



Microsatellites from sponge genomes: the number necessary for detecting genetic structure in *Hemimycale columella* populations

Janina González-Ramos, Gemma Agell, Maria J. Uriz*

Centre d'Estudis Avançats de Blanes (CEAB-CSIC), C/ Accés a la Cala St. Francesc, 14, 17300 Blanes (Girona), Spain

ABSTRACT: Using next-generation sequencing technology, we designed a pool of microsatellite primers and amplified them in 2 physically isolated populations of the Atlanto-Mediterranean sponge *Hemimycale columella*, which exhibits particular evolutionary, biological, and ecological features. The species has contrasting life cycles in shallow and relatively deep waters, and releases small, predictably low-dispersal larvae. This study experimentally evaluated how many polymorphic microsatellites would be necessary to detect genetic structure in *H. columella* populations, and whether or not the use of compound loci is advisable. By sequencing 1/2 454 GS-FLX run, we obtained a total of 4208 sequences enclosing microsatellite motifs. We selected 20 microsatellites, 15 of which proved adequate for the genetic study of the sponge populations. The 2 analyzed populations were genetically structured with all microsatellite combinations assayed, and the values of the D_{est} and F_{ST} statistics did not increase with increasing number of loci. A weak signal of genetic structure, however, was shown in bar plots representing membership coefficients for each individual to each sampling location even with all loci. We conclude that a few polymorphic loci can detect structure in *H. columella* populations, but using 12 or more loci notably enhances the power of the analyses. The study also describes a low-cost protocol for obtaining microsatellites by the hundreds from non-model, ecologically relevant species, which can be used to provide information about population isolation, adaptation, and vulnerability.

KEY WORDS: SSRs · Pyrosequencing · Population genetics · Genetic differentiation · Porifera · *Hemimycale columella* · Conservation genetics

INTRODUCTION

Sponges are one of the foremost groups in terms of diversity and abundance in marine benthic ecosystems (Gili & Coma 1998), where they play poorly studied but critical structural and trophic roles (Bell 2008). Sponges typically produce bioactive secondary metabolites, which are involved in species-specific interactions and promote species coexistence, and thus biodiversity (e.g. Becerro et al. 2003, Thom & Schupps 2007, Pawlik 2011). Moreover, they are key organisms in transforming dissolved organic matter from water into particulate form, both directly (de Goeij et al. 2013) and through their associated

microbiomes (Mohamed et al. 2010, Ribes et al. 2012). Temporal changes in biological and ecological traits of sponge populations have often been attributed to environmental perturbations, such as anomalously high temperatures (e.g. Cebrian et al. 2011), while the contribution of population genetics to the observed changes has been mostly disregarded (but see Guardiola et al. 2012). However, it is agreed that genetic descriptors such as allele diversity, gene flow, effective population size, kinship, inbreeding, and clonality (among others) may provide information about the genetic fitness of sponge populations and their capacity to face environmental changes (Reusch 2000, Hämmerli & Reusch 2003). Studies on

*Corresponding author: iosune@ceab.csic.es

population genetics of sponge species with a key structural role in marine habitats may provide valuable information for predicting the species' local success, vulnerability, or extinction (McCallum et al. 2013), as well as subsequent cascading effects on the ecosystem (e.g. Menge 1995). However until recently, difficulties in obtaining suitable markers (i.e. with sufficient intra-species variation) to characterize sponge populations at ecological time-scales have hindered this type of research.

Microsatellites are commonly used in the study of animal and plant population genetics (Csilléry 2009) because they are among the most variable types of DNA sequences in the genome (Weber 1990). Widely used molecular markers such as the Folmer partition of the mitochondrial cytochrome oxidase (COI) gene have showed little (if any) intra-species variation in sponges (Duran et al. 2004c, De Biasse et al. 2010, Uriz & Turon 2012). Other variable markers, such as Internal Transcriber Spacers (ITSs) have the drawback of intragenomic variation (Duran et al. 2004b), which makes them unsuitable for use in the study of population genetics.

Thus, in basal living metazoans such as sponges and cnidarians, the development of hypervariable markers for population genetics research at fine ecological scales (Guichoux et al. 2011) is crucial. The major downside in the recent past for the development of sponge microsatellites has been the cumbersome effort required to generate microsatellite-enriched libraries (Zane et al. 2002). As a consequence, only 5 Mediterranean sponges (*Paraleucilla magna*, *Crambe crambe*, *Scopalina lophyropoda*, *Spongia officinalis*, and *Spongia lamella*) have been studied so far using these types of markers (Uriz & Turon 2012), although the trend is changing with the current use of next-generation sequencing (NGS) technologies (e.g. Giles et al. 2013, Riesgo et al. 2014).

Microsatellites are considered appropriate markers for studies of genetic differentiation where only small samples are available (Haasl & Payseur 2011), which is often the case for sponge species due to the typically small size of their populations (Duran et al. 2004a, Blanquer et al. 2009, Blanquer & Uriz 2010, Dailianis et al. 2011, Guardiola et al. 2012, Pérez-Portela et al. 2014). On the other hand, the number of independently segregating loci that are necessary to achieve any reasonable precision in studies of population genetics is a matter of discussion (Koskinen et al. 2004, Hale et al. 2012). Empirical studies and power analyses are required (e.g. Koskinen et al. 2004, Hale et al. 2012), which have not been previ-

ously attempted for sponges but are now possible thanks to NGS techniques.

