Development and characterization of microsatellite loci for the haploid–diploid red seaweed *Gracilaria vermiculophylla*

Nicole M. Kollars¹,², Stacy A. Krueger-Hadfield¹,*, James E. Byers², Thomas W. Greig³, Allan E. Strand¹, Florian Weinberger⁴ and Erik E. Sotka¹

¹ Grice Marine Laboratory and the Department of Biology, College of Charleston, Charleston, SC, USA
² Odum School of Ecology, University of Georgia, Athens, GA, USA
³ Center for Coastal Environmental Health and Biomolecular Research, National Oceanic and Atmospheric Administration, Charleston, SC, USA
⁴ Helmholtz-Zentrum fü r Ozeanforschung Kiel (GEOMAR), Kiel, Germany
⁵ Current affiliation: Center for Population Biology, University of California, Davis, CA, USA

* These authors contributed equally to this work.

ABSTRACT

Microsatellite loci are popular molecular markers due to their resolution in distinguishing individual genotypes. However, they have rarely been used to explore the population dynamics in species with biphasic life cycles in which both haploid and diploid stages develop into independent, functional organisms. We developed microsatellite loci for the haploid–diploid red seaweed *Gracilaria vermiculophylla*, a widespread non-native species in coastal estuaries of the Northern hemisphere. Forty-two loci were screened for amplification and polymorphism. Nine of these loci were polymorphic across four populations of the extant range with two to eleven alleles observed. Mean observed and expected heterozygosities ranged from 0.265 to 0.527 and 0.317 to 0.387, respectively. Overall, these markers will aid in the study of the invasive history of this seaweed and further studies on the population dynamics of this important haploid–diploid primary producer.

Subjects Ecology, Evolutionary Studies, Genetics, Marine Biology

Keywords Complex life cycles, Biological invasions, Seaweed, Microsatellites, Haploid-diploid, *Gracilaria vermiculophylla*

INTRODUCTION

In the last decade, genetic approaches to answering evolutionary and ecological questions have become less expensive and more easily applied to non-model species (*Allendorf, Hohenlohe & Luikart, 2010; Guichoux et al., 2011*). Microsatellites, or tandem repeats of two to six nucleotides, are popular molecular markers due to their resolution in distinguishing individual genotypes (*Selkoe & Toonen, 2006*) and their ability to describe patterns of population connectivity across landscapes (*Manel et al., 2003*) and seascapes (*Galindo, Olson & Palumbi, 2006*). Much of the literature focuses on organisms with single free-living diploid stages (i.e., animals and higher plants). Yet, there are many species with both
haploid and diploid stages in the same life cycle in which both ploidies undergo somatic development and live as independent, functional organisms.

While theory predicts that selection should favor either diploidy or haploidy (Mable & Otto, 1998), Hughes & Otto (1999) demonstrated the maintenance of both haploid and diploid stages when the two stages occupy different ecological niches. However, there are relatively few empirical tests of these alternative hypotheses (but see Destombe et al., 1992; Thornber & Gaines, 2004; Guillemin et al., 2013), and for isomorphic species in which ploidy is not easily identified through morphological traits, molecular markers will be essential to advance research in this field. These same markers can additionally be used to understand connectivity and demographic history in haploid–diploid populations. Among marine haploid–diploid macroalgae, relatively few microsatellites have been developed to address any of these issues (but see Table 1).
Understanding the consequences of biphasic life cycles and land- or seascape features on population structure is particularly relevant in light of the increasing frequency of biological introductions. There are numerous examples of widespread, and putatively invasive species, that have free-living haploid and diploid stages, including macroalgae (e.g., Asparagopsis spp.; Andreakis et al., 2007), ferns (e.g., Lygodium spp.; Lott et al., 2003) and mosses (e.g., Campylopous introflexus; Schirmel, Timler & Buchholz, 2010). Macroalgae, or seaweeds, account for approximately 20% of the world’s introduced marine species (Andreakis & Schaffelke, 2012) and a subset of these invasions are by species that are exploited in their native range, either for the phycocolloid industry or as food products (Williams & Smith, 2007).

