Supplementary Material: Using genetics to inform restoration and predict resilience in declining populations of a keystone marine sponge

Sarah M. Griffiths, Evelyn Taylor-Cox, Donald C. Behringer, Mark J. Butler IV & Richard F. Preziosi

Corresponding author: Sarah Griffiths, Manchester Metropolitan University.

griffiths.sarahm@gmail.com

Methods

Microsatellite development

We collected a single *Spheciospongia vesparium* tissue sample from Long Key, Florida, and preserved it in 95% ethanol. We studied the sample under a stereomicroscope to remove any contaminating invertebrates, and processed the tissue following the methods of Freeman and Thacker (2011) to separate eukaryote and prokaryote cells as far as possible. We extracted DNA from the resulting pellet of 'eukaryote' cells using the Qiagen DNeasy® Blood and Tissue Kit, and subsequently concentrated the DNA by vacuum centrifugation. We performed paired-end library construction using 50 ng of DNA and the Nextera® DNA Sample Preparation Kit (Illumina), before paired-end sequencing (2 x 250 bp) in half a flow cell lane on the MiSeq platform (Illumina) (i.e. the lane was shared with one other sample unrelated to this study) at the University of Manchester Core Genomics Facility. 2 x 3,051,330 reads were produced by the sequencing run, which we then processed using the Galaxy Palfinder bioinformatics pipeline, a customized Galaxy instance hosted by the University of Manchester Bioinformatics Core Facility (Griffiths et al 2016). We quality filtered reads using Trimmomatic v.0.32 (Bolger et al 2014) (SLIDING WINDOW: WINDOW SIZE = 4 bp, QUALITY = 20; LEADING = 3; TRAILING = 3; MINLEN = 50). Following this, 2 x 2,960,928 reads remained. To locate microsatellite regions with at least eight repeat units and suitable primer binding regions, we used pal_finder v.0.02.04 (Castoe, Poole & Koning, 2012), and designed primers when possible using Primer3 v.2.0.0 (Koressaar and Remm 2007;

Untergasser et al 2012). Thirty-six loci were tested; twelve could be successfully amplified and scored, and were thus subsequently used in this study.

References

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. doi: 10.1093/bioinformatics/btu170

Castoe T, Poole A, de Koning A, et al (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. PLoS One 7:e30953. doi: 10.1371/journal.pone.0030953

Freeman CJ, Thacker RW (2011) Complex interactions between marine sponges and their symbiotic microbial communities. Limnol Oceanogr 56:1577–1586. doi: 10.4319/lo.2011.56.5.1577

Griffiths SM, Fox G, Briggs PJ, et al (2016) A Galaxy-based bioinformatics pipeline for optimised, streamlined microsatellite development from Illumina next-generation sequencing data. Conserv Genet Resour 8:481–486. doi: 10.1007/s12686-016-0570-7

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–91. doi: 10.1093/bioinformatics/btm091

Untergasser A, Cutcutache I, Koressaar T, et al (2012) Primer3- new capabilities and interfaces. Nucleic Acids Res 40:e115. doi: 10.1093/nar/gks596

Table S1: Characterisation of 12 polymorphic microsatellite loci and two multiplexes for *Spheciospongia vesparium*.

Locus name	Motif	Primer sequences (5'→ 3')	MP	L	Na	Size range (bp)	N	GenBank accession no.
Vesp15	TTC	F: AGAAGGGTTTAAAAGAAGCAGCAGAAGGG R: TATTGTGAGATATCACTTCCACGACCAGC	A	1	17	223-300	0.012	KX758634
Vesp23	TTC	F: CTAGAAGATCAACTCCTTGACCTTGGGC R: TGAGGATGATTTCGATGAAGTACCG	A	2	4	202-238	0.201	KX758641
Vesp35	AGG	F: ACCCCAGTCCGAGTACATCATCAGG R: ATGATTCCCGAACAGAAGTGAGTGC	A	2	14	447-468	0.040	KX758643
Vesp3	AAC	F: TATTATGCTGCAGTGTATTCAGCATCTCC R: CTCTTCCCTTTGGCTCACAGTATCCC	A	1	10	380-411	0.027	KX758633
Vesp27	ACC	F: TTCTTACACAATCTACCAATCCTTGCAGC R: CACACTGTGATCTATTTTAATGTCCCTCC	A	2	25	291-391	0.252	KX758642
Vesp1	ATAC	F: TGGTTCATAATTGTAGCAACTAATCCCGC R: AAGTATGCGTTTGAGCAAGTCTGAAAAGG	B	2	14	174-255	0.232	KX758638
Vesp30	ACGC	F: GGATCATCAAGATGTTTCTCAAGGTCAGC R: TTTGGTCCTGTTACACACAAATTGTAGCC	B	2	27	278-404	0.188	KX758636
Vesp17	AGTG	F: CTAACTTTAGAATGCACTGCAGCAGAAGG R: ATAGTGAGCCTACTACACTGCTGACCTGC	B	1	17	391-445	0.027	KX758635
Vesp19	TTG	F: CTTAGGGTGCGTCTGTTACCCATTACG R: CCATACGCTTAGCGAAACTTCATTCTACG	B	1	10	330-354	0.122	KX758639
Vesp22	ATAC	F: CTAGTATGTGTGATCCTGATATTGTACTGC R: GTTATTGCTATGTTATTACCCTGAGGTGG	B	1	19	228-286	0.273	KX758640
Vesp36	ATG	F: GGCCACGGACACTAACAGAAAATGG R: TGGAGTTACGAAAGAATCTCACTTTGTTGG	-	1	6	110-131	0.127	KX758644
Vesp9	TCC	F: ACCATCACTTCCTCCACCTCCC R: TCAGTCAAA GCAAAACCTAGACTGAGGG	-	1	12	273-324	0.192	KX758637

