MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 4532–4548 doi: 10.1111/mec.12398

Ecological speciation in anemone-associated snapping shrimps (*Alpheus armatus* species complex)

C. HURT,*¶ K. SILLIMAN,* A. ANKER†** and N. KNOWLTON‡§

*Cox Science Center, University of Miami, 1301 Memorial Drive, Miami, FL 33146, USA, †Instituto de Ciências do Mar – Labomar, Universidade Federal do Ceará, Avenida da Abolição 3207, CEP 60.165-081, Fortaleza, Brazil, ‡Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, MRC 163, PO Box 37012, Washington, DC 20013-7012, USA, §Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, University of California, La Jolla, San Diego, CA 92093-0202, USA

Abstract

Divergent natural selection driven by competition for limited resources can promote speciation, even in the presence of gene flow. Reproductive isolation is more likely to result from divergent selection when the partitioned resource is closely linked to mating. Obligate symbiosis and host fidelity (mating on or near the host) can provide this link, creating ideal conditions for speciation in the absence of physical barriers to dispersal. Symbiotic organisms often experience competition for hosts, and host fidelity ensures that divergent selection for a specific host or host habitat can lead to speciation and strengthen pre-existing reproductive barriers. Here, we present evidence that diversification of a sympatric species complex occurred despite the potential for gene flow and that partitioning of host resources (both by species and by host habitat) has contributed to this diversification. Four species of snapping shrimps (*Alpheus armatus*, *A. immaculatus*, *A. polystictus* and *A. roquensis*) are distributed mainly sympatrically in the Caribbean, while the fifth species (*A. rudolphi*) is restricted to Brazil. All five species are obligate commensals of sea anemones with a high degree of fidelity and ecological specificity for host species and habitat. We analysed sequence data from 10 nuclear genes and the mitochondrial COI gene in 11–16 individuals from each of the Caribbean taxa and from the only available specimen of the Brazilian taxon. Phylogenetic analyses support morphology-based species assignments and a well-supported Caribbean clade. The Brazilian *A. rudolphi* is recovered as an outgroup to the Caribbean taxa. Isolation–migration coalescent analysis provides evidence for historical gene flow among sympatric sister species. Our data suggest that both selection for a novel host and selection for host microhabitat may have promoted diversification of this complex despite gene flow.

Keywords: *Alpheus*, Brazil, Caribbean, host-shift, isolation migration, sea anemone, symbiosis, sympatric speciation

Received 19 July 2012; revision received 29 April 2013; accepted 3 May 2013

Introduction

Sympatrically distributed sister species are a common phenomenon in marine, benthic invertebrate communities (Knowlton 1993; Hellberg 1998; Taylor & Hellberg 2005; Faucci *et al.* 2007; Bird *et al.* 2011; Krug 2011), suggesting that reproductive barriers may accumulate despite the potential for gene flow. Marine invertebrates with highly dispersive planktonic larvae, large population sizes and lack of absolute barriers to gene flow should have few opportunities for speciation, yet coral reef invertebrate communities are spectacularly diverse. These patterns raise questions regarding the factors promoting diversification in marine organisms and the geographical mode of speciation in the sea (Palumbi

Correspondence: Carla Hurt, Fax: (305) 284-3039;
E-mail: hurtc@bio.miami.edu.
¶Present address: Pennebaker Hall, Tennessee Technological University, Cookeville, TN 38505, USA.
**Present address: Department of Biological Sciences, TMSI, National University of Singapore, Kent Ridge Road, Singapore City 119260, Singapore

1994). Ecological factors can facilitate speciation when preference for different niches results in disruptive selection and hybrids are unfit in parental habitats (Schluter 2009). In symbiotic organisms, host and habitat specificity when coupled with host fidelity (mating inside/on or near the host) may facilitate sympatric speciation. Symbiotic associations occur in all marine ecosystems, but are particularly common and diversified on coral reefs. Evidence that disruptive selection for host resources may drive speciation in coral reef communities is beginning to emerge and has been proposed across diverse taxonomic groups, including fishes (Munday *et al.* 2004), nudibranchs (Faucci *et al.* 2007), barnacles (Tsang *et al.* 2009) and snapping shrimps (Morrison *et al.* 2004).

Snapping shrimps of the genus *Alpheus* Fabricius, 1798 are an ideal system for examining the role that host and habitat specialization has played in the diversification of marine invertebrates. *Alpheus* is one of the most speciose genera of marine decapod crustaceans, with about 300 described species. Members of this genus occur in a wide range of soft and hard bottom habitats, both intertidal and subtidal, and many are facultative or obligate associates of a wide range of animal hosts (Anker *et al.* 2006). Symbiotic animals associated with *Alpheus* include gobies (Karplus 1987), sponges (Banner & Banner 1982), echiurans (Anker *et al.* 2005, 2007), corals (Banner & Banner 1982) and sea anemones (Knowlton & Keller 1983, 1985).

The western Atlantic *Alpheus armatus* (Rathbun 1901) species complex is particularly well suited for addressing possible host-mediated speciation in sympatry. This complex is comprised of five species, four in the Caribbean Sea and one in Brazil, all of which have obligate symbiotic associations with sea anemones. Until 1983, all five species of the *A. armatus* complex were assigned to a single species, *A. armatus* Rathbun 1901. All species are morphologically very similar and can be distinguished only by slight differences in the shape or proportions of the rostro-orbital area and uropodal spine (Knowlton & Keller 1983, 1985; Almeida and Anker 2011). However, in life they are easily distinguished by their colour patterns (Knowlton & Keller 1985; Fig. 1, Table 1).

Observations of strict assortative mating by colour morphs, and in one colour morph a different host species led to the recognition in the Caribbean of four mostly sympatrically distributed species: *A. armatus* (Fig. 1A), *A. immaculatus* Knowlton & Keller 1983 (Fig. 1B), *A. polystictus* Knowlton & Keller 1985 (Fig. 1C), and *A. roquensis* Knowlton & Keller 1985 (Fig. 1D). The first three of these are associated with the cork-screw anemone, *Bartholomea annulata* (Le Sueur, 1817) (Aiptasiidae) (Fig. 1E), whereas *A. roquensis* is associated with a different, although related sea anemone known as the curly-cue or knobby anemone, *Ragactis lucida* (Duchassaing and Michelotti 1860) (formerly *Heteractis lucida* and often assigned to the Aiptasiidae). The fifth species of the *A. armatus* complex, *A. rudolphi* Almeida and Anker, 2011, was recently described from Brazil, based on a single museum specimen collected off Alagoas in the 1990s (Almeida and Anker 2011). Although little is known about this species, several underwater photographs of Brazilian anemone-associated snapping shrimps suggest that *A. rudolphi* may be associated with the sea anemone *Bellactis ilkalyseae* Dube 1983 and probably also has a distinctive colour pattern (Fig. 1F). Although the genus *Bellactis* is often placed in the family Sagartiidae, Almeida and Anker (2011) describe the anemone host of *A. rudolphi* as an aiptasiid (and thus a relative of the hosts of the Caribbean anemone-associated *Alpheus* species), an assessment that is supported by the photograph (Fig. 1F). Although the affiliations of these anemones remain uncertain given the confused state of aiptasiid taxonomy (D. Fautin, personal communication), the associations of shrimp themselves with specific species of anemones is clear.

The present study will focus primarily on the four Caribbean members of the *A. armatus* complex, which exhibit three critical features consistent with ecological divergence with gene flow (as defined by Pinho and Hey 2010): (i) contemporary sympatric geographical distributions; (ii) partitioning of host resources by species and microhabitat; and (iii) host fidelity. Below we review details of these three features for this species complex.

First, the geographical distributions of the four Caribbean species are broadly overlapping (Fig. 1G). *Alpheus armatus* has the widest distribution, occurring throughout the Caribbean Sea and extending north to the southwestern Gulf of Mexico, southern Florida and the Bahamas (Fig. 1G-1). *Alpheus immaculatus* overlaps with *A. armatus* throughout most of its range (A. Anker, personal observation), although it has not yet been reported from the Bahamas (Fig. 1G-2). *Alpheus polystictus* is much more abundant in the southern Caribbean Sea (Panama, Venezuela), but may also be encountered in the northern Caribbean (e.g. uncommonly in Haiti and Jamaica) and southern Florida (Fig. 1G-3) (N. Knowlton and A. Anker, personal observation). Finally, *A. roquensis* has the most restricted geographical range, occurring only on reefs of the island archipelagos off northern Venezuela, where all four species occur sympatrically. Furthermore, all members of the *A. armatus* complex seem to have an extended larval development, and some larvae are capable of at least moderate dispersal because recruitment occurs on isolated patch reefs lacking reproductive adults (Knowlton & Keller 1986).

