Eusociality in snapping shrimps is associated with larger genomes and an accumulation of transposable elements

Solomon T. C. Chak[a,b,1], Stephen E. Harris[a,c], Kristin M. Hultgren[d], Nicholas W. Jeffery[e,f], and Dustin R. Rubenstein[a]

[a]Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY 10027; [b]Department of Biological Sciences, State University of New York College at Old Westbury, Old Westbury, NY 11568; [c]Department of Biology, State University of New York Purchase College, Purchase, NY 10577; [d]Department of Biology, Seattle University, Seattle, WA 98122; [e]Bedford Institute of Oceanography, Fisheries and Oceans Canada, Dartmouth, NS B2Y 4A2, Canada; and [f]Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada

Edited by Joan E. Strassmann, Washington University in St. Louis, St. Louis, MO, and approved April 11, 2021 (received for review December 6, 2020)

Despite progress uncovering the genomic underpinnings of sociality, much less is known about how social living affects the genome. In different insect lineages, for example, eusocial species show both positive and negative associations between genome size and structure, highlighting the dynamic nature of the genome. Here, we explore the relationship between sociality and genome architecture in *Synalpheus* snapping shrimps that exhibit multiple origins of eusociality and extreme interspecific variation in genome size. Our goal is to determine whether eusociality leads to an accumulation of repetitive elements and an increase in genome size, presumably due to reduced effective population sizes resulting from a reproductive division of labor, or whether an initial accumulation of repetitive elements leads to larger genomes and independently promotes the evolution of eusociality through adaptive evolution. Using phylogenetically informed analyses, we find that eusocial species have larger genomes with more transposable elements (TEs) and microsatellite repeats than noneusocial species. Interestingly, different TE subclasses contribute to the accumulation in different species. Phylogenetic path analysis testing alternative causal relationships between sociality and genome architecture is most consistent with the hypothesis that TEs modulate the relationship between sociality and genome architecture. Although eusociality appears to influence TE accumulation, ancestral state reconstruction suggests moderate TE abundances in ancestral species could have fueled the initial transitions to eusociality. Ultimately, we highlight a complex and dynamic relationship between genome and social evolution, demonstrating that sociality can influence the evolution of the genome, likely through changes in demography related to patterns of reproductive skew.

crustacean | ddRAD | sociality | genome architecture | transposon

Eusociality is defined by a reproductive division of labor where some individuals forego independent breeding to cooperatively rear others' young (1). Recent advances in molecular biology have enabled researchers to begin to uncover the genomic underpinnings of eusociality by identifying associated genetic variants and pathways (2–4) and highlighting the importance of recombination (5), gene regulation (6), novel genes (7), genetic accommodation (8), epigenetics (9, 10), and developmental plasticity (11–13) in the transition toward social complexity. Yet, more recently there has been a shift toward also considering the phenotypic (14, 15), demographic (16–18), and genomic consequences (17, 19) of living in complex social groups. Indeed, cooperative group living, or sociality, has been hypothesized to influence the architecture of the genome, including its size and structure, since changes in demography resulting from increased reproductive skew and a reproductive division of labor (20) will impact effective population sizes, recombination rates, and the strength of purifying selection (17, 19).

The genome is dynamic, changing in both size and structure over evolutionary time (21–23). Genome size, the total amount of DNA contained within a haploid chromosome set, is a basic property of every genome that can vary considerably among species, even closely related ones (24). Because eukaryotic genomes often have large and varying quantities of noncoding and repetitive DNA (25), genome size in eukaryotes is generally unrelated to organismal complexity (26) and is instead associated with the abundance of repetitive regions like transposable elements (TEs) (27). Several studies in insects have suggested a relationship between eusociality and genome architecture, including both genome size and structure. For example, ants, which are all eusocial, tend to have smaller genomes than other insects (28). Similarly, across 131 Hymenoptera, eusocial and parasitoid species have smaller genomes than solitary and nonparasitoid species, with the eusocial honey bees *Apis mellifera* and *Apis cerana* having some of the smallest genomes among all Hymenoptera (29). Yet, the relationship between eusociality and genome size is quite different in termites, where despite the fact that eusocial species also have smaller genomes than their noneusocial relatives (30), the genome size of the socially more

Significance

Despite great progress in uncovering the genomic underpinnings of advanced forms of social organization like eusociality, much less is known about how eusociality feeds back to drive genome evolution. Using snapping shrimps that exhibit multiple origins of eusociality and extreme interspecific variation in genome size, we show that eusocial species have larger genomes with more repetitive elements. Although our results support the idea that eusociality influences the accumulation of repetitive elements and an increase in genome size through changes in demography, there is also some evidence that repetitive elements could have also helped fuel the transition to eusociality in some lineages. Our work highlights a fluid relationship between genome and social evolution, demonstrating how eusociality can influence genome evolution and architecture.

Author contributions: S.T.C.C., S.E.H., K.M.H., N.W.J., and D.R.R. designed research; S.T.C.C., S.E.H., K.M.H., and N.W.J. performed research; S.T.C.C. contributed new reagents/analytic tools; S.T.C.C., K.M.H., N.W.J., and D.R.R. analyzed data; and S.T.C.C., S.E.H., K.M.H., N.W.J., and D.R.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: chakt@oldwestbury.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2025051118/-/DCSupplemental.

Published June 7, 2021.

complex *Macrotermes natalensi* is twice that of the socially less complex *Zootermopsis nevadensis* (31). This difference between Hymenoptera and termites in the relationship between social organization and genome size extends to genome structure, where social complexity is associated with a reduced abundance and diversity of TEs in Hymenoptera (32), but an increased abundance of TEs in termites (31). Some of these TEs have been hypothesized to play a causal role in gene family expansion that is associated with the transition to eusociality in termites (33), as they have in other forms of adaptive evolution (34–37). Ultimately, these contrasting patterns in different eusocial insect lineages suggest that the relationship between social organization and genome architecture may be a complex and fluid one that we do not yet fully understand.

Untangling the linkages between social organization and genome architecture requires a group of organisms that shows interspecific variation in both of these traits. Species of sponge-dwelling snapping shrimps in the genus *Synalpheus* not only exhibit great social diversity, they also show some of the most extreme interspecific variation in genome size of any animal group studied. Social organization in *Synalpheus* ranges from pair-living to communal breeding (multiple mating pairs in the same sponge) to eusociality (one or a few queens and a larger number of nonsterile workers of both sexes) (14, 38–40). Although *Synalpheus* shrimps in the West Atlantic *gambarelloides* group represent a relatively young lineage that radiated between ~5 and 7 Mya (41), eusociality has evolved at least four times within this group (40), and eusocial and communal breeding species both evolved independently from pair-living ancestors (42). Furthermore, the genus *Synalpheus* shows a more than fivefold difference in genome size across species (43), ranging from roughly 4 Gb to more than 20 Gb, and their genomes—particularly those of eusocial species—harbor many repetitive elements (44, 45).

Here, we examine the dynamic relationship between eusociality and genome architecture—both genome size and structure—across 33 *Synalpheus* species using phylogenetically informed analyses. First, we determine whether eusocial shrimp species have smaller or larger genomes than noneusocial species (pair-living and communal breeding combined). Next, we use double-digest restriction site-associated DNA sequencing (ddRAD-seq) to extract the abundance of repetitive elements (the proportions of TEs and microsatellite repeats) from the genomes of each species to explore the relationships among social organization, genome structure, and genome size. This reduced-representation approach has been shown to accurately estimate the relative proportions of TEs in the genomes of nonmodel species by comparing this method to whole-genome assemblies and simulated ddRAD-seq markers across arthropods (46). We then use phylogenetic path analysis and ancestral state reconstruction to examine the relative importance of alternative causal relationships linking these traits. Our goal is to determine whether eusociality leads to an accumulation of repetitive elements and an increase in genome size, potentially through changes in demography resulting from a reproductive division of labor (20), or whether an initial accumulation of repetitive elements (i.e., TEs) leads to larger genomes and independently promotes the evolution of eusociality through adaptive evolution (33). Ultimately, exploring the relationship between eusociality and genome architecture will not only help elucidate the genomic consequences of living in complex societies, it will also improve our understanding of the interacting relationship between genome evolution and social evolution.

Results

Eusocial *Synalpheus* Species Have Larger Genomes. After finding moderate but significant phylogenetic signal in genome size among 33 *Synalpheus* species (Blomberg's $K = 0.38$, $P = 0.049$), we fit three alternative models of continuous trait evolution: a Brownian motion model that assumes a phylogenetic random walk under Brownian motion (47), an Ornstein–Uhlenbeck model that assumes a random walk with a central tendency with an attraction strength (48, 49), and a white-noise model that disregards phylogenetic signal and assumes that trait values are random. Akaike information criterion (AIC) comparison indicated that the Ornstein–Uhlenbeck model was better supported than the Brownian motion (ΔAIC = 9.28) or the white-noise models (ΔAIC = 2.82). Since the higher support of the Ornstein–Uhlenbeck model is unlikely to be due to the structure of our data (*SI Appendix*, Fig. S1), factors other than phylogenetic history appear to underlie interspecific genome size variation in *Synalpheus*.

