SNP identification and validation in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*)

L. Peñarrubia, J. Viñas, N. Sanz, B. L. Smith, J. R. Alvarado Bremer, C. Pla, O. Vidal

Peñarrubia, L., Viñas, J., Sanz, N., Smith, B. L., Alvarado Bremer, J. R., Pla, C., Vidal, O., 2019. SNP identification and validation in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*). *Animal Biodiversity and Conservation*, 42.1: 65–68.

Abstract

*SNP identification and validation in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*).* The development of affordable massive parallel sequencing (MPS) has reduced both time and costs of SNP identification for use in conservation and population genetic studies. After MPS, a second validation is usually required. High resolution melting analysis (HRMA) is a fast and simple method for mutation scanning, and thus a suitable validation protocol, particularly in non–model species. We present a set of nine novel polymorphic SNPs identified by MPS and validated with HRMA in two invasive species (the zebra mussel *Dreissena polymorpha* and the Asian clam *Corbicula fluminea*). These SNPs can be used in genetic studies to accurately assess and understand past and future invasion events.

Key words: *Corbicula fluminea*, *Dreissena polymorpha*, Invasive species, Massive parallel sequencing, Single nucleotide polymorphisms

Resumen

*Identificación y validación de PNU en dos especies invasoras: el mejillón cebra (*Dreissena polymorpha*) y la almeja asiática (*Corbicula fluminea*).* El desarrollo de las plataformas asequibles de secuenciación masiva en paralelo (SMP) ha reducido el coste y el tiempo en la identificación de marcadores de polimorfismos de nucleótido único (PNU) para su uso en estudios de genética de poblaciones y de conservación. Tras la SMP, suele ser necesaria una segunda validación. El análisis de las curvas de fusión a alta resolución (HRMA en su sigla en inglés) es un método rápido y sencillo para escanear mutaciones y, por tanto, es un protocolo adecuado de validación de dichos marcadores, especialmente en especies no modelo. En este trabajo se presenta un juego de nueve marcadores polimórficos de PNU nuevos identificados mediante SMP y validados con el HRMA en dos especies invasoras (el mejillón cebra *Dreissena polymorpha* y la almeja asiática *Corbicula fluminea*), que pueden utilizarse en estudios de genética de poblaciones para evaluar y entender correctamente los episodios de invasión pasados y los que podrían ocurrir en el futuro.

Palabras clave: *Corbicula fluminea*, *Dreissena polymorpha*, Especies invasoras, Secuenciación masiva en paralelo, Polimorfimos de nucleótido único

Received: 02 III 18; Conditional acceptance: 26 IV 18; Final acceptance: 04 VII 18

Luis Peñarrubia, Jordi Viñas, Nuria Sanz, Carles Pla, Oriol Vidal, Lab. d'Ictiologia Genètica, Dept. of Biology, Univ. de Girona, LEAR Building, Girona, 17073, Catalonia, Spain.– Brad L. Smith, Dept. of Natural Sciences, Brigham Young Univ. Hawaii, Laie, HI 96762, USA.– Jaime R. Alvarado Bremer, Dept. of Marine Biology, Texas A&M Univ. at Galveston, OCSB 3029 room #216, P. O. Box 1675, Galveston, TX 77553, USA and Dept. of Wildlife and Fisheries Sciences, Texas A&M Univ., TAMU 2258, College Station, TX 77840, USA.

Corresponding author: Oriol Vidal. E–mail: oriol.vidal@udg.cat

ISSN: 1578–665 X
eISSN: 2014–928 X

Rapid developments in massive parallel sequencing (MPS) technologies have facilitated the use of single nucleotide polymorphisms (SNPs) in population genetic studies (Morin et al., 2009). SNPs have many advantages, including low–scoring error rates, high abundance, functional relevance, easy high–throughput genotyping (Liu et al., 2005), and more accurate estimates of population differentiation (Morin et al., 2009).

SNP identification in non–model species can be performed using MPS technologies without a reference genome (Everett et al., 2011). However, after the SNP calling step, a subsequent validation is usually required. High resolution melting analysis (HRMA) is a relatively new and inexpensive technology. It is based on highly precise and accurate measures of melting temperatures (Tm) of PCR–amplified DNA achieved by recording the fluorescence of a saturating DNA dye (Wittwer et al., 2003; Reed et al., 2007). As differences in melting curve profiles are diagnostic of SNPs, homozygote and heterozygote genotypes can be distinguished (Montgomery et al., 2007). This technique has been successfully used for SNP validation in several species, such as swordfish (Smith et al. 2010) and chum salmon (Seeb et al., 2011).