Second, strict host-species specificity has been well documented in all four Caribbean species and is likely to be genetically determined, given that the larvae are

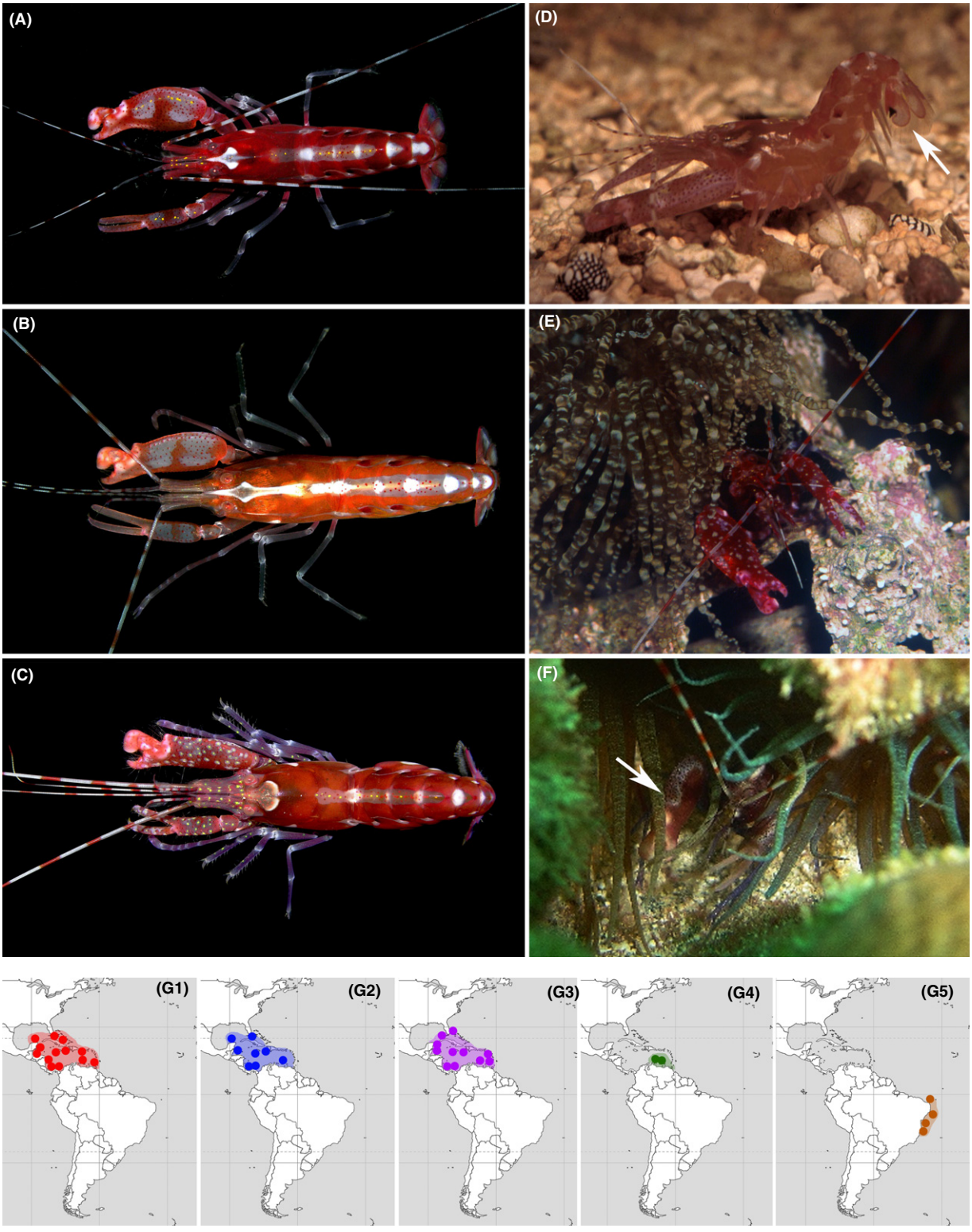

Fig. 1 General appearance, colour pattern and geographical range of species of the *Alpheus armatus* complex: (A) *A. armatus*, (B) *A. immaculatus*, (C) *A. polystictus*, (D) *A. roquensis* (arrow pointing to large uropod spine), (E) *A. polystictus* associated with *B. annulata* and (F) *A. rudolphi* (arrow pointing to claw of individual *in situ*). (G) Distribution maps of the five species based on confirmed records and colour photographs: *A. armatus* (G-1), *A. immaculatus* (G-2), *A. polystictus* (G-3), *A. roquensis* (G-4) and *A. rudolphi* (G-5).

Table 1 Main distinguishing features (morphology, colour, ecology, distribution) of five species of the *Alpheus armatus* complex. For more detailed morphological comparison, see Knowlton & Keller (1983, 1985) and Almeida & Anker (2011). For detailed geographical ranges for all species, refer to Fig. 1G.

Features	*A. armatus*	*A. immaculatus*	*A. polystictus*	*A. roquensis*	*A. rudolphi*
Rostrum width and length	Broad at base, relatively short	Narrow at base, long	Broad at base, relatively short	Broad at base, relatively short	Moderately broad at base, long
Rostral margins	Broadly concave	Shallowly concave	Broadly concave	Broadly concave	Straight
Adrostral furrows	Broad	Narrow	Broad	Broad	Very narrow
Adrostral teeth	Far from margin	Submarginal	Far from margin	Far from margin	Marginal
Setae on rostral margin	Along entire margin	Along entire margin	Along entire margin	Along entire margin	On distal half only
Distolateral spine on uropodal exopod	Narrow	Narrow	Narrow	Very broad	Narrow
Colour of chelipeds	Marbled pale red with white, dark red spots	Marbled pale red with white, dark red spots	Marbled dark red with white, white spots	Marbled pale red with white, dark red spots	Possibly with dark red and white patch, dark red spots
Neon-yellow spots on claws, antennules, body	Present	Absent	Present	Present	Unknown (not visible in UW photographs)
Host	*Ba. annulata*	*Ba. annulata*	*Ba. annulata*	*R. lucida*	Possibly *Be. ilkalyseae*
Depth range	1–12 m	10–25 m	1–12 m	5–12 m	single specimen at 49 m
Geographical range	Caribbean, Florida, Bahamas	Caribbean, Florida	Caribbean, Florida	S Caribbean	NE and E Brazil

released into the plankton and must find a host anemone after settlement. Although in aquaria all four species are capable of adapting either to *Bartholomea annulata* or to *Ragactis lucida*, they exhibit a statistical preference for their typical host when given a choice, and field observations documented a strict host-species relationship (Knowlton & Keller 1983). Furthermore, the three species associated with *B. annulata* occur at different depths and in slightly different environments. For instance, *A. immaculatus* is typically found in deeper water (15 m or below) on fore-reefs, whereas *A. armatus* and *A. polystictus* are more commonly found in back-reef and near-mangrove rubble habitats at shallower depths (typically <12 m, sometimes as shallow as 1 m) (Knowlton & Keller 1983, 1985; A. Anker, personal observation).

Third, all four Caribbean species of the *A. armatus* complex maintain strict host fidelity when mating, as they only rarely leave their hosts, especially when the shrimp is large (Knowlton 1980). As with many other alpheids, the anemone-associated snapping shrimps are typically found as adults in male–female pairs, with a single male–female pair occupying a large anemone or anemone cluster (juveniles are typically found occupying single, smaller anemones). Unoccupied hosts are scarce, and large anemones or clusters of anemones almost always shelter a pair of shrimps. Mating occurs on the host, and eggs laid by females without a male present do not develop (Knowlton 1980), indicating the absence of sperm storage. Host fidelity, when coupled with host or microhabitat specificity, is recognized as an important mechanism for promoting diversification in sympatry and has been shown to facilitate differentiation in a number of terrestrial insects (Bush 1969; Via 1999; Dorchin *et al.* 2009). In such cases, host fidelity acts as an automatic isolating trait (AIT), a trait that is under disruptive selection and simultaneously increases assortative mating (Bird *et al.* 2012). In sympatry, gene flow and recombination prevent differentiation by breaking down the association between genes under diversifying selection and genes causing assortative mating. AITs create an automatic link between ecological preference and mate selection, thereby leading to the formation of reproductive barriers as the pleiotropic consequence of genetic preference for a specific host species or host microhabitat (Bush 1966, 1969; Gavrilets 2003; Krug 2011; Servedio *et al.* 2011).

Thus, present-day biogeography, morphology and ecology of the Caribbean anemone snapping shrimps are consistent with a scenario of ecological speciation despite potential gene flow. However, to differentiate between alternative speciation scenarios, several critical questions about their evolutionary history need to be addressed. First, molecular evidence for the reproduc-

tive isolation of morphologically defined species is lacking. Second, the recent discovery of a Brazilian member of the *A. armatus* complex (Almeida and Anker 2011) raises the possibility that the four Caribbean sympatric forms are not monophyletic and that contemporary sympatry of Caribbean anemone shrimp could be the result of a recent recolonization of the Caribbean from geographically distant (and genetically isolated) populations. Therefore, evidence for reciprocal monophyly of the clade consisting of the four Caribbean taxa (*A. armatus*, *A. immaculatus*, *A. polystictus* and *A. roquensis*) with respect to the Brazilian taxon (*A. rudolphi*) is needed. Third, relationships among the Caribbean anemone shrimp need to be resolved to determine the role that host and microhabitat specificity may have played in the radiation of this group. Finally, molecular evidence for gene flow among incipient species is needed to determine whether gene flow between incipient species did occur following their initial divergence.

Here, we analyse a multilocus sequence data set to reconstruct the speciation history of the *Alpheus armatus* species complex. Molecular data were used to test the hypothesis that the five currently recognized taxa, four of them occurring in sympatry, represent reproductively isolated species. We employed both gene-tree and species-tree methods to determine whether the four Caribbean taxa are monophyletic with respect to the Brazilian taxon and therefore had the potential for gene flow during the incipient stages of speciation. An isolation–migration model was then used to test for historical gene flow among sympatric species. Ecological and biogeographical traits are discussed and interpreted in the context of phylogenetic results.

Materials and methods

Sample collections

Specimens of *Alpheus armatus*, *A. immaculatus* and *A. polystictus* were collected from three locations: Discovery Bay, Jamaica; Los Roques, Venezuela; and Isla Grande, Panama. *Alpheus roquensis* samples were only collected from Los Roques. All samples were frozen in liquid nitrogen and then stored at −80 °C. Only one specimen of *A. rudolphi*, from Ceará, Brazil, was available for molecular analyses. This specimen, preserved in 75% ethanol, is deposited as a voucher specimen in the collections of the Oxford University Museum of Natural History, Oxford, UK (OUMNH).

