Development and characterization of microsatellite markers in two agarophyte species Gracilaria birdiae and G. caudata (Gracilariaceae, Rhodophyta) using next-generation sequencing

Ligia M. Ayres-Ostrock, Stéphane Mauger, Estela Maria Plastino, Mariana C. Oliveira, Myriam Valero, Christophe Destombe

Development and characterization of microsatellite markers in two agarophyte species *Gracilaria birdiae* and *G. caudata* (Gracilariaceae, Rhodophyta) using next-generation sequencing

L.M. Ayres-Ostrock [a], S. Mauger [b, c], E.M. Plastino [a], M.C. Oliveira [a], M. Valero [b, c], C. Destombe [b, c]

[a] Instituto de Biociências, Universidade de São Paulo, Rua do Matão 277, CEP: 05508-090. São Paulo, SP, Brazil.

[b] Sorbonne Universités, Université Pierre et Marie Curie, UPMC, UMI 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France.

[c] Centre National de la Recherche Scientifique, CNRS, UMI 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France.

Correspondence: Christophe Destombe, Fax:+33 298292324; E-mail: destombe@sb-roscoff.fr

Abstract The two species *Gracilaria birdiae* Plastino and E.C.Oliveira and *G. caudata* J.Agardh are the most important natural sources of agar in Brazil. Using the 454 sequencing system, we identified 464 and 487 perfect microsatellite loci in 6,908 and 9,602 sequences/contigs from *G. birdiae* and *G. caudata*, respectively. After a conservative removal of potentially problematic loci, 144 loci were tested (72 from each species). A total of 25 polymorphic microsatellite loci were defined (13 loci for *G. birdiae* and 17 loci for *G. caudata*, including five loci common to both species). The five microsatellite loci that cross-amplified in both species showed species-specific differences in allele size. Polymorphic microsatellite loci were used to assess the genetic diversity of both species in their main harvest and cultivation areas on the Brazilian coast. Gene diversity was similar in *G. birdiae* and *G. caudata*. However, significant heterozygote deficiency was observed in *G. birdiae*, whereas heterozygote excess occurred in *G. caudata*, suggesting that these two related species differ in their mating system. These results also raised new questions on their biology in the field and on their patterns of genetic structure across their geographical ranges. In addition, the 20 loci developed in this study proved successful in identifying each individual in the field as a unique multilocus genotype, and will be useful for studying lineage sorting, breeding programs or conservation issues.

Keywords: microsatellite, NGS, cross-amplification, diagnostic loci, seaweed, conservation

Introduction

Microsatellite markers are short tandem repeats (STRs) of 2-6 nucleotide motifs randomly distributed throughout the genome; they are highly polymorphic due to their high mutation rate (Tautz and Schlötterer 1994; Jarne and Lagoda 1996). Given their high level of variability, single-locus inheritance, and codominance, microsatellite markers are used for parentage analyses (Avise 2004; Pemberton 2009) and genetic mapping (Jarne and Lagoda 1996; Röder et al. 1998). In red algae, microsatellite markers have been used to retrace paternity in intertidal rocky shores in *Gracilaria gracilis* (Stackhouse) M. Steentoft, L.M. Irvine & W.F. Farnham (Engel et al. 1999) and in *Chondrus crispus* Stackhouse (Krueger-Hadfield et al. 2015). These markers are also efficient for describing intraspecific genetic diversity, and are used to estimate population size or connectivity for conservation genetics purposes (for review see Sunnucks 2000). For example, microsatellites have been used to assess the effects of domestication and human activities on the genetic diversity of the cultivated *G. chilensis* C.J. Bird, McLachlan & E.C. Oliveira, revealing that the reduction in genetic diversity is driving this species to extinction (Guillemin et al. 2008; Guillemin et al. 2014). In *Asparagopsis armata* Harvey, STRs have been used to evaluate the genetic similarities of populations from the Mediterranean, Hawaii, and California (Andreakis et al. 2007). In *Furcellaria lumbricalis* (Hudson) J.V. Lamouroux, neutral and EST-derived microsatellite markers have been developed to examine the genetic structure of northern European populations inhabiting different salinity conditions (Kostamo et al. 2012).

The need for *de novo* development using costly and time-consuming methods, as well as the low success rate of cross-species transfer, have been a major obstacle to the application of these markers in poorly studied taxonomic groups (Squirrell et al. 2003). Moreover, cross-amplified markers often show lower allelic diversity and higher frequency of null alleles than in the source species (Selkoe and Toonen 2006). With the advent of next-generation sequencing (NGS) technologies, the identification and characterization of microsatellites in non-model species have become more feasible. In red algae, polymorphic microsatellite markers have been reported for only 16 species from seven orders (Pardo et al. 2014), including only three species of *Gracilaria*: *G. changii* (B.M.Xia & I.A.Abbott) I.A.Abbott, J.Zhang & B.M.Xia (Song et al. 2013), *G. chilensis* (Guillemin et al. 2005), and *G. gracilis* (Wattier et al. 1997; Luo et al. 1999; Song et al. 2013).

Gracilaria birdiae Plastino and E.C. Oliveira and *G. caudata* J. Agardh are important sources of agar in Brazil (Plastino and Oliveira 2002). The commercial exploitation of *Gracilaria* spp. in Brazil began in the 1960s, with the direct harvesting of seaweeds in natural beds (Câmara-Neto 1987). However, in the 1970s, it became clear that intensive and uncontrolled exploitation of natural beds was leading to a significant decline of native populations (Hayashi et al. 2014). Hence, the cultivation of selected agarophyte species was proposed (Oliveira 1984). Today, artisanal mariculture of *G.*

birdiae and *G. caudata* in the Northeast State of Ceará, Brazil offers an important source of income for the local fishing community (Araujo and Rodrigues 2011; Costa et al. 2011).

In the present study, we report the development of microsatellite markers in the *G. birdiae* and *G. caudata*. The polymorphism of these markers was assessed in the Ceará region of northeastern Brazil, where both species are probably highly affected by human activities.