The red seaweed Gracilaria vermiculophylla (Omhi) Papenfuss is native to the northwest Pacific and, in the last 30–40 years, has spread throughout high to medium salinity estuaries of the eastern North Pacific (Saunders, 2009), the western North Atlantic (Byers et al., 2012) and the eastern North Atlantic (Weinberger et al., 2008; Guillemin et al., 2008a). G. vermiculophylla transforms the ecosystems into which it is introduced through negative impacts on native species (e.g., direct competition, Hammann et al., 2013), the addition of structural complexity to soft-bottom systems (e.g., Nyberg, Thonsen & Wallentinus, 2009; Wright et al., 2014) and the alteration of community structure, species interactions and detrital pathways (e.g., Byers et al., 2012). Previous studies of the population genetics of G. vermiculophylla focused on the mitochondrial gene cytochrome b oxidase I (Kim, Weinberger & Boo, 2010; Gulbransen et al., 2012), but mitochondrial genetics do not necessarily predict the population genetics of the nuclear genome and cannot assess patterns of ploidy and mating system. Thus, we developed nine polymorphic microsatellite loci for G. vermiculophylla.

MATERIALS AND METHODS
A library of contigs for G. vermiculophylla was generated using the 454 next-generation sequencing platform (Cornell University Life Sciences Core Laboratory Center) from a single individual collected from Charleston, South Carolina, USA. For library preparation, DNA was extracted using CTAB (Eichenberger, Gugerli & Schneller, 2000) and library construction followed Hamilton et al. (1999). Dimeric to hexameric microsatellite repeats were identified with the program MSATCOMMANDER, ver 1.0.8 (Faircloth, 2008) and primers were designed using PRIMER 3 (Rozen & Skalesty, 2000) for contigs with at least four sequences present in the library. Bioinformatics of these sequences was facilitated by the APE package (Paradis, Claude & Strimmer, 2004) in R (R Development Core Team, 2014).

Total genomic DNA was isolated using 120 µL of a 10% Chelex solution (BioRad Laboratories, Hercules, California, USA) in which approximately 1 cm of dried algal tissue was heated at 95 °C for 30 min and vortexed intermittently (Walsh, Metzger & Higuchi, 1991). Loci were amplified on a thermocycler (BioRad) as follows: 10 µL final volume, 2 µL of stock DNA template, 0.5 units of GoTAQ Flexi-DNA Polymerase (Promega, Madison, Wisconsin, USA), 1X buffer, 250 µM of each dNTP, 1.5 nM of MgCl₂,
Table 2 Characteristics of nine polymorphic microsatellite loci developed for Gracilaria vermiculophylla. Acc. No., GenBank accession number; locus; motif; primer sequences; allele range; avg. error: TANDEM (Matschiner & Saltzburger, 2009) rounding errors for each microsatellite locus (the authors of TANDEM suggest that good loci have an average rounding error which is below 10% of the repeat size); N_tall, total number of alleles. All loci showed one-locus genetic determinism.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Acc. No.</th>
<th>Motif</th>
<th>Primer sequence</th>
<th>Allele range</th>
<th>Avg. Error</th>
<th>N_tall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gverm_5276</td>
<td>KT232089</td>
<td>(AC)_10</td>
<td>F: GGAGAGCAGCACGTTTTAGG R: CTGCTTAGTCACAGCTGCA</td>
<td>282–316</td>
<td>0.14</td>
<td>11</td>
</tr>
<tr>
<td>Gverm_6311</td>
<td>KT232090</td>
<td>(AG)_9</td>
<td>F: GCGTCATTCCACTGAATGTG R: GATGAAACCTCAATGCGCCTG</td>
<td>203–223</td>
<td>0.17</td>
<td>6</td>
</tr>
<tr>
<td>Gverm_8036</td>
<td>KT232091</td>
<td>(AC)_12</td>
<td>F: GCCCTTTTTAGGATGCAACA R: GGGTAAAACGACCACAGAGA</td>
<td>213–251</td>
<td>0.14</td>
<td>5</td>
</tr>
<tr>
<td>Gverm_3003</td>
<td>KT232092</td>
<td>(AG)_11</td>
<td>F: CATCTTGGCTTCTGCTTCC R: TTGAAGGCGGGAATTTATCG</td>
<td>198–230</td>
<td>0.11</td>
<td>4</td>
</tr>
<tr>
<td>Gverm_1203</td>
<td>KT232093</td>
<td>(AAG)_8</td>
<td>F: CTCCCTGGACACAGCAATA R: ACATCTCGCGACCTTCTTT</td>
<td>284–308</td>
<td>0.12</td>
<td>4</td>
</tr>
<tr>
<td>Gverm_1803</td>
<td>KT232094</td>
<td>(AC)_11</td>
<td>F: GCCTGCGACAGTCTCACC R: GACAGCAACAGTGGGTTT</td>
<td>352–356</td>
<td>0.07</td>
<td>3</td>
</tr>
<tr>
<td>Gverm_804</td>
<td>KT232095</td>
<td>(AAG)_8</td>
<td>F: TGACTGAGTTGCTCTGCTTTG R: GAGGCTGCGCAACAAAATCAT</td>
<td>182–188</td>
<td>0.16</td>
<td>3</td>
</tr>
<tr>
<td>Gverm_10367</td>
<td>KT232096</td>
<td>(AG)_8</td>
<td>F: GCTGAAATAAGAAGCCAAGG R: GCAAACCTGGCTTGGTTT</td>
<td>198–200</td>
<td>0.07</td>
<td>2</td>
</tr>
<tr>
<td>Gverm_2790</td>
<td>KT232097</td>
<td>(AATGC)_5</td>
<td>F: GAAACATGCCCAGGAAAACATT R: GGAAGAGCTCAAAACAGCA</td>
<td>262–267</td>
<td>0.16</td>
<td>2</td>
</tr>
</tbody>
</table>