In addition to designing microsatellites for a sponge using NGS, the current study focuses on *Hemimycale columella*, an intriguing common Atlanto-Mediterranean sponge with a contrasting life span according to its habitat. The sponge releases small, predictably low-dispersal larvae (Pérez-Porro et al. 2012), which suggests population differentiation at short distances due to restricted gene flow and genetic drift. Thus, the rationale for this study was to use a species with predictably structured populations to explore the minimum number and type of microsatellite loci necessary to detect the purported population structure.

Frequent discussions have addressed the number of microsatellites required for precise detection of the genetic structure of a population. It is generally accepted that the power to resolve population structure in part depends on the polymorphism of the markers, where the conservative conclusion has been that 'more are better' (e.g. Koskinen et al. 2004). However, this might not be always the case, and we consider it worthwhile to optimize further studies on the population genetics of *H. columella* by ascertaining the minimal number and type of polymorphic microsatellite loci that would be required to detect significant genetic differentiation in sponge populations without losing accuracy. For this purpose, we investigated the number of loci/alleles needed to capture the genetic diversity of a population, and whether the use of imperfect or compound microsatellites (which have been considered undesirable for these studies because of a higher probability of presenting unnoticed size homoplasy; Angers et al. 2000) produced biased results.

By selecting the type and number of microsatellites that optimize costs and accuracy, this study also provides guidelines for microsatellite-based approaches for studying population genetics of other genetically non-model marine invertebrates targeted for ecological or biotechnological reasons.

MATERIALS AND METHODS

Sampling and preservation

Two populations of *Hemimycale columella*, located approximately 1.3 km apart near of Arenys de Mar, Iberian Peninsula (NW Mediterranean), were sampled by SCUBA diving during the winter of 2010. The 2 sampled areas, called 'Aquari' and 'Castell' (hereafter HA and HC, respectively), were located at

41° 31' 21.61" N, 2° 35' 0.26" E and 41° 31' 50.41" N, 2° 34' 18.29" E, respectively. They consisted of approximately 1 to 2 m high, 10 to 15 m wide, 35 to 40 m long rocky bars at 28 to 30 m depth, running parallel to the coastline, surrounded by sandy bottoms (Fig. 1). In total, 64 samples were collected: 29 individuals from HA and 35 individuals from HC. Collected samples were placed in hermetic plastic bowls while underwater and transported to the laboratory in a cooler, where they were cleaned of associated organisms under a stereomicroscope, preserved in absolute ethanol, and stored at -20°C until DNA extraction.

Sponge pyrosequencing and design of microsatellite primers

The DNA of *H. columella* was extracted for pyrosequencing with QIAmp DNA stool kit (Qiagen), which proved to be the best method for this particular sponge. DNA concentration and quality were assessed using a Qubit fluorometer (Invitrogen) and a 2100 Bioanalyzer (Agilent Technologies), respectively. Sequencing (1/2 run) was performed in a 454 GS-FLX sequencer (Roche) at the Scientific and Technological Center of the University of Barcelona (CCiTUB).

Sequences were analyzed with the open-access program QDD v.2.1 (Megléczy et al. 2010), which is a useful tool for the discovery and selection of microsatellites, and for primer design. The QDD package integrates the software programs BLAST, ClustalW (Larkin et al. 2007) and Primer3-1.1.4 (Rozen &

Skaletsky 2000), and works in 3 steps: sequence cleaning and microsatellite recognition, detection of sequence similarity, and primer design. From the primer pairs that were designed (see Table S1 in the Supplement at www.int-res.com/articles/suppl/b024p025_supp.pdf), the best microsatellites were selected taking into account product size (100 to 350 bp), suitable flanking region, self-complementarity, GC (guanine–cytosine) content (50 to 60%), and GC clamp regions (as indicated by Primer3).

DNA extraction, amplification, and microsatellite genotyping

DNA was extracted from the 64 individual sponges using the QIAmp DNA stool kit (Qiagen). The forward primers of each locus were labeled with fluorescent dyes (NED, PET, VIC, 6-FAM; Applied Biosystems) for screening. A total of 20 microsatellites were selected, however, 3 of them failed to amplify. The DNA samples were then amplified for the remaining 17 polymorphic microsatellites (*Hemi_1* to *Hemi_17*) in a final volume of 25 µl, which contained 1 µl of DNA template, 0.25 µM of each primer, 0.25 mM of dNTP mix (Sigma-Aldrich), 1 U of *Taq* polymerase, 1× reaction buffer, 4 mM of MgCl₂ (all from Bioron), and 30 µM of BSA (Bio-Rad).

PCR was performed in a thermal cycler (MyCycler, Bio-Rad) with the following parameters: 4 min denaturation at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at a locus-specific annealing temperature (see Table S1 in the Supplement), 30 s at 72°C, followed by an extension cycle of 4 min at 72°C. The resulting PCR products were then visualized on 1.3% agarose gel stained with GelRed (Biotium) and then genotyped in an ABI Prism 3700 automated sequencer (Applied Biosystems) at the CCiTUB.

The length of the PCR products was estimated relative to the internal size standard GeneScan 500LIZ and determined using GeneMapper software (Applied Biosystems). The raw data generated by GeneMapper were reviewed using AutoBin v.0.9 (Excel macro written in Microsoft Visual Basic, F. Salin unpubl.), in order to automatically detect relevant gaps in size, and consequently, bin allele size. Moreover, 3 independent readers checked the AutoBin results to ensure a lack of scoring errors.