Materials and methods

Data pre-processing and microsatellite discovery

To develop and characterize microsatellites loci in these species, we employed the 454 pyrosequencing system (Roche, Basel Switzerland) (Margulies et al. 2005). Genomic DNA was extracted and purified from haploid individuals of *G. birdiae* from Ceará State (03°23'87"S 39°00'67"W), and of *G. caudata* from São Paulo State (23°63'87"S 45°38'33"W), using around 500 mg of dried sample. DNA extraction, library preparation, amplification, and sequencing were carried out by Ecogenics GmbH (Zürich-Schlieren, Switzerland). A sample of DNA from each species was used to prepare their respective GS-FLX Titanium general libraries (454 Life Sciences) according to the manufacturer's protocols and included quality control steps. Each DNA sample was appropriately tagged using two different multiplex identifiers (MIDs) to enable pooling for simultaneous sequencing. DNA library fragments were then captured onto beads and clonally amplified within individual emulsion droplets (emPCR). Finally, amplified fragments from both species were evenly mixed and sequenced using Titanium GS-FLX chemistry (454 Life Sciences). Each read was automatically assigned to its correct library using GS-FLX software (454 Life Sciences) based on the sample-specific MID sequence. The 454 sequencing data received was screened for SSRs using SPUTNIK (Abajian 1994). Results were then parsed to Primer3 for locus-specific primer design (Koressaar and Remm 2007; Untergrasser et al. 2012). The resulting tab-delimited files were converted to spreadsheet files for use in Microsoft Office Excel 2007 (Microsoft Corporation, Redmond, WA, USA). Microsatellite frequencies for each species were expressed as the total number of identified microsatellites per million base pairs (Mbp), and subsequent data analysis included sorting according to microsatellite category (i.e., di-, tri-, tetra-, and pentanucleotides), specific motifs, and number of tandem repeated units. Moreover, PCR product size (< 400 bp) and melting temperature (T_m) of 60 °C were also

considered during primer selection. Reverse-complement repeat motifs and translated or shifted motifs were grouped together so that there were 144 microsatellite loci for *G. birdiae* and *G. caudata* (72 each), with 12 dinucleotide repeats, 24 trinucleotides, 24 tetranucleotides, and 12 pentanucleotides. To increase the chance of obtaining polymorphic loci, we selected motifs with the highest number of repetitions.

Test for polymorphism and PCR amplification

A total of 144 primer pairs were designed and tested for amplification on 15 DNA samples (7 individuals of *G. birdiae* and 8 of *G. caudata*) with indirect labelling markers. A subset of the primer pairs that amplified correctly (32 pairs for each species) was tested for polymorphism on 48 individuals of both species from nine locations covering seven regions along the Brazilian coast, according to the distribution range of each species (for more details see Table 1).

PCR amplifications of the direct labeling markers were performed in 10 µL, containing 2µL of template DNA diluted 1:50, 150 µM dNTPs (Thermo Fisher Scientific Inc., Waltham, MA, USA), 30 pmol forward fluorescent-labelled primer FAM, Yakima Yellow (VIC), ATTO550 (NED) or ATTO565(PET) (Eurofins MWG Operon, Ebersberg, Germany), 30 pmol reverse primer (Eurofins MWG Operon, Ebersberg, Germany), 2.5 mM $MgCl_2$ (Promega Corp.), 1 X GoTaq® Flexi Buffer (Promega Corp.), and 0.35 U GoTaq® DNA polymerase (Promega Corp.). PCR amplifications of the indirect labeling markers using the standard M13 universal sequence were performed in 10 µL, containing 2 µL of DNA template diluted 1:50, 150 µM of dNTPs (Thermo Fisher Scientific Inc., Waltham, MA, USA), 30 pmol forward primer, 30 pmol reverse primer, 30 pmol fluorescent-labelled M13(-21) primer FAM, Yakima Yellow (VIC), ATTO550 (NED) or ATTO565(PET) (Eurofins MWG Operon, Ebersberg, Germany), 2.5 mM $MgCl_2$ (Promega Corp.), 1 X GoTaq® Flexi Buffer (Promega Corp.), and 0.35 U GoTaq® DNA polymerase (Promega Corp.). The 5' end of the forward primer from each pair was modified with a M13 (-21) universal sequence tag (5'-TGTAAAACGACGGCCAGT-3') to enable the incorporation of the universal fluorescent labelled M13 (-21) primer for detection on an ABI3130 XL DNA Analyzer (Life Technologies Corporation). Amplifications were carried out separately for each primer pair on a DNA Engine Peltier Thermal Cycler (Bio-Rad, Corporation). For the direct labeling markers, PCR cycling consisted of an initial denaturation at 95 °C for 5 min followed by 30 cycles: denaturation of 95 °C for 30 s, annealing at 54 °C for 30 s, and extension at 72 °C for 30 s, with a final extension of 72 °C for 10 min. For the indirect labeling marker cycling conditions included an initial denaturing at 95 °C for 5 min followed by 10 cycles: denaturation of 95 °C for 30 s, annealing at 58 °C for 30 s and extension at 72 °C for 30 s;

followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 53 °C for 30 s, and extension at 72 °C for 30 s, with a final extension of 72 °C for 10 min. Each microsatellite amplification was diluted with distilled H_2O (1:10), and 2 μL were mixed with 9.5 μL of Hi-Di formamide (Life Technologies Corporation) and 0.5 μL of SM594 size standard (Mauger et al. 2012).

Fragments were separated using an Applied Biosystems 3130XL Genetic Analyzer (Life Technologies Corporation). Genotypes were scored manually using GeneMapper® v4.0 software (Life Technologies Corporation) and microsatellite loci selected for subsequent population analyses according to their polymorphic information content (PIC), calculated with Cervus software (Marshall et al. 1998; Kalinowski et al. 2007).

