

Microsatellite Characterization and Marker Development from Public EST and WGS Databases in the Reef-Building Coral *Acropora millepora* (Cnidaria, Anthozoa, Scleractinia)

SHI WANG*, LINGLING ZHANG*, AND MIKHAIL MATZ

From the Section of Integrative Biology, School of Biological Sciences, University of Texas at Austin, 1 University Station C0930, Austin, TX 78712.

*These authors contributed equally to this work.

Address correspondence to S. Wang at the address above, or e-mail: swang@mail.utexas.edu.

Abstract

Mining for microsatellites (also called simple sequence repeats [SSRs]) in public sequence databases of a common Indo-Pacific coral *Acropora millepora* identified 191 SSRs from 10 258 expressed sequence tag (EST) and 618 SSRs from 14 625 whole-genome shotgun (WGS) sequences. In contrast to other animals, trinucleotide repeats, rather than dinucleotide repeats, are dominant in the WGS-SSRs, and AAT is the most frequent trinucleotide motif in EST-SSRs. We successfully developed 40 polymorphic markers from EST-SSRs and WGS-SSRs. Both EST- and WGS-SSRs show high levels of polymorphism within corals from the same reef patch. Interestingly, markers *WGS079* and *WGS227* revealed SSR duplications in a few individuals, suggesting recent duplication events. Genotypic linkage disequilibrium was identified in 5 pairs of SSR markers, which will be invaluable for high-resolution studies of genetic admixture in natural populations of *A. millepora*. Transferability analysis showed that 25 of these markers can be successfully amplified in one of the most ubiquitous Indo-Pacific corals *Acropora hyacinthus*. The marker collection reported here is the largest ever developed for any reef-building coral. It holds great potential for addressing coral reef connectivity across the Indo-Pacific with an unprecedented precision, especially taking into account the cross-species transferability of a substantial number of markers.

Key words: connectivity, genome mapping, linkage disequilibrium, locus duplication, population genetics, transferability

Coral reefs are among the most productive and diverse ecosystems on Earth. Unfortunately, they have suffered long-term decline due to a range of anthropogenic disturbances and are now also under threat from climate change (Hughes et al. 2003). Understanding the pathways of coral connectivity and mechanisms of adaptation is therefore critically important to devise scientifically sound management and remediation practices (Palumbi 2003; Van Oppen and Gates 2006). Multilocus genotyping is currently the most powerful methodology to address connectivity (Pritchard et al. 2000). To date, most coral genetic connectivity studies have employed only a few loci for multilocus genotyping (Baums et al. 2006; Van Oppen and Gates 2006; Carlon and Lippe 2008). Utilization of more

than 20 loci would make the analysis dramatically more precise in detecting individual migrants, recent tendencies in population differentiation, or older admixture events (Pritchard et al. 2000). Addition of the genome mapping information for the genotyped loci during connectivity studies would provide substantial extra power in detecting migration and population admixture by making it possible to track individual haplotypes (chromosome regions) across populations (Falush et al. 2003). A high-quality linkage map will also serve as a basis for genome-wide association studies of coral physiological variation, which is the focus of our ongoing project aiming at molecular and evolutionary mechanisms of coral adaptation to climate change. Genomic mapping is not possible with only a few markers that are

known at the moment. Even in the well-studied coral *Acropora millepora*, only 10 microsatellite markers have been developed thus far based on a genomic DNA library (Van Oppen et al. 2007).

Microsatellites or simple sequence repeats (SSRs) are small tandem repeated sequences that are widely dispersed in eukaryotic genomes (Toth et al. 2000). Currently, SSRs remain the markers of choice for genetic analyses in most nonmodel organisms due to their codominant and highly polymorphic nature. The traditional way of developing SSR markers is costly and time consuming (Zane et al. 2002; Squirrell et al. 2003). However, with the advent of the genomics age, data mining of public DNA databases has been proved as an alternative way to get large amounts of SSR markers efficiently (Serapion et al. 2004; Ellis and Burke 2007; Sharma et al. 2007). *Acropora millepora* is an emerging coral molecular biology model for molecular ecology studies focusing on the evolutionary adaptation of corals to climate change. As of January 2008, there were both expressed sequence tag (EST) (10 258 entries) and whole-genome shotgun (WGS) (14 625 entries) sequences of *A. millepora* available in the National Center for Biotechnology Information (NCBI) database. Recently, a large body of sequence data obtained by sequencing the *A. millepora* larval transcriptome using 454 technology was made available by our laboratory, (http://www.bio.utexas.edu/research/matz_lab/matzlab/454.html). These database resources are valuable for microsatellite mining. Investigation and comparison of SSRs distribution and constitution in the genome can provide a glimpse into genome evolution of *A. millepora*. In this study, we focused on the following: 1) identification and characterization of SSRs in EST and WGS sequences of *A. millepora*, 2) marker development from identified EST and WGS-SSRs, 3) marker polymorphism evaluation and genotypic linkage disequilibrium (LD) detection, and 4) marker transferability to *Acropora hyacinthus* (Dana 1846).

Materials and Methods

SSR Mining from EST and WGS Sequences

EST and WGS sequences of *A. millepora* were downloaded from NCBI databases and then analyzed with the SSR EXPERT 1.0 program (Zhan et al. 2005) to identify sequences containing SSRs. The search was conducted for sequences that showed more than 7 repetitions for dinucleotides, 5 repetitions for tri- and tetranucleotides, and 4 repetitions for penta- and hexanucleotides. EST and WGS sequences that contained SSRs were then vector trimmed and clustered using the SEQMAN program, part of the DNASTAR software package (DNASTAR, Madison, WI).

