Received: 14 June 2017 | Revised: 16 July 2017 | Accepted: 25 July 2017

DOI: 10.1111/1755-0998.12705

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

Development of genome- and transcriptome-derived microsatellites in related species of snapping shrimps with highly duplicated genomes

Kaitlyn M. Gaynor[1,2] | Joseph W. Solomon[1] | Stefanie Siller[1] | Linnet Jessell[1] | J. Emmett Duffy[3] | Dustin R. Rubenstein[1,4]

[1]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, USA

[2]Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, CA, USA

[3]Tennenbaum Marine Observatories Network, Smithsonian Institution, Washington, DC, USA

[4]Department of Ornithology, Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA

Correspondence
Dustin Rubenstein, Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, USA.
Email: dr2497@columbia.edu

Funding information
Directorate for Biological Sciences, Grant/Award Number: DEB 92-01566, DEBDEB 98–15785, IBN-0131931, IOS-1121435, IOS-1121716, IOS-1257530, IOS-1439985; US National Science Foundation

Abstract

Molecular markers are powerful tools for studying patterns of relatedness and parentage within populations and for making inferences about social evolution. However, the development of molecular markers for simultaneous study of multiple species presents challenges, particularly when species exhibit genome duplication or polyploidy. We developed microsatellite markers for *Synalpheus* shrimp, a genus in which species exhibit not only great variation in social organization, but also interspecific variation in genome size and partial genome duplication. From the four primary clades within *Synalpheus*, we identified microsatellites in the genomes of four species and in the consensus transcriptome of two species. Ultimately, we designed and tested primers for 143 microsatellite markers across 25 species. Although the majority of markers were disomic, many markers were polysomic for certain species. Surprisingly, we found no relationship between genome size and the number of polysomic markers. As expected, markers developed for a given species amplified better for closely related species than for more distant relatives. Finally, the markers developed from the transcriptome were more likely to work successfully and to be disomic than those developed from the genome, suggesting that consensus transcriptomes are likely to be conserved across species. Our findings suggest that the transcriptome, particularly consensus sequences from multiple species, can be a valuable source of molecular markers for taxa with complex, duplicated genomes.

KEYWORDS
genome duplication, microsatellites, molecular markers, social evolution, *Synalpheus*, transcriptome

1 | INTRODUCTION

Molecular markers are valuable tools for understanding the ecology and evolution of social species (Garant & Kruuk, 2005; Sunnucks, 2000). Highly polymorphic markers such as microsatellites, which consist of repeating base pair motifs that are widespread throughout the genome, allow for the examination of parentage and relatedness within populations and social groups (Ellegren, 2004; Selkoe & Toonen, 2006). Single nucleotide polymorphisms (SNPs) have begun to replace microsatellites in many studies of relatedness and parentage because they are easier to score and becoming cheaper to develop (Morin, Luikart, & Wayne, 2004; Städele & Vigilant, 2016; Vignal, Milan, SanCristobal, & Eggen, 2002; Weinman, Solomon, & Rubenstein, 2014). However, in species with complicated genomes characterized by extensive gene duplication (including polyploidy), the presence of paralogs makes identifying SNPs challenging

(Christensen et al., 2013; Dufresne, Stift, Vergilino, & Mable, 2013). In contrast, identifying polysomic microsatellite markers is easier because multiple alleles can be determined visually in genotyping output (David, Blum, Feldman, Lavi, & Hillel, 2003; Huang, Cipriani, Morgante, & Testolin, 1998). While SNP technology continues to improve (Blischak, Kubatko, & Wolfe, 2016; Waples, Seeb, & Seeb, 2015), microsatellite markers are likely to remain important for studies of parentage and relatedness in species with highly duplicated genomes. Yet, genome duplication still presents significant challenges for the development of disomic microsatellite markers, particularly markers to be used for multiple species simultaneously (Ashley et al., 2003; Buteler, Jarret, & LaBonte, 1999).

Few molecular markers exist for West Atlantic *Synalpheus gamberloides* sponge-dwelling snapping shrimps, which provide a unique opportunity to study the evolution of sociality (Duffy & Macdonald, 2010; Rubenstein, McCleery, & Duffy, 2008). The ~45 species in the *S. gamberloides* group include species that range in social organization from pair-living, to communal breeding (i.e., multiple breeding pairs living together in the same sponge), to eusociality (i.e., one or more queens living together in a sponge with a variable of nonbreeding individuals; Hultgren, Duffy, & Rubenstein, 2017). Attempts to study the role of kin selection in the evolution of such extreme social diversity within snapping shrimp have been hindered by challenges of marker development (Rubenstein et al., 2008). Many species have large genomes with a high degree of partial genome duplication, although they do not appear to be polyploid (Jeffery, Hultgren, Chak, Gregory, & Rubenstein, 2016; Rubenstein et al., 2008). Moreover, genome sizes within some *Synalpheus* species vary by up to 10 Gb, and even individuals from the same sponge can have different genome sizes (Jeffery et al., 2016).

Assessing kin structure across the genus *Synalpheus* and the role of kin selection in the evolutionary transition to eusociality require the development of molecular markers that can be used to quantify parentage and relatedness for multiple species representative of a range of forms of social organization. When possible, adoption of the same molecular approaches across closely related species with different traits can elucidate divergent evolutionary trajectories (Butler, Siletti, Oxley, & Kronauer, 2014). However, the specificity of molecular markers can limit cross-species utility and hinder comparative studies of multiple species simultaneously. Moreover, developing a panel of shared markers in a group where the genomes sizes are large (up to ~20 Gb) but vary more than fivefold across the genus presents additional challenges in identifying shared markers (Baisakh et al., 2009). With partial genome duplication and associated variation in genome size, many markers are not conserved among species, and flanking sequences undergo mutation following duplication events (Guyomarc'h, Sourdille, Edwards, & Bernard, 2002; Whitton & Rieseberg, 1997). Furthermore, a large percentage of markers are likely to be polysomic and exhibit more than two alleles per individual, presenting challenges for conventional analysis of relatedness (Huang, Ritland, Guo, Shattuck, & Li, 2014).

To address the obstacles associated with genome duplication, we used multiple approaches to develop primarily disomic microsatellite markers for 25 species of *Synalpheus gamberloides* shrimp spanning the entire range of social organization from pair-forming species that occasionally form groups, to communal breeding and eusocial species. We developed markers based upon the genomes of four different target species representing the major clades in the genus, as well as a consensus transcriptome from individuals of two different species. We then screened these markers on 21 additional *Synalpheus* species that varied in social organization and genome size. For 16 of these species, we developed complete panels to use for genotyping and validated these markers by calculating relatedness among known parent–offspring pairs. These markers can be used in future studies to examine and compare kin structure of multiple *Synalpheus* species that differ in social system.

As we expected that markers would be conserved among closely related species within the genus, we predicted that markers developed for a given species would have the highest amplification success among other species in its clade (Primmer, Moller, & Ellegren, 1997). We also expected to find a positive relationship between genome size and the proportion of polysomic markers for a given species, given that both should arise from genome duplication. Finally, because we did not know whether the partial genome duplication in *Synalpheus* was in the coding or noncoding regions (or both; Toth, 2000; Chistiakov, Hellemans, & Volckaert, 2006), we compared the utility of marker development from genomic and transcriptomic data. We expected that markers developed from the transcriptome would be more likely to work across species than those developed from the genome, given that the transcriptome should be more conserved (Boscari et al., 2015; Postolache et al., 2013).

2 | MATERIALS AND METHODS

2.1 | Genomic marker development

We developed genomic microsatellite markers for three target species from different clades encompassing all of the social species within the *Synalpheus gamberloides* species group: *S. dardeaui* ("SD" markers; Longicarpus clade), *S. belizensis* (formerly *S. paraneptunus*; "SP" markers; Paraneptunus clade) and *S. regalis* ("SR" markers; Rathbunae clade). We also used markers that had been previously developed for *S. brooksi* ("ZB" markers; Brooksi clade; Rubenstein et al., 2008), for a total of four sets of genomic markers.