Sampling and DNA extraction for population analyses

Fifty diploid individuals of *G. birdiae* and 50 of *G. caudata* were sampled in the region of Ceará (CE), northeastern Brazil, at two locations 1.45 km apart: Pedra Rachada and Naufrágio, Paracurú (Table 1). These locations were situated at about 30.97 km from the nearest cultivated farms in Trairi (CE). Samples were collected along a 20 m transect placed perpendicularly to the shoreline during low tide. Reproductive diploid individuals (tetrasporophytes) identified under stereoscopic microscope (American Optical FORTY) were selected for genotyping.

DNA was extracted and purified using the NucleoSpin® 96 Plant II kit (Macherey-Nagel, Düren, Germany). Apical fragments of approximately 8 mg of silica-dried sample were powdered using mechanical tissue disruption with 3 mm diameter steel beads. Extraction was carried out according to the manufacturer's instructions, except for incubation in cell lysis buffer: samples were left at room temperature overnight rather than heating to 65 °C for 30 minutes. PCR amplifications and genotype scoring were carried out using the procedures described above.

Population Data Analyses

Prior to analysis and for each species, the frequency of null alleles was estimated using MicroChecker software (van Oosterhout et al. 2004). The power of the developed markers to discriminate all the multilocus microsatellite genotypes (MLG) present in the sample was tested using the Monte Carlo procedure implemented in GenClone 2.0 (Arnaud-Haond and Belkhir 2007). Genotype diversity was calculated as R= (G-1)/(N-1), where G is the number of distinct genotypes identified

and N is the number of individuals (Dorken and Eckert 2001). Linkage disequilibrium was assessed using a single multilocus measurement of LD that is provided by the association index $\bar{r}_d$ (Brown et al. 1980 modified by Agapow and Burt 2001) and was computed using Multilocus ver. 1.2 (Agapow and Burt 2001). Significance tests were based on the comparisons of the observed values with those of randomized data sets generated from 1000 permutations (Burt et al. 1996). Single and multilocus estimates of gene diversity were calculated as the mean number of alleles per locus (*Na*), expected heterozygosity (*He*, sensus Nei 1978) and observed heterozygosity (*Ho*) using GENETIX 4.05 software (Belkhir et al. 1996-2004)). Single and multilocus estimates of deviation from random mating (F_{is}) were calculated according to Weir and Cockerham (1984), and significant departure from panmixia was tested by running 1000 permutations of alleles among individuals within samples, using GENETIX 4.05 software (Belkhir et al. 1996-2004). To test for differences in gene diversity between species, one-way analyses of variance (ANOVA) were performed on *Na*, *He*, *Ho*, and F_{is} values considering each locus as random repeat of these estimates, using Minitab® (2009).

Results

Development of microsatellites

In *G. birdiae* of the 14,358 reads obtained, 464 (3.2%) contained (perfect) microsatellite inserts. Once the redundant sequences have been removed, there were 318 microsatellite loci and they contained dinucleotide (48.7%) or trinucleotide (32.7%) motifs. The other motifs were less frequent (8.8% tetranucleotides, 9.7% pentanucleotides). Of the 318 potentially amplifiable loci, 72 primer pairs were tested and 32 were selected according to their amplification. Twenty-two primer pairs were validated on a panel of samples, i.e. produced specific PCR products of the expected size, and tested for polymorphism. Thirteen of these 22 loci (59.1%) proved to be polymorphic in populations. These loci were dominated by tetranucleotide (50%) and trinucleotide (25%) motifs followed by pentanucleotide (18.7%) and dinucleotide (6.2%) repeats.

In *G. caudata*, of the 24,757 reads obtained, 487 (1.9%) contained (perfect) microsatellite inserts. Of these, there were 403 microsatellite loci and they contained dinucleotide (43.9%) or trinucleotide (34.2%) motifs. The other motifs were less frequent (12.4% tetranucleotides and 9.4% pentanucleotides). Of the 403 potentially amplifiable loci, 72 primer pairs were tested, and 32 were selected according to their amplification. Twenty-seven primer pairs were validated on a panel of

samples, i.e. produced specific PCR products of the expected size, and were tested for polymorphism. Seventeen of these 27 loci (63.0%) proved to be polymorphic in populations. These loci were dominated by tetranucleotide (39.1%) and trinucleotide (26.1%) motifs followed by dinucleotide (21.7%) and pentanucleotide (13.0%) repeats.

Overall, five microsatellite loci cross-amplified in both species, eight were diagnostic loci for *G. birdiae* and 12 for *G. caudata* (Table 2). Clear differences in allele size according to the species were found for the five loci that cross-amplified (Figure 2).

Genotypic and genetic diversity within the Ceará state

As shown above, the entire data set included 25 loci (13 loci for *G. birdiae* and 17 loci for *G. caudata*, including the 5 common loci). We characterized 107 alleles across the 13 microsatellite loci in *G birdiae*, whereas only 69 alleles were found across the 17 loci in *G. caudata* over the whole data set (i.e. 50 individuals sampled in the Ceará state for both species, Table 3).

The frequency of null alleles was significant for six loci in *G. birdiae* (GraB_03, GraB_05, GraB_07, GraB_08, GraBC_03, and GraBC_04) and for three loci in *G. caudata* (GraC_06, GraC 09, and Gra C11) (Table 3). These loci were therefore eliminated from subsequent analyses. The final data sets consisted of 50 individuals genotyped at 7 loci for *G. birdiae* and 50 individuals genotyped at 12 loci for *G. caudata*. The combinatory procedure implemented in GenClone 2.0 indicated that the discrimination power of our loci set was adequate in both species. In fact, even after excluding the locus suspected to harbor null alleles, a combination of all 7 loci in the case of *G. birdiae* (Figure 1a) and 10 loci in the case of *G. caudata* (Figure 1b) was enough to distinguish among all genotypes. All the individuals analyzed in both species had unique multilocus genotypes leading to a value of clonal richness (R) equal to 1.