DNA Extraction and Whole-Genome Amplification

Tissue samples from 24 *A. millepora* colonies were collected in November 2007 at Magnetic Island (MI) in the Great Barrier Reef, Australia. To inspect the applicability of our markers in other populations, 7 more colonies were tested in

this study, of which 4 came from Orpheus and Pelorus Islands (OPI) and 3 from Keppels Islands (KI), the Great Barrier Reef locations separated by more than 650 km. The tissue from a small coral branch was removed using an airbrush, fixed in RNAlater solution (Ambion, Foster City, CA), and stored at -20°C . Genomic DNA was extracted by a phenol/chloroform extraction method (Sambrook et al. 1989). In order to make sufficient DNA templates for multiple polymerase chain reaction (PCR) amplifications, we used REPLI-g UltraFast Mini Kit (Qiagen, Valencia, CA) for whole-genome amplification (WGA) of each sample. The average product length was typically greater than 10 kb and could therefore be reliably used for downstream SSR analysis (Paez et al. 2004).

Primer Design, PCR Amplification, Fragment Analysis, and Estimation of Error Rate

EST-SSRs and WGS-SSRs with sufficient flanking sequences were chosen for primer design. Primers were designed based on several principles as described by Matz (2003) so that all PCR amplifications could be achieved at the same annealing temperature. Forward primers were 5'-end fluorescently labeled with 6 carboxyfluorescein (FAM) or hexachlorofluorescein (HEX). PCR amplifications were set up in a 10- μl volume composed of ~ 30 ng DNA template (equivalent to 0.12 μl of WGA product), 0.1 μM each primer, 2 mM MgCl_2 , 0.3 mM deoxynucleoside triphosphate, 1 \times PCR buffer, and 0.5 U *Taq* DNA polymerase (NEB, Ipswich, MA) in a DNA Engine Tetrad 2 Thermal Cycler (Bio-Rad, Hercules, CA). All cycling began with an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 40 s, 60°C for 30 s, and 72°C for 30 s. Two microliters of PCR product was run on a 0.5 \times Tris-Boric Acid-EDTA and 1.5% agarose gel to determine the success of the PCR. PCR product was diluted 1:45 to 1:150 and run on the ABI 3130XL capillary sequencer along with the rhodamine-labeled size standard. In order to estimate possible allele scoring errors, we repeated our experiment in 4 *A. millepora* individuals from DNA extraction to allele scoring for all polymorphic markers, and error rate was then calculated by the number of incorrect alleles divided by the total number of alleles (Hoffman and Amos 2005; Selkoe and Toonen 2006).

Evaluation of SSR Markers

All genetic statistics were carried out based on the genotyping data from MI population. For each marker, the number of alleles, observed heterozygosity (H_o), expected heterozygosity (H_e) (Nei 1978), and neutrality test were calculated using the POPGENE program (Yeh and Boyle 1997). Hardy-Weinberg equilibrium (HWE), heterozygote deficiency (HD), and genotypic LD among SSR markers were detected using the GENEPOP 4.0 program (Raymond and Rousset 1995), which performed exact tests for HWE and a log-likelihood ratio statistic (G test) for genotypic LD detection. For HWE, HD, and Genotypic LD tests, Bonferroni correction was applied in order to avoid type I error.

Marker Transferability to *A. hyacinthus*

All 40 SSR markers were tested in 20 individuals of *A. hyacinthus* for transferability. Two specimens of *A. hyacinthus* were acquired from private aquaria in Austin, TX, and 18 were kindly provided by Thomas A. Oliver (Stanford University, CA), which had been collected from 2 lagoon pools at Ofu Island, AS. Genomic DNA of these individuals was extracted and amplified as described above.

Results

EST- and WGS-SSRs in *A. millepora*

In total, 10 258 EST sequences and 14 625 WGS sequences of *A. millepora* that were available in NCBI database were included in the analysis. The SSR EXPERT program identified 291 SSRs in 256 of EST sequences, as well as 652 SSRs in 533 WGS sequences. Fifty-six EST sequences containing AG repeats were discarded because they were likely derived from the oligo(dT) linker primer, which was used in the cDNA library construction (Technau et al. 2005). After cluster analysis, there remained 191 SSRs in 157 EST contigs and 618 SSRs in 499 WGS contigs. As shown in Table 1, trinucleotide repeats were the most abundant type in both EST- and WGS-SSRs, accounting for 66.0% of EST-SSRs and 57.9% of WGS-SSRs, respectively. AAT was the most frequent trinucleotide motif in both sequence types. Dinucleotide repeats were the second most abundant type in both kinds of SSRs, accounting for 21.5% of EST-SSRs and 23.1% WGS-SSRs, respectively. AG was the most frequent dinucleotide motif in EST-SSRs and AT in WGS-SSRs. Distributions among tetra-, penta-, and hexanucleotide repeat types were different in EST and WGS collections. EST-SSRs contained more hexanucleotide repeats but less tetra- and pentanucleotide repeats than WGS-SSRs. TTAGGG was the most frequent hexanucleotide motif in WGS-SSRs, which had been recognized as telomere repeat motif in Cnidaria (Traut et al. 2007).