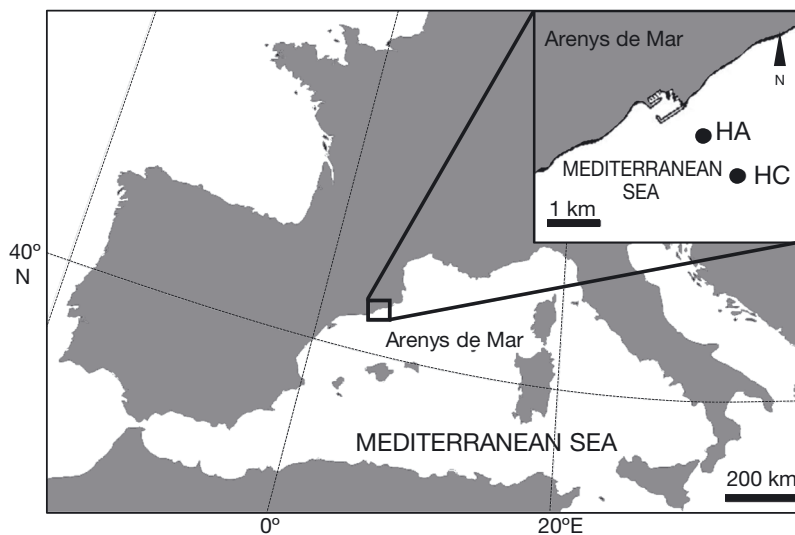


Fig. 1. Sampling locations near Arenys de Mar, Iberian Peninsula. HA = Aquari; HC = Castell

Data analysis

The linkage disequilibrium was first analyzed for the 17 selected loci by GENEPOP v.4.0 (Raymond & Rousset 1995, Rousset 2008). Significant linkage disequilibrium after Bonferroni sequential correction (Rice 1989) was analyzed for all microsatellites. The presence of null alleles was assessed with MicroChecker v.2.2.3 (Van Oosterhout et al. 2004).

Determination of the allele frequencies, the expected (H_e) and observed heterozygosity (H_o), the number of alleles per loci (N_A), the number of private alleles, the inbreeding coefficient (F_{IS}) (Weir & Cockerham 1984) for each locus individually, as well as for all loci combined, and the departure from the Hardy-Weinberg equilibrium was analyzed after Bonferroni corrections using GENEPOP. We considered informative alleles in the populations those showing allele frequencies higher than 0.05 (Hale et al. 2012).

Genetic differentiation and structure, and molecular variance

The genetic structure of the 2 populations was assessed by both the differentiation index D_{est} , which is independent of within-population heterozygosity (Jost 2008) and the F_{ST} statistic, which measures allele fixation and depends on within-population heterozygosity (Weir & Cockerham 1984). Pair-wise F_{ST} values and their statistical significance were calculated on the allele frequencies adjusted for the downward bias resulting from the presence of null alleles using MicroChecker software. F_{ST} values and power analyses with several loci sets were computed using POWSIM software (10 000 permutations) (Ryman & Palm 2006). D_{est} values, p-values, and confidence intervals (CIs) of pair-wise D_{est} comparisons were calculated with DEMETics v.0.8-2. This software, which is implemented within the statistical package R v.2.12.1, estimates the p-values and CIs according to Manly (1997) with a bootstrap method that distributes either the alleles (populations in Hardy-Weinberg equilibrium for this locus) or the genotypes (populations in Hardy-Weinberg disequilibrium) randomly among populations (Goudet et al. 1996) for a specific locus.

The analyses were performed with several microsatellite sets: perfect (7 loci), compound (8 loci), and perfect plus compound (15 loci) groups with the 6, 9, and 12 most polymorphic microsatellites, and a random selection (by using a random number generator) of 6, 9, or 12 microsatellites (10 times), disregarding

whether they were compound or perfect. The relationship between the number of loci (and the number of alleles) used in the analyses and the differentiation values that were obtained was assessed using Spearman's correlation with IBM SPSS Statistics v.20.0. A hierarchical analysis of molecular variance (AMOVA) was performed with ARLEQUIN (Excoffier et al. 2005) to assess variance partition between populations, within populations, and within individuals with the microsatellite sets.

The number of genetically homogeneous clusters (K) was inferred using a Bayesian algorithm with the software STRUCTURE v.2.3.3, which assigns genotypes to clusters by minimizing linkage disequilibrium and deviations from Hardy-Weinberg within clusters (Pritchard et al. 2000, Falush et al. 2003, 2007, Hubisz et al. 2009), with the following parameters: 20 runs for a range of $K = 1$ to 5, with 10^5 Markov Chain Monte Carlo (MCMC) repetitions each and a burn-in of 10 000 iterations, and with 2 location priors (the 2 sampling sites). Admixed populations were assumed based on previous studies on sponges (e.g. Blanquer & Uriz 2010, Guardiola et al. 2012). The most likely value for K based on the STRUCTURE output was determined by plotting the natural log mean probability, $\ln(K)$, of the data over 20 runs for K from 1 to 5 as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). Results of the 20 runs were merged with CLUMPP (Jakobsson & Rosenberg 2007), and DISTRUCT (Rosenberg 2004) was then used for a bar plot representation of the membership coefficients for each individual in each population (i.e. HA and HC).

