**Development and characterization of twenty-nine microsatellite markers for
Ligumia nasuta (Bivalvia: Unionidae) using an Illumina sequencing approach.**

Mariah W. Scott^{1*}, Jordan R. Hoffman¹, Trevor L. Hewitt^{1,2}, Rochelle R. Beasley³,
Stacey L. Lance³, Kenneth L. Jones⁴, Todd J. Morris⁵, & David T. Zanatta¹

¹ Central Michigan University, Institute for Great Lakes Research, Biology Department,
Brooks Hall 217, Mount Pleasant, MI 48859, U.S.A.

² University of Michigan Museum of Zoology, Division of Mollusks, 1013 Ruthven, Ann
Arbor, MI 48109, U.S.A. (Current Address)

³ Savannah River Ecology Laboratory, University of Georgia, Aiken, SC 29802, U.S.A.

⁴ Department of Biochemistry and Molecular Genetics, University of Colorado School of
Medicine, Aurora, CO 80045, U.S.A.

⁵ Fisheries and Oceans Canada, Great Lakes Laboratory for Fisheries and Aquatic
Sciences, 867 Lakeshore Rd., Burlington ON L7S 1A1, Canada

*Corresponding author, email: scott2mw@cmich.edu

Abstract

Ligumia nasuta (Say, 1817; Eastern Pondmussel) is an imperiled freshwater mussel (Unionidae) in eastern North America. Population declines in the Laurentian Great Lakes resulting from the introduction of dreissenid mussels and habitat destruction in the 20th Century have greatly reduced and limited its distribution. To properly inform restoration and management efforts for *L. nasuta*, fine-scale genetic analyses must be performed on the remnant populations. This study used Illumina paired-end shotgun sequencing to identify potential microsatellite loci for *L. nasuta*, utilizing two samples to develop the Illumina paired-end shotgun library. Forty-eight primer pairs were tested on the remaining 24 samples. Twenty-nine of the 48 microsatellite primer sets screened were successfully amplified using 24 *L. nasuta* samples collected from the Great Lakes watershed. The estimated fragment size ranged from 167-445 base-pairs (bp) and the number of alleles per locus varied between 5 and 16 (mean = 9.7). Only five of the loci deviated significantly from Hardy-Weinberg expectations after Bonferroni corrections. The development of these new microsatellite loci will greatly facilitate future genetic studies on *L. nasuta*.

Highlights

- Microsatellite loci identified from two specimens using Illumina sequencing
- 48 primer pairs tested for amplification success on 24 *L. nasuta* specimens
- 29 loci were successfully screened and characterized

Keywords

Conservation genetics, SSR, population genetics

Introduction

Ligumia nasuta (Say, 1817; Eastern Pondmussel) is a regionally imperiled freshwater bivalve species from the order Unionoida (unionids). *Ligumia nasuta* inhabits the Atlantic Slope and Great Lakes region of North America (Sepkoski and Rex, 1974; Strayer and Jirka, 1997; Metcalfe-Smith et al., 1998). Along the Atlantic Slope, *L. nasuta* occurs in rivers draining into the Atlantic Ocean from South Carolina to New Hampshire (NatureServe, 2015). In the Great Lakes region, *L. nasuta* was historically widespread and common in nearshore and coastal areas of the Great Lakes, low gradient tributaries, and some inland lakes. In western Lake Erie and Lake St. Clair, *L. nasuta* was an abundant and common species (Nalepa et al., 1991; Nalepa et al., 1996). The species was also found in eastern Lake Erie, coastal Lake Ontario, and parts of the Lake Huron and Lake Michigan watersheds (Masteller et al., 1993; COSEWIC, 2007; Michigan Natural Features Inventory, 2015). The introduction of dreissenid mussels [*Dreissena polymorpha* (Pallas, 1771) and *D. rostriformis bugensis* (Andrusov, 1897)] in the late 1980s and early 1990s lead to precipitous declines of *L. nasuta*, as the habitats it typically occupied were among the most heavily infested areas by these invasive mussels (Schloesser and Nalepa, 1994; Nalepa et al., 1996; Schloesser et al., 1996; COSEWIC, 2007). The current distribution in the Great Lakes region now consists of small, isolated populations within the historical boundaries, many of which are under continued threat by dreissenid mussels (Nalepa et al., 1996; COSEWIC, 2007; Zanatta et al., 2015). As a result of these declines, *L. nasuta* is listed as endangered in Canada (Species at Risk Public Registry, 2016; NatureServe, 2015). In the United States, *L. nasuta* is ranked as