Although *Synalpheus* species exhibit three forms of social organization (pair-living, communal breeding, and eusociality), we were primarily interested in how genome size and repetitive element abundance differ between eusocial and noneusocial species. Therefore, our analyses focused on a priori contrasts to test the differences between 1) eusocial and pair-living species and 2)

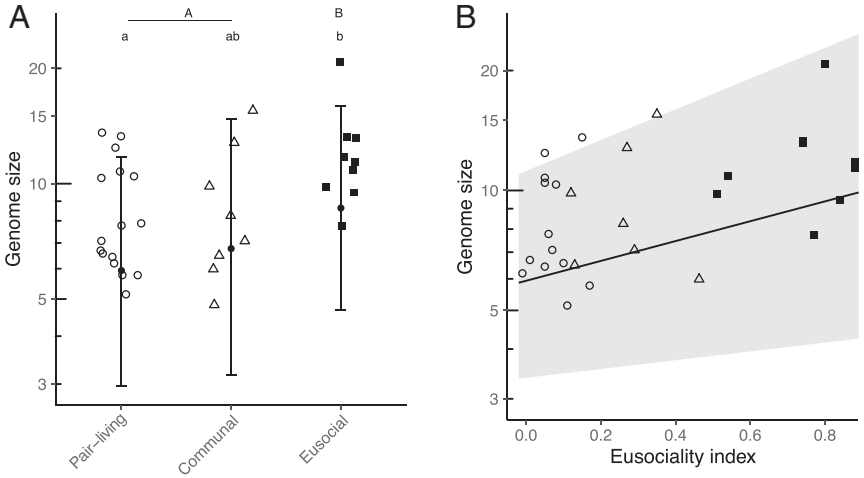


Fig. 1. Relationships between genome size and categorical forms of social organization (*A*) and the eusociality index (*B*). In *A*, shapes represent raw values, closed circles and error bars represent posterior means and 95% CI. In *B*, solid line and gray shade represents the predicted linear relationship and 95% CI. The models control for the effect of egg and body size. In *A*, letters indicate significant differences in the posterior means in pair-wise comparisons between the three forms of social organization observed in *Synalpheus* snapping shrimps (a and b) and planned comparisons between eusocial and noneusocial species (pair-living and communal species combined) (A and B) (pMCMC < 0.05). Symbols represent pair-living (○), communal breeding (△), and eusocial species (■).

eusocial and noneusocial species (i.e., pair-living and communal breeding combined). The comparisons between eusocial and pair-living species were especially important because previous work suggested that eusociality and communal breeding evolved independently from pair-living ancestors along different evolutionary trajectories (42). In addition, we also examined social diversity using the eusociality index, a continuous measure of social organization that captures the reproductive skew within a group (42, 50, 51), making no a priori assumption of social organization. Ultimately, we found that, while controlling for egg and body size, eusocial species had larger genomes than pair-living ($P = 0.009$) and noneusocial species ($P = 0.033$), and that genome size increased as reproductive skew within colonies increased ($P = 0.002$) (Fig. 1*A* and *SI Appendix*, Table S2).

The Genomes of Eusocial *Synalpheus* Species Have More Repetitive Elements. To determine whether an increase in repetitive elements was the mechanism of genome size increase among *Synalpheus* species, we generated ddRAD-seq data (52) and extracted the proportions of TEs and microsatellite repeats in different subclasses and families using the pipeline TERAD (extraction of TE composition from RAD-seq data), which has been shown to be as efficient in estimating the relative proportions of TEs across species as low-coverage whole-genome sequencing (46). Our final dataset included over 500 million paired-end ddRAD reads from 178 samples across 33 species ($n = 2$ to 10 samples per species, median = 4 samples) (*SI Appendix*, Table S1). After quality filtering, each sample had between 10,837 and 14,853,546 reads (median = 2.3 million reads). Since the median number of

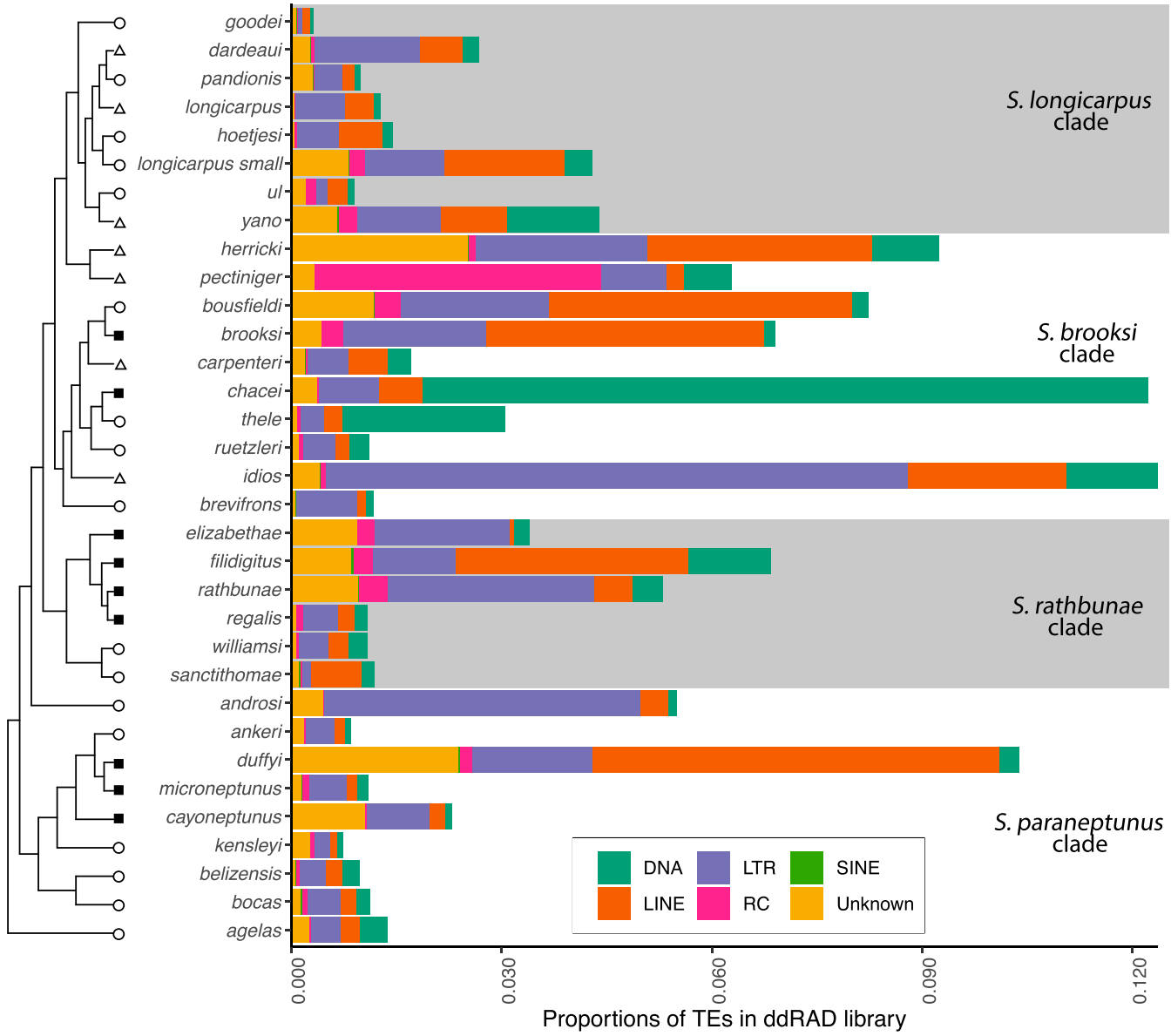


Fig. 2. Phylogenetic distribution of the proportions of TEs in *Synalpheus* species from the four primary groups that independently evolved eusociality. TE subclasses include DNA transposons (DNA), LTR retrotransposons, LINEs, RCs, SINEs, and TEs with unknown subclass. Eusocial species (■) tended to have more TEs than pair-living species (○) and communal breeding species (△). The increase in the relative abundance of TEs is influenced by different TE subclasses in different species. The phylogenetic tree [based on Chak et al. (42)] on the left depicts the relationships among species. Colors in the bar graphs represent major TE subclasses that are defined in the *Inset*. The gray boxes separate the four major *Synalpheus* clades, those with pair-living and eusocial species (the *paraneptunus* and *rathbunae* clades), with pair-living and communal breeding species (*longicarpus* clade), and with pair-living, communal breeding, and eusocial species (*brooksi* clade). *SI Appendix*, Fig. S3 depicts the interspecific difference within each TE subclass.

ddRAD reads per species was not affected by genome size ($F_{1, 31} = 0.17$, $P = 0.69$), the observed variability among samples was likely due to differences in DNA quality or library preparation. The median number of ddRAD reads per species was also not affected by social organization ($F_{2, 30} = 0.54$, $P = 0.59$).