PCR Amplification, Polymorphism, and Error Rate

In total, 29 EST-SSRs and 46 WGS-SSRs were chosen for primer design. Twenty (69.0%) EST-SSRs and 22 (47.8%) WGS-SSRs could be amplified successfully in the MI population (Table 2) as well as OPI and KI populations, and the others were discarded due to high PCR failure rate. One EST-SSR and one WGS-SSR were monomorphic and were therefore not included in Table 2. Across all individuals and polymorphic 40 loci, there were 5 failed amplifications in the MI population and 4 in the OPI and KI populations. Among 40 loci, PCR failure was found in 6 WGS-SSRs and 1 EST-SSR (*WGS005*, *WGS107*, *WGS116*, *WGS145*, *WGS205*, and *WGS229* and *EST171*). Both EST- and WGS-SSRs showed high levels of polymorphism in MI population. The average number of alleles was 7.7 for EST-SSRs, ranging from 2 to 16, and 9 for WGS-SSRs, ranging from 5 to 18. The average H_e was 0.66 for EST-SSRs,

ranging from 0.22 to 0.94. The average H_e for WGS-SSRs was 0.77, ranging from 0.57 to 0.93. High levels of polymorphism did not correlate with particular repeat types, but there was an observed tendency of higher repeat numbers that show higher polymorphism. Interestingly, a few individuals showed 3 or 4 alleles in *WGS079* and *WGS227*, which implies possible locus duplication events. In the error-rate estimation experiment, wrong scoring was only found in 2 alleles across all 40 markers, and the error rate was then estimated as 0.625%.

Neutrality, HWE, HD, and LD Tests

All EST-SSRs and WGS-SSRs passed the neutrality test at the $\alpha = 0.05$ level. For the HWE test, only 3 loci (*EST043*, *WGS005*, and *WGS145*) were not in accord with HW proportions after Bonferroni correction ($P < 0.01/40$). For the HD test, only 6 loci (*EST043*, *EST164*, *WGS005*, *WGS035*, *WGS051*, and *WGS145*) were significant after Bonferroni correction ($P < 0.01/40$). Highly significant LD ($P < 10^{-6}$) was identified for 5 pairs of markers *EST098* and *EST014*, *EST171* and *EST032*, *EST171* and *EST196*, *WGS092* and *EST122*, and *WGS152* and *WGS005*. Because all our markers were developed from independent EST or WGS contigs, the LD detected between them cannot be attributed to their colocalization within the same contig.

SSR Transferability

Twenty-five of the 40 primer pairs resulted in successful amplification in more than 19 individuals. Among these 25 loci, there were only 4 failed amplifications across all the individuals, and these failed amplifications were found in 3 WGS-SSRs (*WGS134*, *WGS152*, and *WGS211*) and 1 EST-SSR (*EST007*). For most transferable markers, allele size distribution was similar between *A. hyacinthus* and *A. millepora*. Of these 25 primer pairs, 15 came from EST-SSRs and 10 came from WGS-SSRs. Many of the amplification products corresponded to different alleles than in *A. millepora*. At some loci, the amplification of certain alleles was considerably weaker than others, which may be due to mild cases of null allelism (Wang and Guo 2007). For *WGS153*, possible SSR duplication was also found in several individuals.

Discussion

Characterization of EST- and WGS-SSRs in *A. millepora*

Microsatellites are thought to play a significant role in genome evolution by creating and maintaining quantitative genetic variation (Tautz et al. 1986; Kashi et al. 1997). It has been shown that different taxa exhibit different preferences for SSR repeat types (Beckmann and Weber 1992; Lagercrantz et al. 1993; Toth et al. 2000). *Acropora millepora* is a member of the basal phylum Cnidaria, which is regarded as the sister group to the Bilateria and is thought to be critical to understand the evolution of metazoan genetic and developmental complexity (Kortschak et al. 2003; Technau et al. 2005). Toth et al. (2000) had surveyed and analyzed

Table 1. Characterization of SSRs in EST and WGS sequences of *Acropora millepora*

Dinucleotide		Trinucleotide		Tetranucleotide		Pentanucleotide		Hexanucleotide											
EST	WGS	EST	WGS	EST	WGS	EST	WGS	EST	WGS										
AG ^a	17 ^b	AT	115	AAT	49	AAT	244	TTTC	1	AAAT	13	AAAAC	1	AAAAC	7	AGACGA	10	TTAGGG	6
AT	14	AG	14	ATG	26	AAC	44	AGAT	1	AGAT	13	ACCCC	1	ATTGT	3	GTCCT	2	TTTTTG	2
AC	9	AC	12	AAC	15	ATG	31			GTAT	9	ATTTG	1	TTTTA	3	TTTTTG	1	AGACGA	1
CG	1	CG	2	AGG	12	AAG	19			TTTG	5	GAAAG	1	TTACA	3	CCACCT	1	GTAACG	1
				AAG	7	ACT	11			CTGA	4			CAAAT	2	ATGCCG	1	ACGGCA	1
				AGC	6	AGC	3			AATC	4			TTTTTC	2	GACCGA	1	ATTTAT	1
				GAC	6	GAC	3			GAAT	3			CAGGT	2	GACAAA	1	TCTATC	1
				ACC	5	ACC	1			ACTT	2			ACTGA	2	TCTTCG	1		
						AGG	1			ATGC	2			TGACA	1				
						CGG	1			CTCA	2			CCTAA	1				
										ACCG	1			TGGAA	1				
										CTGG	1			ATTAA	1				
										TTAA	1			ATATC	1				
										GTAC	1			AATAT	1				
										GTTG	1			TTTAG	1				
										AGGT	1			GGTTT	1				
										CACG	1			TAGTA	1				
										AGCT	1			TACGT	1				
														CATAA	1				
														ATAGA	1				
														CCACG	1				
														GTAGA	1				
														AGGTA	1				
Total	41(21.5%) ^c	143(23.1%)	126(66.0%)	358(57.9%)	2(1.0%)	65(10.5%)	4(2.1%)	39(6.3%)	18(9.4%)	13(2.1%)									

^a Repeat motif type.^b Total number of a given motif type.^c Percentage of a given motif type in EST- or WGS-SSRs.

