Supplementary Document 1:
Microsatellite Development Narrative

INTRODUCTION

Gonidea angulata (I. Lea 1838) or the Rocky Mountain ridged mussel is one of four major freshwater mussel lineages in the western United States (Chong et al. 2008). With an overall decline in Unionids across their range and their importance in ecological and cultural roles (Bogan 2008, Lopes-Lima et al. 2014, Haag 2012, Blevins 2017), the use of cost-effective genetic markers can aid in monitoring populations and inform future conservation efforts. The importance of these approaches is further illustrated by the global vulnerable conservation status listing by both NatureServe (G3, 2007) and the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (2017).

Here, we develop the first set of species-specific microsatellites for *G. angulata* (Rocky Mountain ridged mussel). Due to their intrinsic characteristics such as co-dominant inheritance, high polymorphism and stability, microsatellites are an ideal tool for assessing population divergence, diversity, gene flow, and detection of inbreeding (Sun et al. 2008). Despite the growing advancements in high-throughput sequencing, microsatellites remain a valuable, low-cost tool to assess population level genomic metrics, especially for studies with smaller sample sizes, which are unable to justify the investment in next generation sequencing (NGS) library preparation and sequencing. This is especially true for well-designed polymorphic loci that can be efficiently multiplexed, scored, and analyzed.

MATERIALS AND METHODS

Development of primers for microsatellite assays for *G. angulata* was broken down into 4 main steps 1) sequencing and discovery, 2) screening and primer design, 3) simplex testing and 4) final microsatellite selection.

Sequencing and Discovery

All samples were preserved in 95% ethanol. Genomic DNA was extracted from each of the subsampled tissues using a Qiagen DNeasy Blood and Tissue kit following the manufacturer protocol, with negative controls in each separate extraction group. Five individuals from different populations in the Okanagan basin in British Columbia, Canada were pooled for library preparation and 'shotgun' genomic sequencing. Populations included in library preparation, sequencing and discovery include KBB, LKB, OKV, RFV, and RLC (Table A). 'Shotgun' genomic sequencing was performed using and Illumina MiSeq® platform, following the manufacturer's instructions. A single run 600-cycle v3 kit with a 2X300 paired-end configuration was used yielding 20 million sequences of varying length. Sequences were assembled, cleaned, and screened using SSR_Pipeline (Miller et al. 2013) resulting in sequences containing simple and complex sequence repeats. SSR_Pipeline (Miller et al. 2013) parameters were set with a minimum 40 base pair (bp) flank length, to allow for primer design, and a minimum number of microsatellite repeats set to 7, 6, and 5 respectively, for di- (2mer, e.g. AG),

tri- (3mer, e.g. TTC), and tetra-nucleotide (4mer, e.g. ATGT) tandem sequence repeats (microsatellites), resulting in 6,185 (3,282 2mer, 1,695 3mer, 1,208 4mer) sequences. Raw sequence data can be found in Documents S2: Raw Sequence Data.

Screening and Primer Design

Microsatellite sequences were further screened using custom scripts (Document S3: Screening Scripts) to contain microsatellites that contained at least one 'G' or 'C' within the motif. Simple motifs and compound motifs (2 or more identical microsatellite repeats separated by a different short sequence) with no more than 2 motif types per sequence were also included in our selection process. Simple sequences repeats were further filtered to contain 10-20 motif repeats (Gardner et al. 2011), compound microsatellites to contain between 5-15 motif repeats each locus, and removal sequences with >1 microsatellite. Subsequently, this resulted in 1,179 candidate sequences (948 2mer, 65 3mer, and 166 4mer) to be used in simplex testing and final microsatellite selection (Document S4: Filtered Sequences).

Sequence filtering was followed by primer design, using the online utility BatchPrimer3 (You et al. 2008). BatchPrimer3 settings were set to 45 - 55% primer GC content, primer melting temperature of 58 – 60 °C with a max temperature difference of 2 °C, product size 100 – 500 bp, primer size of 18 – 25 bp, and a 3' GC clamp (i.e. last two 3' nucleotides contain one G or C) (Castoe et al. 2012). The above conditions produced 577 primer pair candidates for use in the subsequent simplex testing (amplification of a single PCR product), polymorphism (within and among populations), and optimization (PCR assay fine-tuning and multiplexing design) steps. Candidate loci in which primer pairs were successfully designed can be found in Document S5: Primer Design.