We identified TEs from most major subclasses, including long-terminal repeat retrotransposons (LTR), long interspersed nuclear elements (LINEs), DNA transposons (DNA), and rolling-circle transposons (RC), as well as a variety of unknown TEs (*SI Appendix*, Fig. S3 and Dataset S1). We found very few short interspersed nuclear elements (SINEs) in our dataset, presumably because of the short sequence lengths of these elements. Across TE subclasses, we identified 142 major TE families. Although none of the TE families were associated only with species exhibiting a specific form of social organization, 24 TE families were each unique to only a single shrimp species. The phylogenetic distribution of TE subclasses (Fig. 2 and *SI Appendix*, Fig. S3) showed that species with large numbers of TEs were not clustered within any of the four primary *Synalpheus* clades. Instead, within each of the four primary *Synalpheus* clades, a few species showed a large abundance of TEs. The three shrimp species that had the greatest abundance of TEs (*Synalpheus chacei*, *Synalpheus idios*, and *Synalpheus duffyi*), each had very different compositions of TEs. In particular, *S. chacei* had primarily DNA transposons, *S. idios* had mostly LTR retrotransposons, and *S. duffyi* had largely LINEs, suggesting that TE accumulation is not associated with specific types of TE subclasses. Furthermore, while *S. chacei* and *S. duffyi* are both eusocial, they have very different TE compositions, highlighting how TE accumulation in eusocial species is also not associated with specific types of TE subclasses.

Next, we examined the relationship between eusociality and the abundance of repetitive elements in the genome, while controlling for egg size. We found that eusocial species had a higher proportion of TEs than pair-living ($P = 0.001$) and noneusocial species ($P = 0.034$), and that the proportion of TEs increased as reproductive skew within colonies increased ($P = 0.006$) (Fig. 3, *SI Appendix*, Table S2, and Dataset S2). Within just LTR retrotransposons and TEs of unknown subclass, we found the same general patterns (all Markov Chain Monte Carlo pMCMC < 0.04, except in LTR retrotransposons, eusocial vs. noneusocial [pMCMC = 0.07]) (*SI Appendix*, Table S2). Additionally, we found that eusocial species had a higher proportion of microsatellite repeats than pair-living ($P = 0.011$) and noneusocial species ($P = 0.036$), and that the proportion of microsatellite repeats increased as reproductive skew within colonies increased ($P = 0.043$) (Fig. 3 and *SI Appendix*, Table S2). Consistent with the prediction that the accumulation of repetitive elements explained the differences in genome size between eusocial and noneusocial species, we found that shrimp species with larger genomes had a higher abundance of TEs ($P = 0.004$) (*SI Appendix*, Fig. S4), but not of microsatellite repeats ($P = 0.10$) (*SI Appendix*, Table S2).

TEs Modulate the Relationship between Social Organization and Genome Size. We used phylogenetic path analysis, which allows for the incorporation of all three variables (i.e., social organization, TE abundance, and genome size) in the same path model to test five alternative hypotheses describing the causal relationships between sociality and genome architecture. When modeling social organization either categorically as eusocial vs. noneusocial species or continuously with the eusociality index (as a measure of reproductive skew), the same two models (models 1 and 2 in Table 1) were consistently and equally supported with ΔC statistic information criterion (ΔCICc) values < 2 and P values strongly above 0.05, indicating that the hypothesized causal models provided a good fit to the data (Table 1). In model 1, social organization influences genome size indirectly through the intermediate factor of TE abundance. In model 2, TE abundance independently influences social organization and genome size. Thus, the two models differed in the direction of the causal relationship between social organization and TE abundance. Although the path analysis could not mathematically differentiate between models 1 and 2, path coefficients are standardized, hence comparable for revealing the relative strength of each causal relationship (53). The path coefficients of social organization influencing TE abundance were always higher than that of TE abundance influencing social organization (*SI Appendix*, Table S3), regardless of whether we modeled social organization as a categorical or continuous variable. Models 1 and 2 also yielded the same path coefficient for TE abundance influencing genome size (*SI Appendix*, Table S3). In addition, the path coefficients from both of these models are higher than expected from datasets where the association between the three variables were permutated (*SI Appendix*, Table S4). Together, these results provide evidence that TEs modulate the relationship between social organization and genome size observed across *Synalpheus* species.

Ancestral Species Have Intermediate Genome Sizes and TE Abundances. Since our phylogenetic path models could not rule out the hypothesis that TEs may have played a causal role in the evolution of eusociality in *Synalpheus*, we used ancestral state reconstruction to explore the genome architecture of ancestral *Synalpheus* species. We found that the ancestral nodes of each of the four major lineages in the *gambarelloides* species group had intermediate genome sizes with moderate levels of TEs (Fig. 4). Furthermore, the ancestors of the three clades that include eusocial species (*paraneptunus*, *rathbunae*, and *brooksi* clades) tended to have large genomes and higher TE abundances than the only clade that does not include eusocial species (*longicarpus* clade), suggesting that moderate levels of TEs predated, and hence could potentially have promoted, the evolution of eusociality. Relative to these ancestral species, extant eusocial species tended to show an increase, and noneusocial species a decrease, in genome size and TE abundance, suggesting that eusociality further leads to an accumulation of TEs and large genomes, a result consistent with our phylogenetic path analyses.

Discussion

Eusociality, often considered the pinnacle of animal social evolution, has convergently arisen at least 17 times in arthropods (54). Although a great deal of research has explored the genomic underpinning of eusociality (5–8, 12, 55), eusociality can also produce feed-back on the genome and influence its architecture, including both size and structure (20). Focusing on a group of snapping shrimps that exhibit multiple independent origins of eusociality (56) and large interspecific variation in genome size (43), we found that eusocial species tended to have larger genomes with more TEs and microsatellite repeats than noneusocial species. Ancestral state reconstruction suggests that ancestral species of *Synalpheus* had intermediate genome sizes with moderate levels of TEs, and that both increased over evolutionary time in eusocial species. In contrast, noneusocial species evolved smaller genomes with fewer TEs or maintained similar genome sizes and structures to their ancestors. Together, these results suggest that the relationship among social organization, genome structure (particularly TE abundance), and genome size in snapping shrimps is a fluid one that may have shifted over the past 5 million y, a scenario that is consistent with the idea that eukaryotic genomes tend to be dynamic and can change in both size and structure over evolutionary time (21, 22).

Why do eusocial snapping shrimp species have more repetitive elements in their genomes than their noneusocial relatives? The accumulation of TEs and microsatellite repeats in extant eusocial *Synalpheus* species is most likely the result of reduced effective population sizes due to their reproductive division of