RESULTS

Sponge pyrosequencing and design of microsatellite primers

Approximately 300 000 reads with an average length of 350 bp were generated. After removing chimeras and redundancies, we obtained 4208 unique sequences containing microsatellite motifs. Of these, 1358 proved to be suitable for primer design (1317 contained perfect, and 41 contained compound microsatellites). Finally, 10 perfect and 10 compound microsatellites were selected at random for the experimental study. The size of microsatellite sequences ranged from 140 to approximately 350 bp (see Table S1 in the Supplement at www.int-res.com/articles/suppl/b024p025_supp.pdf); microsatellite accession numbers are listed in Table S1.

Table 1. Statistics for the *Hemimycale columella* microsatellites for each population. *Hemi_1* and *Hemi_8* were unsuitable for analysis. Location designations: HA = Aquari; HC = Castell. N_A : number of alleles; H_e : expected heterozygosity; H_o : observed heterozygosity; F_{IS} : inbreeding coefficient; HWE: probability of departure from Hardy-Weinberg equilibrium.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Locus	HA	HC	Tot. N_A locus ⁻¹	Locus	HA	HC	Tot. N_A locus ⁻¹
<i>Hemi_2</i>				<i>Hemi_11</i>			
N_A	5	4	5	N_A	6	6	6
H_e	0.737	0.697		H_e	0.712	0.701	
H_o	0.654	0.633		H_o	0.591	0.714	
F_{IS}	0.115*	0.092**		F_{IS}	0.386	0.33*	
<i>Hemi_3</i>				<i>Hemi_12</i>			
N_A	3	5	5	N_A	2	2	2
H_e	0.222	0.389		H_e	0.275	0.119	
H_o	0.160	0.067		H_o	0.321	0.000	
F_{IS}	0.284	0.831**		F_{IS}	-0.174	1**	
<i>Hemi_4</i>				<i>Hemi_13</i>			
N_A	6	4	8	N_A	3	4	4
H_e	0.518	0.434		H_e	0.515	0.369	
H_o	0.560	0.357		H_o	0.536	0.286	
F_{IS}	-0.082	0.181		F_{IS}	-0.041*	0.229**	
<i>Hemi_5</i>				<i>Hemi_14</i>			
N_A	3	4	4	N_A	7	7	8
H_e	0.592	0.591		H_e	0.618	0.622	
H_o	0.966	0.875		H_o	0.375	0.667	
F_{IS}	-0.649*	-0.493*		F_{IS}	0.398**	-0.074	
<i>Hemi_6</i>				<i>Hemi_15</i>			
N_A	4	2	4	N_A	7	6	7
H_e	0.727	0.489		H_e	0.808	0.776	
H_o	0.607	0.290		H_o	0.536	0.353	
F_{IS}	0.167**	0.41*		F_{IS}	0.341**	0.549**	
<i>Hemi_7</i>				<i>Hemi_16</i>			
N_A	6	5	6	N_A	5	6	6
H_e	0.738	0.714		H_e	0.661	0.727	
H_o	0.740	0.594		H_o	0.423	0.677	
F_{IS}	0.297*	0.198**		F_{IS}	0.364**	0.069*	
<i>Hemi_9</i>				<i>Hemi_17</i>			
N_A	4	5	6	N_A	4	5	5
H_e	0.680	0.766		H_e	0.567	0.700	
H_o	0.714	0.739		H_o	0.519	0.480	
F_{IS}	-0.053*	0.036		F_{IS}	0.088*	0.318*	
<i>Hemi_10</i>				Mean N_A			
N_A	9	9	12	pop. ⁻¹	5	5	
H_e	0.816	0.806		Mean H_e	0.661	0.697	
H_o	0.480	0.095		Mean H_o	0.536	0.48	
F_{IS}	0.416	0.884*		F_{IS}	0.136*	0.239*	
				HWE	***	***	

Population genetic estimators

Hemi_1 and *Hemi_8* showed linkage disequilibrium in both populations after Bonferroni correction, and were discarded from further analyses. Once these 2 loci were removed, the mean number of alleles per locus for the 15 microsatellites used was 5 per

population (HA and HC). Mean H_o was 0.536 (HA) and 0.48 (HC) (Table 1). The compound microsatellites showed 6 ± 2.92 (mean \pm SD) alleles while the perfect microsatellites showed 5 ± 1.38 alleles. Differences in the mean number of alleles between perfect and compound loci were not significant (t -test, $p > 0.05$). The total number of alleles per locus ranged from 2 for *Hemi_12* (compound) to 12 for *Hemi_10* (compound) (Table 1), and the total number of informative alleles (i.e. those with higher frequencies than 0.05) was 53 and 51 for HA and HC populations, respectively.

Overall F_{IS} values were positive and significant ($p < 0.001$) for both populations, indicating heterozygote deficiency and inbreeding. The exact tests for Hardy-Weinberg equilibrium confirmed these results showing a significant deviation for both populations (Table 1). MicroChecker analysis indicated the presence of null alleles for 7 of the 15 loci, which could account for the homozygote excess found (see Table S2 in the Supplement).