Molecular methods

Total nucleic acid (combined DNA and RNA) was extracted from frozen tissue using the SV Total RNA Isolation System (Promega) following a modification of the manufacturer's protocol (Regier 2007). Extracted samples included between 10 and 20 individuals from each of the four Caribbean species of the *Alpheus armatus* complex (*A. armatus*, *A. immaculatus*, *A. polystictus* and *A. roquensis*), the single specimen of the Brazilian *A. rudolphi*, and from two individuals of three outgroup taxa: *A. formosus* Gibbes, 1850; *A. malleator* Dana, 1852; and *A. websteri* Kingsley, 1880. The outgroup taxa were selected based on both molecular and morphological evidence, suggesting that these species belong to lineages closely positioned to the *A. armatus* complex (all members of clade III in Williams *et al*. 2001).

An RT–PCR method was used to amplify mitochondrial COI and partial coding regions from the following 10 nuclear genes: tetrahydrofolate synthase (THS), alanyl-tRNA synthetase (ATS), glucose phosphate dehydrogenase (GPDH), elongation factor Iα locus-1 (EF1-α_1) (EF1-α_1 failed to amplify in outgroups *A. formosus*, *A. malleator* and *A. websteri*), elongation factor Iα locus-2 (EF1-α_2), elongation factor 2 (EF2), putative GTP-binding protein (GBP), glucose-6-phosphate isomerase (GPI), phosphoenolpyruvate carboxykinase (PEPCK) and phosphogluconate dehydrogenase (PGDH). Although we were not able to obtain sequence from the nuclear gene EF1-α_1 from *A. formosus*, *A. malleator* or *A. websteri*; however, we were able to use EF1-α_1 sequence from GenBank (Williams *et al.* 2001) for *A. saxidomus* Holthuis, 1980, also a member of clade III, which was used to root this gene tree.

Sample sizes and amplicon lengths are listed in Table 3. In the case of COI, DNase I was used to destroy contaminating DNA prior to amplification. This approach reduced the risk of amplification of nuclear COI pseudogenes, previously shown to be pervasive within the genus *Alpheus* (Williams & Knowlton 2001). With respect to nuclear loci, RT–PCR allowed us to accurately predict the size of the desired homologous amplicon and increased the likelihood of sequencing single-copy functional genes (Regier 2007). First-strand synthesis of cDNA was performed using MuLV reverse transcriptase (Applied Biosystems), RNase inhibitor (Applied Biosystems) and sequence-specific reverse primers. The resulting cDNA was used as template in a polymerase chain reaction (PCR) that included a sequence-specific forward primer and used thermocycler conditions described by Regier (2007). For several loci, PCR products were re-amplified using a nested or hemi-nested amplification reaction. All amplification strategies and primer sequences are listed in Table 2. PCR products of the correct size were gel excised on a 1% (w/v) low-melt agarose gel and extracted using the Wizard SV Gel and PCR Clean-UP System (Promega), following manufacturers' instructions. For nine of the

loci, gel-purified PCR products were directly sequenced on an ABI3130xl Genetic Analyzer. For EF1-α_1 and EF1-α_2 sequences, gel-purified PCR products were cloned into JM109 cells using the Promega P-GemT easy Vector System II (Promega) following the manufacturer's protocol. Between 10 and 15 colonies were selected and PCR amplified using M13 forward and reverse primers. PCR products were cleaned enzymatically using exonuclease I/shrimp alkaline phosphatase prior to sequencing.

Sequences were edited and aligned using the ClustalW alignment algorithm as performed by BioEdit (Hall 1999). Alignments contained no insertions or deletions and were unambiguous. For nuclear loci that were sequenced directly from PCR product, heterozygous sites were identified as double peaks, and all sequences were verified by sequencing both strands. Most sequences contained one or zero heterozygous sites so that haplotypic phase was unambiguous. For sequences with more than one heterozygous site, the Bayesian program PHASE version 2.1 (Stephens *et al.* 2001) was used to reconstruct 'best guess' haplotypes from nuclear genotypic sequence data. The default values for number of iterations (100), burn-in (100) and thinning interval (1) were used for all runs. Each data set was run a minimum of three times with different seed values to test for consistency in haplotype reconstruction as recommended by the author. Genotypes that resulted in inconsistent haplotype reconstruction across runs were eliminated from further analysis. Six of the 11 nuclear loci showed evidence of recombination, and the largest nonrecombining blocks of DNA were used in the IMA2 analysis. Complete data sets were used for phylogenetic reconstructions. The lengths of the data sets before and after removal of recombining regions are shown in Table 3.

Phylogenetic analyses

Single-gene trees. Phylogenetic reconstructions of individual gene trees were performed using Bayesian analysis as implemented in MRBAYES version 3.1.2 (Ronquist & Huelsenbeck 2003). A best-fit model of nucleotide substitution was selected using the BIC as implemented in MEGA version 5.0 (Tamura *et al.* 2011). Data were partitioned by codon, and parameters were unlinked between partitions. Each gene-tree analysis ran for 10^7 generations sampling every 1000 steps. The first 1000 trees were discarded as burn-in, and the remaining 9000 trees were used to construct a majority rule consensus tree. Trace plots of log-likelihood values were viewed to confirm burn-in values using TRACER version 1.5, and convergence was assessed by monitoring the standard deviation of split frequencies (<0.01). Minimum-spanning trees (MST) were calculated using ARLEQUIN version 3.5 (Excoffier & Lischer 2010) and drawn using the software HAPSTAR version 0.6 (Teacher & Griffiths 2011). Partial sequences were removed before MST reconstructions. Diversity indices and tests for neutrality were also performed using ARLEQUIN version 3.5.

Species trees. Individual gene trees often misrepresent their underlying species due to the stochasticity inherent in the evolutionary processes, that is, lineage sorting. Short internal branch lengths coupled with large population sizes, a scenario that fits most rapid radiations of marine invertebrates, will increase the likelihood that a given gene tree and species trees disagree (Degnan & Rosenberg 2006; Kubatko & Degnan 2007). To address this, we have used the species-tree method *BEAST (*Bayesian Evolutionary Analysis Sampling Trees; Heled & Drummond 2010) to analyse our 11-loci data set in addition to a traditional phylogenetic reconstruction of the concatenated data set. *BEAST is a coalescent-based method that estimates both the individual gene trees and the shared underlying species tree while accounting for lineage sorting and uncertainty in the gene trees themselves (Heled & Drummond 2010). This method is able to accommodate different numbers of individuals per gene; therefore, all phased alleles were included in the data set and all outgroups were used to root the tree. Allele copies ranged from 4 to 40 copies per species per gene. Each gene was assigned the best model of nucleotide substitution as determined by the BIC criterion, and all genes were partitioned by codon. Three final Markov chain Monte Carlo (MCMC) runs were performed; each chain ran for 5×10^7 generations sampling every 1000 generations, and the first 5000 trees (10%) were discarded as burn-in. Resulting trees were compared for topological congruence, and the mean posterior probabilities across independent runs were calculated. Convergence was assessed by examining trace plots and histograms of logged output files in TRACER 1.5. and by examining effective sample size (ESS) values for all parameters (Rambaut & Drummond 2007).

Concatenation. In addition to species-tree analysis, we also performed a phylogenetic analysis of the concatenated data set that included sequences from all 11 loci (6299 bp) from a subset of three individuals belonging to each of the four ingroup taxa and from the one outgroup, *A. formosus*, which had the smallest number of missing genes (only EF1-α_1). These individuals were selected to maximize data representation across loci. The final alignment included the mtDNA COI haplotype and one of the phased nuclear alleles selected at random from each locus. Phylogenetic reconstructions

Table 2 List of primer sequences, sources and amplification strategies used to amplify all loci