Microsatellites were identified using pooled DNA from seven to 10 individuals per species, which were extracted from ethanol-preserved pereopods using Qiagen DNAeasy Tissue Kits (Qiagen, Hilden, Germany). As described in Rubenstein et al. (2008), we used a universal SNX linker ligation enrichment protocol (Hamilton, Pincus, Di Fiore, & Fleischer, 1999) with modifications (Barnett, Stenzler, Ruiz-Gutierrez, Bogdanowicz, & Lovette, 2008; Grant & Bogdanowicz, 2006). Enriched fragments were made double-stranded by polymerase chain reaction (PCR), digested with NheI, and ligated into XbaI-digested, dephosphorylated pUC19 (New England BioLabs). We used streptavidin beads to capture digested gDNA that was probed with single-stranded biotinylated dimeric, trimeric and tetrameric

repeats (dimers: GT_8, $TC_{9.5}$; trimers: TTA_{12}, GAT_7, $GTT_{6.33}$, $GTA_{8.33}$, TTC_7, GCT_8, $GTG_{4.67}$, GTC_{12}, TCC_5; tetramers: $TTTA_{8.5}$, $GAAT_{5.5}$, $GATA_7$, $GATT_{5.5}$, $GTAT_{6.25}$, $GTTA_{6.25}$, $GTTT_{8.5}$, $TTAC_{6.75}$, $TTTC_6$, $GATG_{4.25}$, $GGTT_4$, $GCTT_{3.75}$, $GTAG_{4.5}$, $GTCA_{4.25}$, $GTCT_{4.25}$).

We then amplified, ligated and sequenced fragments and multiplexed each species library using the Roche 454 Titanium system (454 Life Sciences, Roche) at the Cornell University Evolutionary Genetics Core Facility. We assembled raw sequence data from *S. belizensis* and *S. regalis* libraries using SEQMAN NGEN v2.0.0 build 29 (DNASTAR, Madison, WI), with average quality of 30, and minimum match percentage of 85. We reassembled raw sequence data from *S. dardaeui* data using SEQMAN PRO v8.1 (DNASTAR) and screened for repeats with MSATCOMMANDER v1.03 (Faircloth, 2008). Ultimately, we produced 1,820 contigs for *S. dardeaui*, 2,686 for *S. belizensis* and 2,125 for *S. regalis*.

To confirm and identify tandem repeats in *S. dardeaui*, *S. belizensis* and *S. regalis* sequences, we used the PHOBOS v3.3.11 (Mayer, 2006) bundled with GENEIOUS v6.05 (Biomatters Ltd.; Kearse et al., 2012). We then used Geneious to design primers. Following Rubenstein et al. (2008), we used dye-labelled forward primers and a universal fluorescent tag labelling protocol (Schuelke, 2000; Waldbieser, Quiniou, & Karsi, 2003). We added a 20-bp forward primer tag (5′-CGAGTTTTCCCAGTCACGAC) to the locus-specific sequence, a universal primer with the same 20-bp sequence fluorescently labelled (6-FAM, NED, PET or VIC) and a reverse primer specific to each locus with the addition of a "pigtail" sequence to reduce stutter (5′GTTTCT) (Brownstein, Carpten, & Smith, 1996).

2.2 | Transcriptomic marker development

We developed transcriptomic microsatellite markers by aligning *S. elizabethae* sequences to a de novo *S. brooksi* transcriptome assembly (see Appendix S1 for software parameters for transcriptome alignment). We extracted total RNA using RNeasy Micro Kits (Qiagen) from the head tissue of six *S. elizabethae* and 16 *S. brooksi* samples. We sequenced *S. elizabethae* RNA in two 2 × 101 bp PE lanes using an Illumina TruSeq mRNA Library Prep Kit on the HiSeq 2000 (Illumina) at the Genomics Core Facility, Huntsman Cancer Institute, University of Utah. For *S. brooksi*, we sequenced TruSeq mRNA libraries in one 2 × 100 bp PE lane on the HiSeq 2500, Genomic Sequencing and Analysis Facility, University of Texas at Austin.

We adapter- and quality-trimmed raw FASTQ sequences using TRIMMOMATIC v0.32 (Bolger, Lohse, & Usadel, 2014) and performed a de novo transcriptome assembly on the XSEDE PSC Blacklight supercomputer using Trinity r2014-07-17 (Grabherr et al., 2011; Haas et al., 2013). We then aligned the Trimmomatic trimmed *S. elizabethae* fasta reads to the *S. brooksi* Trinity transcriptome assembly using BOWTIE2 v2.2.3 (Langmead & Salzberg, 2012) and derived 4,972 mapped consensus contigs for tandem repeat calls using PHOBOS v3.3.11. We designed primers in GENEIOUS v8.1 (Kearse et al., 2012), and we used the same 20-bp forward primer tag and universal primer protocol described for genomic loci above.

2.3 | PCR optimization

For each locus, we identified the optimal annealing temperature and $MgCl_2$ concentration for the PCR using DNA from the target species. PCR consisted of 0.5–1 U JumpStart *Taq* DNA polymerase (Sigma), 0.8 μl 1× PCR buffer (Sigma), 1–4 mm $MgCl_2$ (Sigma), 0.2 mm dNTPs (Invitrogen), 0.15 μm of forward, universal, and reverse primers (Integrated DNA Technologies), and approximately 5–50 ng of genomic DNA. Cycling conditions were as follows: one cycle of 94°C for 1 min; 35 cycles of 94°C for 50 s, *Ta* for 1 min, 72°C for 1 min; and a final extension of 72°C for 4 min 30 s. We then ran PCR product on agarose gel and identified the temperature and $MgCl_2$ concentration that produced the clearest bands in the expected product size range. If PCR consistently failed across temperature and $MgCl_2$ gradients, we did not test the marker further for the target species or for any other *Synalpheus* species.

2.4 | Marker screening

We first screened markers for allele variability in the target species for which the marker was developed using DNA from at least 16 individuals from three to five unique colonies. To genotype these samples, we loaded PCR product onto HiDi™ formamide plates with GeneScan™ 500 LIZ™ dye size standard and ran plates on an ABI 3730XL DNA Analyzer (Applied Biosystems) at the Sackler Institute for Comparative Genomics at the American Museum of Natural History. We scored alleles using the microsatellite plug-in of GENEIOUS v6.0 through v8.1 (Kearse et al., 2012). We then optimized annealing temperatures and $MgCl_2$ and conducted variability screens on all 25 species (Table S1). First, we ran PCRs with a temperature and $MgCl_2$ gradient on at least two individual samples per species, preferentially from different geographic locations. If bands were visible on agarose gels, we ran variability screens on eight individuals, as described above for each target species.

Based upon the results from the optimizations and variability screens, we classified each marker as follows: disomic (one to two alleles per individual); polysomic (at least one individual with ≥3 alleles); homozygous (only one allele per individual); or failed during PCR (no clear alleles). A few markers were later reclassified during genotyping if genotyping revealed individuals with more alleles than in the variability screen (e.g., some disomic markers turned out to be polysomic in genotyping). To minimize cost and labour, not all markers were tested on all species. We preferentially tested markers that were variable and easy to score for the target species. Markers that repeatedly failed for the first round of nontarget species were not tested on all species. For a given species, we continued testing markers until we developed enough markers to proceed with genotyping or abandoned genotyping efforts due to repeated failures or sample limitations.

2.5 | Genotyping

Ultimately, there were 17 of 25 species for which we had a sufficient number of samples and developed enough variable, disomic

markers to genotype (Table S2). For ease of scoring and analysis, we used only disomic markers for genotyping (with the exception of *S. duffyi*, for which most markers were polysomic, with up to four alleles per locus). We excluded some disomic markers from genotyping due to low variability, null alleles or difficulty in scoring alleles (Selkoe & Toonen, 2006). We arranged final genotyping into multiple genotyping panels for each species according to dye type and product size to avoid cross-loci interference (Table 1). We used DNeasy Tissue Kits (Qiagen) to extract a plate of DNA from adults taken from at least three sponge colonies, and queens with eggs to generate relatedness values and confirm marker reliability (Table S2). We used an expected mother–egg relatedness value of 0.5 to confirm that our panels of markers were working reliably for a given species, and genotyping followed the protocol described above.

2.6 | Relatedness analysis

We analysed the genotypes for each species using CERVUS v3.0.3 (Kalinowski, Taper, & Marshall, 2007) to derive loci and allele statistics (i.e., number of loci used, total number of alleles for each locus, allele size range for each locus, observed and expected heterozygosity). We then used STORM v2.0 (Frasier, 2008) to calculate internal relatedness (IR) and homozygosity by loci (HL) for all individuals, which we then averaged for each colony. For *S. duffyi*, which required using some polysomic markers, we used GENODIVE v2.0b27 (Meirmans & van Tienderen, 2004) to calculate observed heterozygosity and SPAGEDI v1.3a (Hardy & Vekemans, 2002) to calculate expected heterozygosity. We then used the Method of Moments estimator in POLYRELATEDNESS v1.5 (Huang et al., 2014) to generate pairwise relatedness values, and as an internal check for allele value validity across loci. We used the algorithm from Huang et al. (2014) because it could also analyse polysomic alleles (for *S. duffyi* only), and because it produced consistent values for disomic queen vs. egg comparisons. Ultimately, we confirmed that the average relatedness of each queen to her eggs was ~0.5 for all species (Table S2).