Simplex Testing

Loci for simplex testing were selected based on amplicon size range and number of motif repeats. Simplex testing was completed with an initial round of 48 loci followed by a second round containing 36 loci. The first round contained 9 loci from each of 5 amplicon size ranges (100 - 175 bp, 176 – 250 bp, 251 – 325 bp, 326 - 400 bp, and 401 – 500 bp) and was tested through the polymorphism stage before the second round of loci were selected (Lepais et al. 2011). 3mer and 4mer loci with the highest number of motif repeats were selected first, anticipating that this would improve selection of loci with high polymorphism, reduce PCR stutter, and increase scoring reliability. Both simplex rounds were initially tested without modification to the designed primers, i.e. without any additional fluorophores or tags (M13, CAG, or PIGtails) (Oetting et al. 1995, Culley et al. 2013, Brownstein et al. 1995), and verified via agarose gel to confirm robust and concise amplification in the appropriate size range with an absence of spurious bands. The second round of simplex testing used the same criteria as round 1 but included additional loci from amplicon size ranges that performed poorly during round 1 testing. This allowed us to optimize multiplexing potential for a final set of loci. Simplex testing was done in 10µL reactions with 1x MyTaq HS Master Mix (Bioline™), 0.5 µM each primer, 0.1 mg/mL BSA, and approximately 15 ng genomic DNA. DNA from eight populations (OL3M, OKV, GPM, GKK, GJH, GCV, GSK, and GJF; Table A) was pooled for simplex testing in duplicate PCR reactions to assure that amplification failure was not due to technician error,

reagent failure, or inhibition. Thermal cycling conditions consist of 95 °C for 3 minutes, 35 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 40 s, followed by 72 °C for 10 minutes.

Table A. List of populations, site codes, geographic location and the development step in which it was used. Code definitions for 'Step Used' are sequencing and discovery (SD), simplex testing (ST), and final microsatellite selection (FS).

Location	Site Code	Step Used			LatitudeDegN	LongitudeDegW
Kin Beach	KBB	SD		FS	50.250238	-119.350013
Summerland	LKB	SD		FS	49.60754	-119.650085
Naramata	OL3M		ST	FS	49.599536	-119.602172
Vaseaux Lake	OKV	SD	ST	FS	49.301025	-119.531299
Fairview Campground	RFV	SD		FS	49.192434	-119.550677
Osoyoos Lakehead campsite	RLC	SD		FS	49.043821	-119.468089
Similkameen River, WA	GSK		ST	FS	48.95341	-119.648716
Chehalis River, WA	GCV		ST	FS	46.981879	-123.470972
John Day, Middle Fork, OR	GJR			FS	44.888068	-119.196832
John Day, Middle Fork, OR	GJH		ST		44.75985	-118.865474
John Day, Middle Fork, OR	GJF		ST		44.738551	-118.847699
Clear Creek, CA	GKK		ST	FS	41.710523	-123.448757
Pit River, CA	GPM		ST		40.96497	-121.790283

Final Microsatellite Selection

Following these two rounds of simplex screening, candidate loci were tested for polymorphism, repeatability, interpretability, allele range, and signal strength using a fluorophore-labeled 3-primer multiplex system (Culley et al. 2013). Multiplex Manager v1.0 software (Holleley and Geerts, 2009) was used to combine loci that passed simplex testing into multiplex groups, allowing for multiple loci to be combined into a single PCR reaction. The forward primer of each locus was then tagged with 1 of 3 fluorophore-specific M13 or M13-modified tags as per Culley et al. (2013). Additionally, to reduce unwanted adenylation and peak stutter, a PIGtail was attached to the reverse primer of each locus as described in Brownstein et al. (1995). A third fluorophore-labeled primer complementary to the forward primer tag was also used to accommodate multiplexing (Culley et al. 2013). PCR products were sent to Eton Biosciences (San Diego, CA) for fragment analysis. Each locus being tested contained four reactions. Reactions (1) and (2) contained the same sample to test for repeatability, interpretability, and signal strength. Reaction (3) contained another separate individual to further test interpretability and signal strength. Reaction (4) contained 12 unique individuals to test for allelic size range as well as the number of unique alleles (Lepais et al. 2011, Cryer et al. 2005). Allelic size range tests included individuals from 4 additional populations in the US (Table A).