Population structure

Several private alleles were found with all microsatellite sets assayed in the 2 populations (Table S2). Their mean frequency per population was 0.087 when only the compound microsatellites were considered, 0.089 when the perfect ones were used and slightly decreased from 0.089 (6 loci) to 0.083 (15 loci) as we added from more to less polymorphic loci to the analyses.

Both populations proved to be differentiated (significant F_{ST} and D_{est}) with groups of 6, 9, 12 and 15 from more to less polymorphic loci (Table 2). There was no significant relationship between the number of loci (groups formed by the most polymorphic loci) used and the F_{ST} values, as revealed by correlation analyses ($r = -0.7$, $p = 0.4$). F_{ST} remained constant with increasing number of loci, and the p -value for the F_{ST} statistic was higher with 12 loci than with 15 loci (Table 2). The power of the analyses for F_{ST} estimation enhanced progressively as the number of loci increased (Fisher test),

Table 2. Pairwise comparisons for genetic differentiation between populations using 4 microsatellite sets with an increasing number of loci (from more to less polymorphic). Location designations: HA = Aquari; HC = Castell; number of loci indicated in parentheses. Upper and lower values in the diagonal cells correspond to D_{est} and F_{ST} , respectively. * $p < 0.05$; ** $p < 0.01$. Loci used in each set are given in *italics*

	HA (6)	HA (9)	HA (12)	HA (15)
HC (6)	0.174**			
<i>Hemi_4,10,11,14,15, 16</i>	0.01*			
HC (9)		0.161**		
Above 6 loci plus <i>Hemi_7,9,17</i>		0.01*		
HC (12)			0.123**	
Above 9 loci plus <i>Hemi_2,3,5</i>			0.01***	
HC (15)				0.111**
<i>All loci</i>				0.01*

although this increase was noteworthy from 6 loci (69%) to 9 (76%) but tended to stabilize from 12 (84%) to 15 (86%). Conversely, a significant negative correlation was found between D_{est} values and the number of loci used ($r = -0.98$, $p < 0.05$). D_{est} values slightly decreased with an increasing number of loci, but in all cases remained statistically significant ($p < 0.05$) (Table 2)

Population structure was also detected with compound or perfect loci separately (Table 3), although the power of the analyses was low in both cases (57 and 54%, respectively).

When microsatellite sets were randomly selected (10 repetitions), D_{est} revealed significant values with all groups assayed (6, 9, and 12 loci), while F_{ST} was significant 100% of the time (6 loci), 83% of the time (9 loci) and 91.6% of the time (12 loci) (Fig. 2). No correlation was found between the number of randomly selected loci and the F_{ST} or D_{est} values ($r = 0.14$

and $r = 0.1$, respectively). The power of the analyses performed with sets of randomly selected loci ranged from 50 to 63% (6 loci), 61 to 76% (9 loci) and 75 to 87% (12 loci).

Hierarchical AMOVA confirmed significant genetic differences between the 2 populations with all the above-mentioned microsatellite sets ($p < 0.001$) (Table S3 in the Supplement). The between-population variation explained from 4.7 to 5.6% of the variance, while variation within populations explained 17.6 to 27%, depending on the microsatellite set considered.

The Bayesian Inference method gave the highest likelihood for 2 ($K = 2$) genetically homogeneous groups (populations) when 9, 12, and 15 loci were used, but failed to

discriminate 2 populations when only 6 loci were considered. The mean log-likelihood of K was higher for $K = 2$ than for $K = 1$, with 12 and 15 loci (Fig. 3). The representation of the estimated membership coefficients for each individual in each of the 2 location priors indicated weak structure in the purported populations with all microsatellite sets assayed (Fig. 3).

Table 3. Pairwise genetic differentiation comparisons between populations using perfect, compound and perfect-plus-compound microsatellite sets (indicated in parentheses). Location designations: HA = Aquari; HC = Castell. Upper and lower values in the diagonal cells correspond to D_{est} and F_{ST} , respectively. * $p < 0.05$; ** $p < 0.01$

	HA (8 compound)	HA (7 perfect)	HA (15)
HC (8 compound)	0.152**		
	0.010*		
HC (7 perfect)		0.065**	
		0.010**	
HC (15)			0.111**
			0.010*

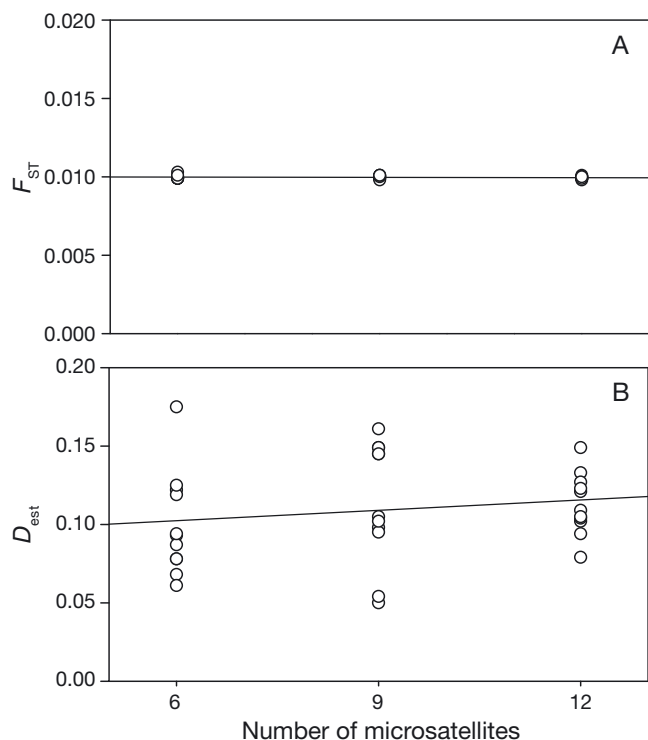


Fig. 2. *Hemimyscale columella*. Graphic representation of the relationship (Spearman's correlation) between (A) D_{est} and (B) F_{ST} values and increasing numbers of randomly selected microsatellites (10 repetitions)