NA: number of alleles per locus; MP: multiplex; L: tail sequence/ florescent label combination (1: 6FAM-GCCTCCCTCGCGCCA; 2: HEX-GCCTTGCCAGCCCGC); bp: base pairs; N: frequency of null alleles.

Table S2: Cumulative average null allele frequency (from lowest to highest individual locus null allele frequency), with global F_{ST} uncorrected for null alleles, and corrected for null alleles using the ENA method (with 1000 bootstrap replicates).

No. loci →	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Global F_{ST}	0.092	0.163	0.124	0.142	0.122	0.105	0.102	0.105	0.101	0.095	0.087	0.087
Global F_{ST} (corrected)	0.091	0.159	0.121	0.140	0.121	0.105	0.099	0.101	0.097	0.093	0.086	0.085
Mean null allele frequency	0.012	0.019	0.022	0.026	0.045	0.059	0.077	0.092	0.104	0.117	0.129	0.141

Table S4: Probability of departure from Hardy-Weinberg Equilibrium for each site and locus (*p* values corrected for multiple tests using Benjamini-Yekutieli correction, significant (<0.05) values in bold)

	Vesp15	Vesp23	Vesp35	Vesp3	Vesp27	Vesp1	Vesp30	Vesp17	Vesp19	Vesp22	Vesp36	Vesp9
PK	1.000	0.075	1.000	-	0.102	0.075	0.062	1.000	0.156	1.000	1.000	**0.035**
SCB	1.000	0.953	**0.048**	1.000	0.109	1.000	**0.019**	1.000	1.000	**0.000**	**0.000**	**0.000**
SB	1.000	0.425	1.000	1.000	**0.000**	**0.000**	1.000	1.000	**0.035**	**0.000**	1.000	**0.000**
FK	1.000	**0.000**	1.000	0.418	**0.000**	**0.000**	1.000	1.000	0.109	**0.000**	1.000	**0.048**
LKB	1.000	**0.000**	0.191	1.000	**0.000**	**0.000**	**0.000**	1.000	0.967	**0.000**	**0.000**	**0.000**
GKB	1.000	**0.048**	1.000	1.000	**0.000**	**0.000**	0.197	1.000	0.167	**0.000**	1.000	1.000
BK	0.833	**0.019**	1.000	1.000	**0.000**	**0.048**	0.965	1.000	**0.000**	**0.000**	1.000	**0.000**
CKA	1.000	**0.019**	1.000	1.000	**0.000**	**0.000**	0.109	0.145	0.334	**0.000**	0.647	0.109
LKA	1.000	**0.035**	1.000	1.000	**0.000**	**0.000**	0.220	1.000	0.487	**0.000**	**0.019**	**0.035**
KC	1.000	0.134	1.000	1.000	**0.000**	**0.000**	0.302	1.000	1.000	**0.000**	1.000	**0.000**
LC	1.000	0.197	1.000	1.000	**0.000**	**0.000**	**0.035**	1.000	1.000	**0.000**	0.102	1.000
WK	1.000	**0.000**	1.000	1.000	0.230	**0.000**	**0.019**	0.593	**0.000**	0.197	0.302	**0.000**
BC	1.000	0.609	1.000	1.000	**0.000**	0.145	0.387	1.000	0.109	**0.000**	1.000	0.062
LP	1.000	**0.048**	1.000	1.000	**0.000**	**0.000**	**0.000**	1.000	1.000	**0.000**	**0.000**	0.075
BH	1.000	**0.035**	1.000	1.000	0.062	1.000	**0.000**	1.000	1.000	**0.000**	1.000	1.000
BAR	1.000	0.251	1.000	1.000	0.109	**0.000**	**0.000**	1.000	0.284	**0.019**	**0.019**	1.000
BZ	1.000	1.000	1.000	1.000	**0.000**	1.000	**0.000**	1.000	1.000	**0.019**	1.000	1.000