Locus	Sequence 5′–3′	Source	Amplification strategy
COI			
COI-59F	5′-GMATAGTAGGMACRGCYCTNA-3′	*Alpheus* alignment	Nested reaction:
COI-1185R	5′-YCCTGTGAATAGGGGGAATC-3′	*Alpheus* alignment	1st amplification
COI-89F	5′-CGAGCTGAAYTMGGWCAACCA-3′	*Alpheus* alignment	COI-59F/COI- 1185R
COI-1152R	5′-DGCAAARATHCCRAATACRG-3′	*Alpheus* alignment	2nd amplification
			COI-89F/COI-1152R
THS			
3017-94F	5′-ATTGATGCCCGAATGTTCC-3′	*Alpheus* alignment	Nested reaction:
3017-635R	5′-CAAATGGACCAGCATGAACA-3′	*Alpheus* alignment	1st amplification
3017-88F	5′-CTCAGATTGATGCCCGAATG-3′	*Alpheus* alignment	3017-94F/3017-635R
3017-596R	5′-ACGTCTGCATGAGGTTAGGG-3′	*Alpheus* alignment	2nd amplification
			3017-88F/3017-596R
ATS			
3070-65F	5′-GYTCRGAAATTCATTWYGACC-3′	*Alpheus* alignment	Nested reaction:
3070-728R	5′-ARYCTCCAAGCCACATCAC-3′	*Alpheus* alignment	1st amplification
3070-101F	5′-ATGCTGCWCATCTYGTGAAT-3′	*Alpheus* alignment	3070-65F/3070-728R
3070-705R	5′-GAATYTCCYTRCWACCTCCT-3′	*Alpheus* alignment	2nd amplification
			3070-101F/3070-705R
GPDH			
3007-1F	5′-AARAARAARATHTAYCC-3′	Regier 2007	Hemi-nested reaction:
69-648R	5′-ARRTGRTTYTGCATNACRTC-3′	Regier 2007	1st amplification
69-612R	5′-ACYTTNACYTTYTCRTC-3′	Regier 2007	69-67F/69-648R
			2nd amplification
			69-67F/69-612R
EFIα			
EFIα-356F	5′-GCACDGARCCCAAGTACTCH-3′	*Alpheus* alignment	Hemi-nested reaction:
EFIα-1240R	5′-TTMACGATGCARGAGTCMCC-3′	*Alpheus* and crab* alignment	1st amplification
EFIα-421F	5′-CAAGAAGGTGGGCTACAACC-3′	*Alpheus* alignment	EFIα-356F/EFIα-1240R
			2nd amplification
			EFIα-356F/EFIα-421F
PGDHβ			
PGDHβ-32F	5′-ATCTGTGAAGCCTACCACCTC-3′	*Alpheus* alignment	Hemi-nested reaction:
PGDHβ-704R	5′-CTGGAGTAGGAATGCCAAGC-3′	*Alpheus* alignment	1st amplification
PGDHβ-698R	5′-TAGGAATGCCAAGCAAGACG-3′	*Alpheus* alignment	PGDHβ-32F/PGDHβ-704R
			2nd amplification
			PGDHβ-32F/PGDHB-698R
PEPCKα			
PEPCKα-454R	5′-TGCTGTAGGTAGTGGCCAAA-3′	*Alpheus* alignment	PEPCKα-1F/PEPCKα-454R
PEPCKα-1F	5′-GTAGGTGACGACATTGCYTGGATGAA-3′	Tsang *et al.* 2008	
GPI			
DS-1097F	5′-AATCTAATGGAAAGTAYGTAAC-3′	Williams *et al.* 2001	Hemi-nested reaction:
DS-1574R	5′-AGCTCAACACCCCACTGATC-3′	Williams *et al.* 2001	1st amplification
DS-1523R	5′-TGGGTGAAAATCTTGTGTTC-3′	Williams *et al.* 2001	DS-1097F/DS-1574R
			2nd amplification
			DS-1097F/DS-1523R
EFII			
EFII-723F	5′-MMAAGYTSTGGGGTGARAAC-3′	*Alpheus* alignment	Nested reaction:
EFII-1587R	5′-AYRATGTGYTCTCCRGAYTC-3′	*Alpheus* alignment	1st amplification
EFII-739F	5′-GAGRGCYTTCAACACCTAYA-3′	Alignment between *Alpheus* and crab†	EFII-723F/EFII-1587R
			2nd amplification
EFII-1499R	5′-ARTCGGAGGGGTTCTTGG-3′	Alignment between *Alpheus* and crab†	EFII-739F/EFII-1499R
GBP			
42fin-1F	5′-GCNGARAAYTTYCCITTYTG-3′	Regier 2007	Hemi-nested reaction:
42fin-2R	5′-ATDATRAANCCYTTYTCRAARTC-3′	Regier 2007	1st amplification
GBP-103F	5′-TGCAAGYAAAGTCCCTGCATT-3′	*Alpheus* alignment	42fin-1F/42fin-2R
GBP-781R	5′-TGCAAGYAAAGTCCCTGCATT-3′	*Alpheus* alignment	2nd amplification
			42fin-1F/GBP-103F

GenBank Accession nos: *U90050; †AY305506.

were performed using both Bayesian and maximum-likelihood (ML) criteria. Bayesian reconstruction of the concatenated gene tree was performed using MRBAYES version 3.1 (Ronquist & Huelsenbeck 2003). The concatenated data set was partitioned by loci, and the best-fit model of nucleotide substitution was assigned to each partition. Three separate analyses were performed, and each MCMC chain ran for 10×10^6 generations sampling every 1000 steps. The first 1000 trees were discarded as burn-in, and the remaining 9000 trees were used to construct a majority rule consensus tree. Trace plots were viewed to confirm burn-in values using TRACER version 1.5, and convergence was assessed by monitoring the standard deviation of split frequencies (<0.01).

Likelihood analysis of the concatenated sequence alignment was performed using GARLI version 2.0 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl 2006). The data set was partitioned by gene, and the best model of nucleotide substitution was assigned to each partition. Model parameters were unlinked, and locus-specific rate multipliers were used for each partition. Starting topologies were determined using the ML stepwise-addition option. Confidence at nodes was assessed using nonparametric bootstrap analysis (100 replicates). We ran three independent runs with different starting seeds to assess the reliability of the topology, and the mean bootstrap support value was calculated for each node.

Isolation migration

The program IMa2 was used to estimate demographic parameters and evaluate the level of gene flow among *A. armatus*, *A. immaculatus*, *A. polystictus* and *A. roquensis* using sequence data from our 11-loci data set. IMa2 uses a MCMC method to sample gene genealogies and estimate joint posterior probability distributions of parameters included in the Isolation–migration model using multilocus sequence data. Estimated parameters for each of 1-k sampled populations include directional migration rates $m_i = M/\mu$ (M is the per generation migration rate per gene copy), population divergence times $t_i = T\mu$ (T is the number of generations since common ancestry) and effective population size $\Theta_i = 4N_e\mu$. All parameter estimates are scaled by the geometric mean of the per locus mutation rate, μ. Posterior probability distributions were used to test the null hypothesis that no migration has occurred following the initial divergence of sister species (i.e. $N_e m = 0$). Nuclear loci were assigned an inheritance scalar of 1, and the COI data set was assigned an inheritance scalar of 0.25. Because IMa2 assumes no recombination within loci, we generated maximally informative blocks of nonrecombining data sets for each locus using the program IMgc (Woerner *et al.* 2007). Tests for selection on individual loci were performed using Tajima's D as performed by ARLEQUIN 3.5. IMa2 supports both the infinite sites (IS) and the Hasegawa–Kishino–Yano (HKY) models of nucleotide substitution. The majority of our loci contained at least one site with more than one substitution, violating the assumptions of the IS model; therefore, the HKY model of substitution was assigned to all loci. Preliminary runs were performed to optimize upper bounds on prior distributions and to optimize heating schemes. Upper bounds for priors were as follows: $\Theta = 15$, $t = 15$ and $m = 15$. The heating scheme used a geometric model with parameters ha = 0.96 and hb = 0.9. The MCMC analysis was performed with 25 heated chains and a burn-in of 100 000 steps. Trendline plots, effective sample sizes (ESSs) and swapping rates were monitored to ensure good mixing and convergence. Three independent runs with different seed numbers were performed to ensure consistency of parameter estimates. Each run included a minimum of 2 000 000 steps, sampling every 100 steps for a total of 20 000 recorded generations. Significance of migration rates was determined using the log-likelihood ratio test of Nielsen & Wakeley (2001). Migration parameter (m_i), divergence time parameter (t_i) and effective population size (Θ) estimates from IMa2 were scaled by the geometric mean of the per locus mutation rate across loci (μ). A mutation rate of 1.7×10^{-7} was used to rescale parameters into demographic units. This rate is based on a calibration from eight of 11 loci used in this study in sister species separated by the Isthmus of Panama approximately 3 Ma (Hurt *et al.* 2009).

Results

Alignments, sequence variation, selection and recombination tests

We obtained sequence data for all 11 loci from each of the four Caribbean taxa. Alignments for all data sets contained no indels and no stop codons. Tests for selection using Tajima's D were not significant. PHASE was sometimes unable to unambiguously reconstruct haplotypes from genotype sequence data for five of 11 nuclear loci. These sequences were not included in the final analyses. Samples sizes given in Table 3 reflect the number of haplotypes after the removal of ambiguously phased alleles.

Individual gene trees

The four Caribbean taxa of the *Alpheus armatus* complex formed reciprocally monophyletic clades in Bayesian analysis of mitochondrial COI sequence data with

Table 3 Summary Statistics. The number of alleles used in the analyses (N) (for nuclear loci this is the number of alleles after using PHASE), length of final sequence alignment before (and after) removal of putative recombining regions, number of haplotypes (H), number of segregating sites (S), average number of pairwise nucleotide differences (π), Tajima's *D* and GenBank Accession nos. Significance of Tajima's *D* is indicated by *

Loci/species	*N*	Sites (nr)	*H*	*S*	π	Tajima's *D*	GenBank nos
COI							
arm	12	967	10	23	6.036	−1.057	KF131481–KF131536
imm	10	967	9	21	6.982	−0.119	
pol	9	967	5	5	1.389	−0.100	
roq	12	967	6	5	1.212	−0.988	
GPDH							
arm	10	492 (426)	7	8	2.422	−0.619	KF131436–KF131480
imm	12	492 (426)	8	9	3.745	0.920	
pol	16	492 (426)	4	9	1.169	−2.021*	
roq	10	492 (426)	4	2	0.711	0.019	
THS							
arm	19	420 (367)	6	4	1.000	−0.420	KF130972–KF131025
imm	9	420 (367)	3	3	1.000	−0.359	
pol	20	420 (367)	20	6	0.868	−1.55*	
roq	20	420 (367)	20	4	1.021	−1.263	
ATS							
arm	26	584 (584)	4	4	0.378	−1.707*	KF131062–KF131125
imm	20	584 (584)	6	5	1.584	0.380	
pol	42	584 (584)	7	6	0.419	−1.846**	
roq	26	584 (584)	2	1	0.508	1.533	
EF1α_1							
arm	22	657 (530)	1	0	0.000	0.000	KF131126–KF131195
imm	16	657 (530)	8	11	3.142	−0.198	
pol	16	657 (530)	9	11	2.442	−0.995	
roq	16	657 (530)	4	4	1.008	−0.512	
EF1 α_2							
arm	8	631 (631)	1	0	0.000	0.000	KF131026–KF131061
imm	12	631 (631)	5	5	1.879	0.497	
pol	8	631 (631)	3	2	1.000	1.104	
roq	8	631 (631)	2	3	1.607	1.601	
EF2							
arm	18	628 (465)	1	0	0.000	0.000	KF131196–KF131252
imm	24	628 (465)	3	6	1.533	−0.140	
pol	20	628 (465)	8	11	2.716	−0.440	
roq	24	628 (465)	2	1	0.228	−0.248	
GBP							
arm	16	602 (506)	5	4	1.100	−0.274	KF131253–KF131306
imm	4	602 (506)	2	1	0.500	−0.612	
pol	8	602 (506)	5	5	1.786	−0.335	
roq	16	602 (506)	1	0	0.000	0.000	
GPI							
arm	26	381 (381)	2	1	0.077	−1.156	KF131352–KF131399
imm	24	381 (381)	1	0	0.000	0.000	
pol	36	381 (381)	3	2	0.111	−1.495*	
roq	28	381 (381)	2	1	0.071	−1.151	