2.7 | Data analysis

To evaluate potential biases introduced by uneven testing of markers across species, we used generalized least squares linear models (GLMs) with phylogenetic Brownian correlation structure to assess the relationship between the percentage of markers tested for a given species and the percentage of markers working and the percentage of polysomic markers, using the MCMCGLMM and APE packages in R v3.3.2 (Hadfield, 2010; Jeffery et al., 2016; Paradis, Claude, & Strimmer, 2004). For all comparative analyses, we used a previously published phylogeny created using three loci (the mitochondrial 16S rRNA locus (16S), the 5′ barcoding end of the mitochondrial cytochrome oxidase I gene (COI), and a region of the 18S nuclear large ribosomal subunit (18S)) and 33 morphological characters (Chak, Duffy, & Hultgren, 2017; Hultgren, Hurt, & Anker, 2014). We also performed an ANOVA in the R program to compare number of markers tested across the four *Synalpheus* clades (R Core Team 2016). Here and in all further analyses, we applied an arcsine square root transformation to all percentage data to meet assumptions of normality.

2.7.1 | Marker retention across species

Each of the 25 species was assigned to one of four clades, based upon phylogenetic grouping, with the exception of *S. agelas*, the most distant congener (Table 2; Hultgren & Duffy, 2011; Chak et al., 2017). We then examined the effect of phylogenetic relatedness on the percentage of markers working (homozygous, disomic or polysomic) of all markers tested for a given species, based upon its relationship to the target species for which the marker was developed. We performed Welch's two-sample *t* tests to compare the success of in-clade markers to success of outside-clade markers for a given species. For each clade, we used *t* tests to compare the success of markers developed for its target species on in-clade nontarget species relative to (i) the success of markers developed for outside-clade target species and (ii) the success of that clade's target species' markers on outside-clade nontarget species. We excluded target species (*S. belizensis*, *S. brooksi*, *S. dardeaui*, *S. regalis*) from these analyses. Additionally, we used only genome-derived markers in this analysis, as the transcriptome-derived markers were based upon a consensus alignment from species from two different clades. For markers used in genotyping, we compared the number of alleles and the observed and expected heterozygosity in target and nontarget species.

2.7.2 | Polysomic markers and genome size

For each species, we calculated the percentage of polysomic markers of all working markers (as described above). To assess the relationship between genome size and the percentage of polysomic loci, we built generalized least squares linear models with phylogenetic Brownian correlation structure using the MCMCGLMM and APE packages in R (Hadfield, 2010; Jeffery et al., 2016; Paradis et al., 2004). We excluded transcriptome-derived markers from this analysis because they were tested only on 12 of 25 species and had overall low percentages of polysomic loci.

2.7.3 | Genome vs. transcriptome

We used Welch's two-sample *t* tests to compare the percentage of genome- vs. transcriptome-derived markers that worked for a given species (as described above). We also used a paired *t* test to compare genomic vs. transcriptomic marker success for only the 11 species for which we tested both types of markers. Finally, we used *t* tests to compare the percentage of genomic vs. transcriptomic markers that were polysomic.

3 | RESULTS

We developed and tested a total of 143 microsatellite markers on 25 *Synalpheus* species (exclusive of markers that did not work for

TABLE 1 Markers for genotyping. Final disomic microsatellite loci for 15 species of *Synalpheus* snapping shrimp, used in genotyping

Locus	Species	*Ta* (°C)	MgCl2 (mM)	Product size range (bp)	No. of alleles per locus	*N* (individuals)	H_O	H_E
SEB-568	*S. agelas*	53	0.8	203–213	9	92	0.39	0.58
SEB-915	*S. agelas*	52	0.8	228–239	8	89	0.29	0.64
SEB-929	*S. agelas*	52	0.8	217–236	11	92	0.59	0.85
SEB-988	*S. agelas*	52	0.8	240–272	10	90	0.53	0.75
SD-1300	*S. agelas*	57	1.2	152–204	17	90	0.61	0.89
SD-1513	*S. agelas*	57	0.8	252–359	23	87	0.30	0.91
SD-31	*S. agelas*	57	1.2	275–293	10	84	0.63	0.85
SD-418	*S. agelas*	57	0.8	123–166	13	92	0.73	0.86
SP-879	*S. agelas*	53	0.8	129–309	33	91	0.67	0.92
SR-1153	*S. agelas*	53	0.8	140–174	11	92	0.58	0.83
SEB-344	*S. bousfieldi*	52	0.8	219–321	16	89	0.17	0.49
SEB-355	*S. bousfieldi*	52	0.8	250–258	6	89	0.16	0.50
SEB-458	*S. bousfieldi*	52	0.8	274–299	13	85	0.73	0.81
SEB-554	*S. bousfieldi*	52	0.8	164–174	4	84	0.060	0.39
SEB-654	*S. bousfieldi*	52	0.8	325–340	4	86	0.047	0.45
SEB-878	*S. bousfieldi*	52	0.8	137–158	10	86	0.27	0.50
SEB-989	*S. bousfieldi*	52	0.8	166–171	5	89	0.090	0.44
SD-1300	*S. bousfieldi*	55	0.8	148–168	13	89	0.43	0.77
SD-1495B	*S. bousfieldi*	53	0.8	128–166	14	85	0.35	0.69
SD-1513	*S. bousfieldi*	55	0.8	330–380	21	72	0.76	0.93
ZB-P1A12	*S. brooksi*	63	1.0	169–199	21	92	0.97	0.93
ZB-P1D4B	*S. brooksi*	51	1.0	208–263	23	92	0.98	0.93
ZB-P1E12	*S. brooksi*	52	2.0[a]	89–108	9	92	0.58	0.67
ZB-P1E4	*S. brooksi*	63	2.5[a]	280–359	25	92	0.63	0.94
ZB-P1G8	*S. brooksi*	53	0.8	211–260	21	92	0.93	0.93
ZB-P1H10	*S. brooksi*	57	1.0	239–270	15	89	0.42	0.85
ZB-P2C12	*S. brooksi*	54	2.5[a]	99–129	7	92	0.83	0.76
ZB-P2H12	*S. brooksi*	53	2.5[a]	154–274	28	92	0.79	0.90
ZB-P3F11	*S. brooksi*	58	4.0[a]	125–268	36	92	0.90	0.96
ZB-P5A8	*S. brooksi*	58	1.0	162–204	14	92	0.90	0.92
ZB-P5C10	*S. brooksi*	57	1.5[a]	369–548	23	92	0.93	0.90
ZB-P6B8	*S. brooksi*	54	2.0[a]	372–419	11	92	0.63	0.78
ZB-P6E12	*S. brooksi*	58	1.5[a]	183–225	24	92	0.89	0.92
ZB-P6G11	*S. brooksi*	57	3.5[a]	113–155	14	92	0.86	0.89
SEB-187	*S. carpenteri*	52	0.8	126–134	4	91	0.49	0.56
SEB-328	*S. carpenteri*	53	0.8	186–207	12	92	0.66	0.72
SEB-354	*S. carpenteri*	52	0.8	160–181	6	91	0.32	0.61
SEB-540	*S. carpenteri*	52	0.8	366–384	10	92	0.36	0.42
SEB-568	*S. carpenteri*	53	0.8	203–212	8	92	0.54	0.62
SEB-927	*S. carpenteri*	53	0.8	228–245	12	92	0.73	0.80
SEB-989	*S. carpenteri*	52	0.8	167–176	10	90	0.66	0.75
SP-879	*S. carpenteri*	57	0.8	229–318	32	83	0.50	0.94
SR-157	*S. carpenteri*	59	0.8	217–428	32	91	0.63	0.87
ZB-P1E12	*S. carpenteri*	50	1.2	72–144	25	92	0.46	0.67
ZB-P3F11	*S. carpenteri*	54	1.8	116–236	38	87	0.52	0.95
SEB-29	*S. cayoneptunus*	53	0.8	208–209	2	75	0.37	0.43

(Continues)

TABLE 1 (Continued)