Table 2. Summary of 40 polymorphic SSR markers in *Acropora millepora*

Locus	Repeat motif	Primer sequence (5'-3') ^a	Size (bp)	Allele					Accession No.	Transferability ^f
				No.	N ^b	H _o ^c	H _e ^d	P ^e		
<i>EST007</i>	(TTTC) ₅	F: FAM-tgcaatggtctgtgcatca R: gatctctttaccgatttacagca	99–107	3	24	0.38	0.47	0.2025	DY587595	✓
<i>EST014</i>	(TCT) ₁₃	F: FAM-cagctccttcatcttcatcct R: agccgaagaggggacagagt	143–173	10	24	0.92	0.88	0.6926	DY586774	✓
<i>EST016</i>	(AAC) ₇	F: FAM-ctatctgtgtatgatcaggacta R: tccatctgttggaaactggt	97–122	7	24	0.67	0.69	0.1383	DY586537	✓
<i>EST032</i>	(TTA) ₂₁	F: FAM-aggcacaagaagtggaaaaaaca R: tgaaggatgtgaagcatggt	138–187	15	24	0.96	0.94	0.7274	DY585386	✓
<i>EST043</i>	(TAT) ₁₀	F: FAM-atcaatcattgttactatggctat R: gtggtagaacatcagtcaga	131–171	13	24	0.25	0.75	0.0000*	DY585059	
<i>EST062</i>	(GAT) ₉	F: FAM-cgagttagtctgttaagatggt R: ctctaagtcctctcttccca	110–126	5	24	0.67	0.71	0.1437	DY584488	✓
<i>EST063</i>	(TC) ₈	F: FAM-tattgtagtcttactaggct R: aacaatcgtgcatactagctca	96–104	5	24	0.29	0.42	0.0116	DY584457	✓
<i>EST097</i>	(TGA) ₇	F: FAM-tgacaacgacatcaatcaggt R: acagcaggagctgtcagcact	123–135	5	24	0.71	0.69	0.0923	DY583334	✓
<i>EST098</i>	(TG) ₁₂	F: FAM-aaaaattgcgctcaagttgatg R: acggctgcgaaggagtctagt	98–118	8	24	0.58	0.68	0.0362	DY583314	✓
<i>EST121</i>	(ATGCCG) ₄	F: FAM-acagttgcagccttgcaga R: gtgggaattgcgacagcagcat	100–112	3	24	0.50	0.51	0.8431	DY582587	✓
<i>EST122</i>	(TTA) ₁₈	F: gttagaatggatattccttcatct R: FAM-ctgtgcctgaaaatcatcatagt	99–141	13	24	0.92	0.91	0.2018	DY582557	✓
<i>EST149</i>	(GAT) ₉	F: FAM-acgtcaaatggattttcacatga R: aggtgcttctcttctcctcaga	118–130	5	24	0.42	0.60	0.0458	DY581661	✓
<i>EST164</i>	(ATAG) ₂₀	F: gcagcttgagatggtgatgta R: FAM-atatcgatgtatctatctctct	147–249	16	24	0.70	0.91	0.0044	DY581172	
<i>EST165</i>	(CAG) ₈	F: gccaaagcaaacagcagcagt R: FAM-ctactcccactctggttctgta	82–106	8	24	0.67	0.57	0.9147	DY581162	
<i>EST171</i>	(GACCGA) ₅	F: gacgagcgaagctaccaata R: FAM-ttcgctcttctggtgctctt	153–189	6	24	0.79	0.77	0.8050	DY581008	
<i>EST181</i>	(ATG) ₁₀	F: FAM-tgattgctgagaagctagagat R: gcctcaccttgccctgtaca	145–157	2	24	0.25	0.22	1.0000	DY580714	✓
<i>EST196</i>	(TAA) ₉	F: gtttggctatctcatgtatagt R: FAM-acaacacatcaacaacagca	117–145	9	24	0.79	0.85	0.0926	DY580091	✓
<i>EST245</i>	(CA) ₁₀	F: FAM-cagaatgatatttctgcagcact R: cgcaatcgagattataggaaga	115–124	5	24	0.21	0.23	0.0870	DY578136	✓
<i>EST254</i>	(CA) ₁₂	F: ggtgaccaatcagagtcttga R: FAM-tacacttgctatagtaactgtct	86–100	8	24	0.75	0.82	0.2802	DY577596	✓
<i>WGS005</i>	(ATT) ₈	F: HEX-ccataattctacgtgacattaca R: gttgctcagataaagtaggagct	185–227	7	22	0.14	0.82	0.0000*	714183505	
<i>WGS035</i>	(GTAT) ₆ (GTTT) ₈	F: HEX-ttgattggtcaatgaagagagta R: tttgtaggtggacaggggtt	162–188	9	24	0.50	0.87	0.0005	714180598	
<i>WGS051</i>	(GATA) ₈	F: gccgaaacttcaactggagca R: HEX-aaacttaactgagacaacacaga	151–216	12	24	0.61	0.86	0.0004	714184394	✓
<i>WGS079</i>	(ATT) ₁₁ N ₆ (ATT) ₅	F: HEX-ccgtacatcgaagaacgctga R: ttagtaggcaccgtccttagt	136–151	6	24	0.55	0.71	0.0825	745001289	
<i>WGS092</i>	(ATT) ₁₂	F: HEX-ctgggcaaatattaccacttga R: aagacaggtatgtatgcaatgat	111–200	18	24	0.79	0.93	0.1645	745002572	✓
<i>WGS101</i>	(TTA) ₁₄ (GTA) ₁₀	F: HEX-agatgtagagctactctgtaga R: gtcagcaaggcaagggtga	134–190	17	24	0.88	0.88	0.3978	745002426	✓
<i>WGS107</i>	(TAA) ₅ N ₃ (TAA) ₉	F: gattctgttcaggatccattct R: HEX-tacatggtcagctgccacgat	152–245	10	23	0.74	0.89	0.0063	745000273	
<i>WGS112</i>	(AAT) ₉	F: HEX-actccactcagctctattacca R: acactccaagagtccttaca	166–184	6	24	0.79	0.73	0.9619	745001340	✓
<i>WGS116</i>	(AT) ₁₄	F: HEX-caattgctggattgggacaga R: tgcactgataggcccgaa	105–135	10	24	0.67	0.88	0.0712	745002493	
<i>WGS131</i>	(AAG) ₁₀	F: HEX-cggtgctgtgatcgctattca R: cgcccttgcaattcagctca	99–119	6	24	0.58	0.65	0.2641	744999704	
<i>WGS134</i>	(GATA) ₆	F: HEX-tgttcggacccaacctgat R: gctgcgcccttcgcaattca	105–133	7	24	0.58	0.67	0.3767	745001492	✓