150 nM of fluorescently-labeled forward primer, 100 nM of unlabeled forward primer and 250 nM of unlabeled reverse primer. The PCR program included 2 min at 95 °C, 30 cycles of 30 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C, and a final 5 min at 72 °C. One μL of each PCR product was added to 10 μL of loading buffer containing 0.35 μL of size standard (GeneScan500 Liz; Applied Biosystems, Foster City, California, USA). Samples were electrophoresed on an ABI 3130xL genetic analyzer equipped with 36 cm capillaries (Applied Biosystems). Alleles were scored manually using GENEMAPPER ver. 4 (Applied Biosystems) and allele sizes were binned with TANDEM ver. 1.08 software (Matschiner & Saltzburger, 2009; Krueger-Hadfield et al., 2013).

We screened a total of 42 primer pairs for amplification and polymorphism in G. vermiculophylla (Table 2, Table S1). For the amplifiable loci that also showed polymorphism (nine total, see “Results and Discussion”), we verified single locus genetic determinism (SGLD). Loci were in SGLD if known haploids produced a single allele and diploids produced either one or two alleles in their homozygous or heterozygous state, respectively. We verified SGLD in a subset of known haploid gametophytes (n = 28) and diploid tetrasporophytes (n = 30) collected at Elkhorn Slough, California, USA (Table 3, Fig. S1). Elkhorn Slough was the only population for which ploidy was determined by reproductive structures and for which we had known haploids and diploids for genotyping.

The frequency of null alleles was estimated in the haploid subpopulation from Elkhorn Slough as well as diploid tetrasporophytes for each of the four populations (Table 3). It is possible to calculate the null allele frequency directly in the haploids based on the
Table 3 Location of the four populations used to test for polymorphism in newly characterized microsatellite loci in *Gracilaria vermiculophylla*. The region, range (native or non-native), latitude, longitude, sampling date, collector* and ploidy determination (using reproductive phenology or microsatellite genotype) are provided.