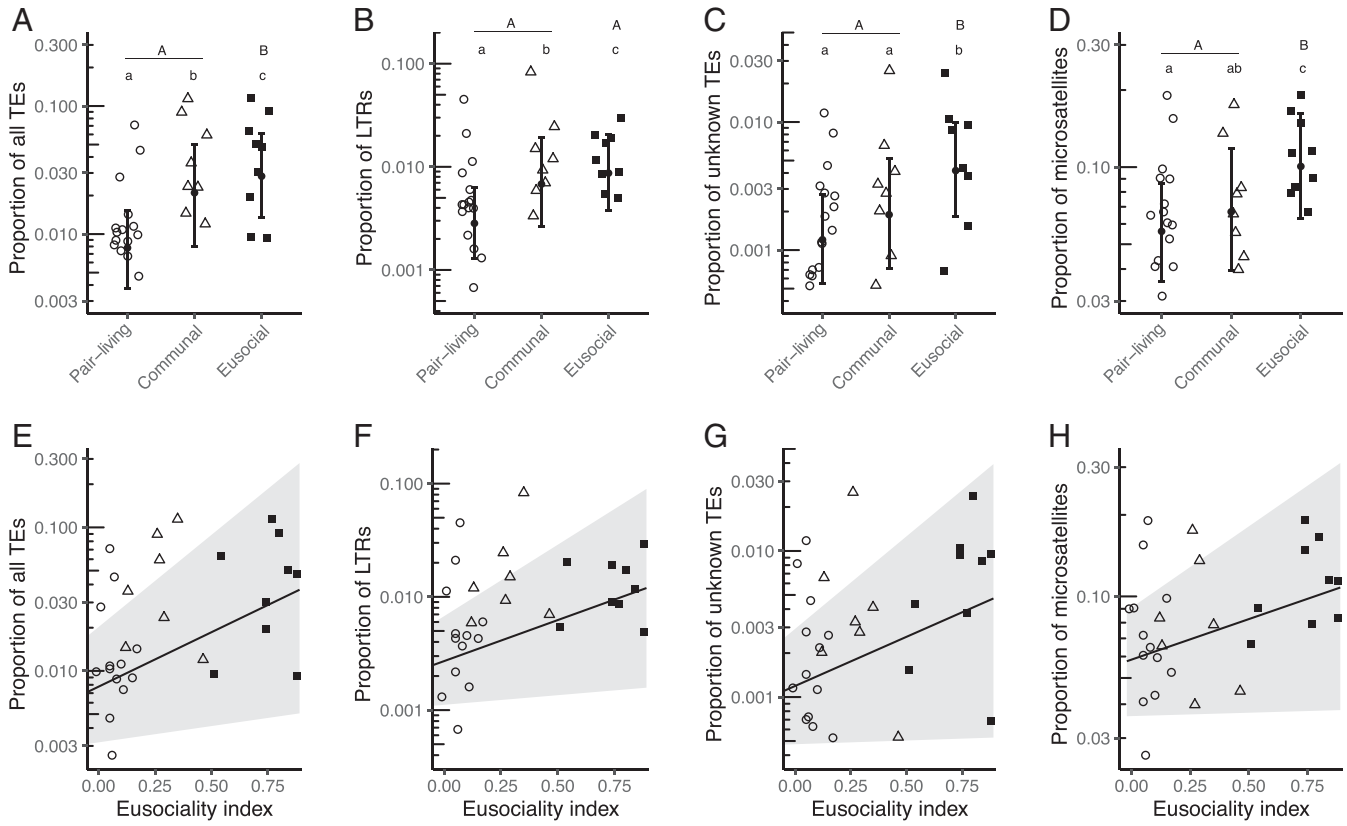


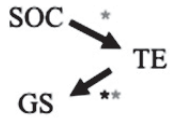
Fig. 3. Relationships between the proportions of repetitive elements across *Synalpheus* species ($n = 33$) with different forms of social organization (*A–D*), and with the eusociality index (*E–H*). The *y* axes are the proportions of all TEs (*A* and *E*), LTRs (*B* and *F*), unknown TE subclass (*C* and *G*), and microsatellite repeats (*D* and *H*). Raw values are shown in shapes (○: pair-living; △: communal breeding; ■: eusocial; jittered along the *x* axes in *A–D*). Posterior means and 95% CI based on phylogenetic mixed models are show as closed circles and error bars in *A–D* and regression lines and shaded areas in *E–H*. In *A–D*, letters indicate significant differences in the posterior means in pair-wise comparisons between forms of social organization (a, b, and c) and planned comparisons between eusocial and noneusocial species (A and B) (pMCMC < 0.05).

labor and high reproductive skew (57–59). Several studies in *Synalpheus* are consistent with the idea that eusocial species have reduced effective population sizes relative to noneusocial species. First, demographic inference using single nucleotide polymorphism data generated from the same ddRAD loci presented here found that eusocial species had lower and more stable effective population sizes over 100,000 generations than noneusocial species (16). Second, eusocial species showed relaxed purifying selection in their mitochondrial protein-coding genes, further indicating that eusocial species had lower effective population sizes than noneusocial species (19). Reduced effective populations have also been similarly inferred in eusocial Hymenoptera (17, 18). The neutral theory of molecular evolution predicts that reduced effective population size lessens the efficiency of selection and allows for the accumulation of deleterious alleles (60), a result that is supported by empirical studies (61, 62). Accordingly, TEs, which generally have weakly deleterious effects, have been predicted and shown to accumulate in species with reduced effective population sizes (62, 63), unless counteracted by TE-suppressing mechanisms (64). Therefore, our finding that eusocial species have larger genomes with more repetitive elements than noneusocial species is consistent with the idea that social living can feed back on the architecture of the genome due to changes in social structure, reproductive division of labor, and demography (20).

Although effective population size is likely to be the driving force behind the patterns we observed, it may not be the sole factor that affects repetitive-element accumulation and genome size, as there are communal breeding species that also have a large number of TEs. For example, *S. idios* is a communal breeding species that has high TE abundance. Intriguingly, it is the only noneusocial species that has relatively large eggs similar in size to eusocial species, which is consistent with our finding that egg size is correlated with genome size. Other factors that affect TE abundance and genome size, including developmental or metabolic rate (65), remain to be further explored in this system.

The finding that eusocial snapping shrimp species have larger genomes with more TEs than noneusocial species is similar to observations in termites (31, 33), but opposite to what has been found in Hymenoptera (29, 32). The pattern of smaller genome sizes in eusocial Hymenoptera has been attributed to the exceptionally high recombination rates in eusocial species (66, 67), which are most pronounced in the honey bee *A. mellifera* (5). Kent and Zayed (68) proposed a conceptual model to explain the relationship between eusociality and high recombination rates, arguing that the early stages in the evolution of eusociality are likely characterized by lower effective population sizes due to reproductive division of labor (57, 58, 69), which in turn leads to increased linkage disequilibrium (70) and an increased frequency of slightly deleterious mutations due to drift (60). Hence, the high recombination rates in eusocial Hymenoptera may be a genomic response to balance these suboptimal conditions because recombination can reduce the interference between linked mutations and enhance the effectiveness of natural selection. Since TE accumulation is predicted to be reduced as recombination rates increase (71), the reduced number of TEs and smaller genomes observed in bees can also be considered a genomic consequence of social living. Because the genome is dynamic and can fluctuate in size over evolutionary time (21, 22), the difference in genome sizes

Table 1. Alternative path models that differ in the direction of causality between eusociality, genome size, and TE accumulation in *Synalpheus* snapping shrimps

	Eusocial vs. non-eusocial				Eusociality index			
Diagram	C statistic	p-value	CIC c	Delta CICc	C statistic	p-value	CIC c	Delta CICc
1. Social organization influences TE abundance, which then influences genome size								
	**1.29**	**0.53**	**13.5**	**0**	**2.86**	**0.24**	**15.5**	**0**
2. TE abundance independently influences social organization and genome size								
	**1.29**	**0.53**	**13.5**	**0**	**2.86**	**0.24**	**15.5**	**0**
3. Social organization influences genome size, which in turn influences TE abundance								
	4.87	0.088	17.1	3.58	4.94	0.085	17.5	2.08
4. Social organization independently influences TE abundance and genome size								
	8.59	0.014	20.8	7.3	5.15	0.076	17.8	2.29
5. Social organization and TE abundance independently influence genome size								
	7.14	0.028	19.4	5.85	7.99	0.018	20.6	5.12

We modeled social organization either categorically as eusocial vs. noneusocial species (pair-living and communal breeding combined) ($n = 33$) or continuously as the eusociality index ($n = 29$). Models 1 and 2 (in bold) were equally supported. C statistics summarize the P values from linear models that reflect the conditional independence statements. A path model was rejected if the P value of the C statistic was below 0.05, meaning that the hypothesized causal model did not provide a good fit to the data. CICc: a modified version of AIC for path models; GS: genome size; SOC: social organization; TE: abundance of transposable elements. Path coefficients and confidence intervals are reported in *SI Appendix*, Table S3. Significant paths are indicated by asterisks (black for models using eusocial vs. noneusocial classifications and gray for models using the eusociality index).

across social lineages may simply reflect different stages of genome size evolution in Hymenoptera, termites, and *Synalpheus*. *Synalpheus* shrimps are likely at an early stage in the evolution of eusociality because of the short evolutionary history of the clade (41) and the lack of a permanently sterile worker caste (14, 38). Yet, eusociality in termites is ancient (72) and many species are in the highly advanced stage of eusociality. Considering that both shrimps and termites are diploid, whereas Hymenoptera are haplodiploid, TEs may be more efficiently purged in haploid males seen in Hymenoptera (73). Although these explanations are not mutually exclusive, they highlight the fact that that sociality can affect the architecture and dynamics of the genome via different routes.