critically imperiled in eight states and imperiled in three, out of 13 ranked states (NatureServe, 2015).

The phylogeographic structure of *L. nasuta* populations was previously analyzed by Scott et al. (2014), using mitochondrial DNA (mtDNA) sequences. Scott et al. (2014) focused on *L. nasuta* in the Great Lakes region, with some samples collected from Atlantic coastal rivers. Two mtDNA regions (Cytochrome Oxidase subunit 1 and NADH Dehydrogenase subunit 1) were sequenced and revealed very low genetic diversity in the Great Lakes region, with only one and two haplotypes, respectively. The results indicated that the Great Lakes region could be treated as a single population for conservation management purposes. Scott et al. (2014) did not find sufficient genetic diversity to distinguish among putative management units using mtDNA sequence data. To properly inform restoration and management efforts for *L. nasuta*, more fine-scale genetic analyses, conducted through comparisons of multiple microsatellite loci, instead of two single-locus mitochondrial genes, must be performed.

Microsatellite loci, also known as simple sequence repeats (SSR), are commonly used in detailed genetic analyses to inform conservation efforts for imperiled species (Baker, 2000; Freeland et al., 2011). The development of highly variable microsatellite markers for *L. nasuta* will allow future studies to analyze the genetic diversity of the species on a finer scale, compared to mtDNA sequences. The purpose of this study was to use Illumina paired-end shotgun sequencing (Lance et al., 2013) to develop and characterize microsatellite loci for *L. nasuta*. These new markers will facilitate future range-wide studies on the genetic diversity and structure of *L. nasuta* and aid in making

sound recovery and management plans guiding population augmentation and artificial propagation programs.

Materials and Methods

Ligumia nasuta were collected from 10 sampling locations in four waterbodies, in the Great Lakes region (n = 26; Scott et al., 2014; Douglas Lake, MI; Paradise Lake, MI; Lake St. Clair, MI and ON; and Presque Isle, PA). Mussels were collected by tactile search with snorkels and/or SCUBA, or by using clam rakes during extensive surveys in 2011 and 2012 (Zanatta et al., 2015). Genetic material was collected from the mussels using non-destructive swabs of the foot (Henley et al., 2006). The swab tip and accumulated mucus from the foot of the mussel were stored in 250 µl of sterile lysis buffer. Each mussel was returned to the substrate, and samples were brought to the lab and placed at -20°C. The DNA from the individual swab samples were incubated in an overnight digestion with 10 µl Proteinase K (20 mg/ml), using alcohol extraction (Sambrook et al., 1989). The genomic DNA was stained with SYBR Green and electrophoresed in a 1.5% agarose gel to confirm the quality of the extractions.

An Illumina paired-end shotgun library was developed for the microsatellite DNA using the two samples of the highest genomic concentration from the 26 samples, following methods described by Lance et al. (2013). The library was prepared by first shearing 1 µg of DNA with a Covaris™ S220, following the standard protocol for the Illumina TruSeq® DNA Library Kit, and using a multiplex identifier adaptor index. Ten million of the resulting reads were analyzed using PAL_FINDER_v0.02.03 (Castoe et al., 2010). Reads containing di-, tri-, tetra-, penta-, and hexanucleotide microsatellites were

identified. PRIMER3 (version 2.0.0) was used to identify primer regions based on the reads containing putative microsatellite regions. During the designing of locus specific primers, a random selection of tetra- and penta-nucleotide microsatellites were used, in order to simplify the process of scoring during genotyping.

