geographic extension to avoid lost of biodiversity and failed of the restocking program.

Next generation sequencing (NGS) technologies have revolutionized genomics research mainly due to the massive amount of sequence data generated, facilitating the understanding of genomes in non-model species through gene discovery and detection of polymorphisms. NGS data can be screened for the presence of molecular markers such as microsatellites (SSR) and single nucleotide polymorphisms (SNPs) (Davey et al. 2011). Furthermore, large expressed sequence tag (EST) databases from transcriptome high-throughput sequencing offers the possibility to detect SNPs within messenger RNA or transcripts. The most important feature of SNPs is their possible association with candidate genes involved in biological adaptation. The main aim of this study was to characterize novel SNP markers in the overexploited marine snail *C. concholepas.*

A total of 601,802 reads were generated using ½ plate of a Roche 454 GS-FLX titanium platform at Macrogen Inc. (Korea). De novo assembly was performed with CLC Genomics Workbench (CLC bio, USA), yielding 27,116 contigs with a mean length of 433 bp (database available at the Dryad Digital depository under the access doi:10.5061/dryad.mf00q). A total of 2,991 sequences containing SNPs

Electronic supplementary material The online version of this article (doi:10.1007/s12686-013-9968-7) contains supplementary material, which is available to authorized users.

C. Gallardo-Escárate (✉) · G. Núñez-Acuña · V. Valenzuela-Muñoz
Laboratory of Biotechnology and Aquatic Genomics, Interdisciplinary Center for Aquaculture Research (INCAR), University of Concepción, Casilla 160-C, Concepción, Chile
e-mail: crisgallardo@udec.cl

Published online: 07 June 2013

Table 1 Summary of the statistical analysis for 18 SNP markers in two populations of *C.concholepas* analyzed

Locus	Populations							
	Coquimbo			Genetic variability indices		Valdivia		
	F_{is}	*Ho*	*He*	HWE	F_{is}	*Ho*	*He*	HWE
Cc_SNP1	−0.130	0.571	0.508	1	−0.467	0.667	0.464	0.217
Cc_SNP2	−0.110	0.500	0.452	1	−0.063	0.538	0.508	1
Cc_SNP3	1	0	0.476	0	0.836	0.083	0.489	0.006
Cc_SNP4	−0.368	0.571	0.423	0.506	−0.151	0.538	0.471	1
Cc_SNP5	1		0.138	0.037	0	0.077	0.077	1
Cc_SNP6	0.304	0.214	0.304	0.348	0.520	0.231	0.471	0.091
Cc_SNP7	0.536	0.214	0.452	0.073	0.762	0.091	0.368	0.038
Cc_SNP8	0.762	0.091	0.368	0.038	0.585	0.167	0.391	0.090
Cc_SNP9	0.316	0.357	0.516	0.316	−0.200	0.615	0.517	0.602
Cc_SNP10	−0.132	0.545	0.485	1	0.651	0.167	0.464	0.044
Cc_SNP11	0.111	0.462	0.517	1	−0.132	0.545	0.485	1
Cc_SNP12	0.111	0.462	0.517	1	−0.091	0.231	0.212	1
Cc_SNP13	0.592	0.200	0.480	0.033	0.814	0.083	0.431	0.012
Cc_SNP14	0.649	0.071	0.198	0.111	1	0	0.271	0.005
Cc_SNP15	0	0.071	0.071	1	0.818	0.077	0.409	0.008
Cc_SNP16	−0.053	0.545	0.519	1	−0.519	0.769	0.517	0.112
Cc_SNP17	0.294	0.231	0.323	0.374	−0.176	0.364	0.312	1
Cc_SNP18	0.268	0.385	0.520	0.577	1	0	0.442	0.003
Mean	0.286	0.305	0.404		0.291	0.291	0.405	

F_{is} inbreeding coefficient, *Ho* observed heterozygosity, *He* expected heterozygosity, P_{HWE} Hardy–Weinberg Equilibrium *p* value

were detected with a frequency of one SNP each 183 bp (Table S1). Seventy percent of contigs containing SNPs were annotated to gene ontology (GO) using Blast2GO (Conesa et al. 2005). Herein, relevant genes were identified (Table S2). From *C. concholepas* SNP-dataset, eighty-four SNPs were evaluated by high-resolution melting analysis (HRMA) (Table S3). Genomic DNA was extracted from mantle of 50 individuals from Caleta Chungungo, Punta de Choros, Coquimbo (29°16′S–71°31′W) and Los Molinos, Valdivia (39°40′S–73°12′W), using rapid salt-extraction method. HRM primers were designed using Primer3 included in Geneious Pro 6.1 software (Biomatters, New Zeland). The PCR was carried out in 10 µl reaction with 13 ng template DNA using Fast EvaGreen® qPCR Master Mix (Biotum, USA). For HRMA, thermal cycling was performed with an ECO Real Time PCR System (Ilumina Inc., USA) as follows: 2 min for enzyme activation, 40 cycles: 95 °C for 5 s, 56 °C for 5 s, 60 °C for 25 s. HRMA data were collected between 60 °C and 95 °C with a temperature interval of 0.3 %. The genotyping were analyzed for the presence of discrete melting curve using the software EcoReal Time System (Illumina Inc., USA).

Eighteen of 84 SNPs were polymorphic (Table 1). The expected and observed heterozygosities estimated using FSTAT averaged 0.517 and 0 on the southern locality and 0.519 and 0.071 on the northern, respectively. Nine SNPs evidenced Hardy–Weinberg equilibrium for both analyzed populations. Average F_{is} estimated were 0.291 and 0.286 for the southern and northern populations, respectively. Pairwise F_{st} between populations was 0.004 indicating low genetic differentiation. Novel SNPs identified in relevant genes such as HSP70, Ferrine, MKNK1, UBE2 and Leucine amino peptidase (Table S4) will be useful to future conservation strategies in *C. concholepas*.

References

Cárdena L, Viard F, Castilla JC (2008) Paleobiogeography and taxonomy of the genus Concholepas (Lamarck, 1801): a review and new evidences. Rev Chil Hist Nat 81(3):417–436

Castilla JC (1999) Coastal marine communities: trends and perspectives from human-exclusion experiments. Trends Ecol Evol 14:280–283

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12(7):499–510

Fernández M, Castilla JC (2005) Marine conservation in Chile: historical perspective, lessons, and challenges. Conserv Biol 19:1752–1762