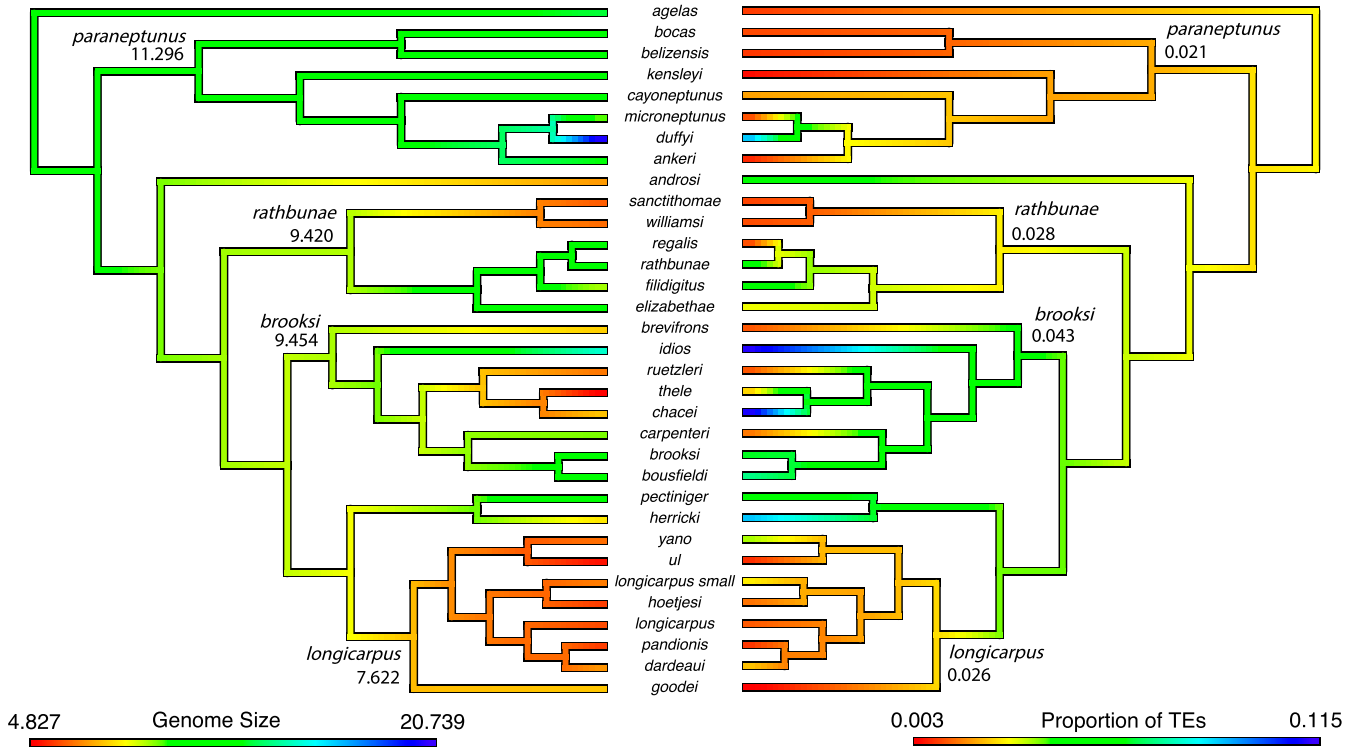


Fig. 4. Ancestral states reconstructions of genome size (*Left*) and TE abundance (*Right*) across 33 *Synalpheus* species. Branch colors indicate reconstructed values according to the scale bars below each phylogeny. Names and values on nodes represent the names and ancestral values of the ancestors of the four major *Synalpheus* clades, those with pair-living and eusocial species (the *paraneptunus* and *rathbunae* clades), with pair-living and communal breeding species (*longicarpus* clade), and with pair-living, communal breeding, and eusocial species (*brooksi* clade).

The presence of large and repetitive genomes in eusocial shrimps and termites (31) shows that eusocial species may be able to tolerate having larger genomes that are more costly to replicate (65) and more TEs that require active defense from the host genome (74). However, since TEs are a source of mutation and genomic rearrangement (75), it has been hypothesized that they could fuel genomic changes that promoted the evolution of eusociality in termites (33). Since the basal branch leading to termites tended to have more TEs in flanking regions of expanded gene families, including ionotropic receptors that are under selection and are differentially expressed between workers and queens, Harrison, et al. (33) hypothesized that TE expansion in the ancestors of termites allowed the evolution of gene families that facilitated their transition to eusociality (33). Indeed, expanded gene families of chemoreceptors have been found to be enriched in TEs in both eusocial termites (33) and ants (76). In fact, there is increasing evidence that TEs are not solely genomic parasites and can instead participate in macroevolutionary events, such as adaptive radiations (36), diversification (37), and speciation (77) in a wide variety of organisms. Since ancestral *Synalpheus* species were likely to have moderate TE abundances, some of these mobile elements could have played an adaptive role in the transition toward eusociality. Ultimately, eusocial snapping shrimps and termites may maintain a high abundance of TEs in their genomes because TEs at least partially fuel and support social evolution. Whether and how TEs are related to the evolution of eusociality in *Synalpheus* remains to be investigated with more knowledge of the complete genomes in these shrimps, which will facilitate studies exploring the locations in the genome of TEs, genes, and TE-derived exons (78, 79).

In *Synalpheus* shrimps, different TE subclasses appear to contribute to TE accumulation in different species. This suggests that the evolution of eusocial species with large genomes resulted from the proliferation of one or a few TE subclasses in noneusocial ancestors with intermediately sized genomes with moderate TE abundances. These TE subclasses broadly differ in the mechanism of replication and enzymology (80). For example, DNA transposons, which are most abundant in the eusocial species *S. chacei*, replicate mostly by "cut-and-paste," whereas LINEs, which are most abundant in another eusocial species *S. duffyi*, replicate by "copy-and-paste" (Fig. 2). The patchy distribution of the predominant TE subclass is consistent with current knowledge of the interacting relationship between TEs and their host genomes. TEs have evolved highly specific targeting mechanisms that allow them to reduce their damage to hosts, while hosts employ various restriction mechanisms to suppress the amplification of TEs (81). Hence, mutations in TEs or hosts that enable TE amplification would generally occur for a specific type of TE. Therefore, deleterious mutations that accumulate in eusocial species due to small effective population sizes may result in massive amplification of a small number of TEs, instead of an increase in all types of TEs, similar to other organisms with genome expansions (82–84). Such a process likely explains why different eusocial species that have high abundances of TEs do not have high proportions of the same TE subclasses.

In addition to the accumulation of TEs in eusocial species, we also observed that eusocial species had more microsatellite repeats than noneusocial species, although their expansion was only marginally related to genome size increase. An association between TEs and microsatellite repeats is not surprising because TEs have been identified to be one of the primary progenitors of microsatellite repeats in crustaceans (85), insects (86, 87), plants (88, 89), and primates (90, 91). Among crustaceans, a similar association between microsatellite repeats and TEs has been suggested to explain the variable flanking sequences of microsatellite repeats that have led to difficult primer design for PCR amplification (92). Similarly, microsatellite loci in *Synalpheus* are

often polysomic, most prominently in species with larger genomes (44, 45). If the primer sites for a microsatellite locus were located inside a TE, then other regions with the same TE might also be amplified, resulting in polysomic loci. Since our TE data are based on short ddRAD sequences, we cannot adequately evaluate the proximity of TEs and microsatellite repeats in shrimps. Nevertheless, the larger genomes in eusocial *Synalpheus* species are primarily the result of TE accumulation and less so the accumulation of microsatellite repeats. Since microsatellite repeats can have a functional impact related to adaptive evolution (85) and sociality (93), their effects on sociality in shrimps remain to be further explored, something that will once again require sequencing whole genomes.

In conclusion, we find that eusocial snapping shrimp species tend to have larger genomes with more repetitive elements than noneusocial species. Eusociality appears to influence the accumulation of TEs in particular and leads to an increase in genome size. Although this relationship is presumably due to reduced effective population sizes in eusocial species resulting from a reproductive division of labor and high reproductive skew (16), the moderate abundances of TEs in ancestral *Synalpheus* species could have also helped fuel the transition to eusociality. Future work should further explore the idea that TEs could play an adaptive role in social evolution, not only in snapping shrimps but also in other eusocial lineages. Ultimately, our work highlights a complex and interacting relationship between genome evolution and social evolution, providing empirical support for the idea that sociality can feed back on genome architecture through changes in demography.

Materials and Methods

***Synalpheus* Species.** *Synalpheus* shrimps were collected from six Caribbean countries (Belize, Barbados, Curaçao, Jamaica, Panama, and the United States) between 1993 and 2016 (*SI Appendix*, Table S1). Details of field-collection protocols have been reported previously (94). Our study used 33 of 45 (73%) *Synalpheus* species in the West Atlantic *gambarelloides* group, including all 9 described eusocial species that represent 4 independent origins of eusociality within this clade (40). Categorization of social organization (pair-living, communal breeding, and eusociality) was based on previous reports (14, 39, 40). We calculated a modified version of Keller and Perrin's eusociality index as 1 − ([2 × number of ovigerous females]/colony size) (in the sense of refs. 42, 50). In the analyses using the eusociality index, we removed four species (three pair-living and one communal breeding) in which only five or fewer colonies were sampled (remaining $n = 29$), since the calculation of the eusociality index has been shown previously to be inaccurate when sample sizes are low (42).