Tests of linkage disequilibrium (LD) revealed an overall P-value of 0.391 in *G. caudata* with $\bar{r}_d$ (i.e. LD estimator) values of 0.0029. In *G. birdiae,* the $\bar{r}_d$ value of 0.0919 was significant with a P-value < 0.001.

The single and multilocus estimates of gene diversity are given in Table 3. The number of alleles per polymorphic locus (*Na*) (excluding loci with null alleles) ranged for *G. birdiae* from 2 (locus GraB_02) to 15 alleles (locus GraBC_05) and for *G. caudata* from 2 (locus GraC_10) to 10 alleles (locus GraC_04). The mean *Na* values were not significantly different between species (ANOVA, $F_{(1,17)}$= 1.80, P= 0.197), although larger in *G. birdiae* (6.86) than in *G. caudata* (4.58). The expected heterozygosity (*He*) ranged for *G. birdiae* from 0.059 (locus GraBC_02) to 0.854 (locus GraBC_05) and for *G.*

caudata from 0.059 (locus GraBC_01) to 0.830 (locus GraC_04). The mean *He* values were similar between *G. birdiae* (0.435) and *G. caudata* (0.420) (ANOVA, $F_{(1,17)}$= 0.02, P= 0.904). The observed heterozygosity (*Ho*) ranged for *G. birdiae* from 0.060 (locus GraBC_02) to 0.580 (locus GraB_01) and for *G. caudata* from 0.060 (locus GraBC_01) to 0.780 (locus GraC_01). The mean *Ho* values were similar between *G. birdiae* (0.342) and *G. caudata* (0.460) (ANOVA, $F_{(1,17)}$= 0.92, P= 0.350).

F_{is} values were highly variable among loci and species. The number of significant F_{is} values per locus was higher in *G. birdiae* (9/13) than in *G. caudata* (6/15) (Table 3). Excluding loci with null alleles, the F_{is} values ranged from -0.205 (locus GraB_02) to 0.5 (locus GraBC_05) in *G. birdiae* and from -0.343 (GraC_03) to 0.112 (GraC_04) in *G. caudata*. The mean F_{is} value was significantly higher in *G. birdiae* (F_{is}= 0.150, SE= 0.103) than in *G. caudata* (F_{is}= -0.083, SE= 0.037) (ANOVA, $F_{(1,17)}$= 6.61, P= 0.020).

Discussion

In the current study, we used 454 pyrosequencing to obtain sequence reads for microsatellite primer development in *Gracilaria birdiae* and *G. caudata.* Our newly developed microsatellite marker set is highly informative. Seven polymorphic microsatellites markers amplified in *G. birdiae* and 12 in *G. caudata*. In our results, only five loci (20%) were able to cross-amplify in both species. These results are in agreement with previous observations in the genus *Gracilaria* that demonstrated that microsatellite loci are often poorly conserved between related species (Wattier et al. 1997; Song et al. 2013).

Based on the primer pairs initially tested, the percentage of amplified markers of the expected size was similar for *G. caudata* (27/72: 37.5%) and *G. birdiae* (22/72: 30.6%), but generally lower than those reported for other organisms using the same method; ranging from 44% to 100% in plants, invertebrates, vertebrates, and fungi (for a review see Schoebel et al. 2013), and in red algae from 57.1% in *Phymatolithon calcareum* (Pallas) W.H.Adey & D.L.McKibbin (Pardo et al. 2014) to 82.6% in *Grateloupia lanceolata* (Okamura) Kawaguchi (Couceiro et al. 2011). However, among the correctly amplified markers, the percentage of polymorphic markers in *G. birdiae* (59.1%) and *G. caudata* (63.0%) were higher than in other red algae (36.8% in *G. lanceolata* and 45.8% in *P. calcareum*). Overall, the rate of success of the current work was thus similar to other studies that employed the same technique.

Excluding loci with null alleles, 7 microsatellite loci in *G. birdiae* and 12 in *G. caudata* were used to assess population genetic diversity of both species in the Ceará region of Brazil, their main area of harvest and cultivation. The mean numbers of alleles per locus did not differ significantly between species, *Na*= 4.6 for *G. caudate* and *Na*= 6.8 for *G. birdiae*; and were similar to the values reported for other red algae species: *Na*= 4.4 for *Gracilaria gracili*s (Luo et al. 1999; *Na*= 4.6 for *Asparagopsis taxiformis* (Delile) Trevisan de Saint-Léo (Andreakis et al. 2007); *Na*=5.1 for *Grateloupia filicina* (J.V.Lamouroux) C.Agardh (Wang et al. 2013); *Na*= 6.6 for *Furcellaria lumbricalis* (Kostamo et al. 2012), with the exception of the Chilean populations of *Gracilaria chilensis* for which the remarkably low diversity (mean allelic richness of 0.95) is attributed to a recent genetic bottleneck due to over-exploitation of natural stocks (Guillemin et al. 2014)). Similarly, the genetic diversity observed for *G. birdiae* (*He*= 0.43) and *G. caudata* (*He*= 0.42) was similar to the values found for other *Gracilaria* species such as *G. gracilis* (from *He*= 0.45 to 0.52) (Luo et al. 1999; Engel et al. 2004), but greater than the over-harvested Chilean populations of *G. chilensis* (*He*= 0.32, Guillemin et al. 2014). A better assessment of the effect of the harvesting of natural stocks in the two Brazilian study species of *Gracilaria* requires further comparison of their genetic diversity with non-harvested natural populations along the Brazilian coast.