Locus	Species	*Ta* (°C)	MgCl2 (mM)	Product size range (bp)	No. of alleles per locus	*N* (individuals)	H_O	H_E
SEB-417	*S. cayoneptunus*	53	0.8	165–168	2	75	0.53	0.50
SEB-568	*S. cayoneptunus*	53	0.8	206–228	6	74	0.66	0.64
SEB-608	*S. cayoneptunus*	53	0.8	127–139	4	74	0.28	0.25
SEB-654	*S. cayoneptunus*	53	0.8	339–340	2	75	0.52	0.46
SEB-678	*S. cayoneptunus*	53	0.8	287–288	2	73	0.40	0.46
SEB-927	*S. cayoneptunus*	53	0.8	234–236	2	75	0.027	0.41
SEB-929	*S. cayoneptunus*	53	0.8	219–223	2	75	0.52	0.43
SEB-971	*S. cayoneptunus*	53	0.8	277–284	2	75	0.01	0.040
SEB-344	*S. chacei*	52	0.8	231–238	6	151	0.42	0.66
SEB-458	*S. chacei*	52	0.8	285–316	7	150	0.087	0.097
SEB-478	*S. chacei*	52	0.8	237–250	5	152	0.020	0.046
SEB-540	*S. chacei*	52	0.8	370–378	5	151	0.23	0.30
SEB-878	*S. chacei*	52	0.8	136–151	5	152	0.22	0.36
SEB-927	*S. chacei*	52	0.8	230–236	4	152	0.01	0.06
SEB-971	*S. chacei*	52	0.8	278–284	3	152	0.026	0.06
SP-1259	*S. chacei*	53	0.8	373–414	4	150	0.088	0.097
SP-467	*S. chacei*	56	0.8	281–291	4	150	0.87	0.57
ZB-P3F11	*S. chacei*	60	1.8	110–181	27	148	0.42	0.86
ZB-P6B2	*S. chacei*	55	1.2	216–284	20	141	0.42	0.85
SD-1026	*S. dardeaui*	60	0.8	268–276	2	43	0.23	0.43
SD-1122	*S. dardeaui*	55	0.8	237–258	6	44	0.50	0.52
SD-1300	*S. dardeaui*	55	0.4	163–180	8	44	0.73	0.72
SD-1364	*S. dardeaui*	57	0.8	177–277	14	44	0.80	0.86
SD-1513	*S. dardeaui*	59	0.8	291–297	3	43	0.35	0.40
SD-183	*S. dardeaui*	59	0.8	124–142	5	44	0.43	0.68
SD-31	*S. dardeaui*	57	0.6	279–323	15	43	0.84	0.90
SD-33	*S. dardeaui*	60	0.4	128–140	3	42	0.048	0.18
SD-334	*S. dardeaui*	59	0.4	135–183	13	44	0.57	0.84
SD-418	*S. dardeaui*	59	0.6	136–193	16	44	0.64	0.88
SD-521	*S. dardeaui*	55	0.8	147–150	2	44	0.16	0.19
SD-531	*S. dardeaui*	59	0.6	115–178	15	44	0.80	0.82
SD-643	*S. dardeaui*	54	1.0	151–181	10	44	0.48	0.82
SD-8	*S. dardeaui*	55	0.6	283–335	12	43	0.76	0.83
SD-88	*S. dardeaui*	59	0.8	154–158	2	44	0.20	0.22
SEB-354	*S. duffyi*	52	0.8	163–170	3	172	0.81	0.50
SEB-476	*S. duffyi*	53	0.8	262–275	5	168	0.54	0.44
SEB-554	*S. duffyi*	52	0.8	164–171	3	171	0.41	0.35
SEB-575	*S. duffyi*	52	0.8	176–178	2	173	0.15	0.13
SEB-854	*S. duffyi*	53	0.8	126–127	2	171	0.49	0.44
SEB-927	*S. duffyi*	52	0.8	213–238	7	171	0.44	0.58
SEB-971	*S. duffyi*	52	0.8	278–284	3	173	0.087	0.13
SEB-328[b]	*S. duffyi*	52	0.8	191–203	9	165	0.86	0.77
SEB-540[b]	*S. duffyi*	52	0.8	364–386	7	166	0.73	0.77
SEB-568[b]	*S. duffyi*	52	0.8	198–252	22	165	0.88	0.89
SEB-69	*S. elizabethae*	52	0.8	208–296	14	91	0.81	0.86
SEB-417	*S. elizabethae*	55	0.8	166–175	3	95	0.36	0.32

(Continues)

TABLE 1 (Continued)

Locus	Species	*Ta* (°C)	MgCl2 (mM)	Product size range (bp)	No. of alleles per locus	*N* (individuals)	H_O	H_E
SEB-464	*S. elizabethae*	55	0.8	251–268	10	89	0.57	0.75
SEB-478	*S. elizabethae*	55	0.8	229–239	4	98	0.27	0.42
SEB-560	*S. elizabethae*	52	0.8	82–283	6	96	0.63	0.65
SEB-568	*S. elizabethae*	55	0.8	208–210	2	95	0.40	0.41
SEB-575	*S. elizabethae*	55	0.8	175–177	2	98	0.22	0.30
SEB-831	*S. elizabethae*	52	0.8	246–250	3	94	0.41	0.53
SEB-927	*S. elizabethae*	55	0.8	228–236	3	97	0.14	0.17
SEB-971	*S. elizabethae*	55	0.8	284–290	3	96	0.13	0.19
SEB-354	*S. filidigitus*	52	0.8	164–181	4	96	0.50	0.47
SEB-417	*S. filidigitus*	52	0.8	169–171	3	96	0.26	0.24
SEB-464	*S. filidigitus*	52	0.8	241–264	12	96	0.93	0.88
SEB-478	*S. filidigitus*	52	0.8	277–286	3	89	0.33	0.37
SEB-540	*S. filidigitus*	52	0.8	364–376	5	93	0.40	0.42
SEB-758	*S. filidigitus*	52	0.8	267–279	3	94	0.61	0.47
SD-1026	*S. herricki*	51	0.8	245–338	12	94	0.50	0.74
SD-1300	*S. herricki*	59	1.2	154–184	12	92	0.62	0.72
SD-31	*S. herricki*	59	0.8	281–290	5	95	0.65	0.73
SP-1323	*S. herricki*	57	0.8	109–112	2	96	0.48	0.48
SP-879	*S. herricki*	61	0.8	228–440	77	93	0.70	0.99
ZB-P1D4B	*S. herricki*	55	1.0	174–268	26	89	0.63	0.93
ZB-P1E12	*S. herricki*	54	1.0	74–98	5	96	0.50	0.51
ZB-P3A8	*S. herricki*	57	0.8	225–279	12	93	0.35	0.66
ZB-P6B8	*S. herricki*	53	1.2	265–363	29	94	0.81	0.90
ZB-P6E12	*S. herricki*	53	1.0	126–136	6	96	0.79	0.79
SD-1344	*S. hoetjesi*	49	0.8	250–301	18	33	0.82	0.94
SD-1364	*S. hoetjesi*	57	0.8	255–368	8	33	0.52	0.73
SD-1495B	*S. hoetjesi*	53	0.8	160–235	12	33	0.64	0.83
SR-425B	*S. hoetjesi*	55	0.8	150–206	19	32	0.66	0.93
ZB-P1D4B	*S. hoetjesi*	59	0.8	179–262	27	33	0.88	0.96
ZB-P4D2	*S. hoetjesi*	55	0.8	282–307	7	33	0.24	0.77
ZB-P5C10	*S. hoetjesi*	49	0.8	283–293	6	30	0.63	0.70
ZB-P6B2	*S. hoetjesi*	57	0.8	200–260	12	33	0.39	0.78
ZB-P6B8	*S. hoetjesi*	49	0.8	277–332	16	30	0.40	0.90
ZB-P6E12	*S. hoetjesi*	53	0.8	170–211	19	33	0.42	0.94
SEB-66	*S. idios*	52	0.8	138–155	3	58	0.03	0.03
SEB-187	*S. idios*	52	0.8	134–139	3	51	0.65	0.63
SEB-355	*S. idios*	52	0.8	260–282	10	58	0.84	0.83
SEB-554	*S. idios*	52	0.8	164–171	2	55	0.055	0.05
SEB-927	*S. idios*	52	0.8	226–237	5	58	0.62	0.65
SEB-971	*S. idios*	52	0.8	276–278	3	58	0.59	0.45
SD-1495B	*S. idios*	53	0.8	113–135	11	57	0.61	0.75
SP-840	*S. idios*	53	0.8	243–268	7	54	0.31	0.32
SR-157	*S. idios*	57	0.8	302–326	6	54	0.80	0.77
ZB-P1F6	*S. idios*	57	0.8	141–176	8	58	0.67	0.73
SD-1495B	*S. microneptunus*	53	1.2	135–136	2	48	0.46	0.50
SD-418	*S. microneptunus*	53	0.8	115–148	10	48	0.92	0.88

(Continues)

TABLE 1 (Continued)