Genome Size Evolution. Genome sizes of 33 *Synalpheus* species were obtained from Jeffery et al. (43). Briefly, the genome size of each species (7 to 45 individuals per species) was quantified as a C value by comparing the absorbance of Feulgen-stained cell nuclei from shrimps to standards (chicken and rainbow trout) of known genome sizes. To determine if there is phylogenetic signal in the variation in genome sizes across *Synalpheus* species, we calculated Blomberg's K (95) using *phylosig* in the R package phytools v0.6-99 (96). Blomberg's K ranges from 0 to infinity, where $K < 1$ means that species resemble each other less than expected from the Brownian motion of continuous trait evolution (47), and $K > 1$ means that species resemble each other more than expected from Brownian motion. We also fit three alternative models of character evolution (Brownian motion, Ornstein–Uhlenbeck, and white noise) using *fitContinuous* in the R package geiger v2.0.6.4 (97). Since model choice based on information criteria can have high error rates due to the underlying structure of the data (98, 99), we compared the log-likelihood ratios ($\sigma = -2(\ln L_{BM} - \ln L_{OU})$) between the observed dataset and 1,000 sets of simulated data under the Brownian motion model [using *rTraitCont* in *ape* (100)]. We used a published Bayesian consensus tree of *Synalpheus* species, constructed with 16S, 18S, and COI sequence data (42).

Since biotic factors, such as body and egg size, may also be correlated with genome size (21, 101), we tested whether differences in genome size were related to carapace length and egg volume, as well as our primary variable of interest: social organization. Although egg volume may change with developmental stage (102), maximum and mean egg volumes were strongly and positively correlated across *Synalpheus* species ($P < 0.001$) (*SI Appendix*, Fig. S2). We used mean egg volume as our estimate of egg size because it was less prone to measurement error. We used Shapiro–Wilk tests (103) to assess normality and log-transformed body, egg, and genome size. We performed phylogenetic mixed-model analysis using *MCMCglmm* v2.29 (104) to control for phylogenetic nonindependence among species. We used a weakly informative prior (variance parameters, $V = 1$, degree of belief, $\nu = 0.002$) and ran 2,000,000 MCMC iterations with 50,000 iterations of burn-in and a thinning interval of 250. Pairwise differences between categorial forms of social organization (eusocial vs. pair-living, eusocial vs. noneusocial) were tested by comparing the posterior distributions of genome size between each group and calculating a pMCMC value as twice the probability that the posterior distribution of the difference is above or below zero (104). When testing the relationship between genome size and social organization using the continuous variable eusociality index as a measure of reproductive skew, the posterior distribution of the slope was used to calculate a pMCMC value.

Repetitive Element Abundances Across Species. We used the TERAD pipeline to estimate the relative abundance of TEs from ddRAD data (46), a method that is ideal for lineages with large genome sizes when low-coverage whole-genome sequencing is uneconomical. This reduced-representation approach has been shown to accurately estimate the relative proportions of TEs by comparing to whole-genome assemblies and simulated ddRAD-seq markers across arthropods (46). For each of 33 *Synalpheus* species, we sampled 2 to 10 individuals per locality. We extracted genomic DNA from several walking legs of alcohol-preserved specimens using Qiagen DNAeasy Tissue Kits (Qiagen) and quantified DNA on a Qubit 3.0 Fluorometer with the dsDNA assay (ThermoFisher Scientific). For ddRAD preparation, we followed the protocol in Peterson et al. (105) using the restriction enzymes EcoRI and MspI, and a wide size selection criterion (338 to 414 bp). Briefly, we digested 1,000 ng of genomic DNA with EcoRI and MspI (New England Biolabs) and cleaned the digested DNA using Agencourt AMPure XP beads (Beckman Coulter Life Sciences). We ligated the double-digested DNA with barcoded adaptors, pooled barcoded samples, and bead-cleaned before size selection using a Pippin Prep and a dye-free cassette (CDF2010, Sage Science). We performed a 10-cycle PCR using a Phusion PCR kit according to manufacturer protocols (New England Biolabs) with multiplexed primers and adjusted PCR products to 10 μM. Genomic libraries were sequenced on either an Illumina HiSeq2500 (125-bp pair-end, New York Genome Center) or an Illumina HiSEq. 3000 (150-bp pair-end, Center for Genome Research and Biocomputing, Oregon State University).

We used *process_radtags* in Stacks v2.1 (106) to demultiplex and clean raw reads. All reads were trimmed to 120 bp. We used the TERAD pipeline (46) to extract the proportions of TEs in different subclasses and families. To identify TEs from *Synalpheus* shrimps, we used a custom repeat database that included the Repbase arthropod database, as well as a database generated using RepeatModeler (107) from three decapods whole-genome assemblies: *Eriocheir sinensis* (108), *Neocaridina denticulate* (109), and *Procambarus virginalis* (110). Due to greater coverage of the genome, species with a higher median number of ddRAD reads had more TE families ($F_{1,\ 31} = 5.75$, $P = 0.02$) but not more major TE classes ($F_{1,\ 31} = 0.96$, $P = 0.11$). Because ddRAD may not sample TEs evenly across samples, especially those with large genomes, there was considerable variation in the proportions of TEs across samples (SDs of the proportions of TE across samples for each species: median = 0.004, range = 0.0005 to 0.028). Therefore, we used the maximum proportion of each major TE subclass across samples of the same species to better reflect the proportions of TEs within a species. Microsatellite repeats were also identified in RepeatMasker using *tandem repeat finder* v4.09 (111) and reported from the TERAD pipeline.

To examine the relationship between eusociality and genome structure, we determined whether the proportions of TEs and microsatellite repeats 1) differed between eusocial and pair-living species, 2) differed between eusocial and noneusocial species (pair-living and communal breeding species combined), and 3) were correlated with the eusociality index (a measure of reproductive skew), while controlling for mean egg volume, using MCMCglmm. We performed analyses for all TEs combined and separately for TEs of each major subclass. Finally, to determine whether species with larger genomes had more repetitive elements, we examined the correlation between genome size and the proportion of TEs or microsatellite repeats, while controlling for mean egg volume in our linear model.

Phylogenetic Path Analysis. To jointly evaluate the relationships among social organization, TE abundance, and genome size in *Synalpheus* shrimps, we used phylogenetic path analysis (53, 112) to describe a set of five models that hypothesized causal relationships between social state and both genome states. In model 1, social organization influences genome size indirectly through the intermediate factor of TE abundance, whereas in model 2, TE

abundance independently influences social organization and genome size. In models 3 to 5, social organization directly influences genome size (Table 1). Although models 1 and 2 differed in directionality between social organization and TE abundance, they generated identical independence statements and linear models. Models 3 to 5 differed in the directionality and in how TE abundance was modeled. We modeled social organization using either a categorical classification of eusocial vs. noneusocial species (pair-living and communal breeding combined) or the eusociality index as a continuous variable.

Based on the hypothesized causal relationship between variables, each model led to a minimum set of conditional independence statements (i.e., "d-separated" statements) that can be formulated and tested as phylogenetic generalized least squares models (53, 112). For each model, Fisher's C statistic, which follows a χ^2 distribution, was calculated to summarize the *P* values of the linear models. A path model was rejected if the *P* value of the C statistic was below 0.05, meaning that the hypothesized causal model did not provide a good fit to the data (i.e., the variables in the "independence statement" are not independent). Models were compared based on information theory using CICc, a modified version of AIC for path models (113). Models with $\Delta CICc < 2$ mean that their implied causal relationships are equally supported. Typically, each path model leads to a minimum set of conditional independence statements (i.e., d-separated statements) that can be formulated and tested as phylogenetic generalized least-squares models (53, 112). However, since models 1 and 2 generate identical independence statements and linear models, we also compared coefficients of the causal paths to determine the relative strength of a causal relationship between variables. For example, a path coefficient of 0.4 from variables *x* to *y* means that an increase by one SD in *x* will lead to a 0.4 increase of SD in *y*. Therefore, the path coefficient is different from the partial coefficient in linear regression, such that the former describes the causal relationship in terms of the correlated variance between two variables, while the latter describes the correlated change in magnitude. We used the R package *phylopath* v1.1.2 (114) to perform phylogenetically controlled path analyses and ran 2,000 parametric bootstraps to obtain confidence intervals (CIs) of the path coefficients for each model.

Because path analysis cannot include a null model with no causal relationships (i.e., paths) between variables, we generated a permutation dataset (n = 5,000) by shuffling the three variables (social organization, TE abundance, and genome size). We ran the path analysis as described above and calculated the 95% CIs of the path coefficients. An observed path coefficient is significantly different from the null model when it lies outside of the CI.