Our results suggest that *G. birdiae* and *G. caudata* differ in their reproductive systems and/or their population structure. The studied population of *G. birdiae* exhibited a significant heterozygote deficiency compared with Hardy-Weinberg equilibrium, whereas the population of *G. caudata* showed an excess of heterozygotes. This difference can be due to the life history of these two species. Red seaweeds typically have complex sexual haploid-diploid life cycles with the alternation of haploid (gametophytic) and diploid (tetrasporophytic) individuals. Variations in the frequencies of haploid and diploid individuals are due to the reproductive system (sexual vs. asexual) and/or to differences in fertility and survivorship (for a review see Kain and Destombe 1997). In our study, the population of *G. caudata* was composed almost entirely of tetrasporophytic (diploid) individuals (90.5%), whereas in *G. birdiae* the haploid:diploid ratio was more balanced (Ayres-Ostrock and Plastino, unpublished data). The low frequency of gametophytes (haploids) and the heterozygote excess found in *G. caudata* highly suggests that asexual reproduction occurs in this species (Guillemin et al. 2008), even though there were no repeated multilocus genotypes. In *G. birdiae,* significant heterozygote deficiency may be explained by the occurrence of null alleles, bi-parental inbreeding or by a Wahlund effect. Additional population genetic studies are needed to test these hypotheses.

In conclusion, the 7 polymorphic microsatellites markers developed for G. *birdiae* and the 12 markers developed for *G. caudata* proved reliable for population genetic studies. The study of genetic variation is the basis of any breeding or

conservation program. Therefore, future studies using these polymorphic microsatellites will lead to a better understanding of the patterns of genetic diversity in *G. caudata* and *G. birdiae*, providing information on population structure and dynamics that can be useful in guiding management strategies, as well as for the exploitation and culture of these agarophyte species in Brazil.

Acknowledgments This research was supported by the São Paulo Research Foundation (FAPESP: 2010/50175-3; 2011/10189-8), the Brazilian National Council for Scientific and Technological Development (CNPq: 300148/93-3; 301491/2013-5), the program of international cooperation USP/COFECUB between the University of São Paulo and the Comité Français d´Evaluation de la Coopération Universitaire avec le Brésil, and the International Research Network "Diversity, Evolution and Biotechnology of Marine Algae" (GDRI CNRS 0803). We are also most grateful to the Biogenouest Genomics core facility for its technical support and to Carolyn Engel-Gautier for the English review of this manuscript.

References

Abajian C (1994) SPUTNIK: http://abajian.net/sputnik/

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1:101–102

Andreakis N, Kooistra WHCF, Procaccini G (2007) Microsatellite markers in an invasive strain of *Asparagopsis taxiformis* (Bonnemaisoniales, Rhodophyta): insights in ploidy level and sexual reproduction. Gene 406:144-151

Araujo GC, Rodrigues JAG (2011) Maricultura da alga marinha vermelha *Gracilaria birdiae* em Icapuí, Ceará. Arq Ciên Mar 44:62-68

Arnaud-Haond S, Belkhir K (2007) GENECLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. Mol Ecol Notes 7:15–17

Avise J (2004) Molecular markers, natural history and evolution. Chapman & Hall, New York-London

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Genome, Populations, Interactions, CNRS UMR 5000, Universite´ de Montpellier II, Montpellier, France. URL:http://www.genetix.univmontp2.fr/genetix/intro.htm

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics 96:523-536

Burt A, Carter DA, Koenig GL, White TJ, Taylor JW (1996) Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. Proc Natl Acad Sci USA 93:770-773

Câmara-Neto C (1987) Seaweed culture in Rio Grande do Norte, Brazil. Hydrobiologia 151/152:363–367

Costa RF, Salles MCT, Matias LGO (2011) Cultivating algae marines and values for the development of a coastal community in the city of Icapuí/CE. Cadernos de Agroecologia 6:1-5

Couceiro L, Maneiro I, Mauger S, Valero M, Ruiz JM, Barreiro R (2011) Microsatellite development in Rhodophyta using high-throughput sequence data. J Phycol 47:1258-1265

Dorken ME, Eckert CG (2001) Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodonverticillatus* (Lythraceae). J Ecol 89:339–350

Engel CR, Destombe C, Valero M (2004) Mating system and gene flow in the red seaweed *Gracilaria gracilis*: effect of haploid–diploid life history and intertidal rocky shore landscape on finescale genetic structure. Heredity 92:289–298

Engel CR, Wattier R, Destombe C, Valero M (1999) Performance of non–motile male gametes in the sea: analysis of paternity and fertilization success in a natural population of a red seaweed, *Gracilaria gracilis*. Proc R Soc Lond B 266:1879-1886

Guillemin ML, Faugeron S, Destombe C, Viard F, Correa JA, Valero M (2008) Genetic variation in wild and cultivated populations of the haploid-diploid red alga *Gracilaria chilensis*: how farming practices favor asexual reproduction and heterozygosity. Evolution 62:1500-19

Guillemin ML, Valero M, Faugeron S, Nelson W, Destombe C (2014). Tracing the trans-Pacific evolutionary history of a domesticated seaweed (*Gracilaria chilensis*) with archaeological and genetic data. PLoS One 9(12): e114039. doi:10.1371/journal.pone.0114039.

Hayashi L, Bulboa C, Kradolfer P, Soriano G, Robledo D (2014) Cultivation of red seaweeds: a Latin American perspective. J Appl Phycol 26:719-727

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. TREE 11:424-429

Kain JM, Destombe C (1995) A review of the life history, reproduction and phenology of *Gracilaria*. J Appl Phycol 7:269-281

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289-1291

Kostamo K, Korpelainen H, Olsson S (2012) Comparative study on the population genetics of the red algae *Furcellaria lumbricalis* occupying different salinity conditions. Mar Biol 159:561-571

Krueger-Hadfield SA, Roze D, Correa JA, Destombe C, Valero M (2015) O father where art thou? Paternity analyses in a natural population of the haploid–diploid seaweed *Chondrus crispus*. Heredity 114:185-194

Luo H, Morchen M, Engel CR, Destombe C, Epplen JT, Epplen C, Saumitou-Laprade P, Valero M (1999) Characterization of microsatellite markers in the red alga *Gracilaria gracilis*. Mol Ecol 8:700-702