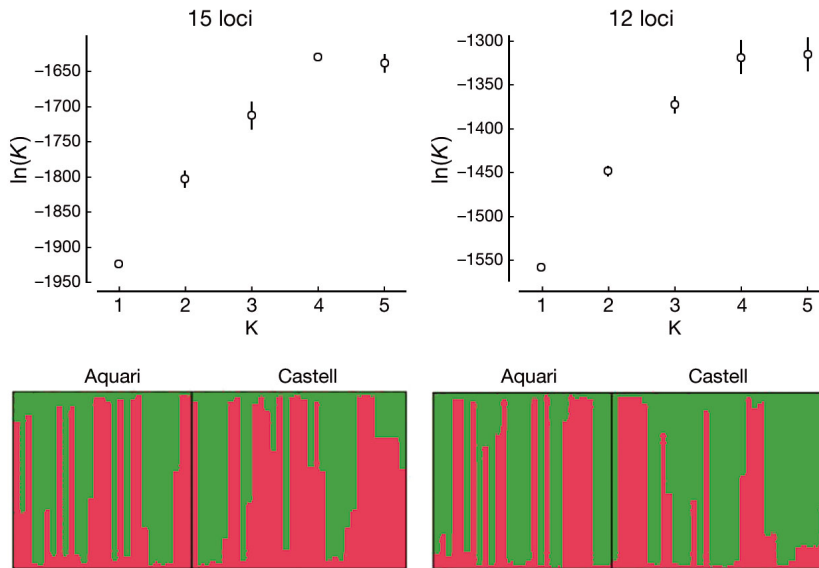


Fig. 3. *Hemimycale columella*. Mean (\pm SD) probability of the number of genetically homogeneous clusters (K) resulting from Bayesian inference (above) and membership of the analyzed individuals to the 2 sampling locations (below), with 15 and 12 loci

DISCUSSION

This study describes an approach to design microsatellite primers using NGS sequencing methods in sponges. Microsatellites are abundant motifs in the animal genome, and massive sequencing of a small part of the *Hemimycale columella* genome has provided a large number of sequences with repeated motifs.

The 2 geographically close populations of *H. columella* proved to be significantly structured (significant F_{ST} and D_{est} , private alleles; AMOVA) independent of the number of loci used, or whether the loci were perfect or compound. The statistical power of the F_{ST} tests was low with 6 loci, but increased with increasing number of loci (up to 12). This agrees with the conclusion of Koskinen et al. (2004) that the power of F_{ST} is low with a small number of loci, and suggests that previous studies on sponge population genetics that have systematically used a low number of microsatellites may have detected population structure by chance alone. However, statistical power did not improve significantly in our experimental study when the number of loci increased from 12 to 15, suggesting that 'more is not better', and that performing empirical tests and power analyses before addressing whole population genetic studies will optimize costs without losing accuracy.

Amazingly, D_{est} slightly decreased with increasing number of loci, when more to less polymorphic loci

were added. The strong dependence of D_{est} on the mutation rate of each particular locus (Whitlock 2011) might have contributed to this pattern, since D_{est} did not decrease but rather slightly increased with increasing number of loci when the loci were selected randomly.

A negative relationship or the absence of any correlation between the number of loci and F_{ST} and D_{est} values indicates that increasing the number of loci above a minimum threshold does not necessarily improve detection of genetic differentiation in this sponge. Moreover, finding private alleles in populations located 1.3 km apart even with only 6 loci highlights the efficiency of the selected markers to detect structure.

Bayesian analyses with STRUCTURE showed a higher probability ($\ln(K)$) for 2 genetically homogeneous clusters than for only one, with 9, 12, and 15 microsatellites. A weak signal of genetic structure was shown by the bar plots representing the membership coefficients for each individual to each location, despite the fact that the F_{ST} and D_{est} statistics and the AMOVA analyses indicated significant population differentiation between samples collected at the 2 locations. This lack of consistency between descriptors population structure (which has been reported previously) has been explained by the considerable amount of genetic evidence necessary to provide strong support for any particular partition of the Bayesian algorithm of STRUCTURE (Hubisz et al. 2009). In any case, the same STRUCTURE output was obtained with 12 and 15 loci, indicating that the number of microsatellites above 9 does not seem to influence the outcome.

Compound loci gave higher values for D_{est} and F_{ST} than perfect loci. Both statistics were statistically significant with the 2 types of microsatellites, although the power of the F_{ST} analyses consistently was low, likely because of a relatively low number of loci (Koskinen et al. 2004).

The compound microsatellites used in the present study mainly contained tri- and tetranucleotide repeats, which may encompass higher mutation rates than dinucleotides since the longer the sequence is, the higher the probability of mutation (Schug et al. 1998). Therefore, a higher number of alleles would be expected for the compound loci than for the

primarily dinucleotide-made-perfect loci. However, both types of loci were similarly polymorphic in the 2 target populations, and thus some homoplasy can be assumed for the compound loci used.