In this study, we used non–coding genomic MPS reads from previous studies (Peñarrubia et al., 2015a, 2015b) to identify and subsequently validate with HRMA new SNPs in two invasive species, the zebra mussel (*Dreissena polymorpha*, Pallas, 1771) and the Asiatic clam (*Corbicula fluminea*, Müller, 1774). Roche 454 GS FLX reads of *D. polymorpha* (GenBank SRA accession number SRP051009) and *C. fluminea* (GenBank SRA accession number SRP073154) were processed using CLC Genomics Workbench version 4.0 (http://www.clcbio.com/) for SNPs as described in Seeb et al. (2011) and Shahin et al. (2012). As a result, 783 sequence variants were found in *D. polymorpha*, of which 721 were single nucleotide polymorphisms (SNPs) (356 transversions and 365 transitions), 20 were multiple nucleotide variants (MNVs) and 42 were insertions–deletions (InDels). Conversely, in *C. fluminea* the analysis produced 446 sequence variants, 417 of which were SNPs (188 transversions and 229 transitions), 11 were MNVs and 18 were InDels.

We selected 46 of these putative SNPs *in D. polymorpha* and 40 SNPs in *C. fluminea* for validation with short amplicon (SA) HRMA assays (Smith et al., 2013) in up to 96 individuals collected for previous studies (Peñarrubia et al., 2015a, 2015b). All the selected positions harbored a single SNP and the length of the amplicon after primer design was limited to less than 65 bp. Melting temperatures of the putative PCR products were pre–checked using uMELT^{SM} v2.0 (https://www.dna.utah.edu/umelt/umelt.html). BLASTN analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was run to identify possible homologies producing non–specific amplifications in the PCR.

HRMA amplifications were conducted in 10 µL reactions containing 25–100 ng of genomic DNA, 1× EconoTaq Plus Master Mix (Lucigen), 1× LC-Green+ (Idaho Technology), and 0.2 µM of each primer. Thermal cycling was performed on a LightCycler 480 Real–Time PCR system (Roche Diagnostics) with an initial denaturation of 10 min at 95 ºC followed by 35 cycles denaturing for 10 s at 95 ºC, annealing at 60 ºC for 30 s, and extension for 10 s at 72 ºC. Reactions were overlaid with 15 µL of mineral oil to ensure that evaporative losses did not affect ionic strength which may affect melting uniformity across samples (Smith et al., 2010). Twenty–five HRMA data acquisitions per ºC were collected with a ramp rate of 0.02 ºC/s between 60 and 95 ºC. All melting curve patterns were analyzed using the LightCycler 480 Gene Scanning Software v. 1.5.0 SP1 (Roche Diagnostics).

SA–HRMA characterization of the 46 putative positions in *D. polymorpha* indicated that five were monomorphic, seven produced patterns that were not consistent with homo– and hetero–duplex curves (e.g., three or more melting peaks, potential primer dimers, etc.), 23 produced heteroduplex curves (i.e., double peaks) indicative of the heterozygous condition in every individual in the sample (n = 15) tested, six failed to amplify even after repeated attempts to optimize PCR (not shown), and five loci (10,87 %) displayed polymorphic SNPs melting curves (table 1). In *C. fluminea*, four were identified as monomorphic, 10 produced non–scorable melting patterns, 18 generated double–peaks in every individual tested, four failed to amplify, and four SNPs (10 %) produced melting curves of polymorphic SNPs. While 10 % may be considered a low success rate in polymorphic yield (validated/polymorphic loci), our results suggest that a complete validation of the initial 783 and 446 sequence variants in both species would generate an increase in the number of SNPs to 40 and 80 respectively for *C. fluminea* and *D. polymorpha*.

Those positions with a heterozygous condition in all the first analyses (23 for *D. polymorpha* and 18 for *C. fluminea*) were genotyped in 96 individuals of each species, and all of them displayed the same result. This pattern may be explained by paralogous sequence variants (PSV) (Smith et al., 2005). PSV are a common trait of the genomes of mollusks because of the highly abundant cryptic repetitive genomic DNA (McInerney et al., 2011).

All the validated SNPs (five in *D. polymorpha* and four in *C. fluminea*) and their flanking regions are available in GenBank (accession numbers KT220181–85 for *D. polymorpha*, and KT220186 and KT220188–90 for *C. fluminea*, table 1). They are the first validated SNPs in the two species and they add to the limited number of currently available loci (Peñarrubia et al., 2015a, 2015b). They can thus be used in further studies with alternative genotyping techniques. Interestingly, even such a limited number of SNPs can be useful to describe population structure. In the invasive mosquito fish (*Gambusia holbrooki*), 5 SNPs were used to further characterize the genetic structure in European populations (Vidal et al., 2012), and in swordfish (*Xiphias gladius*) the same number of SNPs was able to detect genetic differentiation in Atlantic and Mediterranean samples (Smith et al., 2013).

Table 1. Name, nucleotide variant (NV), amplicon size (As), forward (F) and reverse (R) primers, and annealing temperature (T^a) for polymorphic SNPs in *D. polymorpha* and *C. fluminea*.

Tabla 1. Nombre, variante nucleotídica (NV), tamaño del amplicón (As), cebador directo (F) e inverso (R), temperatura de unión (T^a) para los PNU polimórficos en D. polymorpha *y* C. fluminea.