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Goodwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in open microfabricated high density picoliter reactors. Nature 437:376–380

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Mauger S, Couceiro L, Valero M (2012) A simple and cost-effective method to synthesize an internal size standard amenable to use with a 5-dye system. Prime Research on Biotechnology 2:40-46

Minitab Inc. (2009) Minitab statistical software, version 16 for Windows, State College, Pennsylvania

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590

Oliveira EC (1984) The cultivation of seaweeds for the production of agar and agaroids in Brazil–actual state and future perspectives. Mems Assoc Latinoam Acuicult 5:431-435

Pardo C, Peña V, Bárbara I, Valero M, Barreiro R (2014) Development and multiplexing of the first microsatellite markers in a coralline red alga (*Phymatolithon calcareum*, Rhodophyta). Phycologia 53:474-479

Pemberton JM (2009) Wild pedigrees: the way forward. P Roy Soc Lond B Bio 275:613–621

Plastino EM, Oliveira EC (2002) *Gracilaria birdiae* (Gracilariales, Rhodophyta), a new species from the tropical South American Atlantic with terere frond and deep spermatangial conceptacles. Phycologia 41:389-396

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A Microsatellite Map of Wheat. Genetics 149:2007-2023

Schoebel CN, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Ricanova S, Segelbacher G, Werth S, Csencsics D (2013) Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. J Evol Biol 26:600-611

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615–629

Song SL, Lim PE, Phang SM, Lee WW, Lewmanomont K, Largo DB, Han NA (2013) Microsatellite markers from expressed sequence tags (ESTs) of seaweeds in differentiating various *Gracilaria* species. J Appl Phycol 25:839-846

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339-1348

Sunnucks P (2000) Efficient genetic markers for population biology. Trends Ecol Evol 15:199-203

Tautz D, Schlötterer C (1994) Simple sequences. Curr Opin Genet Dev 4:832–837

Untergrasser A, Cutcutache I, Koressaar TYJ, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Res. doi:10.1093/nar/gks596

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Wang J, Peng C, Liu Z, Tang Z, Yang G (2013) Isolation and characterization of microsatellites of *Grateloupia filicina*. Conserv Genet Resour 5:763-766

Wattier R, Dallas JF, Destombe C, Saumitou-Laprade P, Valero M (1997) Single locus microsatellites in Gracilariales (Rhodophyta): high level of genetic variability within *Gracilaria gracilis* and conservation in related species. J Phycol 33:868–880

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370

Fig. 1 Plots describing the genotypic resolution of microsatellites for all 50 samples of (a) *Gracilaria birdiae* and (b) *G. caudata*, giving the average clonal diversity R (±SE) for each loci. Clonal diversity was estimated by R=G-1/N-1, where G is the number of genotypes and N is the sample size

Fig. 2 Allele frequency histogram of three microsatellite loci that cross-amplified in both *Gracilaria birdiae* and *G. caudata* without null alleles. A. GraBC_01. B. GraBC_02. C. GraBC_05

Table 1 Geographic locations of *Gracilaria birdiae* and *G. caudata* populations and number of individuals used in amplification (amp test) and polymorphism test (poly test).

Table 2 Characteristics of 25 polymorphic microsatellite loci (8 for *Gracilaria birdiae*, 12 for *G. caudata* and 5 common to both species): PCR conditions: annealing temperature (T) used in PCR (start T (^{o}C) ↓ final T (^{o}C))

Table 3 Genetic variability within populations of *Gracilaria birdiae* and *G. caudata* from the Ceará State, for 13 and 16 microsatellite loci, respectively. Null allele frequency (*Nf*) obtained with MicroChecker software (van Oosterhout et al. 2004), locus showing significant frequencies of null alleles are shown in bold. Number of individuals genotyped (*n*); number of alleles per locus (*Na*); expected heterozygosity (*He* sensus Nei 1978); observed heterozygosity (*Ho*) (SE); standard error computed over loci. F_{is}: single locus and multilocus estimates of deviation from random mating (Weir and Cockerham 1984). F_{is} values significantly different from zero are shown in bold

Table 1

Click here to download table: Table 1.docx

Region (State)	Location	Coordinates		*G. birdiae*		*G. caudata*	
				amp test	poly test	amp test	poly test
Rio Grande do Norte (RN)	Rio do Fogo	05°15'41.0"S	35°23'11.0"W	2	12	2	12
Ceará (CE)	Pedra Rachada*	03°23'87.0"S	39°00'67.0"W	2	12	-	-
	Naufrágio*	03°23'50.5"S	39°00'38.8"W	-	-	1	6
Bahia (BA)	Emissário	12°44'28.0"S	38°09'01.0"W	2	12	1	6
Pernambuco (PE)	Muro Alto	08°26'13.0"S	34°58'47.0"W	1	12	-	-
	Boa Viagem	08°07'58.9"S	34°53'57.3"W	-	-	1	6
Paraíba (PB)	Carapibus	07°17'62.9"S	34°48'08.5"W	-	-	1	6
Espírito Santo (ES)	Parati	20°48'31.0"S	40°36'39.0"W	-	-	1	6
São Paulo (SP)	Cigarras	23°55'01.0"S	46°19'16.8"W	-	-	1	6
	Total			7	48	8	48