Locus	Species	*Ta* (°C)	MgCl2 (mM)	Product size range (bp)	No. of alleles per locus	*N* (individuals)	H_O	H_E
SP-1259	*S. microneptunus*	63	0.8	277–291	5	47	0.26	0.58
SP-1323	*S. microneptunus*	57	0.8	120–145	6	48	0.71	0.70
SP-1855	*S. microneptunus*	61	0.8	239–248	4	46	0.59	0.49
SP-669	*S. microneptunus*	53	0.8	192–204	5	47	0.64	0.70
SP-840	*S. microneptunus*	61	0.8	269–296	10	48	0.83	0.83
SP-879	*S. microneptunus*	57	0.8	266–327	4	48	0.71	0.62
ZB-P1D4B	*S. microneptunus*	55	0.8	216–268	18	47	0.68	0.92
ZB-P1E12	*S. microneptunus*	50	1.2	105–130	8	48	0.73	0.83
ZB-P5C10	*S. microneptunus*	53	0.8	283–290	5	48	0.40	0.42
ZB-P6B2	*S. microneptunus*	58	0.8	208–216	3	48	0.19	0.17
SEB-328	*S. rathbunae*	52	0.8	187–202	6	96	0.66	0.76
SEB-464	*S. rathbunae*	52	0.8	252–264	7	96	0.82	0.82
SEB-540	*S. rathbunae*	52	0.8	373–376	3	96	0.49	0.55
SEB-568	*S. rathbunae*	52	0.8	207–229	7	96	0.57	0.63
SP-1323	*S. rathbunae*	57	0.8	110–195	14	96	0.73	0.79
SR-1226	*S. rathbunae*	55	0.8	262–285	7	96	0.75	0.64
SR-1573	*S. rathbunae*	59	1.2	159–167	3	96	0.60	0.61
SR-405	*S. rathbunae*	62	0.8	159–219	14	96	0.79	0.74
SR-704B	*S. rathbunae*	62	0.8	265–282	6	96	0.51	0.72
ZB-P1D4B	*S. rathbunae*	55	0.8	253–310	17	95	0.75	0.88
SEB-187	*S. regalis*	53	0.8	128–134	4	96	0.15	0.16
SEB-355	*S. regalis*	53	0.8	251–255	3	96	0.31	0.27
SEB-554	*S. regalis*	53	0.8	166–172	2	96	0.05	0.05
SEB-878	*S. regalis*	53	0.8	135–148	5	96	0.74	0.66
SR-122	*S. regalis*	55	1.0	145–163	5	89	0.47	0.65
SR-1226	*S. regalis*	52	1.0	265–285	4	96	0.32	0.48
SR-1406	*S. regalis*	57	0.8	226–238	3	96	0.47	0.45
SR-1517	*S. regalis*	59	0.8	407–455	9	88	0.49	0.88
SR-405	*S. regalis*	60	0.8	152–200	14	94	0.84	0.88
SR-425B	*S. regalis*	55	0.8	133–186	12	92	0.75	0.83
SR-573B	*S. regalis*	54	0.8	146–173	10	96	0.82	0.85
SR-704B	*S. regalis*	58	0.8	263–279	8	94	0.67	0.79
ZB-P1D4B	*S. regalis*	51	1.2	246–266	9	84	0.64	0.86
SEB-464	*S. thele*	52	0.8	227–249	7	51	0.49	0.46
SEB-478	*S. thele*	52	0.8	244–249	5	51	0.49	0.53
SEB-568	*S. thele*	52	0.8	204–207	4	51	0.16	0.15
SD-1495B	*S. thele*	53	0.8	113–116	2	51	0.020	0.020
SD-1513	*S. thele*	59	0.8	289–301	5	51	0.35	0.49
SP-467	*S. thele*	61	0.8	281–293	7	47	0.55	0.81
SP-879	*S. thele*	61	0.8	243–484	40	41	0.59	0.97
SR-157	*S. thele*	57	0.8	242–338	26	50	0.68	0.96
ZB-P1E12	*S. thele*	49	1.2	74–92	3	51	0.37	0.35
ZB-P3F11	*S. thele*	57	1.0	97–219	36	50	0.64	0.97

[a]Reported $MgCl_2$ concentrations from Rubenstein et al., 2008. Later PCR runs on eggs were optimal at lower MgCl2 concentrations (0.8–1.0 mM).
[b]Polysomic locus.

TABLE 2 Marker screening summary. Microsatellite markers were screened on 25 species of *Synalpheus*. For those species that were genotyped, sample sizes reflect the number of regions and colonies sampled in genotyping and marker screening

Species	Clade	Genome size (pg)	No. geographic regions	No. colonies	No. markers tested	No. markers working	No. and per cent disomic markers	No. and per cent polysomic markers	No. and per cent[a] homozygous markers	Genotyped
S. agelas	None	13.6	3	17	89	41	33 (80%)	1 (2%)	7 (17%)	Y
S. bousfieldi	Brooksi	10.46	3	10	120	56	33 (59%)	9 (16%)	14 (25%)	Y
S. brooksi	Brooksi	10.88	1	5	32	32	25 (78%)	7 (22%)	0	Y
S. carpenteri	Brooksi	9.84	1	12	121	51	27 (53%)	9 (18%)	15 (29%)	Y
S. chacei	Brooksi	7.74	1	8	124	50	22 (44%)	4 (8%)	24 (48%)	Y
S. idios	Brooksi	15.52	2	4	88	47	23 (49%)	4 (9%)	20 (43%)	Y
S. thele	Brooksi	4.83	1	4	65	32	18 (56%)	7 (22%)	7 (22%)	Y
S. dardeaui	Longicarpus	7.09	2	10	22	22	16 (73%)	6 (27%)	0	Y
S. goodei	Longicarpus	7.78	1	4	54	18	9 (50%)	3 (17%)	6 (33%)	N
S. herricki	Longicarpus	8.25	3	10	87	27	15 (56%)	5 (19%)	7 (26%)	Y
S. hoetjesi	Longicarpus	5.77	3	7	54	23	13 (57%)	2 (9%)	8 (35%)	Y
S. pectiniger	Longicarpus	12.79	1	5	85	23	11 (48%)	5 (22%)	7 (30%)	N
S. ul	Longicarpus	5.14	1	7	58	16	8 (53%)	5 (33%)	3 (19%)	N
S. yano	Longicarpus	6.49	1	3–5[b]	86	23	13 (57%)	3 (13%)	7 (30%)	N
S. belizensis	Paraneptunus	10.35	3	10	38	25	13 (52%)	4 (16%)	8 (32%)	N
S. bocas	Paraneptunus	12.42	2	5	59	23	10 (43%)	5 (22%)	8 (35%)	N
S. cayoneptunus	Paraneptunus	13.15	1	6	33	29	15 (52%)	0	14 (48%)	Y
S. duffyi	Paraneptunus	20.74	2	13	126	50	16 (32%)	23 (46%)	11 (22%)	Y
S. kensleyi	Paraneptunus	10.76	1	5	54	19	8 (42%)	5 (26%)	6 (32%)	N
S. microneptunus	Paraneptunus	9.8	1	7	85	32	14 (44%)	4 (13%)	14 (44%)	Y
S. androsi	Rathbunae	7.09	2	6	87	25	11 (44%)	2 (8%)	12 (48%)	N
S. elizabethae	Rathbunae	13.28	1	11	139	47	30 (64%)	2 (4%)	15 (32%)	Y
S. filidigitus	Rathbunae	9.47	1	8	88	35	14 (40%)	2 (6%)	19 (54%)	Y
S. rathbunae	Rathbunae	11.38	1	9	95	35	20 (57%)	3 (9%)	12 (34%)	Y
S. regalis	Rathbunae	11.76	3	16	53	38	27 (71%)	3 (8%)	8 (21%)	Y

[a]Percentages calculated based upon all working markers (disomic, polysomic and homozygous).
[b]Two different sets of *S. yano* samples were used for marker testing. Some markers were screened on samples from three colonies, and others on samples from five colonies.

any species; Table S3). Of these, 97 were genome-derived (32 "ZB" markers from *S. brooksi*, 19 "SP" markers from *S. belizensis*, 22 "SD" markers from *S. dardeaui* and 24 "SR" markers from *S. regalis*), and the remaining 46 were derived from the *S. elizabethae/brooksi* transcriptome. We tested a mean of 78 ± 32 markers per species (±*SD*; $n = 25$ species; Table 2). A mean ± *SD* of 33 ± 12 (47% ± 21%, $n = 25$) markers successfully amplified for each species, with 18 ± 8 (54% ± 12%) of these markers disomic, 5 ± 4 (16% ± 10%) polysomic and 10 ± 6 (30% ± 13%) homozygous. Although there was a strong negative correlation between the number of markers tested and the number of markers working (homozygous, disomic or polysomic; $\beta = -0.0059$, $SE = 0.0014$, $p < .001$), there was no relationship between the number of markers tested and the percentage of polysomic markers ($\beta = -0.00015$, $SE = 0.0010$, $p = .89$); the number of markers tested per species did not differ across clades (ANOVA: $F_3 = 1.24$, $p = .32$).