<table>
<thead>
<tr>
<th>Population</th>
<th>Region</th>
<th>Range</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Date</th>
<th>Collector</th>
<th>Ploidy determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akkeshi, Japan</td>
<td>NW Pacific</td>
<td>Native</td>
<td>43.04774</td>
<td>144.9498</td>
<td>25 Aug 10, 31 Jul 12</td>
<td>NMK, KH, KM, AP, MS</td>
<td>Genotype</td>
</tr>
<tr>
<td>Elkhorn Slough, CA</td>
<td>NE Pacific</td>
<td>Non-native</td>
<td>36.50447</td>
<td>−121.4513</td>
<td>3 Nov 13</td>
<td>SAKH, BFK, TDK, BH</td>
<td>Genotype, phenology</td>
</tr>
<tr>
<td>Fort Johnson, SC</td>
<td>NW Atlantic</td>
<td>Non-native</td>
<td>32.7513</td>
<td>−79.900</td>
<td>11 Dec 13</td>
<td>CEG</td>
<td>Genotype</td>
</tr>
<tr>
<td>Nordstrand, Germany</td>
<td>North Sea</td>
<td>Non-native</td>
<td>54.454571</td>
<td>8.874846</td>
<td>24 Mar 10</td>
<td>MH</td>
<td>Genotype</td>
</tr>
</tbody>
</table>

Notes. Collector abbreviations: NMK, NM Kollars; KH, K Honda; KM, K Momota; AP, A Pansch; MS, M Sato; SAKH, SA Krueger-Hadfield; BFK, BF Krueger; TDK, TD Krueger; BH, B Hughes; CEG, CE Gerstenmaier; MH, M Hammann.

number of non-amplification events, after discounting technical errors. For diploid tetrasporophytes, we used a maximum likelihood estimator (ML-NullFreq: *Kalinowski & Taper, 2006*).

Next, we screened loci for short allele dominance (*Wattier et al., 1998*). The presence of short allele dominance is rarely tested during microsatellite development, even though it can result in artificial heterozygote deficiencies. In contrast to null alleles, primer binding is successful, but the larger allele is not amplified due to the preferential amplification of the smaller allele. *Wattier et al. (1998)* demonstrated an analytical method to detect short allele dominance using linear models. If a regression of allele-specific $F_s$ (inbreeding coefficient) statistics on allele size reveals a significant negative slope, then short allele dominance may be expected. We determined three to four allele size classes per locus and performed linear regressions using the STATS package in *R* (*R Development Core Team, 2014*).

To provide preliminary assessment of the genotypic and genetic diversity one can gain from these loci, we genotyped diploid tetrasporophytes from one native and three non-native populations of *G. vermiculophylla* (Table 3). Diploids were identified based either on reproductive phenology (Elkhorn) or microsatellite genotype (after assuring SGLD) if at least one locus was heterozygous (Akkeshi, Fort Johnson and Nordstrand, Table 3).

We calculated expected allelic richness using rarefaction in order to account for differences in sample size (HP-Rare; *Kalinowski, 2005*). Observed ($H_O$) and expected heterozygosities ($H_E$) were calculated using GenAlEx, ver. 6.501 (*Peakall & Smouse, 2006; Peakall & Smouse, 2012*). Tests for Hardy–Weinberg equilibrium and $F$-statistics were performed in FSTAT, ver. 2.9.3.2 (*Goudet, 1995*). $F_s$ was calculated for each locus and over all loci according to *Weir & Cockerham (1984)* and significance (at the adjusted nominal level of 0.001) was tested by running 1,000 permutations of alleles among individuals within samples. We also tested for linkage disequilibrium in each population using GENEPOP, ver. 4.2.2 (*Rousset, 2008*), with 1,000 permutations followed by Bonferroni correction for multiple comparisons (*Sokal & Rohlf, 1995*).
RESULTS AND DISCUSSION

Of the 42 loci screened, 16 did not amplify for *G. vermiculophylla* even after several PCR modifications (Table S1). Of the remaining 26 loci, four loci exhibited multi-peak profiles and were discarded from further use, 13 loci were considered monomorphic (Table S1), and nine loci showed polymorphism (Table 2). The nine polymorphic loci exhibited SLGD in which known haploids always exhibited one allele. The low number of polymorphic loci revealed from this screening process is consistent with previous efforts to develop microsatellite loci for some seaweeds (e.g., Varlea-Álvarez et al., 2011; Arnaud-Haond et al., 2013).