Forty-eight primer pairs were tested on the remaining 24 samples. Polymerase Chain Reaction (PCR) amplifications were performed in a 12.5 μ L reaction volume (1X Buffer, 25 μ g/mL BSA, 0.4 μ M unlabeled primer, 0.04 μ M tag labeled primer, 0.36 μ M 6-FAM florescent dye-labeled CAG universal primer (5' - CA GTC GGG CGT CAT CA - 3'), 3 mM $MgCl_2$, 0.8 mM dNTPs, 0.5 U AmpliTaq Gold® polymerase (Applied Biosystems), 20 ng DNA template). All of the samples were amplified with a touchdown thermal cycling program, with a span of annealing temperatures ranging between 65°C and 55°C. The touchdown cycling parameters consisted of an initial denaturation step at 95°C for 5 min, followed by 20 cycles of 95°C for 30 s, 65°C (decreased by 0.5°C per cycle) for 30 s, and 72°C for 30 s; then 20 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. All loci were run separately on an ABI3130XL sequencer and sized with ROX labeled size standard prepared as described in DeWoody et al. (2004). The resulting alleles were analyzed using GENEMAPPER version 3.7 (Applied Biosystems).

For the successful primers, estimated fragment length, number of successfully genotyped sequences, randomization tests for linkage disequilibrium, and number of observed alleles were assessed using GENALEX version 6.5 (Peakall and Smouse, 2012). Additionally, the observed versus the expected heterozygosity, based on Hardy-Weinberg equilibrium (HWE) expectations after Bonferroni corrections, and the

probability of identification in Genalex version 6.5 (Peakall and Smouse, 2012) were calculated using GENALEX version 6.5.

Results and Discussion

Of the 19,230 Illumina reads with tetra- and penta-nucleotide repeats, 48 putative loci were selected at random for screening and primers were ordered. Twenty-nine of the 48 microsatellite primer sets screened were successfully amplified from the 24 *L. nasuta* samples (Table 1). While many of the loci had the same repeat motif, each locus selected had unique flanking regions. Between 21 and 24 of the *L. nasuta* samples were successfully genotyped at each of the 29 loci. Estimated fragment size at each locus was between 167 and 445 bp (base-pairs), including the 16 bp CAG tag on the forward primer. The number of alleles per locus varied between 5 and 16 (mean = 9.7). Only five of the loci displayed heterozygosity levels that significantly deviated from HWE expectations after Bonferroni corrections. These deviations from HWE could be the result of a Wahlund effect, given that the specimens genotyped were from several different regions and waterbodies. Randomization tests found nine out of 406 pairs had significant linkage disequilibrium among the samples between different pairs of loci. With only 24 samples included, the statistical power of the tests were not strong, but important to consider in future population genetic studies for the species. The probability of identity for each individual locus, indicating the chance for two random individuals to share a common genotype, varied between 0.021 and 0.510 (mean = 0.092).

The development of 29 loci using the Illumina pair-end shotgun sequencing approach was a success, similar to other studies using next-generation sequencing

approaches for developing and characterization new microsatellite loci. Other recent efforts using next-generation sequencing techniques with unionids have produced between nine and 16 new characterized microsatellite loci per species (O'Bryhim et al., 2012; Froufe et al., 2013; Popa et al., 2015). The allelic richness at each locus described in this study is similar to the richness found at the characterized microsatellite loci (between 2 and 25 alleles per locus) in other unionid species (O'Bryhim et al., 2012; Froufe et al., 2013; Popa et al., 2015). With 29 polymorphic microsatellite loci now characterized and detailed, fine-scale genetic-diversity studies, including parentage analysis, will be possible for *L. nasuta*. These newly developed microsatellite markers will be useful for facilitating future range-wide studies on the genetic diversity and structure of *L. nasuta* and aid in making sound recovery and management plans guiding population augmentation and artificial propagation programs.