Each genomic marker was tested on a mean ± *SD* of 16 ± 6 species, including the target species for which it was developed. Transcriptomic markers were tested on a mean ± *SD* of 7 ± 4 species. Genomic markers worked for a mean ± *SD* of 37% ± 27% ($n = 24$) of the species on which they were tested (excluding target species, 31% ± 29%, $n = 21$), and transcriptomic markers worked for a mean ± *SD* of 77% ± 24% ($n = 11$) of species (excluding target species, 76% ± 24%, $n = 10$). Of the 133 markers tested on multiple species, 14% ($n = 18$) worked only for the target species.

We created complete panels of disomic polymorphic markers for 16 species, and a panel of disomic and polysomic polymorphic markers for *S. duffyi*, which had insufficient disomic markers. Each panel had a mean ± *SD* of 10.65 ± 2.03 markers (range: 6–15). A total of 87 markers were used in genotyping across species (34 transcriptome-derived and 56 genome-derived). For genotyped species, mean queen–egg relatedness ranged from 0.41 (*S. idios*) to 0.62

(*S. microneptunus*), with a mean ± *SD* relatedness of 0.49 ± 0.06 across species.

Each marker was used to genotype a mean ± *SD* of 2.08 ± 1.40 (range: 1–6) of the 17 species. For the 23 markers used to genotype the target and at least one nontarget species, the number of alleles did not differ significantly for target and nontarget species (Welch *t* test: $t = -0.094$, $df = 22$, $p = .93$), nor did observed (Welch *t* test: $t = 0.86$, $df = 22$, $p = .40$) or expected (Welch *t* test: $t = 0.40$, $df = 22$, $p = .69$) heterozygosity. Observed heterozygosities for all markers used in genotyping ($n = 181$) ranged from 0.01 to 0.98, with a mean ± *SD* of 0.51 ± 0.25. We did not conduct tests of Hardy–Weinberg equilibrium because individuals were sampled from multiple colonies with a likelihood of genetic subdivision among colonies (Duffy, 1996).

3.1 | Marker retention across species

For a given species, genomic markers developed for a closely related species worked better than markers developed for more distant relatives (Figure 1; Welch *t* test: $t = 4.39$, $df = 23.30$, $p < .001$). The mean ± *SD* success of within-clade markers was 50% ± 19%, compared to 29% ± 8% for outside-clade markers ($n = 23$ species). Although this pattern is consistent across all four clades, the difference was significant only for the Longicarpus (Welch *t* test: $t = 2.82$, $df = 5$, $p = .04$, $n = 6$ species) and Paraneptunus clades (Welch *t* test: $t = 8.77$, $df = 3$, $p = .003$, $n = 4$ species), and not for the Brooksi (Welch *t* test: $t = 2.08$, $df = 4$, $p = .11$, $n = 5$ species) or Rathbunae clades (Welch *t* test: $t = 1.58$, $df = 3$, $p = .21$, $n = 4$ species) (Figure 1). Similarly, markers developed for a given species tended to have higher amplification success for other species in the same clade than for those outside that clade (Figure 1). The difference was significant for ZB (Brooksi clade; paired *t* test: $t = 2.85$, $df = 9.30$, $p = .02$) and SP markers (Paraneptunus clade; paired *t* test: $t = 6.45$, $df = 6.51$, $p < .001$), but not for SD (Longicarpus clade; paired *t* test: $t = 1.86$, $df = 10.07$, $p = .09$) or SR markers (Rathbunae clade; paired *t* test: $t = 2.02$, $df = 3.28$, $p = .13$).

3.2 | Polysomic markers and genome size

Despite a nearly fivefold difference in genome size among the species in our analysis (Jeffery et al., 2016), there was no relationship between genome size and the percentage of polysomic markers for a given species (Figure 2). The relationship was very weakly positive (GLM: $\beta = 0.0079$, $SE = 0.0080$, $p = .34$), but was driven largely by *S. duffyi*, which had the largest genome size and the greatest percentage of polysomic markers. Removing this outlier species resulted in a weakly negative relationship between genome size and polysomic markers (Figure 2; GLM: $\beta = -0.011$, $SE = 0.0087$, $p = .22$). Thus, there was no evidence for a relationship between genome size and the number of polysomic markers.

3.3 | Genome vs. transcriptome

Markers developed from the *S. elizabethae* and *S. brooksi* consensus transcriptome sequence had higher overall amplification rates (75% ± 16%, $n = 12$) than markers developed from genome sequences (41% ± 21%, $n = 24$; Welch *t* test: $t = 4.20$, $df = 27.43$, $p < .001$; Figure 3). This difference remained when target species were excluded (78% ± 14% transcriptomic, $n = 11$; 32% ± 6% genomic, $n = 20$; Welch *t* test: $t = -7.79$, $df = 10.87$, $p < .001$), as well as when including only the 11 species for which both genomic and transcriptomic markers were tested (74 ± 16% transcriptomic; 36 ± 12% genomic; paired *t* test: $t = 6.12$, $df = 10$, $p < .001$, $n = 11$). Additionally, the percentage of polysomic markers (of all working markers) was greater for genome-derived markers (19 ± 11%, $n = 24$) than for transcriptome-derived markers

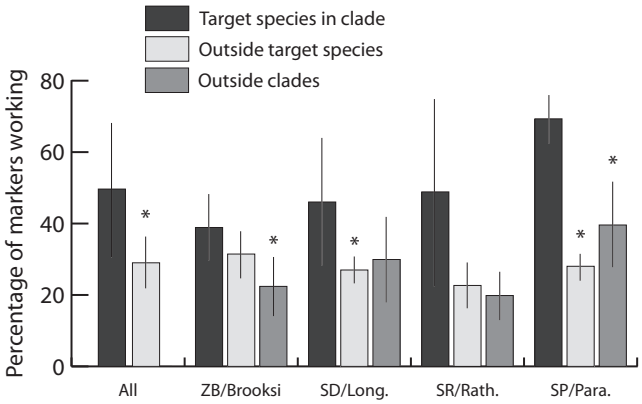

FIGURE 1 Comparison of marker success within and across clades. For each clade, we compared the success of genomic markers developed for and tested on species in the same clade (dark grey), markers developed for target species in other clades (light grey), and markers developed for that clade, tested on species from other clades (medium grey). An asterisk represents a significant difference ($p < .05$) from the "target species in clade" percentage

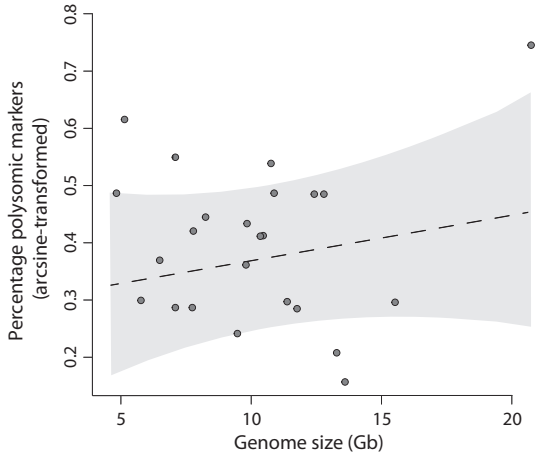

FIGURE 2 Polysomy and genome size. There was no significant relationship between genome size and the percentage of polysomic markers (GLM: $\beta = 0.0079$, $SE = 0.0080$, $p = .34$). The weakly positive trend was largely driven by *Synalpheus duffyi*, an outlier for both genome size and polysomic markers

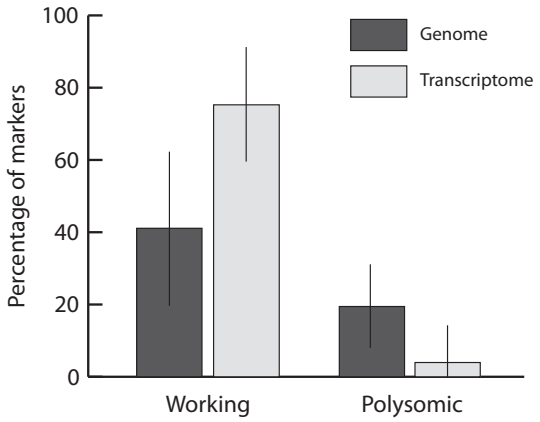

FIGURE 3 Comparison of genome- and transcriptome-derived markers. The mean and standard deviation of the percentage of screened markers that were working and working markers that were polysomic for a given species

(4% ± 10%, $n = 11$; Welch t test: $t = 5.25$, $df = 18.55$, $p < .001$; Figure 3). The relationship remained when considering only the 11 species for which both sets of markers were tested (20 ± 15% genomic; 4% ± 15% transcriptomic; paired t test: $t = 4.42$, $df = 10$, $p = .0013$, $n = 11$).