The frequency of null alleles was zero at all loci except Gverm_1803 and Gverm_2790 in which the frequencies were both 0.019 in the haploids at Elkhorn Slough (Table S2). The only evidence of null alleles in the diploids from Elkhorn Slough was at locus Gverm_1803, with a maximum likelihood estimated frequency of 0.115. The discrepancy between the haploid and diploid estimates is likely due to assumptions underlying the maximum likelihood estimators implemented in software like HP-Rare (Kalinowski, 2005), such as random mating. Krueger-Hadfield et al. (2013) demonstrated a strong bias in the estimates of null allele frequency when using these maximum likelihood estimators in macroalgal populations that have undergone non-random mating. The higher frequencies of null alleles (0.115–0.207) in the Akkeshi diploid subpopulation were most likely driven by a violation of these assumptions as well, though empirical estimates in haploid subpopulations are warranted. Nevertheless, the low frequency of null alleles and lack of evidence for short-allele dominance (all regression *p*-values were >0.2, Table S3), suggest that observed heterozygote deficiencies using these loci will be due to the mating system or spatial substructuring (Guillemin et al., 2008b; Krueger-Hadfield et al., 2011; Krueger-Hadfield et al., 2013).

Previous studies have used microsatellite loci to distinguish among individual clones and to describe the genetic diversity and the mating systems of seaweed populations despite low levels of polymorphism (e.g., Guillemin et al., 2008b; Arnaud-Haond et al., 2013). In the current study, the nine polymorphic markers described genetic variability in four populations sampled across the extant distribution of *G. vermiculophylla*. Overall, there was little evidence for linkage disequilibrium after Bonferroni correction (Table S4). Additionally, allelic diversity was comparable among the one native and three non-native sites we sampled, but *F*<sub>s</sub> varied considerably (summary in Table 4; per locus statistics in Table S5). Together, these results suggest that unique demographic and evolutionary processes could be operating between native and non-native ranges and within each population, but more detailed sampling is needed to address these patterns.

In summary, we have developed and characterized microsatellite markers for the haploid–diploid red seaweed *G. vermiculophylla*. These loci have the resolution to distinguish individual thalli and will aid studies on the invasive history of *G. vermiculophylla*, as well as the evolutionary ecology of rapidly spreading populations and mating system shifts in organisms that have biphasic life cycles with free-living haploid and diploid stages (i.e., macroalgae, ferns, mosses and some fungi).
Table 4 Genetic features of four populations of *Gracilaria vermiculophylla*. These include: the sample size, $N$; the diploid genotypic richness, $N_A$; + standard error (SE); mean allelic richness, $A_E$, based on the smallest sample size of 46 alleles (23 diploid individuals) + SE; mean observed heterozygosity, $H_O$, + SE; mean expected heterozygosity, $H_E$, + SE; inbreeding coefficient, $F_{is}$, multilocus and per locus estimates.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Akkeshi</th>
<th>Elkhorn slough</th>
<th>Fort Johnson</th>
<th>Nordstrand</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>31</td>
<td>30</td>
<td>38</td>
<td>23</td>
</tr>
<tr>
<td>$N_A$</td>
<td>3.2 ± 0.5</td>
<td>2.2 ± 0.4</td>
<td>2.0 ± 0.2</td>
<td>1.9 ± 0.2</td>
</tr>
<tr>
<td>$A_E$</td>
<td>3.1 ± 0.4</td>
<td>2.2 ± 0.3</td>
<td>2.0 ± 0.2</td>
<td>1.9 ± 0.2</td>
</tr>
<tr>
<td>$H_O$</td>
<td>0.265 ± 0.060</td>
<td>0.311 ± 0.089</td>
<td>0.520 ± 0.110</td>
<td>0.527 ± 0.125</td>
</tr>
<tr>
<td>$H_E$</td>
<td>0.374 ± 0.079</td>
<td>0.317 ± 0.084</td>
<td>0.387 ± 0.077</td>
<td>0.352 ± 0.079</td>
</tr>
<tr>
<td>$F_{is}$</td>
<td>0.294 *</td>
<td>0.017</td>
<td>−0.350 *</td>
<td>−0.512 *</td>
</tr>
</tbody>
</table>