Ancestral State Reconstruction. To directly examine the evolutionary changes in TE abundance and genome size across *Synalpheus* shrimps, we reconstructed the ancestral states of these variables across the phylogeny. We used the anc.ML method implemented in the R package *phytools* (96) to calculate the maximum-likelihood estimates based on the Ornstein–Uhlenbeck model, and then we visualized the estimates on the *Synalpheus* phylogeny (115). Ancestral values based on the Brownian motion model were not numerically different from those of the Ornstein–Uhlenbeck model. We were primarily interested in the ancestral states of, and the changes within, the four major *Synalpheus* clades, which led to either clades with pair-living and eusocial species (the *paraneptunus* and *rathbunae* clades), with pair-living and communal breeding species (*longicarpus* clade), and with pair-living, communal breeding, and eusocial species (*brooksi* clade).

Data Availability. The sequence reported in this paper has been deposited in the National Center for Biotechnology Information's Sequence Read Archive, https://www.ncbi.nlm.nih.gov/sra (accession no. PRJNA560035) (52). All other study data are available in the main text and supporting information.

ACKNOWLEDGMENTS. S.T.C.C. was supported by the Simons Foundation via the Life Sciences Research Foundation. S.E.H. was supported by the Frontiers of Science fellowship at Columbia University. D.R.R. was supported by the US NSF (IOS-1121435, IOS-1257530, IOS-1439985). K.M.H. was supported by a Natural Sciences grant from the Murdock Charitable Trust. This work made use of the high-performance computing resources from Columbia University.

1. E. O. Wilson, *The Insect Societies* (Belknap Press of Harvard University Press, Cambridge, MA, 1971).
2. A. J. Berens, J. H. Hunt, A. L. Toth, Comparative transcriptomics of convergent evolution: Different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* **32**, 690–703 (2015).
3. V. Chandra *et al*., Social regulation of insulin signaling and the evolution of eusociality in ants. *Science* **361**, 398–402 (2018).
4. M. R. Warner, L. Qiu, M. J. Holmes, A. S. Mikheyev, T. A. Linksvayer, Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. *Nat. Commun.* **10**, 2651 (2019).
5. C. F. Kent, S. Minaei, B. A. Harpur, A. Zayed, Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18012–18017 (2012).
6. D. F. Simola *et al*., Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* **23**, 1235–1247 (2013).
7. B. A. Harpur *et al*., Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2614–2619 (2014).
8. B. M. Jones, C. J. Kingwell, W. T. Wcislo, G. E. Robinson, Caste-biased gene expression in a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of eusociality. *Proc. Biol. Sci.* **284**, 20162228 (2017).
9. D. F. Simola *et al*., Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* **351**, aac6633 (2016).
10. H. Yan *et al*., Eusocial insects as emerging models for behavioural epigenetics. *Nat. Rev. Genet.* **15**, 677–688 (2014).
11. K. M. Kapheim *et al*., Developmental plasticity shapes social traits and selection in a facultatively eusocial bee. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13615–13625 (2020).
12. N. Terrapon *et al*., Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* **5**, 3636 (2014).
13. K. M. Kapheim, Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Curr. Opin. Insect Sci.* **13**, 24–32 (2016).
14. S. T. C. Chak, J. E. Duffy, D. R. Rubenstein, Reproductive skew drives patterns of sexual dimorphism in sponge-dwelling snapping shrimps. *Proc. Biol. Sci.* **282**, 20150342 (2015).
15. D. R. Rubenstein, I. J. Lovette, Reproductive skew and selection on female ornamentation in social species. *Nature* **462**, 786–789 (2009).
16. S. T. C. Chak, S. E. Harris, J. E. Duffy, K. M. Hultgren, D. R. Rubenstein, Demographic inference provides insights into the extirpation and ecological dominance of eusocial snapping shrimps. *bioRxiv* [Preprint] (2020). https://www.biorxiv.org/content/10.1101/2020.09.07.283994v1 (Accessed 9 September 2020).
17. A. Weyna, J. Romiguier, Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and pollinating bees. *bioRxiv* [Preprint] (2020). https://www.biorxiv.org/content/10.1101/2020.04.14.038893v1.full (Accessed 15 April 2020).
18. M. A. Imrit, K. A. Dogantzis, B. A. Harpur, A. Zayed, Eusociality influences the strength of negative selection on insect genomes. *Proc. Biol. Sci.* **287**, 20201512 (2020).
19. S. T. C. Chak, J. A. Baeza, P. Barden, Eusociality shapes convergent patterns of molecular evolution across mitochondrial genomes of snapping shrimps. *Mol. Biol. Evol.* **38**, 1372–1383 (2021).
20. D. R. Rubenstein *et al*., Coevolution of genome architecture and social behavior. *Trends Ecol. Evol.* **34**, 844–855 (2019).
21. T. R. Gregory, "Genome size evolution in animals" in *The Evolution of the Genome*, T. R. Gregory, Ed. (Academic Press, Burlington, MA, 2005), chap. 1, pp. 3–87.
22. A. Kapusta, A. Suh, C. Feschotte, Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1460–E1469 (2017).
23. J. Blommaert, Genome size evolution: Towards new model systems for old questions. *Proc. Biol. Sci.* **287**, 20201441 (2020).
24. T. R. Gregory, *The Evolution of the Genome* (Elsevier Science, 2011).
25. T. R. Gregory, Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* **76**, 65–101 (2001).
26. C. A. Thomas Jr, The Genetic Organization of Chromosomes, The genetic organization of chromosomes. *Annu. Rev. Genet.* **5**, 237–256 (1971).
27. M. G. Kidwell, Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49–63 (2002).
28. N. D. Tsutsui, A. V. Suarez, J. C. Spagna, J. S. Johnston, The evolution of genome size in ants. *BMC Evol. Biol.* **8**, 64 (2008).
29. A. M. Ardila-Garcia, G. J. Umphrey, T. R. Gregory, An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. *Insect Mol. Biol.* **19**, 337–346 (2010).
30. S. Koshikawa, S. Miyazaki, R. Cornette, T. Matsumoto, T. Miura, Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). *Naturwissenschaften* **95**, 859–867 (2008).
31. J. Korb *et al*., A genomic comparison of two termites with different social complexity. *Front. Genet.* **6**, 9 (2015).
32. K. M. Kapheim *et al*., Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143 (2015).
33. M. C. Harrison *et al*., Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat. Ecol. Evol.* **2**, 557–566 (2018).
34. A. Koga, A. Iida, H. Hori, A. Shimada, A. Shima, Vertebrate DNA transposon as a natural mutator: The medaka fish Tol2 element contributes to genetic variation without recognizable traces. *Mol. Biol. Evol.* **23**, 1414–1419 (2006).
35. S. E. Staton, J. M. Burke, Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. *BMC Genomics* **16**, 623 (2015).
36. N. Feiner, Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proc. Biol. Sci.* **283**, 20161555 (2016).
37. L. Kang *et al*., Genomic divergence and adaptive convergence in *Drosophila simulans* from Evolution Canyon, Israel. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11839–11844 (2019).