Acknowledgements

Specimens collected for this project were collected under permits obtained from the Canadian government, Ontario, Michigan, and Pennsylvania. Fisheries and Oceans Canada provided the funding for this project. Manuscript preparation was also partially supported by the DOE under award number DE-FC09-07SR22506 to the University of Georgia Research Foundation. Bioinformatics support came from the University of Colorado Cancer Center Bioinformatics Shared Resources, which is supported in part by grant #P30-CA046934 from the National Cancer Institute. This article is contribution #XXX of the Central Michigan University Institute for Great Lakes Research.

197 Literature Cited

- 198 Baker, A., 2000. Molecular Methods in Ecology. Blackwell Science, Malden, Maryland.
 199 Castoe, T.A., Poole, A.W., Gu, W., de Koning, A.P.J., Daza, J.M., Smith, E.N., Pollock,
 200 D.D., 2010. Mol. Ecol. Resour. 10, 341.
 201 COSEWIC, 2007. COSEWIC Assessment and Status Report on the Eastern Pondmussel,
 202 *Ligumia nasuta*, in Canada. COSEWIC, Ottawa, ON.
 203 DeWoody, J.A., Schupp, J., Kenefic, L., Busch, J., Murfitt, L., Keim, P., 2004.
 204 BioTechniques 37, 348.
 205 Freeland, J.R., Kirk, H., Petersen, S.D., 2011. Molecular Ecology, Second Edition. John
 206 Wiley & Sons, Ltd., West Sussex, UK.
 207 Froufe, E., Sobral, C., Teixeira, A., Lopes, A., Sousa, R., Varandas, S., Lopes-Lima, M.,
 208 2013. Aquat. Conserv. 23, 619.
 209 Henley, W.F., Grobler, P.J., Neves, R.J., 2006. J. Shellfish Res. 25, 975.
 210 Lance, S.L., Love, C.N., Nunziata, S.O., O'Bryhim, J.R., Scott, D.E., Flynn, R.W., Jones,
 211 K.L., 2013. PLOS One 8, 1.
 212 Masteller, E.C., Maleski, K.R., Schloesser, D.W., 1993. J. Penn. Acad. Sci. 67, 120.
 213 Metcalfe-Smith, J.L., Staton, S.K., Mackie, G.L., Lane, N.M., 1998. J. Great Lakes Res.
 214 24, 845.
 215 Michigan Natural Features Inventory. Rare Species Explorer. Retr. Oct. 2015,
 216 <http://mnfi.anr.msu.edu/explorer/search.cfm>
 217 Nalepa, T.F., Manny, B.A., Roth, J.C., Mozley, S.C., Schloesser, D.W., 1991. J. Great
 218 Lakes Res. 17, 214.
 219 Nalepa, T.F., Hartson, D.J., Gostenik, G.W., Fanslow, D.L., Lang, G.A., 1996. J. Great
 220 Lakes Res. 22, 354.
 221 NatureServe Explorer: An Online Encyclopedia of Life. *Ligumia nasuta*. Retr. Oct. 2015,
 222 <http://explorer.natureserve.org/>
 223 O'Bryhim, J., Chong, J.P., Lance, S.L., Jones, K.L., Roe, K.J., 2012. Conserv. Genet.
 224 Resour. 4, 787.
 225 Peakall, R., Smouse P.E., 2012. Bioinformatics 28, 2537.
 226 Popa, O.P., Bartakova, V., Bryja, J., Reichard, M., Popa, L.O., 2015. Biochem. Syst.
 227 Ecol. 60, 234.
 228 Sambrook J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual,
 229 2nd Ed. Cold Spring Harbor Laboratory Press, Plainview, New York.
 230 Schloesser, D.W., Nalepa, T.F., 1994. Canadian Journal of Fisheries and Aquatic
 231 Sciences 51: 2234.
 232 Schloesser, D.W., Nalepa, T.F., Mackie, G.L., 1996. Am. Zool. 36, 300.
 233 Scott, M.W., Begley, M.T., Krebs, R.A., Zanatta, D.T., 2014. Walkerana 17, 60.
 234 Sepkoski, J.J., Rex, M.A., 1974. Syst. Zool. 23, 165.
 235 Species at Risk Public Registry. Species Profile: Eastern Pondmussel. Retr. Apr. 2016,
 236 http://www.sararegistry.gc.ca/species/speciesDetails_e.cfm?sid=961
 237 Strayer, D.L., Jirka, K.J., 1997. The Pearly Mussels of New York State. New York State
 238 Museum, Albany, New York.
 239 Zanatta, D.T., Bossenbroek, J.M., Burlakova, L.E., Crail, T.D., de Szalay, F., Griffith,
 240 T.A., Kapusinski, D., Karatayev, A.Y., Krebs, R.A., Meyer, E.S., Paterson, W.L.,