Undetected size homoplasy, which is expected to be higher for compound loci (Angers et al. 2000, Adams et al. 2004), is the main drawback associated with the use of microsatellites, as it may hide mutations and thus underestimate genetic differentiation; however, the purported homoplasy of our compound loci did not prevent them from revealing population structure. In fact, homoplasy may not be a problem with moderate population sizes (Estoup et al. 2002), as only a slight underestimation of genetic differentiation (1 to 2%) is attributed to detectable microsatellite homoplasy (Adams et al. 2004, Curtu et al. 2004). Thus, although the use of compound microsatellites is often not recommended because of homoplasy (Megléczy et al. 2010), where population sizes are small (as in most sponge species and other sessile short-dispersal invertebrates; Combosch & Vollmer 2011), compound microsatellites should not be summarily discarded.

The revealed structure between geographically close, deep populations of *H. columella* indicates reduced gene flow. Population structure and a level of genetic isolation compatible with a scenario of local adaptation (Shine 2012) might also be expected between shallow and deep populations. Local adaptation in *H. columella* might explain, for instance, the annual versus multiannual life cycles of shallow and deep populations, respectively (authors' pers. obs.). The sponge aquiferous system is drastically reduced during embryo incubation (Pérez-Porro et al. 2012), which results in a notable lowering of the sponges' capacity to capture food, and even in a cessation of its filtering activity after larval release (Turón et al. 1999). Any gene selection related to metabolic savings or sponge starvation during this critical life state might help deep sponges to survive after the reproductive period. Moreover, genetic differentiation between geographically close populations is indicative of the species' poor dispersal, and suggests weakness when facing potential environmental changes (Frankham 1998).

Microsatellite markers, which now can easily be uncovered from non-model organisms thanks to NSG technologies, could be systematically used for accurate genetic characterization of sponge populations where species preservation is pursued, since they can reveal information about population vulnerability in the face of both anthropogenically induced and natural environmental changes.

Acknowledgements. We thank A. Blanquer and M. Guardiola for help with sampling and software use, respectively. We thank the 3 anonymous reviewers, who greatly helped to improve the manuscript. The study was funded by Mar-SymBiotics project ref. MINECO, I+D+I of Excellence, CTM2013-43287-P and 2014-SGR120 Benthic Ecology Consolidate Group from the Generalitat of Catalunya, to M.J.U.

LITERATURE CITED

- Adams RI, Brown KM, Hamilton MB (2004) The impact of microsatellite electromorph size homoplasy on multi-locus population structure estimates in a tropical tree (*Corythophora alta*) and an anadromous fish (*Morone saxatilis*). *Mol Ecol* 13:2579–2588
- Angers B, Estoup A, Jarne P (2000) Microsatellite size homoplasy, SSCP, and population structure: a case study in the freshwater snail *Bulinus truncatus*. *Mol Biol Evol* 17: 1926–1932
- Becerro MA, Thacker RW, Turón X, Uriz MJ, Paul VJ (2003) Biogeography of sponge chemical ecology: comparisons of tropical and temperate defenses. *Oecologia* 135:91–101
- Bell JJ (2008) Functional roles of marine sponges. *Estuar Coast Shelf Sci* 79:341–353
- Blanquer A, Uriz MJ (2010) Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC Evol Biol* 10:13
- Blanquer A, Uriz MJ, Caujapé-Castells J (2009) Small-scale spatial genetic structure in *Scopalina lophyropoda*, an encrusting sponge with philopatric larval dispersal and frequent fission and fusion events. *Mar Ecol Prog Ser* 380:95–102
- Cebrian E, Uriz MJ, Garrabou J, Ballesteros E (2011) Sponge mass mortalities in a warming Mediterranean Sea: Are cyanobacteria-harboring species worse off? *PLoS ONE* 6: e20211
- Combosch DJ, Vollmer SV (2011) Population genetics of an ecosystem-defining reef coral *Pocillopora damicornis* in the tropical Eastern Pacific. *PLoS ONE* 6:e21200
- Csilléry K (2009) Statistical inference in population genetics using microsatellites. PhD dissertation, University of Edinburgh
- Curtu AL, Finkeldey R, Gailing O (2004) Comparative sequencing of a microsatellite locus reveals size homoplasy within and between European oak species (*Quercus* spp.). *Plant Mol Biol Rep* 22:339–346
- Dailianis T, Tsigenopoulos CS, Dounas C, Voultsiadou E (2011) Genetic diversity of the imperiled bath sponge *Spongia officinalis* Linnaeus, 1759 across the Mediterranean Sea: patterns of population differentiation and implications for taxonomy and conservation. *Mol Ecol* 20: 3757–3772
- De Biasse MB, Richards VP, Shivji MS (2010) Genetic assessment of connectivity in the common reef sponge, *Callyspongia vaginalis* (Demospongiae: Haplosclerida) reveals high population structure along the Florida reef tract. *Coral Reefs* 29:47–55
- de Goeij JM, van Oevelen D, Vermeij MJA, Osinga R, Middelburg JJ, de Goeij AFPM, Admiraal W (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* 342:108–110
- Duran S, Pascual M, Estoup A, Turón X (2004a) Strong population structure in the marine sponge *Crambe crambe* (Poecilosclerida) as revealed by microsatellite markers.