Table 2

Click here to download table: Table 2.docx

Locus	Repeat motif	Primer sequences (5'-3')	PCR conditions	Allele size range (bp)	
				G. birdiae	*G. caudata*
GraB_01	$(ACTTC)_4$ ACTT	F: CGGAGTAGGAGCAAGGAGGT R: AGATTTGGCGACGGTGAG	58 ↓ 53	(85-295)	-
GraB_02	$(CACCG)_2$ CACCA $(CACCG)_1$	F: TCATTGGAACTGGCAACTATG R: GTGGTGGTGTGTCTTTCGTG	58 ↓ 53	(170-250)	-
GraB_03	$(CGTTC)_3$	F: CCTCCCAGACACCATCCTAA R: CGGTTCCCGTTCCCGTTC	58 ↓ 53	(178-288)	-
GraB_04	$(CAAA)_3$	F: AAGCCATCTTCGTTGTGTCTC R: GGTATTAGCGGAATCGTGG	58 ↓ 53	(100-170)	-
GraB_05	$(CCA)_3$ TCA $(CCA)_6$	F: CCACACGCAACAAACTAAACA R: TGAGCAAGGGAAATGAGT	58 ↓ 53	(94-274)	-
GraB_06	$(CCA)_3$ TCA $(CCA)_6$	F: CCACACGCAACAAACTAAACA R: TGAGCAAGGGAAATGAGT	58 ↓ 53	(135-218)	-
GraB_07	$(TCAG)_3$ TTAG $(TCAG)_1$ TCAA $(TCAG)_2$	F: CTTATCCTCAACAGCAGCGAC R: ACAGGGCGTAGTGAAACAGAA	58 ↓ 53	(160-378)	-
GraB_08	$(GTAG)_5$	F: GGCTTGCGGTGGGCGTAA R: TTCGCTTCGTCGCTTCTC	58 ↓ 53	(83-204)	-
GraBC_01	$(ATG)_{21}$	F: TGCGTCAAAGAGAGACAAG R: CACCTCGTGTTCAGGGTCA	54	(322-355)	(232-331)
GraBC_02	$(TCT)_6$	F: GACCCATTCTCTCCGTTTATG R: AGACCCGATGAAGCCCAC	54	(265-274)	(264-310)
GraBC_03	$(TGCG)_2$ TGCT $(TGCG)_2$	F: TCAAATGCCCTTACTGCTCTC R: CACTCTTCACCGCCTCTCTT	58 ↓ 53	(93-231)	(231)
GraBC_04	$(TAAC)_3$ AAAC $(TAAC)_1$	F: ACAGAACCCAACCATACTCCC R: GCAAATGACAGAGAGCGAAA	58 ↓ 53	(153-324)	(162-184)
GraBC_05	$(GCAC)_5$	F: ACGCACGCAAGCACTCAC R: TTGTTCTCAAATGCCCTTACTG	58 ↓ 53	(87-272)	(232-334)
GraC_01	$(AC)_{37}$	F: CACAACACACGGACGAACTAC R: TACCTATTGCGGTTTGAGGG	58 ↓ 53	-	(122-342)
GraC_02	$(ACACA)_6$ AA (ACACA) CACA $(ACACA)_2$	F: ATCGCAAGGTCACTGGTTC R: TGTGTGGTGTGTTGTGTTGTG	58 ↓ 53	-	(130-260)
GraC_03	$(CACG)_7$	F: GCCTCTTCTGTCTTCCAG R: CGTGTGTGTATGAGTGCGTG	54	-	(161-209)
GraC_04	$(GT)_{23}$	F: CCTACGGTCAGCCACTTTCTT R: ACGACTCCTTCCATCCCTTC	58 ↓ 53	-	(178-242)
GraC_05	$(TTG)_4$ TTT $(TTG)_3$ T $(TTG)_4$ G $(TTG)_2$ T $(TTG)_6$ T $(TTG)_2$ G $(TTG)_4$	F: TTGGACGCAGGAGGAAGTAG R: AATGACAAGCACAACTGACGG	58 ↓ 53	-	(286-293)
GraC_06	$(AAC)_6$ AATA $(AAC)_5$	F: ATCTGTGGGTGGAGAAGTGGT R: ATTGTGTTTGGGTGCTCGTT	58 ↓ 53	-	(204-269)
GraC_07	$(GCAC)_4$	F: GCGATACAGCGACAAAGTAAA R: GCTATGGAGACGAACACTGATG	58 ↓ 53	-	(136-158)
GraC_08	$(CACG)_4$	F: TGGTAATCTTGGAATCGTCATC R: ATTGAGGGTGAGAGCAGGAA	58 ↓ 53	-	(180-184)
GraC_09	$(AGGA)_4$	F: AAAGAACACGAACAAGCCAAC R: TCTCTCTCTCTCTCTCTCTCT	58 ↓ 53	-	(194-212)
GraC_10	$(GCGT)_4$	F: ATTGAGGGTGAGAGCAGGAA R: TGGTAATCTTGGAATCGTCATC	58 ↓ 53	-	(200)
GraC_11	$(CGGTA)_3$ A $(CGGTA)_3$	F: TGTGTGTGTGTGTGTGTGTC R: ACCGTAACCTAACCTAACC	58 ↓ 53	-	(135-140)
GraC_12	$(AACCG)_5$	F: TTGCTGCTGTCGTAATCTGTC R: ACTTCTTTGTTGTGGCGGTT	58 ↓ 53	-	(272-282)