241 Prescott, T.J., Rowe, M.T., Schloesser, D.W., Walsh, M.C., 2015. Northeast. Nat.
242 22, 223.

Table 1. Details for 29 polymorphic microsatellite loci developed for *Ligumia nasuta*, with unique flanking regions. The size indicates the range of observed alleles in base pairs and includes the length of the CAG tag on forward primer sequences (5' - CA GTC GGG CGT CAT CA - 3'); *N* is number of individuals successfully genotyped, of the twenty-four individuals amplified; *N_a* is number of alleles observed; *H_o* and *H_e* are observed and expected heterozygosity, respectively; *PI* is the probability of identity for each locus.

Locus	Primer Sequence 5' --> 3'	GenBank Accession Number	Repeat motif	Size (bp)	N	N _a	H _o	H _e	PI
Lina3	F: GCTGCAACAATGTTACTACCCG R: TTTGTGGAGCGTGTGACG	KU561833	ATAC	316- 332	24	5	0.667	0.769	0.095
Lina4	F: TCCAAATTCCGAGCTGGC R: TCTTTCTTATCATCAGGAAATTGTGC	KU561834	ATCT	252- 328	22	11	0.545	0.876	0.028
Lina6	F: TTACCACAACATAACACAAGATATGCC R: GACCATGGAACCATCCAAGC	KU561835	ATAC	167- 211	24	8	0.625	0.823	0.055
Lina8	F: AATGTCGCGATCTAGACTTTGC R: AGGACTGGTGCGAATGACG	KU561836	ATCT	304- 364	22	13	0.727	0.848	0.038
Lina10*	F: TGTAATGATAATACTGAGGATGTAATTTTCG R: GAATAAGAAGTGATGTGTGCG	KU561837	ATAC	284- 356	24	13	0.833	0.889	0.022
Lina11	F: GCTTGGGAACTGGAATTGTAAGG R: GCGCGTGAGACATATCTATAATCC	KU561838	ATCTT	250- 290	24	5	0.458	0.630	0.20
Lina15*	F: TGGCCTAATTTAATGCACAACG R: TGTAATTATATACTGCTAATTTGTCCAGCC	KU561839	ATAC	330- 390	24	11	0.417	0.819	0.053
Lina17	F: CCTATCTTCCTACCCGCCG R: TTTCCATTAGCAGATTTTCATTGC	KU561840	ATAC	338- 374	24	9	0.542	0.765	0.080
Lina18*	F: CCTGCAGATGGCTTAGACTCC R: CATCCCAGTATTAAGGCAGATATGG	KU561841	ATAC	231- 275	24	10	0.542	0.800	0.062