Name	NV	As	Primer sequences (5' → 3')	T^a (ºC)	GenBank Accesion
D. polymorpha					
Dp292	A/T	60	F: TGCAACCGAGTTTACCAACGGCT	57	KT220181
			R: TGCTGTTCAAATGAACCGGAGCAG		
Dp367	T/G	60	F: TCGCCTTGCAAGTCTCGTGCT	57	KT220182
			R: GCAATTGTTCTTGCAGTAATGTCCCGC		
Dp452	T/G	60	F: GCTGCCTGAAACGTTCAGTGGT	57	KT220183
			R: CCTCCGGGATCGGCCCACTT		
Dp467	A/G	54	F: TGCGTGGAGCCTTTCCACCG	57	KT220184
			R: TGGCAAGAACAAAGCAGACCGC		
Dp501	C/T	55	F: GTGTGAAATCTTGAAAGCGCCTTGT	57	KT220185
			R: GGCTGCTGGTAAATAAATGGGCTCCG		
C. fluminea					
Cf46	C/G	54	F: CGAAAGCTGCGCATTTCTGCGA	57	KT220186
			R: ACCTGCGGATGGATCATTACCGA		
Cf132	T/G	59	F: TGTAGGCGGCCACCCCATGT	60	KT220188
			R: GGTCTTCACTGACGGGCGGC		
Cf190	T/A	60	F: AGCTTACAGTTTGCCCACTTACCTCT	57	KT220189
			R: AGATGCGAATTGGCCCCGGT		
Cf270	T/A	60	F: GTAATGTCCGTCTGCGTATCAGATTCA	57	KT220190
			R: TGCCGGGGTGTCTTGTTTGTCG		

Acknowledgements

We are indebted to the following colleagues for animal samples: R. Miranda (Universidad de Navarra, Spain), J. Checa (Equipos y Suministros S. L., Zaragoza, Spain), I. Planas and R. Puig (Nyctibios Association, Barcelona, Spain), and C. Robertson (Texas Parks and Wildlife Department, TX, USA). This research was carried out within the scope of the research project CGL200909407 of the Ministry of Sciences and Innovation (MICINN) of the Spanish Government. LP received economic support with a PhD fellowship from the Spanish MICINN with reference BES–2010037446. LP also received economic support from the Spanish MICINN for a fellowship with reference EEBB–I–12–05670.

References

Everett, M. V., Grau, E. D., Seeb, J. E., 2011. Short reads and nonmodel species: exploring the complexities of next–generation sequence assembly and SNP discovery in the absence of a reference genome. *Molecular Ecology Resources*, 11: 93–108.

Liu, N., Chen, L., Wang, S., Oh, C., Zhao, H., 2005. Comparison of single–nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, 6: S26.

McInerney, C. E., Allcock, A. L., Johnson, M. P., Bailie, D. A., Prodöhl, P. A., 2011. Comparative genomic analysis reveals species–dependent complexities that explain difficulties with microsatellite marker development in molluscs. *Heredity*, 106: 78–87.

Montgomery, J., Wittwer, C. T., Palais, R., Zhou, L., 2007. Simultaneous mutation scanning and genotyping by high–resolution DNA melting analysis. *Nature Protocols*, 2: 59–66.

Morin, P. A., Martien, K. K., Taylor, B. L., 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, 9: 66–73.

Peñarrubia, L., Araguas, R. M., Pla, C., Sanz, N., Viñas, J., Vidal, O., 2015a. Identification of 246

microsatellites in the Asiatic clam (*Corbicula fluminea*). *Conservation Genetics Resources*, 7: 393–395.

Peñarrubia, L., Sanz, N., Pla, C., Vidal, O., Viñas, J., 2015b. Using Massive Parallel Sequencing for population genetics markers development, validation and application in the invasive bivalve zebra mussel (*Dreissena polymorpha*). *PLOS One*, 10: e0120732.

Reed, H. G., Kent, J. O., Wittwer, C. T., 2007. High–Resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics*, 8: 597–608.

Seeb, J. E., Pascal, C. E., Grau, D. E., Seeb, W., Templin, W. D., Harkins, T., Roberts, S. B., 2011. Transcriptome sequencing and high–resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources*, 11: 335–348.

Shahin, A., van Gurp, T., Peters, S. A., Visser, R. G., van Tuyl, J. M., Arens, P., 2012. SNP markers retrieval for a non–model species: a practical approach. *BMC Research Notes*, 29: 5–79.

Smith, B. L., Lu, C. P., Alvarado Bremer, J. R., 2010. High–resolution melting analysis (HRMA): a highly sensitive inexpensive genotyping alternative for population studies. *Molecular Ecology Resources*, 10: 193–196.

Smith, B. L., Lu, C. P., Alvarado Bremer, J. R., 2013. Methodological streamlining of SNP discovery and genotyping via high–resolution melting analysis (HRMA) in non–model species. *Marine Genomics*, 9: 39–49.

Smith, C. T., Elfstrom, C. M., Seeb, L. W., Seeb, J. E., 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology*, 14: 4193–4203.

Vidal, O., Sanz, N., Araguas, R. M., Fernandez–Cebrián, R., Díez del Molino, D., García–Marín, J. L., 2012. SNP diversity in introduced populations of the invasive *Gambusia holbrooki*. *Ecology of Freshwater Fish*, 21: 100–108.

Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., Pryor, R. J., 2003. High–resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry*, 49: 853–860.