Table 2. Continued

Locus	Repeat motif	Primer sequence (5'-3') ^a	Size (bp)	Allele					Accession No.	Transferability ^f
				No.	N ^b	H _o ^c	H _e ^d	P ^e		
<i>WGS145</i>	(AAT) ₁₃ N ₃ (AAT) ₉	F: gaaaattaagtcgacctacagta R: HEX-ctctgaaagggtgcgttttcgt	152–205	13	23	0.39	0.92	0.0000*	714181198	
<i>WGS152</i>	(AT) ₉	F: gcctattacaatgcatagcacta R: HEX-cgctgggtcctatctatct	87–110	7	24	0.54	0.76	0.0241	714180564	✓
<i>WGS153</i>	(AATC) ₇	F: HEX-ttccaagttgctgtgagtaga R: cgggtgctaagcttgctcaa	106–126	5	24	0.63	0.64	0.9458	714176682	✓
<i>WGS189</i>	(ATCT) ₇	F: HEX-aaatgagcgcctgtgcacga R: gagcatgaaactctgagtagca	158–194	9	24	0.58	0.75	0.0420	714180544	✓
<i>WGS196</i>	(ATAC) ₆	F: HEX-ttcagtcttcgggctgcaagt R: cctacccggtgtattacacta	128–234	14	24	0.63	0.78	0.0126	745000571	✓
<i>WGS205</i>	(TTG) ₈	F: HEX-tcaaaacttctcatgtgcagaact R: tgtaggggaaggtcggcacgta	118–134	5	23	0.43	0.57	0.0625	714185213	
<i>WGS211</i>	(TAA) ₈	F: HEX-tgacgacgaaacgttgectat R: agaccgtttccttaaccagaa	181–199	5	24	0.75	0.61	0.4021	714178565	✓
<i>WGS217</i>	(ATT) ₉	F: HEX-atccgaaaaagtaaaagtctgcaa R: actatgaattctgtgcagcgaa	160–191	6	24	0.75	0.63	0.7026	745000533	
<i>WGS227</i>	(ATA) ₅ N ₂ (AAT) ₁₂	F: HEX-catttgaggaaaggtgacacat R: cattaagccactattgggtgat	125–158	11	24	0.74	0.83	0.4448	714176770	
<i>WGS229</i>	(TTA) ₉	F: HEX-aggggattaagaccaataacgtgta R: ggtgaaagaccgttgctccat	129–169	6	24	0.54	0.70	0.0192	745001248	

^a “FAM” or “HEX” at the 5'-end of the primer indicate FAM- or HEX-labeled primer.

^b Number of individuals with successful amplification.

^c Observed heterozygosity.

^d Expected heterozygosity.

^e P value from exact tests of HWE (* shows significant departure from HWE after Bonferroni correction, $P < 0.01/40$).

^f Transferability in *Acropora hyacinthus*.