Table 3 *Continued*

Loci/species	*N*	Sites (nr)	*H*	*S*	π	Tajima's *D*	GenBank nos
PEPCK							
arm	16	471 (471)	4	3	0.483	−1.349	KF131307–KF131351
imm	16	471 (471)	3	2	0.250	−1.498	
pol	18	471 (471)	4	5	0.477	−1.956**	
roq	16	471 (471)	1	0	0.000	0.000	
PGDHB							
arm	22	590 (549)	1	0	0.000	0.000	KF131400–KF131435
imm	18	590 (549)	10	18	4.706	−0.389	
pol	12	590 (549)	8	16	6.955	1.350	
roq	8	590 (549)	1	0	0.000	0.000	

*$P < 0.05$.
**$P < 0.01$.

100% posterior probability support (Fig. 2A). There was strong support for two sister relationships: *A. armatus*/*A. immaculatus* (pp = 100) and *A. polystictus*/*A. roquensis* (pp = 97), the four taxa together forming a well-supported Caribbean clade. The Brazilian *A. rudolphi* was recovered as a sister to the Caribbean clade with 100% posterior probability support. Bayesian analysis of individual nuclear genes did not result in species-level monophyly for any of the loci examined (Appendix S1). Alleles of *A. roquensis* were reciprocally monophyletic with respect to other morphologically defined species in six of the 10 nuclear genes, whereas those of *A. armatus*, *A. immaculatus* and *A. polystictus* were polyphyletic in the majority of the nuclear gene phylogenies. Visual inspection of MST for individual nuclear loci (Fig. 3) reveals that *A. armatus*/*A. immaculatus* haplotypes tend to group together and *A. roquensis*/*A. polystictus* haplotypes tend to group together for the majority of the nuclear genes. There was some sharing of nuclear haplotypes across the four Caribbean species. *Alpheus armatus* and *A. immaculatus* shared haplotypes at five loci. *Alpheus polystictus* and *A. roquensis* shared haplotypes at three loci. *Alpheus polystictus* and *A. immaculatus* shared a haplotype at the GDBP locus. Finally, *A. polystictus*, *A. immaculatus* and *A. roquensis* shared the same common haplotype at the GPDH locus (Fig. 3).

Species tree and concatenated gene tree

The *BEAST species-tree (Fig. 2C) and Bayesian and ML analysis of the concatenated data set (Fig. 2B) all produced identical, well-supported topologies consistent with results from the analysis of the COI data set. All phylogenetic reconstructions supported sister relationships between *A. armatus*/*A. immaculatus* and *A. polystictus*/*A. roquensis*. Terminal branch lengths in the *A. armatus*/*A. immaculatus* clade were shorter than branch lengths in the *A. polystictus*/*A. roquensis* clade in both the concatenated gene tree and the species tree.

Effective population sizes, divergence times and migration rates

Isolation–migration analysis of more than two populations (IMA2) requires an input topology with a known order of speciation events. We used the topology recovered from the concatenated gene-tree analysis and the species-tree analysis, with the split between *A. armatus* and *A. immaculatus* being more recent than the split between *A. polystictus* and *A. roquensis*. The isolation–migration model, peak posterior probability estimates and confidence intervals for Θ are given in Fig. 4. Population size estimates for all four species were quite large. *Alpheus immaculatus* and *A. polystictus* had the largest effective population sizes ($N_e = 1.6 \times 10^6$ and $N_e = 1.3 \times 10^6$, respectively). The estimated population size of *A. armatus* was 1.0×10^6. *Alpheus roquensis* was substantially smaller ($N_e = 2.9 \times 10^5$). Population sizes for ancestral species A and B were similar to that for *A. roquensis* ($N_e = 4.4 \times 10^5$ and $N_e = 7.4 \times 10^5$, respectively), although confidence intervals for these ancestral population sizes were broad. IMa2 estimates for the size of ancestral population C were not reliable, as posterior distributions were flat across the range of possible values.

Posterior probability distributions, peak posterior point estimates and 95% confidence intervals of divergence time parameters are shown in Fig. 5. The peak posterior point estimates for the split between *A. armatus* and *A. immaculatus* were consistent across runs ($t = 1.73$), corresponding to a divergence time of 9.9 Ma. The peak posterior point estimate for the split between *A. polystictus* and *A. roquensis* ranged from 2.0 to 3.2 across runs with the first run showing a double

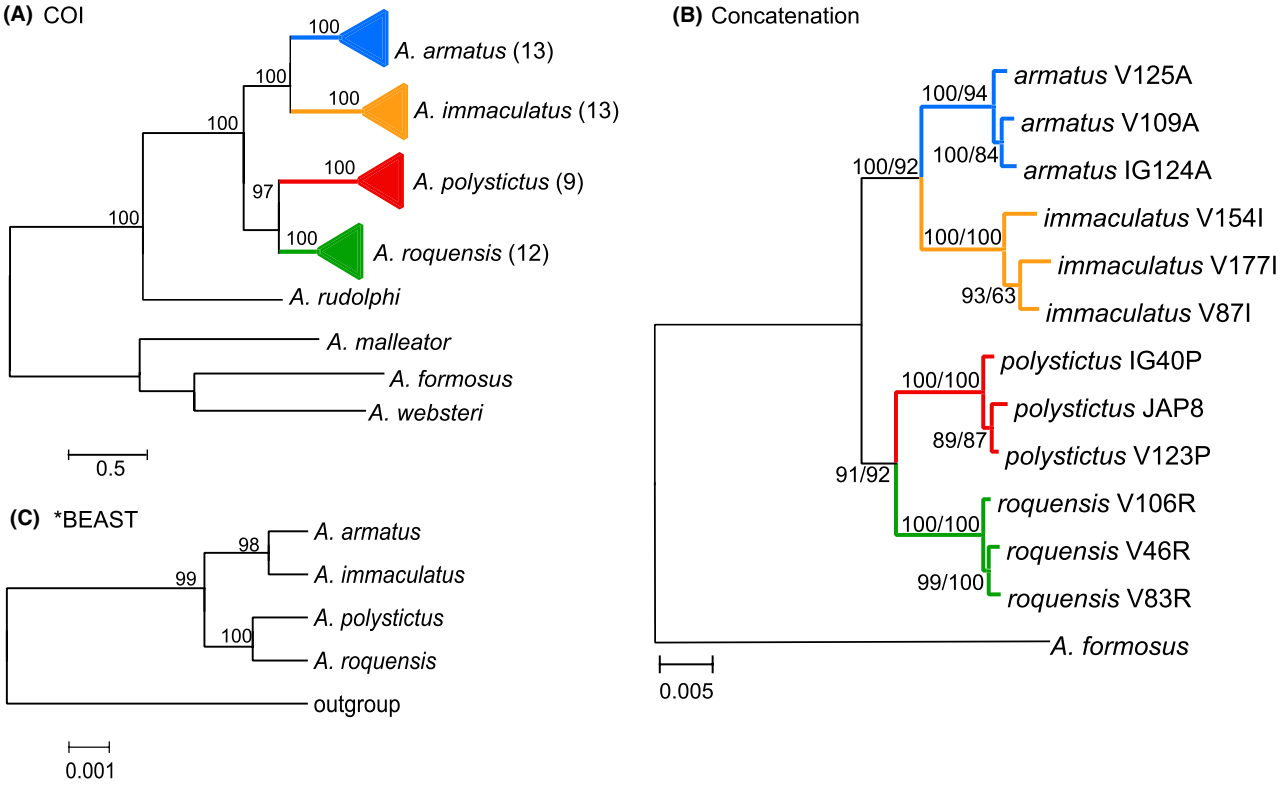

Fig. 2 Phylogenetic reconstructions for the *Alpheus armatus* complex. (A) Bayesian phylogenetic reconstruction of COI data set. Reciprocally monophyletic clades are collapsed by species, and the number in parentheses indicates the number of sequences represented in each clade. Support values indicate posterior probabilities. (B) Bayesian phylogenetic reconstruction of concatenated data set for three individuals per species. Tree rooted with *A. formosus* sequence. Support values are posterior probabilities/maximum-likelihood bootstrap values. (C) *BEAST species tree with posterior probability support at nodes.

peak at 2.0 and 2.8, corresponding to a divergence time of 11.5–18.4 Ma. Peak posterior point estimates of the divergence time between ancestral species A and species B ranged from 3.2 to 3.8, corresponding to a divergence time between 18.9 and 21.9 Ma, but this estimate may not be reliable, as 95% confidence intervals were very broad.