4 | DISCUSSION

We successfully developed molecular markers to be used for cross-species analysis of kin structure in 25 species of *Synalpheus* shrimp, a genus with a high degree of partial genome duplication and interspecific variation in genome size. By developing markers based upon the genomes of four species from distinct clades, in addition to a consensus transcriptome sequence from two species, we addressed some of the challenges associated with cross-species marker utility and genome duplication. We finalized complete panels of 6–15 disomic polymorphic markers for 16 species (and both disomic and polysomic markers for one species, *S. duffyi*), which were sufficient to genotype individuals within colonies. Our analyses of relatedness between queens (mothers) and eggs confirmed that these panels produced reliable genotypes, and the high degree of heterozygosity suggests that the markers will be useful for future studies of kin structure within shrimp colonies.

Although the markers did not perform consistently across the genus, some were conserved across multiple species that varied considerably in genome size. As predicted and as has been observed across other plant and animal taxa, amplification success for a given marker–species combination was greater if the species was more closely related to the target species for which the marker was developed, likely due to higher conserved sequence variation (Gemmell, Allen, Goodman, & Reed, 1997; Primmer et al., 1997; Wilson, Massonnet, & Simon, 2004). However, there were many exceptions to this pattern, with marker performance varying greatly within a clade, and cases where markers successfully amplified across distant congeners despite failing in more closely related species, similar to patterns in polyploid plant taxa (Huang et al., 1998; Whitton & Rieseberg, 1997). These inconsistencies are likely to be attributed to high genome variability across *Synalpheus* species, and associated genome duplication and loss, even in closely related species.

The greater success of the transcriptome-derived markers, when compared to genome-derived markers, suggests that the transcriptome may be a better source of molecular markers for multiple species with varying degrees of genome duplication. While microsatellite frequency is higher in the noncoding region of the genome than in the coding region (Chistiakov et al., 2006; Li, Korol, Fahima, & Beiles, 2002), the coding region and the expressed genes that it contains are more likely to be conserved, explaining why transcriptome-derived markers displayed greater utility across species in *Synalpheus* (Postolache et al., 2013; Wang et al., 2014). Furthermore, the transcriptome yielded a higher percentage of disomic markers than the genome, suggesting that the partial genome duplication in *Synalpheus* may occur primarily in the noncoding region. Compared to the many polysomic markers derived from the genome, the disomic markers are easier to score and to use in conventional relatedness and parentage analyses (Dufresne et al., 2013).

While the transcriptome is a promising source of disomic, conserved markers in taxa with a high degree of genome duplication, these markers may be less likely to meet assumptions of neutrality given that the transcriptome contains only the expressed portion of the genome (but see Wang et al., 2004; Broadley et al., 2008). Markers may be adjacent to genes that are under selection, and the distribution of allele frequencies may therefore fail to meet assumptions of neutrality and provide biased estimates of relatedness (Li et al., 2002; Selkoe & Toonen, 2006). To avoid introducing bias, studies that derive markers from the transcriptome should ensure that allele distributions do not violate assumptions of subsequent analyses (Selkoe & Toonen, 2006).

In this study, the greater cross-species utility of the transcriptome-derived markers can also be attributed to the utilization of a consensus transcriptome sequence of two distantly related congeners for microsatellite identification and primer design. Microsatellites in the consensus sequence may have been more likely to be conserved throughout the genus, compared to those in the single-species genome sequences. For the development of molecular markers within a group of related species, particularly when rapid genome duplication and evolution has occurred, consensus sequences generated from the transcriptomes or genomes of multiple species may yield more robust markers that can be used across species (Buteler et al., 1999; Grivet, Heinze, Vendramin, & Petit, 2001).

Although we hypothesized that genome duplication is the source of both large genome sizes and polysomic markers in *Synalpheus*, we did not find a positive relationship between genome size and the percentage of polysomic loci as we originally predicted. One species, *S. duffyi*, had both the largest genome size and the highest proportion of polysomic markers, but this trend did not follow for other species. In fact, when *S. duffyi* was excluded from analysis, there was a weakly negative (though nonsignificant) relationship between

genome size and the percentage of polysomic loci. Studies of other organisms with genome duplication have also failed to find a positive or linear relationship between genome size and ploidy levels (McIntyre, 2012; Pellicer et al., 2010; Wong & Murray, 2012), possibly explaining why polysomic markers were not consistently more prevalent in shrimp species with larger genomes.

In summary, our findings suggest that microsatellites remain important molecular markers for species with a high degree of genome duplication. The transcriptome may be a valuable source of disomic loci for taxa with rapid evolution and high amounts of genome duplication, particularly when markers are to be used across related species. In this case, markers should be identified from the genomes or transcriptomes of as many different species as possible, and ideally from consensus sequences from several species. While challenges remain in the identification of disomic markers for species with complex genomes, our approach suggests a way forward to minimize time and labour costs and to systematically develop microsatellite markers for multiple species.

ACKNOWLEDGEMENTS

We thank the staff of the Sackler Institute for Comparative Genomics at the American Museum of Natural History for assistance with genotyping, and S. Bogdanowicz for 454 sequencing and assistance with marker development. K. Hultgren supplied samples, S. Chak provided assistance with statistical analyses. N. Bailey, M. Cohen, H. D'Angelo, B. Eckersely, J. Hollarsmith, E. Lopez, K. Nolan, J. Pilowsky, and L. Pollack helped with marker screening and development. J.E.D. (DEB 92-01566, DEB 98–15785, IBN-0131931, IOS-1121716) and D.R.R. (IOS-1121435, IOS-1257530, IOS-1439985) were supported by the US National Science Foundation.

AUTHOR CONTRIBUTIONS

D.R.R. and J.E.D. designed research and supplied samples, K.M.G., J.W.S., S.S. and L.J. performed research, K.M.G. and J.W.S. analysed data, and K.M.G., J.W.S. and D.R. wrote the manuscript.

DATA ACCESSIBILITY

Microsatellite-containing DNA sequences for *S. dardeaui*, *S. belizensis* and *S. regalis*: GenBank Accession nos. MF094823–MF094887. Microsatellite-containing DNA sequences for *S. brooksi*: GenBank Accession nos. EU135992–EU136000, EU136007–EU136012, EU136014–EU136019, EU136023–EU136025. Microsatellite-containing mRNA (cDNA) transcriptome reference sequences for *S. brooksi*: GenBank Accession nos. MF114246–MF114291. Alignments of raw (adapter- and quality-trimmed) Illumina transcriptome sequences for *S. elizabethae* to *S. brooksi* transcriptome reference sequences: SRA Accession no. SRP107140. Online Supporting Information includes primer sequences, PCR profiles, species metadata, and Trinity, Trimmomatic and Bowtie2 commands.

ORCID

Kaitlyn M. Gaynor http://orcid.org/0000-0002-5747-0543
Dustin R. Rubenstein http://orcid.org/0000-0002-4999-3723

REFERENCES

Ashley, M. V., Wilk, J. A., Styan, S. M. N., Craft, K. J., Jones, K. L., Feldheim, K. A., … Ashman, T. L. (2003). High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana* Mill. (Rosaceae). *Theoretical and Applied Genetics*, *107*, 1201–1207.

Baisakh, N., Subudhi, P. K., Arumuganathan, K., Parco, A. P., Harrison, S. A., Knott, C. A., & Materne, M. D. (2009). Development and interspecific transferability of genic microsatellite markers in Spartina spp. with different genome size. *Aquatic Botany*, *91*, 262–266.

Barnett, J. R., Stenzler, L. M., Ruiz-Gutierrez, V., Bogdanowicz, S. M., & Lovette, I. J. (2008). Isolation and characterization of microsatellite markers from the white-ruffed manakin *Corapipo altera* (Aves, Pipridae). *Molecular Ecology Resources*, *8*, 215–218.

Blischak, P. D., Kubatko, L. S., & Wolfe, A. D. (2016). Accounting for genotype uncertainty in the estimation of allele frequencies in autopolyploids. *Molecular Ecology Resources*, *16*, 742–754.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Boscari, E., Vidotto, M., Martini, D., Papetti, C., Ogden, R., & Congiu, L. (2015). Microsatellites from the genome and the transcriptome of the tetraploid Adriatic sturgeon, *Acipenser naccarii* (Bonaparte, 1836) and cross-species applicability to the diploid beluga sturgeon, *Huso huso* (Linnaeus, 1758). *Journal of Applied Ichthyology*, *31*, 977–983.

Broadley, M. R., White, P. J., Hammond, J. P., Graham, N. S., Bowen, H. C., Emmerson, S. T., … May, Z. F. (2008). Evidence of neutral transcriptome evolution in plants. *New Phytologist*, *180*, 587–593.