SSRs in exons and noncoding regions in animals such as various vertebrates and the fruitfly. Despite the narrow taxonomic range of this survey, their data provide an opportunity to investigate evolutionary differences of SSRs between *A. millepora* and at least some of the Bilaterian animals. For dinucleotide repeats, although they are relatively redundant in the EST-SSRs of *A. millepora*, most of them seem to be located in 5'- or 3'-untranslated regions (UTRs) according to the BLASTX results (data not shown). This is consistent with the phenomenon that dinucleotide repeats are rarely found in the exons of Bilaterian animals. AT is the most frequent dinucleotide motif in WGS-SSRs of *A. millepora*. Interestingly, this motif is the most common dinucleotide motif in plants but not vertebrates and invertebrates, where AC is the most abundant dinucleotide motif (Lagercrantz et al. 1993; Toth et al. 2000; Cruz et al. 2005). Unlike Bilaterian animals, trinucleotide repeats, rather than dinucleotide repeats, are dominant in the WGS-SSRs in *A. millepora*. Because different SSR identification programs may produce significantly different results in some cases (Leclercq et al. 2007), we rechecked the motif distribution using another program, SSRIT (Temnykh et al. 2001). This program could only find perfect SSRs but reported similar motif distribution under the default search settings (data not shown). So, our SSR distribution results are robust to the choice of SSR mining software. In fact, similar phenomena

were also observed in yeast and other fungi (Kim, Booth, et al. 2008). Although mechanisms of repeat instability are still unclear, it is well known that trinucleotide repeats form the largest component of a broader category of repeat-associated disorders in humans (Pearson et al. 2005). The AAT is the most frequent trinucleotide motif in EST-SSRs of *A. millepora*. Notably, the same motif has been previously found to be the most common in a Caribbean species *Acropora palmata* (Baums, Hughes, and Hellberg 2005). Most of the AAT motifs are located in 3'-UTRs according to the BLASTX results (data not shown). As a contrast, for Bilaterian animals, (G + C)-rich motifs (e.g., CCG and AGC) dominate mainly in both exons and 5'-UTRs (Toth et al. 2000; Wren et al. 2000). It has been shown that AAT motif has a higher slippage rate than 4 other motifs of trinucleotide repeats in yeast (Kruglyak et al. 2000). Such motifs may affect mRNA stability, representing binding sites for translation factors, as has been described for an AU-rich sequence in the 3'-UTR of mRNA for human plasminogen activator inhibitor type 2 (Maurer et al. 1999).

Marker Development and Polymorphism

This is the first time that EST- and WGS-SSRs have been developed in corals. The rate of amplification success in EST-SSRs is higher than that in WGS-SSRs, which may

result from lower levels of sequence variation at the priming sites in EST-SSRs. EST-SSRs have 2 major advantages over WGS-SSRs. First, as EST-SSRs are associated with functional genes, they are more useful for gene mapping and comparative genomic analysis (Vignal et al. 2002). Second, EST-SSRs have higher cross-species transferability than WGS-SSRs (Ellis and Burke 2007). One common concern is that EST-SSRs may be much less polymorphic than genomic SSRs because they are located within genes (Ellis and Burke 2007). However, our data indicate that this is not an issue in *A. millepora* because H_e calculated from EST-SSRs is only slightly lower than that in WGS-SSRs. Indeed, just like WGS-SSRs, all EST-SSRs could pass neutrality tests in this study. Therefore, there is no evidence for strong positive selection on EST-SSRs. A similar phenomenon was also reported in rice (Cho et al. 2000), bread wheat (Gupta et al. 2003), pines (Liewlaksaneeyanawin et al. 2004), barley (Chabane et al. 2005), sunflower (Pashley et al. 2006), and beetles (Kim, Ratcliffe, et al. 2008). Null alleles are nonamplifying alleles, which are often caused by mutations at the priming sites (Pemberton et al. 1995). Because there were only 9 failed amplifications of SSR markers across all the individuals and 40 loci and only 6 loci showed significant homozygote excess, null alleles must exist with relatively low frequency in our 40 loci. This speculation is also supported by the fact that 25 of these loci could be amplified in another species, *A. hyacinthus*. For both *A. millepora* and *A. hyacinthus*, PCR failure was mostly found in WGS-SSRs, which suggests that the frequency of null alleles in WGS-SSRs is slightly higher than that in EST-SSRs and is possibly attributable to higher degree of conservation of sequences flanking EST-SSRs. High levels of polymorphism were revealed by both EST- and WGS-SSRs, which is consistent with a previous report (Van Oppen et al. 2007). Based on current results, we estimate that about 200–300 polymorphic markers can be efficiently developed from EST and WGS sequences. In addition, when searching our 454 data (44 569 contigs assembled together with publicly available *A. millepora* ESTs comprising a total assembled sequence length of 20.4 Mb plus 70 833 singleton reads; http://www.bio.utexas.edu/research/matz_lab/matzlab/454.html) under the same search criteria, we could identify 798 potential SSRs. Therefore, sufficient SSR markers can be generated from these resources for our mapping purpose.

SSR Duplication

In addition to SSR duplication, occurrence of more than 2 alleles per locus in corals can also be explained by fusion of primary polyps during gregarious settlement (Barki et al. 2002) and retention of a polar body during fertilization (Baums, Hughes, and Hellberg 2005). However, those alternatives are unlikely to account for our case because they would result in more than 2 alleles detected at most loci rather than at 2 out of 40. SSR duplication therefore seems to be the most plausible explanation of our data. Duplicated SSRs have shown great potential as selectively neutral markers for studying gene and genome duplication in both

theoretical and experimental studies (Balaesque et al. 2003; David et al. 2003; Antunes et al. 2006; Zhang and Rosenberg 2007). In this study, SSR duplications were only detected in a few individuals, which may reflect recent SSR duplication events. Moreover, it seems that both the original and duplicated loci can be polymorphic (data not shown). Because gene and genome duplications are important mechanisms for evolving genetic novelty (Lynch and Conery 2000; Zhang 2003), discovery of more duplicated SSRs may provide insight into genome evolution of *A. millepora* and therefore helps us gain a better understanding of basic mechanisms responsible for coral speciation.