Posterior probability distributions for migration rates are shown in Fig. 6. IMa2 migration rate parameter estimates are scaled by the mutation rate μ so that $M = m/\mu$. The effective number of migrants into population *i* from population *j* ($4N_i m_{i>j}$) can be obtained by multiplying $M_{i>j}$ by Θ_i. Highly significant unidirectional migration was inferred within sister species pairs: *A. armatus* to *A. immaculatus* ($4N_e m = 1.05$, $P < 0.001$) and *A. polystictus* to *A. roquensis* ($4N_e m = 0.13$, $P < 0.001$). Less significant unidirectional migration was inferred between sister groups: *A. immaculatus* to *A. polystictus* ($4N_e m = 0.07$, $P < 0.05$) and *A. armatus* to *A. roquensis* ($4N_e m = 0.02$, $P < 0.05$). Estimates for migration parameters among ancestral populations had flat distributions and wide confidence intervals and were therefore not considered reliable (results not shown).

Discussion

Support for morphologically defined species and monophyly of Caribbean taxa

Results from phylogenetic analyses of COI data support morphologically defined species definitions as all species were recovered with high bootstrap/posterior probability support. When considered in the context of previous studies (both field observations and laboratory mate choice trials have demonstrated strict assortative mating by phenotype) (Knowlton & Keller 1985), the evidence for reproductive isolation of the four Caribbean species of the *Alpheus armatus* complex is substantial. In contrast to the COI gene tree, nuclear gene phylogenies were not monophyletic with respect to species. The time to reciprocal monophyly of species is typically much longer for nuclear alleles than for mitochondrial genes, primarily because of the larger effective size of nuclear markers and does not necessarily reflect contemporary gene flow (Palumbi *et al.* 2001).

Placement of the Brazilian *A. rudolphi* outside of (sister to) the Caribbean clade is strongly supported, indicating

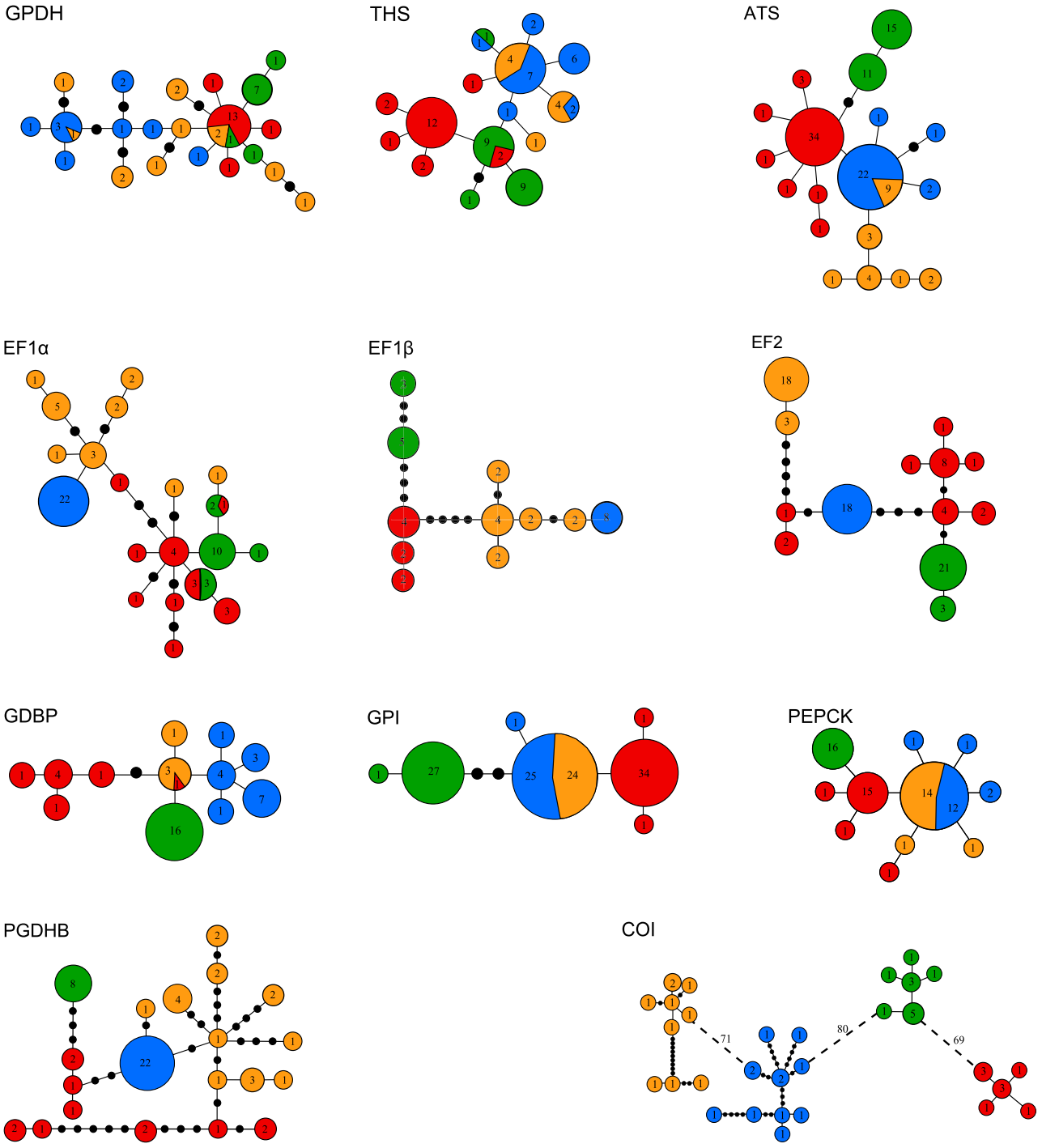

Fig. 3 Minimum-spanning trees (MST) for all ten nuclear genes and mitochondrial COI. Colours indicate species as follows: blue = *A. armatus*, orange = *A. immaculatus*, red = *A. polystictus* and green = *A. roquensis*. Numbers in circles indicate haplotype frequency. Solid black circles indicate the number of mutational differences between haplotypes, and lines without circles indicate single mutational differences. For the mitochondrial COI MST, species are separated by dashed lines, and the number of mutational differences is indicated numerically.

that speciation among the four Caribbean taxa likely occurred within the Caribbean Sea. The species-tree reconstruction and the Bayesian and ML concatenated gene-tree reconstructions all strongly support identical topologies, producing a symmetric tree topology containing two sister species pairs: *A. armatus*/*A. immaculatus* and *A. polystictus*/*A. roquensis*. Branch lengths in both the species tree and gene tree suggest that *A. polystictus*/

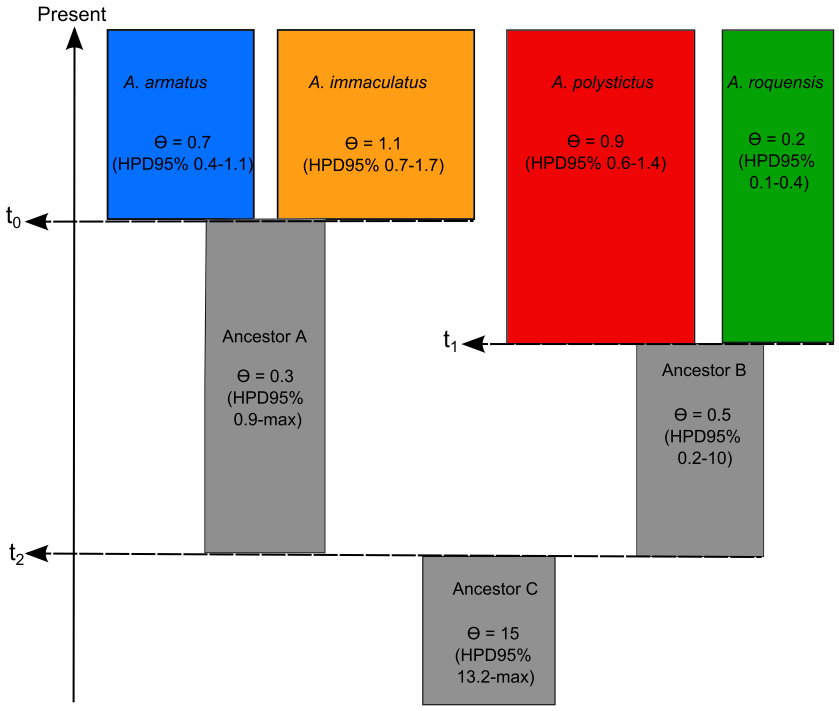

Fig. 4 Four-species IMa2 model for the *Alpheus armatus* complex with estimated effective population sizes and 95% HPD intervals. Width of boxes corresponds to estimated population sizes. Length of boxes corresponds to estimated divergence times.

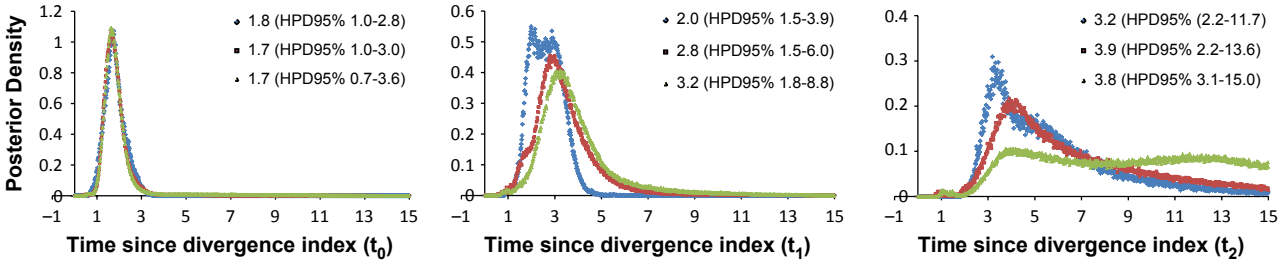

Fig. 5 Posterior probability densities for divergence time parameters (*t*) as outlined in the IMa2 model (Fig. 4). In each graph, red, green and blue traces show results from the three independent runs. Values above graphs indicate marginal peak probability estimates for each parameter and 95% confidence intervals. Divergence times correspond to those defined in Fig. 4.