Lina19	F: TGGGAAGAGAAGGTAGTTCAGG R: TCGATGCACTACGAGAGTTCG	KU561842	ATAC	269-365	24	13	0.833	0.859	0.035
Lina20	F: CAACACTAATATTGCATACACTGGATAGG R: GGTACACTAACATCGAGCCACC	KU561843	ATAAC	342-378	24	10	0.583	0.832	0.047
Lina21	F: AAATATGTGACTATGTCCTTTCAAGC R: GCATTGGTATTAAGGACGTTAGG	KU561844	ATATT	318-383	24	6	0.292	0.297	0.51
Lina22	F: AAGACTGCGTCTTGAAAGTTGG R: AAATGTGGGTCTTCATTTACG	KU561845	ATAC	242-282	24	9	0.625	0.747	0.098
Lina23	F: AGTTTGAATCTGTGCCCACG R: TCTTTCCCAGTTATATGTTATACCGC	KU561846	ATAC	341-401	24	14	0.917	0.892	0.021
Lina24	F: TGAACAAGAAAGGTGATTGGG R: TTGTTGAGGAAAGAGGGTTCG	KU561847	ATAC	375-407	24	8	0.625	0.806	0.064
Lina25	F: ATGTGAATAAGCCGGCAAGG R: TGTACGCACTCACACACCTCC	KU561848	ATAC	262-322	24	10	0.583	0.731	0.11
Lina26	F: TGTATTCTTGACACATCCATGC R: TTTGCCTGAGACAATAAGAAGGG	KU561849	ATAC	365-397	24	8	0.583	0.647	0.15
Lina28	F: AACAGTATATTAGCAAACCTTCTGTGCC R: CACAAAGAACAGTTTGAAATCATCG	KU561850	ATCT	271-391	23	16	0.739	0.890	0.021
Lina31	F: GCGATGCAAGAAATAATGAAATACC R: CTCACCTTCCATTTGTTACG	KU561851	ATAC	293-389	23	12	0.652	0.778	0.075
Lina32	F: TCATGAAAGACACAGACCTTGC R: AGTGAAATGCACAAGTCGGG	KU561852	ATAC	338-402	23	10	0.609	0.807	0.055
Lina35	F: GGATGCGTGAATTACGAGTGG R: ACATAGATCCATTTCCGCCG	KU561853	ATAC	214-246	22	6	0.591	0.553	0.24

Lina36*	F: ATGCCTGATTTCATAAACATTACG R: CGCTTGCTATTACACACATGC	KU561854	ATATT	300-445	21	9	0.286	0.780	0.078
Lina38*	F: CAGGACCTCAGTTCCACGC R: GATCAACATACGCACGACCG	KU561855	ATAC	331-403	23	10	0.391	0.826	0.051
Lina39	F: CACAACCTCTGTAGGACATTTGC R: CAGTAGTGTACATAACGGGATGGG	KU561856	ATAC	198-226	24	8	0.667	0.703	0.11
Lina41	F: GCAAATACCATTTCTTCCAAACTACG R: AAAGCAGCCAGGTAAATCGC	KU561857	ATAC	289-337	24	11	0.750	0.822	0.050
Lina43	F: TCCGAAATACAAACAAGGGACC R: TAAACATTAGCAATGCCGCC	KU561858	ATCTT	377-442	23	8	0.652	0.700	0.12
Lina44	F: TTACCCACAGACACGTGAGC R: TGTTCATATACGTAATCTTTGCCC	KU561859	ATAC	311-339	24	8	0.708	0.812	0.061
Lina45	F: CCTGAATGTATTAAAGAACCAGAACG R: TTGCTCATTAGACAAGTAGGCG	KU561860	ATAC	245-285	24	11	0.792	0.845	0.042
Lina46	F: GACCTTCCGCATCCCAGG R: AAACCGCGGAATTGTTTGG	KU561861	ATAC	278-310	24	9	0.708	0.748	0.094

* Indicates significant deviations from HWE expectations after Bonferroni correction.