- Mol Ecol 13:511–522
- Duran S, Giribet G, Turon X (2004b) Phylogeographical history of the sponge *Crambe crambe* (Porifera, Poecilosclerida): range expansion and recent invasion of the Macaronesian islands from the Mediterranean Sea. Mol Ecol 13:109–122
- Duran S, Pascual M, Turon X (2004c) Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe* (Poecilosclerida). Mar Biol 144:31–35
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361
- Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604
- Excoffier L, Laval G, Schneider S (2005) Arlequin 3.01: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164:1567–1587
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: codominant markers and null alleles. Mol Ecol Notes 7:574–578
- Frankham R (1998) Inbreeding and extinction: island populations. Conserv Biol 12:665–675
- Giles EC, Saenz-Agudelo P, Berumen ML, Ravaso T (2013) Novel polymorphic microsatellite markers developed for a common reef sponge *Stylissa carteri*. Mar Biodiv 43: 237–241
- Gili JM, Coma R (1998) Benthic suspension feeders: their paramount role in littoral marine food webs. Trends Ecol Evol 13:316–321
- Goudet J, Raymond M, de Meeus T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144: 1933–1940
- Guardiola M, Frotscher J, Uriz MJ (2012) Genetic structure and differentiation at a short-time scale of the introduced calcareous sponge *Paraleucilla magna* to the western Mediterranean. Hydrobiologia 687:71–84
- Guichoux E, Lagache L, Wagner S, Léger P, Petit RJ (2011) Two highly validated multiplex (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). Mol Ecol Resour 11:578–585
- Haasl RJ, Payseur BA (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106:158–171
- Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7:e45170
- Hämmerli A, Reusch TBH (2003) Flexible mating: cross-pollination affects sex-expression in a marine clonal plant. J Evol Biol 16:1096–1105
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806
- Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026
- Koskinen MT, Hirvonen H, Landry P, Primmer CR (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. Hereditas 141:61–67
- Larkin MA, Blackshields G, Brown NP, Chenna R and others (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948
- Manly BFJ (1997) Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, Boca Raton, FL
- McCallum KP, Guerin GR, Breed MF, Lowe AJ (2013) Combining population genetics, species distribution modelling and field assessments to understand a species vulnerability to climate change. Aus Ecol 39:17–28
- Megléc E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26:403–404
- Menge BA (1995) Indirect effects in marine rocky intertidal interaction webs: patterns and importance. Ecol Monogr 65:21–74
- Mohamed NM, Saito K, Tal Y, Hill RT (2010) Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. ISME J 4:38–48
- Pawlik JR (2011) The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. Bioscience 61:888–898
- Pérez-Porro AR, González J, Uriz MJ (2012) Reproductive traits explain contrasting ecological features in sponges: the sympatric poecilosclerids *Hemimycale columella* and *Crella elegans* as examples. Hydrobiologia 687:315–330
- Pérez-Portela R, Noyer C, Becerro M (2014) Genetic structure and diversity of the endangered bath sponge *Spongia lamella*. Aquat Conserv, doi:10.1002/aqc.2423
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): a population genetics software for exact test and ecumenicism. J Hered 86:248–249
- Reusch TBH (2000) Five microsatellite loci in eelgrass *Zostera marina* and a test of cross-species amplification in *Z. noltii* and *Z. japonica*. Mol Ecol 9:371–373
- Ribes M, Jiménez E, Yahel G, López-Sendino P and others (2012) Functional convergence of microbes associated with temperate marine sponges. Environ Microbiol 14: 1224–1239
- Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225
- Riesgo A, Blasco G, Erwin PM, Pérez-Portela R, López-legentil S (2014) Optimization of 14 microsatellite loci in a Mediterranean demosponge subjected to population decimation, *Ircinia fasciculata*. Conserv Genet Resour 6: 301–303
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol 13:137–138
- Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resour 8:103–106
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386
- Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Mol Ecol 15:600–602

- Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol Biol Evol* 15:1751–1760
- Shine R (2012) Invasive species as drivers of evolutionary change: cane toads in tropical Australia. *Evol Appl* 5: 107–116
- Thom C, Schupps P (2007) Chemical defense strategies in sponges: a review. In: Custódio MR, Lôbo-Hajdu G, Hajdu E, Muricy G (eds) *Porifera research: biodiversity, innovation and sustainability*. Museo Nacional, Rio de Janeiro, p 627–637
- Turon X, Uriz MJ, Willenz P (1999) Cuticular linings and remodeling processes in *Crambe crambe* (Demospongiae: Poecilosclerida). *Mem Queensl Mus* 44:617–662
- Uriz MJ, Turon X (2012) Sponge ecology in the molecular era. *Adv Mar Biol* 61:345–410
- Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538
- Weber JL (1990) Informativeness of human (dC-dA)_n/(dG-dT)_n polymorphisms. *Genomics* 7:524–530
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358–1370
- Whitlock MC (2011) *G*_{ST} and *D* do not replace *F*_{ST}. *Mol Ecol* 20:1083–1091
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1–16

Editorial responsibility: Nikolaos Schizas,
Mayagüez, Puerto Rico, USA

Submitted: June 11, 2014; Accepted: February 26, 2015
Proofs received from author(s): May 5, 2015