A. roquensis probably diverged earlier than *A. armatus*/*A. immaculatus*, a result that is supported by the IMa2 analysis.

Comparing molecular results with morphology and colour patterns

Morphological differences between the species of the *Alpheus armatus* complex are rather subtle (Table 1). The Brazilian *A. rudolphi* appears to be the most morphologically distinctive taxon of the *A. armatus* complex, for example, in the general shape of the orbito-rostral area (Almeida & Anker 2011), supporting its sister position to the four Caribbean taxa. Among the Caribbean species, *A. immaculatus* has the longest and narrowest rostrum (Knowlton & Keller 1983; Almeida & Anker 2011), and is also the only species lacking bright neon-yellow spots (Fig. 1A–D). *Alpheus immaculatus* also appears to be the most similar to the Brazilian species in several features of the orbito-rostral area (Table 1), as well as tending to be found in deeper water (as is *A. rudolphi*). *Alpheus polystictus* resembles both *A. armatus* and *A. roquensis* morphologically, but can be easily distinguished by the white-spotted major and minor chelae (Fig. 1A–C, Table 1), and from *A. roquensis* by the much narrower uropodal spine in males, which in the latter species is typically extremely stout and broad (Fig. 1D). In sum, these diagnostic features of the Caribbean species

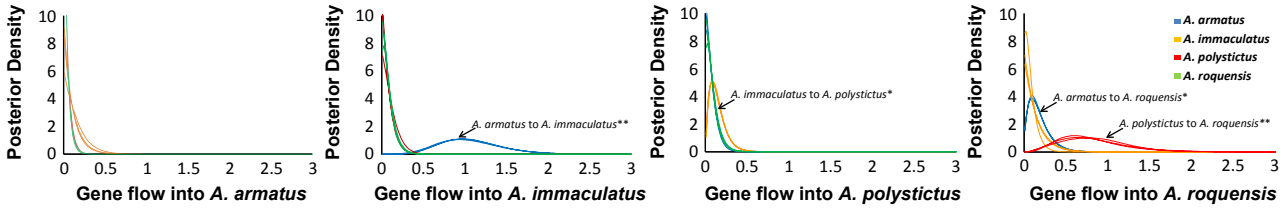

Fig. 6 Posterior probability densities for migration rate parameters (m) as outlined in the IMa2 model (Fig. 4). Plots include results from three independent IMa2 runs. Blue, orange, red and green traces indicate gene flow from *A. armatus*, *A. immaculatus*, *A. polystictus* and *A. roquensis*, respectively. Statistically significant migration rate parameters are indicated on each plot. Values above graphs indicate marginal peak probability estimates for each parameter and 95% confidence intervals. *$P < 0.05$, **$P < 0.001$.

are autapomorphic and therefore phylogenetically noninformative within the clade, neither supporting nor contradicting the molecular results.

Divergence time

Despite their great similarity in appearance (morphology, colour patterns, ecology), the estimated timing of the speciation within the *Alpheus armatus* complex may at first appear as surprisingly old. IMa2-estimated divergence times place the radiation of this complex within the Miocene (between 10 and 20 Ma), prior to the final closure of the Isthmus of Panama, approximately 3 Ma (Keigwin 1978). Their early divergence is, however, consistent with geological evidence, suggesting that the emergence of the Isthmus of Panama caused oceanographic changes in the southern Caribbean long before its closure was complete (Collins *et al.* 1996). During this time, the southern Caribbean, which is the centre of diversity for the *Alpheus armatus* complex, experienced a significant increase in carbonate content, which may have promoted diversification in reef-associated invertebrates (Collins *et al.* 1996). As evidence of this, the timing of this divergence is also consistent with radiations in several other Caribbean invertebrates, including the distantly related sponge-dwelling snapping shrimps of the *Synalpheus gambarelloides* Nardo, 1847 species group (Morrison *et al.* 2004), foraminiferans (Coates *et al.* 1992) and marine gastropods of the genus *Strombina* (Jackson *et al.* 1993) and coincides with accelerated speciation rates in Caribbean reef corals (Collins *et al.* 1996; Budd & Johnson 1999).

Ecological speciation

Competition for limited habitat (i.e. available hosts) can drive both resource specialization and exploitation of novel ecological opportunities and is a major component of ecological speciation theory (Schluter 2009). In regard to the *Alpheus armatus* complex, host anemones are probably the major limiting resource for these snapping shrimps, and in fact unoccupied sea anemones (particularly the large clusters used by reproductive pairs) are rarely seen in the field (Knowlton 1980; A. Anker, personal observation). Both ecological opportunity (an unoccupied anemone host species) and specialization for host microhabitat (depth/substrate/exposure) may have contributed to the Caribbean radiation of the anemone-associated snapping shrimps. Our phylogenetic results indicate that the common ancestor of the present Caribbean taxa was most likely associated with *Bartholomea annulata*. A switch in host from *B. annulata* to *R. lucida* may have triggered the divergence of *A. roquensis* from *A. polystictus*. An unoccupied host species could have provided a valuable ecological opportunity, as *R. lucida* is generally not occupied by snapping shrimps, except where *A. roquensis* occurs. Because their reproductive cycle is completely linked to their host, diversifying selection for host species would automatically lead to premating reproductive isolation.

Host-associated sympatric speciation may be quite common among alpheids and symbiotic marine invertebrates in general. Host-mediated speciation has been documented in the closely related alpheid genus *Synalpheus*. Duffy (1996) reported four cryptic species of *Synalpheus* that are obligate affiliates of different host sponges and displayed host-species specificity and host fidelity. Tsang *et al.* (2009) found evidence for host-mediated speciation in the coral barnacle *Wanella milleporae* (Darwin 1854) based on preference differences in coral growth forms. Faucci *et al.* (2007) provided evidence of host-mediated sympatric speciation in the nudibranch genus *Phestilla*, resulting from specificity on their coral hosts. More evidence of host-shift speciation will likely accumulate as integrative (morphological + molecular) studies continue to disentangle sympatric cryptic species complexes.

In the case of *A. armatus* and *A. immaculatus*, each anemone provides a distinct habitat unit, mating is restricted to the host, and the host is fixed to a substrate; microhabitat preference prevents recombination and leads to reproductive isolation. Under these conditions, a heritable preference for host microhabitat is as effective at

promoting speciation as preference for host species. Both *A. armatus* and *A. immaculatus* partition their host resource based on depth and exposure: *A. immaculatus* typically occurs at depths between 13 and 25 m and occurs in coarse-sand fore-reef habitats, whereas *A. armatus* is generally found at depths <10 m and in protected, usually more silty back-reef areas (Knowlton & Keller 1983, 1985). In marine systems, sympatric sister species often segregate by depth. Evidence for bathymetric segregation of species has been found in many marine groups including corals (*Montastraea*, Knowlton *et al.* 1992; *Favia*, Carlon & Budd 2002), gastropods (*Tegula*, Hellberg 1998), fishes (*Sebastes*, Ingram 2010), limpets (*Cellana*, Bird *et al.* 2011) and snapping shrimps (*Synalpheus*, Macdonald *et al.* 2006; *Alpheus*, Anker 2012). This type of ecological speciation may be more common than speciation resulting from a host shift as there are many more opportunities for microhabitat specialization (particularly for depth and substrate) than for specialization on a novel host species.

Conclusions

Here, we present evidence that ecological factors initiated or at least promoted the diversification of the four Caribbean species of the *Alpheus armatus* complex despite gene flow. Specifically, we present phylogenetic evidence that the sympatrically occurring morphotypes are reproductively isolated. The Caribbean clade, comprising *A. armatus*, *A. immaculatus*, *A. polystictus* and *A. roquensis*, is here shown to be monophyletic with respect to the Brazilian *A. rudolphi*, and no obvious geographical barriers exist among sister species (*A. armatus* and *A. immaculatus*, *A. polystictus* and *A. roquensis*). Coalescent-based estimates of historical demographic parameters infer that gene flow did occur among sister species and to a lesser extent between the two sister clades (*A. armatus* + *A. immaculatus* and *A. polystictus* + *A. roquensis*). However, given the present data, it is not possible to rule out the role of geographical isolation in the speciation of members of the Caribbean clade of the *A. armatus* complex. Although three species (*A. armatus*, *A. immaculatus* and *A. polystictus*) now occur in sympatry throughout the Caribbean Sea and overlap with the fourth species (*A. roquensis*) in the archipelagos to the north of Venezuela, a temporary allopatric phase in the evolutionary history of this shrimp group cannot be discounted (particularly between the ancestor of *A. armatus* + *A. immaculatus* and that of *A. polystictus* + *A. roquensis*, given that the latter two species are notably more abundant in the southern Caribbean). Even between the sister species pairs, one possible scenario is that they diverged initially in allopatry and that gene flow occurred after secondary contact was established. In this case, partitioning of host and habitat resources may have evolved in response to reduced fitness of hybrid offspring, that is, reinforcement. Reconstructing the geographical distributions of the *A. armatus* complex at the time of speciation may be particularly challenging as the estimated timing of speciation (>10 Ma) is much older than initially assumed. Furthermore, coding sequences used in this study are less variable and therefore less sensitive for reconstructing demographic histories than noncoding sequences. Thus, large amounts of data will be needed to adequately test alternative hypotheses regarding geographical origin of this complex. Further research utilizing fine-scale population level sampling, a large number of neutral loci generated from next-generation sequencing and specific model-based analytical methods, such as approximate Bayesian computational methods (Beaumont *et al.* 2002; Hickerson *et al.* 2006), could be used to identify the historical geographical distributions of species and the colonization history in the *A. armatus* complex.