After reviewing the first round of primers we opted to continue with a second round to target a minimum of 15 high-quality and variable loci that could be identified and multiplexed efficiently. From this point the second round of loci was selected and subjected to simplex and

polymorphism testing. The best-performing loci were selected from each round and regrouped via Multiplex Manager for a final step of fine-tuning multiplex designs.

In all, 18 loci contained in 5 multiplex groups were selected for final optimization (GenBank accessions MT209995-MT210012). Additionally, each group was tested with both bovine serum albumin (BSA) and dimethyl sulfoxide (DMSO), and each with BSA and DMSO separately. In all cases, no significant improvement of the reactions was detected. Final amplification reactions contained 1x MyTaq HS Master Mix (Bioline™), 0.2 µM of each PIGtailed reverse primer, 0.05 µM each forward tailed primer, 0.2 µM fluorophore-tagged primer (M13(-21), T7term, or M13modA; Culley et al. 2013), and approximately 15 ng genomic DNA in 10 µL total reaction volume. The reaction was denatured at 95°C for 3 min, followed by 35 cycles of 95 °C for 30 s, 59°C (Mix A) or 62°C (all other mixes) for 40 s, 72°C for 50 s, with a final extension step of 72°C for 10 minutes. Amplified fragments were analyzed at Eton Biosciences (San Diego, CA) and alleles subsequently scored using GeneMarker™ v.2.6.2 (SoftGenetics, LLC) software. The final set of 18 microsatellite loci developed and optimized is presented in Table S2.

Literature Cited

Blevins, E., Jepsen, S., Brim Box, J., and Nez, D. 2016. *Gonidea angulata* (errata version published in 2017). The IUCN Red List of Threatened Species 2016: e.T173073A114146359.

Bogan, A. E. 2008. Global diversity of freshwater mussels (Mollusca, Bivalvia) in freshwater. Hydrobiologia 595: 139–147.

Brownstein, M. J, Carpten, J. D., and Smith, J. R. 1995. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: Primer modifications that facilitate genotyping. BioTechniques 20:1004–1010.

Chong, J. P., Brim Box, J. C., Howard, J. K., Wolf, D., Myers, T. L., and Mock, K. E. 2008. Three deeply divided lineages of the freshwater mussel genus Anodonta in western North America. Conservation Genetics 9:1303–1309.

Cryer, N. C., Butler, D. R., and Wilkinson, M. J. 2005. High throughput, high resolution selection of polymorphic microsatellite loci for multiplex analysis. Plant Methods 1:3.

Culley, T. M., Stamper, T. I., Stokes, R. L., Brzyski, J. R., Hardiman, N. A., et al. 2013. An efficient technique for primer development and application that integrates fluorescent labeling and multiplex PCR. Applications in Plant Sciences 1:1300027.

Haag, W. R. 2012. North American Freshwater Mussels: Natural History, Ecology, and Conservation. Cambridge University Press, Cambridge, England. 288 pp.

Holleley, C. E., Geerts, P. G. 2009. Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. Biotechniques 46:511–517.

Lepais, O., Bacles, C. F. 2011. Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. Ex Benth. 2011. Molecular Ecology Resources 11:711–724.

Lopes-Lima, M., Teixeira, A., Froufe, E., Lopes, A., Varandas, S., and Sousa, R. 2014. Biology and conservation of freshwater bivalves: past, present and future perspectives. Hydrobiologia 735:1–13.

Miller, M. P., Knaus, B. J., Mullins, T. D., and Haig, S. M. 2013. SSR_pipeline: A bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data. Journal of Heredity 104:881–885.

Oetting, W. S., Lee, H. K., Flanders, D. J., Wiesner, G. L., Sellers, T. A., and King, R. A. 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30:450–458.