$F_{is}$ per locus

<table>
<thead>
<tr>
<th></th>
<th>Akkeshi</th>
<th>Elkhorn slough</th>
<th>Fort Johnson</th>
<th>Nordstrand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gverm_5276</td>
<td>0.484 *</td>
<td>0.120</td>
<td>−0.209</td>
<td>−0.492</td>
</tr>
<tr>
<td>Gverm_6311</td>
<td>0.435 *</td>
<td>0.140</td>
<td>−0.267</td>
<td>−0.048</td>
</tr>
<tr>
<td>Gverm_8036</td>
<td>0.334</td>
<td>NA</td>
<td>−0.445 *</td>
<td>−0.217</td>
</tr>
<tr>
<td>Gverm_3003</td>
<td>0.529</td>
<td>−0.121</td>
<td>−0.138</td>
<td>−0.553</td>
</tr>
<tr>
<td>Gverm_1203</td>
<td>−0.15</td>
<td>−0.206</td>
<td>−0.310</td>
<td>−0.508</td>
</tr>
<tr>
<td>Gverm_1803</td>
<td>0.569 *</td>
<td>0.460</td>
<td>−0.696 *</td>
<td>NA</td>
</tr>
<tr>
<td>Gverm_804</td>
<td>−0.278</td>
<td>−0.206</td>
<td>−0.310</td>
<td>−0.508</td>
</tr>
<tr>
<td>Gverm_10367</td>
<td>−0.017</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Gverm_2790</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>−0.913 *</td>
</tr>
</tbody>
</table>

Notes.  
* $p < 0.001$, adjusted nominal value.

ACKNOWLEDGEMENTS

Thanks to TM Bell, E Buchanan, CE Gerstenmaier, M Hammann, K Honda, B Hughes, BF Krueger, TD Krueger, K Momota, M Nakaoka, A Pansch, T Roth and M Sato for field and laboratory support.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This project was supported by NSF OCE-1057707 to JEB, OCE-1057713 to EES and OCE-1357386 to EES, SAKH and AES, as well as College of Charleston Graduate Research Grants, a Phycological Society of America Grants-in-Aid-of-Research Fellowship, and the Zostera Experimental Network Graduate Research Fellowship (NSF OCE-1031061 to JE Duffy) to NMK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
College of Charleston Graduate Research.
National Science Foundation: OCE-1057707, OCE-1057713, OCE-1357386.
Phycological Society of America Grants-in-Aid-of-Research Fellowship.
Zostera Experimental Network Graduate Research Fellowship: NSF OCE-1031061.
Competition Interests
The scientific results and conclusions, as well as any opinions expressed herein, are those of the author(s) and do not necessarily reflect the views of NOAA or the Department of Commerce. The mention of any commercial product is not meant as an endorsement by the Agency or Department. We have no competing interest.

Author Contributions
• Nicole M. Kollars conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables.
• Stacy A. Krueger-Hadfield performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables.
• James E. Byers and Allan E. Strand conceived and designed the experiments, reviewed drafts of the paper.
• Thomas W. Greig contributed reagents/materials/analysis tools, reviewed drafts of the paper.
• Florian Weinberger contributed samples and reviewed the paper.
• Erik E. Sotka conceived and designed the experiments, wrote the paper.

DNA Deposition
The following information was supplied regarding the deposition of DNA sequences:
GenBank accession numbers can be found in Table 2.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.1159#supplemental-information.

REFERENCES