Genotypic LD among EST- and WGS-SSRs

LD describes a situation in which some combinations of alleles occur more or less frequently in a population than would be expected from a random formation. Markers in LD are particularly valuable for detecting most recent admixture and migration events (Goldstein and Weale 2001). Two or more markers in LD form a haplotype, which persist in an admixed population for a few generations until recombination destroys the marker association. Finding such “unbroken” haplotypes originating elsewhere provides the most conclusive evidence of the recent population admixture (Zhou et al. 2007). Previous study on genetic connectivity of *A. millepora* was based on allozymes (Smith-Keune and van Oppen 2006). Compared with allozymes, SSRs are more polymorphic and can therefore be more powerful for detecting LD in a given population (Baums, Miller, and Hellberg 2005). In this study, genotypic LD was detected in 5 pairs of SSR markers, which will be invaluable for high-resolution studies of genetic admixture in natural populations of *A. millepora*.

SSR Transferability in *A. hyacinthus*

In this study, 78.9% of EST-SSRs and 47.6% of WGS-SSRs could be successfully amplified in *A. hyacinthus*. It is expected that EST-SSRs have a higher transferability than WGS-SSRs. Compared with previous studies (Chagne et al. 2004; Gutierrez et al. 2005), this study showed remarkably high transferability in WGS-SSRs. Because many SSR assays turned out to be transferable between species, it is possible to use sequence resources existing for one or a few corals to develop SSR markers applicable to a much wider range of species. Our results indicate that such an approach will be effective at least within the genus *Acropora*.

Acknowledgments

We thank Dr Aibin Zhan (University of Guelph, Canada) for helpful suggestions in experimental design and manuscript preparation. We thank Bette Willis and David Abrego (James Cook University, Townsville, Australia) for fieldwork support and help with permitting. We are especially grateful to the 2 volunteers, Vetea Liao and Helena LaGrande, who collected and preserved the *Acropora millepora* specimens. We are also grateful to the private aquarists Brian Jackson and Dustin Holden (Austin, Texas) and Thomas Oliver (Stanford University, CA) who provided specimens of *Acropora hyacinthus* and Madeleine Van Oppen (Australian

Institute of Marine Science, Australia) who provided OPI and KI samples of *A. millepora*. We thank Dr Elisha Meyer (University of Texas at Austin, TX) for proofreading the manuscript. We would also like to thank 2 anonymous reviewers for their valuable comments on this paper.

References

- Antunes A, Gharbi K, Alexandrino P, Guyomard R. 2006. Characterization of *transferring*-linked microsatellites in brown trout (*Salmo trutta*) and Atlantic salmon (*Salmo salar*). *Mol Ecol Notes* 6:547–549.
- Balaresque P, Toupance B, Heyer E, Crouau-Roy B. 2003. Evolutionary dynamics of duplicated microsatellites shared by sex chromosomes. *J Mol Evol*. 57(Suppl):S128–S137.
- Barki Y, Gateno D, Graur D, Rinkevich B. 2002. Soft-coral natural chimerism: a window in ontogeny allows the creation of entities comprised of incongruous parts. *Mar Ecol Prog Ser*. 231:91–99.
- Baums IB, Hughes CR, Hellberg MH. 2005. Mendelian microsatellite loci for the Caribbean coral *Acropora palmata*. *Mar Ecol Prog Ser*. 288:115–127.
- Baums IB, Miller MW, Hellberg ME. 2005. Regionally isolated populations of an imperiled Caribbean coral, *Acropora palmata*. *Mol Ecol*. 14:1377–1390.
- Baums IB, Miller MW, Hellberg ME. 2006. Geographic variation in clonal structure in a reef-building Caribbean coral, *Acropora palmata*. *Ecol Monogr*. 76:503–519.
- Beckmann JS, Weber JL. 1992. Survey of human and rat microsatellites. *Genomics* 12:627–631.
- Carlson DB, Lippe C. 2008. Fifteen new microsatellite markers for the reef coral *Favia fragum* and a new *Symbiodinium* microsatellite. *Mol Ecol Res*. 8:870–873.
- Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ. 2005. EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genet Resour Crop Evol*. 52:903–909.
- Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio J-M, Echt C, et al. 2004. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor Appl Genet*. 109:1204–1214.
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet*. 100:713–722.
- Cruz F, Perez M, Presa P. 2005. Distribution and abundance of microsatellites in the genome of bivalves. *Gene*. 346:241–247.
- David L, Blum S, Feldman MW, Lavi U, Hillel J. 2003. Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Mol Biol Evol*. 20:1425–1434.
- Ellis JR, Burke JM. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity* 99:125–132.
- Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
- Goldstein DB, Weale ME. 2001. Population genomics: linkage disequilibrium holds the key. *Curr Biol*. 11:R576–R579.
- Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics* 270:315–323.
- Gutierrez MV, Patto MCV, Huguet T, Cubero JI, Moreno MT, Terres AM. 2005. Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor Appl Genet*. 110:1210–1217.
- Hoffman JJ, Amos W. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Mol Ecol*. 14:599–612.
- Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C, Grosberg R, Hoegh-Guldberg O, Jackson JBC, Kleypas J, et al. 2003. Climate change, human impacts, and the resilience of coral reefs. *Science* 301:929–933.
- Kashi Y, King D, Soller M. 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet*. 13:74–78.
- Kim KS, Ratcliffe ST, French BW, Liu L, Sappington TW. 2008. Utility of EST-derived SSRs as population genetics markers in a beetle. *J Hered*. 99:112–124.
- Kim TS, Booth JG, Gauch HG Jr, Sun Q, Park J, Lee YH, Lee K. 2008. Simple sequence repeats in *Neurospora crassa*: distribution, polymorphism and evolutionary inference. *BMC Genomics* 9:31.
- Kortschak RD, Samuel G, Saint R, Miller DJ. 2003. EST analysis of the Cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr Biol*. 13:2190–2195.
- Kruglyak S, Durrett R, Schug MD, Aquadro CF. 2000. Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Mol Biol Evol*. 17:1210–1219.
- Lagercrantz U, Ellegren H, Andersson L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res*. 21:1111–1115.
- Leclercq S, Rivals E, Jarne P. 2007. Detecting microsatellites within genomes: significant variation among algorithms. *BMC Bioinformatics* 8:125.
- Liewlaksaneeyanawin C, Ritland CE, El-kassaby YA, Ritland K. 2004. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theor Appl Genet*. 109:361–369.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicated genes. *Science* 290:1151–1155.
- Matz MV. 2003. Amplification of representative cDNA pools from microscopic amounts of animal tissue. *Methods Mol Biol*. 221:103–116.
- Maurer F, Tierney M, Medcalf R. 1999. An AU-rich sequence in the 3'-UTR of plasminogen activator inhibitor type (PAI-2) mRNA promotes PAI-2 mRNA decay and provides a binding site for nuclear HuR. *Nucleic Acids Res*. 27:1664–1673.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.
- Paez JG, Lin M, Beroukhi R, Lee JC, Zhao X, Richter DJ, Gabriel S, Herman P, Sasaki H, Altshuler D, et al. 2004. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res*. 9:71.
- Palumbi SR. 2003. Population genetics, demographic connectivity, and the design of marine reserves. *Ecol Appl*. 13:S146–S158.
- Pashley CH, Ellis JR, McCauley DE, Burke JM. 2006. EST databases as a source for molecular markers: lessons from *Helianthus*. *J Hered*. 97:381–388.
- Pearson CE, Edamura KN, Cleary JD. 2005. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet*. 6:729–742.
- Pemberton JM, Slate J, Bancroft DR, Barrett JA. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol Ecol*. 4:249–252.
- 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): population genetics software for exact tests and ecumenicism. *J Hered*. 86:248–249.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Selkoe KA, Toonen RJ. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett*. 9:615–629.