38. S. T. C. Chak, D. R. Rubenstein, J. E. Duffy, Social control of reproduction and breeding monopolization in the eusocial snapping shrimp *Synalpheus elizabethae*. *Am. Nat.* **186**, 660–668 (2015).
39. S. T. C. Chak, J. E. Duffy, "Crustacean social evolution" in *Reference Module in Life Sciences*, J. C. Choe, Ed. (Elsevier, 2017), pp. 641–650.
40. K. M. Hultgren, J. E. Duffy, D. R. Rubenstein, "Sociality in shrimps" in *Comparative Social Evolution*, D. R. Rubenstein, P. Abbot, Eds. (Cambridge University Press, Cambridge, 2017), pp. 224–249.
41. C. L. Morrison, R. Ríos, J. E. Duffy, Phylogenetic evidence for an ancient rapid radiation of Caribbean sponge-dwelling snapping shrimps (*Synalpheus*). *Mol. Phylogenet. Evol.* **30**, 563–581 (2004).
42. S. T. C. Chak, J. E. Duffy, K. M. Hultgren, D. R. Rubenstein, Evolutionary transitions towards eusociality in snapping shrimps. *Nat. Ecol. Evol.* **1**, 0096 (2017).
43. N. W. Jeffery, K. Hultgren, S. T. C. Chak, T. R. Gregory, D. R. Rubenstein, Patterns of genome size variation in snapping shrimp. *Genome* **59**, 393–402 (2016).
44. D. R. Rubenstein, B. V. McCleery, J. E. Duffy, Microsatellite development suggests evidence of polyploidy in the social sponge-dwelling snapping shrimp *Zuzalpheus brooksi*. *Mol. Ecol. Resour.* **8**, 890–894 (2008).
45. K. M. Gaynor *et al*., Development of genome- and transcriptome-derived microsatellites in related species of snapping shrimps with highly duplicated genomes. *Mol. Ecol. Resour.* **17**, e160–e173 (2017).
46. S. T. C. Chak, D. R. Rubenstein, TERAD: Extraction of transposable element composition from RADseq data. *Mol. Ecol. Resour.* **19**, 1681–1688 (2019).
47. J. Felsenstein, Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973).
48. T. F. Hansen, Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).
49. M. A. Butler, A. A. King, Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *Am. Nat.* **164**, 683–695 (2004).
50. J. E. Duffy, K. S. Macdonald, Kin structure, ecology and the evolution of social organization in shrimp: A comparative analysis. *Proc. Biol. Sci.* **277**, 575–584 (2010).
51. L. Keller, N. Perrin, Quantifying the level of eusociality. *Proc. Biol. Sci.* **260**, 311–315 (1995).
52. S. T. C. Chak, ddRAD sequences from snapping shrimps in the genus Synalpheus. National Center for Biotechnology Information Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA560035. Deposited 13 August 2019.
53. A. Gonzalez-Voyer, A. Von Hardenberg, "An introduction to phylogenetic path analysis" in *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology*, L. Z. Garamszegi, Ed. (Springer, 2014), pp. 201–229.
54. R. Crozier, P. Pamilo, *Evolution of Social Insect Colonies Sex Alocation and Kin Selection* (Oxford University Press, New York, NY, 1996).
55. B. J. Fischman, S. H. Woodard, G. E. Robinson, Molecular evolutionary analyses of insect societies. *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 2), 10847–10854 (2011).
56. J. E. Duffy, C. L. Morrison, R. Ríos, Multiple origins of eusociality among sponge-dwelling shrimps (*Synalpheus*). *Evolution* **54**, 503–516 (2000).
57. J. Romiguier *et al*., Population genomics of eusocial insects: The costs of a vertebrate-like effective population size. *J. Evol. Biol.* **27**, 593–603 (2014).
58. L. Bromham, R. Leys, Sociality and the rate of molecular evolution. *Mol. Biol. Evol.* **22**, 1393–1402 (2005).
59. M. Lynch, J. S. Conery, The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
60. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).
61. D. Willemsen, R. Cui, M. Reichard, D. R. Valenzano, Intra-species differences in population size shape life history and genome evolution. *eLife* **9**, e55794 (2020).
62. T. Lefébure *et al*., Less effective selection leads to larger genomes. *Genome Res.* **27**, 1016–1028 (2017).
63. M. Lynch, *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA, 2007).
64. S. Lamichhaney *et al*., A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011649118 (2021).
65. F. Dufresne, N. Jeffery, A guided tour of large genome size in animals: What we know and where we are heading. *Chromosome Res.* **19**, 925–938 (2011).
66. L. Wilfert, J. Gadau, P. Schmid-Hempel, Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98**, 189–197 (2007).
67. A. Sirviö, J. S. Johnston, T. Wenseleers, P. Pamilo, A high recombination rate in eusocial Hymenoptera: Evidence from the common wasp Vespula vulgaris. *BMC Genet.* **12**, 95 (2011).
68. C. F. Kent, A. Zayed, Evolution of recombination and genome structure in eusocial insects. *Commun. Integr. Biol.* **6**, e22919 (2013).
69. R. E. Chapman, A. F. G. Bourke, The influence of sociality on the conservation biology of social insects. *Ecol. Lett.* **4**, 650–662 (2001).
70. P. D. Keightley, S. P. Otto, Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
71. E. S. Dolgin, B. Charlesworth, The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**, 2169–2177 (2008).
72. M. S. Engel, P. Barden, M. L. Riccio, D. A. Grimaldi, Morphologically specialized termite castes and advanced sociality in the early cretaceous. *Curr. Biol.* **26**, 522–530 (2016).
73. M. Beye *et al*., Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* **16**, 1339–1344 (2006).
74. D. Elsner, K. Meusemann, J. Korb, Longevity and transposon defense, the case of termite reproductives. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5504–5509 (2018).
75. G. Bourque *et al*., Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
76. S. K. McKenzie, D. J. C. Kronauer, The genomic architecture and molecular evolution of ant odorant receptors. *Genome Res.* **28**, 1757–1765 (2018).
77. A. Serrato-Capuchina, D. R. Matute, The role of transposable elements in speciation. *Genes (Basel)* **9**, 254 (2018).
78. N. Sela, B. Mersch, A. Hotz-Wagenblatt, G. Ast, Characteristics of transposable element exonization within human and mouse. *PLoS One* **5**, e10907 (2010).
79. J. Schmitz, J. Brosius, Exonization of transposed elements: A challenge and opportunity for evolution. *Biochimie* **93**, 1928–1934 (2011).
80. T. Wicker *et al*., A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
81. H. L. Levin, J. V. Moran, Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12**, 615–627 (2011).
82. C. Sun *et al*., LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **4**, 168–183 (2012).
83. C. Vitte, O. Panaud, H. Quesneville, LTR retrotransposons in rice (Oryza sativa, L.): Recent burst amplifications followed by rapid DNA loss. *BMC Genomics* **8**, 218 (2007).
84. S. E. Staton *et al*., The sunflower (Helianthus annuus L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* **72**, 142–153 (2012).
85. J. Yuan *et al*., Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp. *Commun. Biol.* **4**, 186 (2021).
86. J. Wilder, H. Hollocher, Mobile elements and the genesis of microsatellites in dipterans. *Mol. Biol. Evol.* **18**, 384–392 (2001).
87. W. T. Tay, G. T. Behere, P. Batterham, D. G. Heckel, Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evol. Biol.* **10**, 144 (2010).
88. L. Ramsay *et al*., Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **17**, 415–425 (1999).
89. H. Akagi, Y. Yokozeki, A. Inagaki, K. Mori, T. Fujimura, Micron, a microsatellite-targeting transposable element in the rice genome. *Mol. Genet. Genomics* **266**, 471–480 (2001).
90. M. Ahmed, P. Liang, Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp. Funct. Genomics* **2012**, 947089 (2012).
91. S. S. Arcot, Z. Wang, J. L. Weber, P. L. Deininger, M. A. Batzer, Alu repeats: A source for the genesis of primate microsatellites. *Genomics* **29**, 136–144 (1995).
92. D. A. Bailie, P. A. Prodöhl, H. Fletcher, High incidence of cryptic repeated elements in microsatellite flanking regions of galatheid genomes and its practical implications for molecular marker development. *J. Crustacean Biol.* **30**, 664–672 (2010).
93. E. A. D. Hammock, L. J. Young, Functional microsatellite polymorphism associated with divergent social structure in vole species. *Mol. Biol. Evol.* **21**, 1057–1063 (2004).
94. K. S. Macdonald, R. Ríos, J. E. Duffy, Biodiversity, host specificity, and dominance by eusocial species among sponge-dwelling alpheid shrimp on the Belize Barrier Reef. *Divers. Distrib.* **12**, 165–178 (2006).
95. S. P. Blomberg, T. Garland Jr, A. R. Ives, Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
96. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
97. M. W. Pennell *et al*., geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218 (2014).
98. C. Boettiger, G. Coop, P. Ralph, Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* **66**, 2240–2251 (2012).
99. N. Cooper, G. H. Thomas, C. Venditti, A. Meade, R. P. Freckleton, A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc. Lond.* **118**, 64–77 (2016).
100. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
101. D. C. Hardie, P. D. N. Hebert, Genome-size evolution in fishes. *Can. J. Fish. Aquat. Sci.* **61**, 1636–1646 (2004).
102. I. S. Wehrtmann, G. Kattner, Changes in volume, biomass, and fatty acids of developing eggs in *Nauticaris magellanica* (decapoda: Caridea): A latitudinal comparison. *J. Crustac. Biol.* **18**, 413–422 (1998).
103. S. S. Shapiro, M. B. Wilk, An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).
104. J. D. Hadfield, MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
105. B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra, Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135 (2012).
106. N. C. Rochette, A. G. Rivera-Colón, J. M. Catchen, Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *bioRxiv* [Preprint] (2019). https://www.biorxiv.org/content/10.1101/615385v1 (Accessed 22 April 2019).
107. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. http://www.repeatmasker.org. Accessed 16 September 2019.
108. L. Song *et al*., Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. *Gigascience* **5**, 5 (2016).
109. N. J. Kenny *et al*., Genomic sequence and experimental tractability of a new decapod shrimp model, *Neocaridina denticulata*. *Mar. Drugs* **12**, 1419–1437 (2014).
110. J. Gutekunst *et al*., Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* **2**, 567–573 (2018).
111. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
112. B. Shipley, *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference* (Cambridge University Press, 2002).
113. B. Shipley, The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology* **94**, 560–564 (2013).
114. W. van der Bijl, *phylopath*: Easy phylogenetic path analysis in R. *PeerJ* **6**, e4718 (2018).
115. L. J. Revell, Two new graphical methods for mapping trait evolution on phylogenies. *Methods Ecol. Evol.* **4**, 754–759 (2013).