Table 3

Click here to download table: Table 3.docx

	G. birdiae						*G. caudata*					
Locus	*Nf*	*n*	*Na*	*He*	*Ho*	F_{is}	*Nf*	*n*	*Na*	*He*	*Ho*	F_{is}
GraB_01	-0.0347	50	7	0.532	0.580	-0.091		-	-	-	-	-
GraB_02	-0.0486	48	2	0.295	0.354	-0.205		-	-	-	-	-
GraB_03*	**0.2056**	19	7	0.606	0.263	**0.572**		-	-	-	-	-
GraB_04	0.0336	49	3	0.099	0.061	0.385		-	-	-	-	-
GraB_05*	**0.2222**	30	12	0.552	0.200	**0.642**		-	-	-	-	-
GraB_06	0.0679	46	13	0.682	0.565	**0.174**		-	-	-	-	-
GraB_07*	**0.1501**	40	12	0.610	0.350	**0.430**		-	-	-	-	-
GraB_08*	**0.1553**	32	14	0.751	0.468	**0.379**		-	-	-	-	-
GraBC_01	0.0966	50	5	0.527	0.400	**0.244**	-0.0011	50	4	0.059	0.060	-0.010
GraBC_02	-0.0013	50	3	0.059	0.060	-0.014	0.001	50	5	0.203	0.200	0.016
GraBC_03**	**0.3515**	20	5	0.714	0.100	**0.863**	0	50	1	0.000	0.000	-
GraBC_04**	**0.2233**	28	9	0.433	0.107	**0.756**	-0.0311	50	3	0.472	0.500	-0.060
GraBC_05	0.2369	45	15	0.854	0.378	**0.560**	0.0003	50	4	0.263	0.260	0.012
GraC_01		-	-	-	-	-	-0.0842	50	7	0.657	0.780	**-0.189**
GraC_02		-	-	-	-	-	-0.1259	45	4	0.568	0.711	**-0.256**
GraC_03		-	-	-	-	-	-0.1424	49	4	0.579	0.776	**-0.343**
GraC_04		-	-	-	-	-	0.0321	46	10	0.783	0.696	0.112
GraC_05		-	-	-	-	-	-0.0407	50	4	0.582	0.640	-0.102
GraC_06*		-	-	-	-	-	**0.1433**	47	4	0.445	0.234	**0.476**
GraC_07		-	-	-	-	-	-0.0229	49	6	0.609	0.612	-0.050
GraC_08		-	-	-	-	-	-0.0238	50	2	0.198	0.220	-0.113
GraC_09*		-	-	-	-	-	**0.0972**	49	6	0.810	0.612	**0.246**
GraC_10		-	-	-	-	-	-0.0022	30	2	0.066	0.067	-0.018
GraC_11*		-	-	-	-	-	**0.1127**	48	2	0.357	0.208	**0.419**
Mean over loci ***			6.857	0.435	0.342	**0.215°**			4.583	0.420	0.460	**-0.097°**
(SE)***			(1.960)	(0.112)	(0.080)				(0.6454)	(0.072)	(0.080)	

figure 1
Click here to download line figure: Fig 1_2015_03_LZW.tif

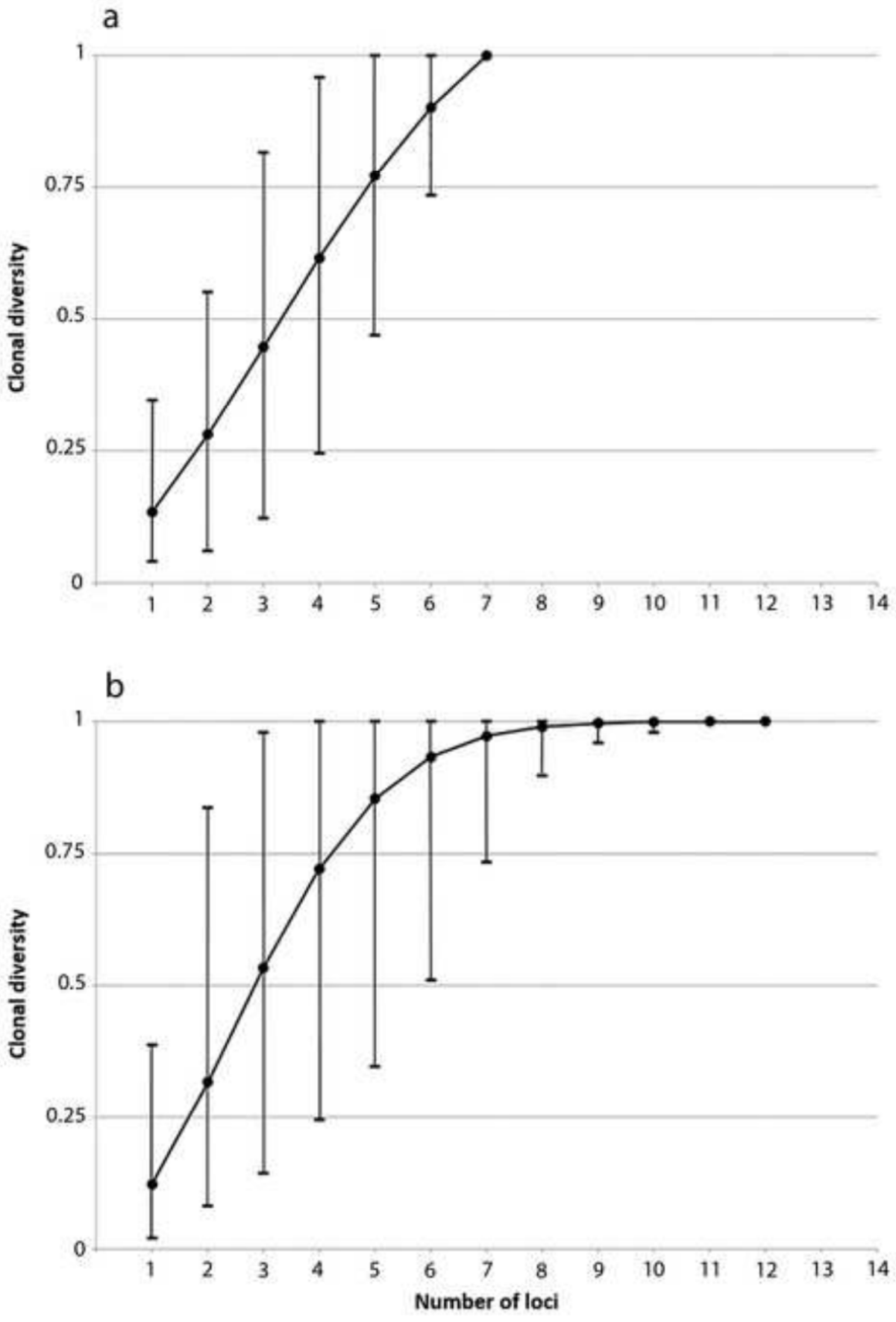

figure 2
Click here to download line figure: Fig 2_2015_03_LWZ.tif