- Serapion J, Kucuktas H, Feng J, Liu Z. 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar Biotechnol.* 6:364–377.
- Sharma PC, Grover A, Kahl G. 2007. Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* 25:490–498.
- Smith-Keune C, van Oppen MJH. 2006. Genetic structure of a reef-building coral from thermally distinct environments on the Great Barrier Reef. *Coral Reefs.* 25:493–502.
- Squirrel 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.
- Tautz D, Trick M, Dover G. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322:652–656.
- Technau U, Rudd S, Maxwell P, Gordon PMK, Saina M, Grasso LC, Hayward DC, Sensen CW, Saint R, Holstein TW, et al. 2005. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet.* 21:633–639.
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome res.* 11:1441–1452.
- Toth G, Gaspari Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10:967–981.
- Traut W, Szczepanowski M, Vitkova M, Opitz C, Marec F, Zrzavy J. 2007. The telomere repeat motif of basal Metazoa. *Chromosome Res.* 15:371–382.
- Van Oppen MJH, Gates RD. 2006. Conservation genetics and the resilience of reef-building corals. *Mol Ecol.* 15:3863–3883.
- Van Oppen MJH, Underwood JN, Muirhead AN, Peplow L. 2007. Ten microsatellites loci for the reef-building coral *Acropora millepora* (Cnidaria, Scleractinia) from the Great Barrier Reef, Australia. *Mol Ecol Notes* 7:436–438.
- Vignal A, Milan D, SanCristobal M, Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol.* 34:275–305.
- Wang Y, Guo X. 2007. Development and characterization of EST-SSR markers in the eastern oyster *Crassostrea virginica*. *Mar Biotechnol.* 9:500–511.
- Wren JD, Forgacs E, Fondon JW 3rd, Pertsemliadis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR. 2000. Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *Am J Hum Genet.* 67:345–356.
- Yeh FC, Boyle TJB. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg J Bot.* 129:157.
- Zane L, Bargelloni L, Patarnello T. 2002. Strategies for microsatellites isolation: a review. *Mol Ecol.* 11:1–16.
- Zhan AB, Bao ZM, Wang XL, Hu JJ. 2005. Microsatellite markers derived from bay scallop *Argopecten irradians* expressed sequence tags. *Fish Sci.* 71:1341–1346.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.
- Zhang K, Rosenberg NA. 2007. On the genealogy of a duplicated microsatellite. *Genetics* 177:2109–2122.
- Zhou X, Burgess TI, De Beer ZW, Lieutier F, Yart A, Klepzig K, Carnegie A, Portales JM, Wingfield BD, Wingfield MJ. 2007. High intercontinental migration rates and population admixture in the sapstain fungus *Ophiostoma ips*. *Mol Ecol.* 16:89–99.

Received July 24, 2008; Revised October 13, 2008;
Accepted November 4, 2008

Corresponding Editor: Jose Lopez