Brownstein, M. J., Carpten, J. D., & Smith, J. R. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques*, *20*, 1004–1010.

Buteler, M. I., Jarret, R. L., & LaBonte, D. R. (1999). Sequence characterization of microsatellites in diploid and polyploid Ipomoea. *Theoretical and Applied Genetics*, *99*, 123–132.

Butler, I. A., Siletti, K., Oxley, P. R., & Kronauer, D. J. C. (2014). Conserved microsatellites in ants enable population genetic and colony pedigree studies across a wide range of species. *PLoS ONE*, *9*, e107334.

Chak, S., Duffy, J. E., & Hultgren, K. M. (2017). Evolutionary transitions towards eusociality in snapping shrimps. *Nature Ecology and Evolution*, *1*, 1–10.

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, *255*, 1–29.

Christensen, K. A., Brunelli, J. P., Lambert, M. J., DeKoning, J., Phillips, R. B., & Thorgaard, G. H. (2013). Identification of single nucleotide polymorphisms from the transcriptome of an organism with a whole genome duplication. *BMC Bioinformatics*, *14*. https://doi.org/10.1186/1471-2105-14-325.

David, L., Blum, S., Feldman, M. W., Lavi, U., & Hillel, J. (2003). Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Molecular Biology and Evolution*, *20*, 1425–1434.

Duffy, J. E. (1996). Resource-associated population subdivision in a symbiotic coral-reef shrimp. *Evolution*, *50*, 360–373.

Duffy, J. E., & Macdonald, K. S. (2010). Kin structure, ecology and the evolution of social organization in shrimp: A comparative analysis. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 575–584.

Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2013). Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology*, *23*, 40–69.

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*, 435–445.

Faircloth, B. C. (2008). msatcommander: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, *8*, 92–94.

Frasier, T. R. (2008). STORM: Software for testing hypotheses of relatedness and mating patterns. *Molecular Ecology Resources*, *8*, 1263–1266.

Garant, D., & Kruuk, L. E. B. (2005). How to use molecular marker data to measure evolutionary parameters in wild populations. *Molecular Ecology*, *14*, 1843–1859.

Gemmell, N. J., Allen, P. J., Goodman, S. J., & Reed, J. Z. (1997). Interspecific microsatellite markers for the study of pinniped populations. *Molecular Ecology*, *6*, 661–666.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652.

Grant, J. B., & Bogdanowicz, S. M. (2006). Isolation and characterization of microsatellite markers from the panic moth, *Saucrobotys futilalis* L. (Lepidoptera: Pyralidae: Pyraustinae). *Molecular Ecology Notes*, *6*, 353–355.

Grivet, D., Heinze, B., Vendramin, G. G., & Petit, R. J. (2001). Genome walking with consensus primers: Application to the large single copy region of chloroplast DNA. *Molecular Ecology Notes*, *1*, 345–349.

Guyomarc'h, H., Sourdille, P., Edwards, K., & Bernard, M. (2002). Studies of the transferability of microsatellites derived from *Triticum tauschii* to hexaploid wheat and to diploid related species using amplification, hybridization and sequence comparisons. *Theoretical and Applied Genetics*, *105*, 736–744.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*. https://doi.org/10.18637/jss.v033.i02.

Hamilton, M. B., Pincus, E. L., Di Fiore, A., & Fleischer, R. C. (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques*, *27*, 500–507.

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Resources*, *2*, 618–620.

Huang, W. G., Cipriani, G., Morgante, M., & Testolin, R. (1998). Microsatellite DNA in *Actinidia chinensis*: Isolation, characterisation, and homology in related species. *Theoretical and Applied Genetics*, *97*, 1269–1278.

Huang, K., Ritland, K., Guo, S., Shattuck, M., & Li, B. (2014). A pairwise relatedness estimator for polyploids. *Molecular Ecology Resources*, *14*, 734–744.

Hultgren, K. M., & Duffy, J. E. (2011). Multi-locus phylogeny of sponge-dwelling snapping shrimp (Caridea: Alpheidae: Synalpheus) supports morphology-based species concepts. *Journal of Crustacean Biology*, *31*, 352–360.

Hultgren, K. M., Duffy, J. E., & Rubenstein, D. R. (2017). Sociality in snapping shrimps. In D. R. Rubenstein & P. Abbot (Eds.), *Comparative social evolution* (pp. 224–249). Cambridge: Cambridge University Press.

Hultgren, K. M., Hurt, C., & Anker, A. (2014). Phylogenetic relationships within the snapping shrimp genus *Synalpheus* (Decapoda: Alpheidae). *Molecular Phylogenetics and Evolution*, *77*, 116–125.

Jeffery, N. W., Hultgren, K., Chak, S. T. C., Gregory, T. R., & Rubenstein, D. R. (2016). Patterns of genome size variation in snapping shrimp. *Genome*, *59*, 393–402.

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*, 1099–1106.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Li, Y. C., Korol, A. B., Fahima, T., & Beiles, A. (2002). Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology*, *11*, 2453–2465.

Mayer, C. (2006). Phobos 3.3.11, 2006–2010, http://www.rub.de/spezzoo/cm/cm_phobos.htm

McIntyre, P. J. (2012). Cytogeography and genome size variation in the *Claytonia perfoliata* (Portulacaceae) polyploid complex. *Annals of Botany*, *110*, 1195–1203.

Meirmans, P. G., & van Tienderen, P. H. (2004). genotype and genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*, 792–794.

Morin, P. A., Luikart, G., Wayne, R. K., & the SNP workshop group. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, *19*, 208–216.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290.

Pellicer, J., Garcia, S., Canela, M. Á., Garnatje, T., Korobkov, A. A., Twibell, J. D., & Vallès, J. (2010). Genome size dynamics in *Artemisia* L. (Asteraceae): Following the track of polyploidy. *Plant Biology*, *12*, 820–830.

Postolache, D., Leonarduzzi, C., Piotti, A., Spanu, I., Roig, A., Fady, B., … Vendramin, G. G. (2013). Transcriptome versus genomic microsatellite markers: Highly informative multiplexes for genotyping *Abies alba* Mill. and Congeneric species. *Plant Molecular Biology Reporter*, *32*, 750–760.

Primmer, C. R., Moller, A. P., & Ellegren, H. (1997). A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology*, *5*, 365–378.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rubenstein, D. R., McCleery, B. V., & Duffy, J. E. (2008). Microsatellite development suggests evidence of polyploidy in the social sponge-dwelling snapping shrimp *Zuzalpheus brooksi*. *Molecular Ecology Resources*, *8*, 890–894.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, *18*, 233–234.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*, 615–629.

Städele, V., & Vigilant, L. (2016). Strategies for determining kinship in wild populations using genetic data. *Ecology and Evolution*, *6*, 6107–6120.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, *15*, 199–203.

Toth, G. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research*, *10*, 967–981.

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. *Genetics Selection Evolution*, *34*, 275–305.

Waldbieser, G. C., Quiniou, S., & Karsi, A. (2003). Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. *BioTechniques*, *35*, 976–979.

Wang, H.-L., Yang, J., Boykin, L. M., Zhao, Q.-Y., Wang, Y.-J., Liu, S.-S., & Wang, X.-W. (2014). Developing conversed microsatellite markers and their implications in evolutionary analysis of the *Bemisia tabaci* complex. *Scientific Reports*, *4*, 6351. https://doi.org/10.1038/srep06351.

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., … Wong, G. K.-S. (2004). Mouse transcriptome: Neutral evolution of "non-coding" complementary DNAs. *Nature*, *431*, 1–2.

Waples, R. K., Seeb, L. K., & Seeb, J. E. (2015). Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in chum salmon (*Oncorhynchus keta*). *Molecular Ecology Resources*, *16*, 17–28.

Weinman, L. R., Solomon, J. W., & Rubenstein, D. R. (2014). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, *15*, 502–511.

Whitton, J., & Rieseberg, L. H. (1997). Microsatellite loci are not conserved across the Asteraceae. *Molecular Biology and Evolution*, *14*, 204–209.

Wilson, A., Massonnet, B., & Simon, J. C. (2004). Cross-species amplification of microsatellite loci in aphids: Assessment and application. *Molecular Ecology*, *4*, 104–109.

Wong, C., & Murray, B. G. (2012). Variable changes in genome size associated with different polyploid events in *Plantago* (Plantaginaceae). *Journal of Heredity*, *103*, 711–719.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Gaynor KM, Solomon JW, Siller S, Jessell L, Duffy JE, Rubenstein DR. Development of genome- and transcriptome-derived microsatellites in related species of snapping shrimps with highly duplicated genomes. *Mol Ecol Resour*. 2017;17:e160–e173. https://doi.org/10.1111/1755-0998.12705