Acknowledgements

We are grateful to Kristen Hultgren (University of Seattle) and Jane Indorf (University of Miami) who reviewed the manuscript and made a number of useful suggestions and corrections. Paulo P.G. Pachelle (Universidade Federal do Ceará) provided a very valuable specimen of *A. rudolphi*. Additional photographs were kindly provided by Frédéric Fasquel (Fig. 1D,E) and Marcio Lisa (Fig. 1F). One of the authors (AA) was supported by CAPES of the Brazilian Government in the form of a postdoctoral fellowship.

References

Almeida AO, Anker A (2011) *Alpheus rudolphi* spec. nov., a new snapping shrimp from northeastern Brazil (Crustacea: Decapoda: Alpheidae). *Zooogische Mededelingen Leiden*, **85**, 1–10.

Anker A (2012) Revision of the western Atlantic members of the *Alpheus armillatus* H. Milne Edwards, 1837 species complex (Decapoda, Alpheidae), with the description of seven new species. *Zootaxa*, **3386**, 1–109.

Anker A, Murina GV, Lira C, Vera Caripe JA, Palmer AR, Jeng MS (2005) Macrofauna associated with echiuran burrows: a review with new observations of the innkeeper worm, Ochetostoma erythrogrammon Leuckart and Rüppel, in Venezuela. *Zoological Studies*, **44**, 157–190.

Anker A, Ahyong ST, Noel PY, Palmer AR (2006) Morphological phylogeny of alpheid shrimps: parallel preadaptation and the origin of a key morphological innovation, the snapping claw. *Evolution*, **60**, 2507–2528.

Anker A, Hurt C, Knowlton N (2007) Three transisthmian snapping shrimps (Crustacea: Decapoda: Alpheidae: *Alpheus*) associated with innkeeper worms (Echiura: Thalassematidae) in Panama. *Zootaxa*, **1626**, 1–23.

Banner DM, Banner AH (1982) The alpheid shrimp of Australia: Part 3: the remaining alpheids, principally the genus *Alpheus*, and the family Ogyrididae. *Records of the Australian Museum*, **34**, 1–357.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **4**, 2025–2035.

Bird CE, Holland BS, Bowen BW, Toonen RJ (2011) Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. *Molecular Ecology*, **20**, 2128–2141.

Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ (2012) Sympatric speciation in the post "Modern Synthesis" era of evolutionary biology. *Evolutionary Biology*, **39**, 158–180.

Budd AF, Johnson KG (1999) Origination preceding extinction during late Cenozoic turnover of Caribbean reefs. *Paleobiology*, **25**, 188–200.

Bush GL (1966) The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera, Tephritidae). *Bulletin of the Museum of Comparative Zoology*, **134**, 431–562.

Bush GL (1969) Mating behavior, host specificity, and ecological significance of sibling species in frugivorous flies of the genus *Rhagoletis* (Diptera-Tephritidae). *American Naturalist*, **103**, 669–672.

Carlon DB, Budd AF (2002) Incipient speciation across a depth gradient in a scleractinian coral? *Evolution*, **56**, 2227–2242.

Coates AG, Jackson JBC, Collins LS *et al.* (1992) Closure of the Isthmus of Panama – the near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin*, **104**, 814–828.

Collins LS, Budd AF, Coates AG (1996) Earliest evolution associated with closure of the Tropical American Seaway. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 6069–6072.

Darwin C (1854) *A monograph on the Sub-class Cirripedia, with Figures of all the Species*. pp. 684. The Ray Society, London.

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **32**, 762–768.

Dorchin N, Scott ER, Clarkin CE, Luongo MP, Jordan S, Abramson WG (2009) Behavioral, ecological and genetic evidence confirm the occurrence of host-associated differentiation in goldenrod gallmidges. *Journal of Evolutionary Biology*, **22**, 729–739.

Duffy JE (1996) Species boundaries, specialization, and the radiation of sponge-dwelling alpheid shrimp. *Biological Journal of the Linnean Society*, **58**, 307–324.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Faucci A, Toonen RJ, Hadfield MG (2007) Host shift and speciation in a coral-feeding nudibranch. *Proceedings of the Royal Society B-Biological Sciences*, **274**, 111–119.

Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 905–920.

Hellberg ME (1998) Sympatric sea shells along the sea's shore: the geography of speciation in the marine gastropod *Tegula*. *Evolution*, **52**, 1311–1324.

Hickerson MJ, Doman G, Moritz C (2006) Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology*, **15**, 209–223.

Hurt C, Anker A, Knowlton N (2009) A multilocus test of simultaneous divergence across the Isthmus of Panama using snapping shrimp in the genus *Alpheus*. *Evolution*, **63**, 514–530.

Ingram T (2010) Speciation along a depth gradient in a marine adaptive radiation. *Proceedings of the Royal Society B*, **22**, 613–618.

Jackson JBC, Jung P, Coates AG, Collins LS (1993) Diversity and extinction of tropical American mollusks and emergence of the Isthmus of Panama. *Science*, **260**, 1624–1626.

Karplus I (1987) The association between gobiid fishes and burrowing alpheid shrimps. *Oceanography and Marine Biology; An Annual Review*, **25**, 507–562.

Keigwin LD (1978) Pliocene closing of the Isthmus of Panama, based on biostratigraphic evidence from nearby Pacific Ocean and Caribbean Sea cores. *Geology*, **6**, 630–634.

Knowlton N (1980) Sexual selection and dimorphism in two demes of a symbiotic, pair-bonding snapping shrimp. *Evolution*, **34**, 161–173.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics*, **24**, 189–216.

Knowlton N, Keller BD (1983) A new sibling species of snapping shrimp associated with the Caribbean sea anemone *Bartholomea annulata*. *Bulletin of Marine Science*, **33**, 353–362.

Knowlton N, Keller BD (1985) Two more sibling species of alpheid shrimps associated with the Caribbean sea anemones *Bartholomea annulata* and *Heteractis lucida*. *Bulletin of Marine Science*, **37**, 893–904.

Knowlton N, Keller BD (1986) Larvae with fall far short of their potential: highly localized recruitment in an alpheid shrimp with extended larval development. *Bulletin of Marine Science*, **39**, 213–223.

Knowlton N, Weil E, Weigt LA, Guzman HM (1992) Sibling species in *Montastraea annularis*, coral bleaching, and the coral climate record. *Science*, **255**, 330–333.

Krug PJ (2011) Patterns of speciation in marine gastropods: a review of the phylogenetic evidence for localized radiation in the sea. *American Malacological Bulletin*, **29**, 169–186.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24.

Macdonald KS, Rios R, Duffy JE (2006) Biodiversity, host specificity, and dominance by eusocial species among sponge-dwelling alpheid shrimp on the Belize Barrier Reef. *Diversity and Distributions*, **12**, 165–178.

Morrison C, Rios R, Duffy E (2004) Phylogenetic evidence for an ancient rapid radiation of Caribbean sponge-dwelling snapping shrimps (*Synalpheus*). *Molecular Phylogenetics and Evolution*, **30**, 563–581.

Munday PL, Van Herwerden L, Dudgeon CL (2004) Evidence for sympatric speciation by host shift in the sea. *Current Biology*, **14**, 1498–1504.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution*, **55**, 859–868.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.

Rambaut A, Drummond AJ (2007) Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer.

Rathbun MJ (1901) The Brachyura and Macrura of Porto Rico. *Bulletin of the United States Fish Commission*, **200**, 1–127.

Regier JC (2007) Protocols, concepts, and reagents for preparing DNA sequencing templates. Version 9/19/07. www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Schluter D (2009) Evidence for ecological speciation. *Science*, **323**, 737–741.

Servedio M, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: 'magic' but not rare? *Trends in Ecology and Evolution*, **26**, 389–397.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Taylor M, Hellberg M (2005) Marine radiations at small geographic scales: speciation in neotropical reef gobies (*Elacatinus*). *Evolution*, **59**, 374–385.

Teacher AGF, Griffiths DJ (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources*, **11**, 151–153.

Tsang LM, Ma KY, Ahyong ST, Chan TY, Chu KH (2008) Phylogeny of Decapoda using two nuclear protein-coding gens: Origin and evolution of the Reptantia. *Molecular Phylogenetics and Evolution*, **48**, 359–368.

Tsang LM, Chan BKK, Shih FL, Chu KH, Chen CA (2009) Host-associated speciation in the coral barnacle *Wanella milleporae* (Cirripedia: Pyrgomatidae) inhabiting the *Millepora* coral. *Molecular Ecology*, **18**, 1463–1475.

Via S (1999) Reproductive isolation between sympatric races of pea aphids: gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. *Molecular Biology and Evolution*, **18**, 1484–1493.

Williams ST, Knowlton N, Weigt LA, Jara JA (2001) Evidence of three major clades within the snapping shrimp genus *Alpheus* inferred from nuclear and mitochondrial gene sequence data. *Molecular Phylogenetics and Evolution*, **20**, 375–389.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, The University of Texas at Austin.

CH and NK designed the experiment. CH and KS generated and analyzed the data. CH and AA drafted the manuscript. CH, AA and NK participated in the interpretation of the results. All authors approved the final manuscript.

Data accessibility

DNA sequences: GenBank Accessions nos KF130972–KF131536 (individual GenBank Accession nos listed in Appendix S2).

IMa2 input files: Dryad doi:10.5061/dryad.d1n31.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1. Bayesian analysis of individual nuclear genes.

Appendix S2. Individual GenBank Accession nos.