doi:10.1111/j.1558-5646.2008.00566.x

# A MULTILOCUS TEST OF SIMULTANEOUS DIVERGENCE ACROSS THE ISTHMUS OF PANAMA USING SNAPPING SHRIMP IN THE GENUS *ALPHEUS*

**Carla Hurt,[1,2] Arthur Anker,[3] and Nancy Knowlton[4]**

[1]*Department of Biology, University of Miami, Coral Gables, Florida 33124*

[2]*E-mail: hurtc@bio.miami.edu*

[3]*Department of Zoology, University of Florida, Gainesville, Florida 32611*

[4]*Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560*

Received June 5, 2008
Accepted September 29, 2008

**The completion of the Panamanian Isthmus is one of the greatest natural experiments in evolution, sending multiple species pairs from a broad range of taxonomic groups on independent evolutionary trajectories. The resulting transisthmian sister species have been used as model systems for examining consequences that accompany cessation of gene flow in formerly panmictic populations. However, variance in pairwise genetic distances of these "geminates" often exceeds expectations, seemingly conflicting with the assumption that separation of populations was contemporaneous with the final closure of the Isthmus. Multilocus datasets and coalescent-based analytical methods can be used to estimate divergence times while accounting for variance in gene divergence that predates isolation, thus removing the need to invoke unequal divergence times. Here we present results from Bayesian analyses of sequence data from seven nuclear and one mitochondrial marker in eight transisthmian species pairs in the snapping shrimp genus *Alpheus*. Divergence times in two species pairs were shown to occur much earlier than the Isthmus final closure, but much of the variance in pairwise genetic distances from cytochrome oxidase I (COI) was explained when ancestral polymorphisms were accounted for. Results illustrate how coalescent approaches may be more appropriate for dating recent divergences than for estimating ancient speciation events.**

**KEY WORDS:** ***Alpheus*, coalescent theory, Isthmus of Panama, molecular evolution, speciation.**

Calibration of a molecular clock requires independent information regarding the timing of species divergence, information that often comes from the fossil record. However, for species for which fossilized material is not available, the presence of a well-dated biogeographic feature can provide a time frame for estimating the rate of DNA evolution. The final closure of the Isthmus of Panama is particularly well suited for inferring the minimum amount of time two species have been separated, as it has served as a virtually impenetrable barrier to migration and gene flow among tropical marine species since its closure, the timing of which has been well documented using multiple lines of evidence and independent surveys. Estimates based on Foraminifera, isotope ratios, and fossilized molluscan shells date this event at between 2.7 and 3.5 mya (Keigwin 1982; Coates et al. 1992; Coates and Obando 1996). Furthermore, this geological event could have resulted in the near simultaneous interruption of gene flow in numerous independent species pairs, thus providing an estimate of the variance associated with the stochastic nature of mutation accumulation. Marine taxa that arose through vicariant speciation as a result of the emergence of the Isthmus, termed "geminate" species (Jordan

1908), have long been recognized for their potential contribution to studies of speciation and diversification.

The utility of transisthmian sister taxa for examining the rate and consistency of a molecular clock has received more attention than any other aspect of geminate divergence; studies of molecular evolution have made use of geminate species for nearly three decades (Lessios 1979; Vawter et al. 1980; Bermingham and Lessios 1993; Knowlton et al. 1993; Collins 1996; Bermingham et al. 1997; Knowlton and Weigt 1998). Thus far, attempts to calibrate DNA substitution rates have focused on mitochondrial genes, proceeding with the assumption that the timing of speciation for all geminates was contemporaneous with the final closure of the Isthmus. However, large variances in genetic distances among species pairs that span the Isthmus have led many to question the reliability of transisthmian-based mutation rates; for example, pairwise genetic distances among geminates in the snapping shrimp genus *Alpheus* vary nearly sevenfold (Knowlton et al. 1993; Knowlton and Weigt 1998). Explanations for this variation have included differences in the actual timing of gene flow cessation among species pairs, failure to identify true sister transisthmian species, and extreme variation in the substitution rate of mitochondrial DNA, even among closely related lineages. Variance in coalescent times due to polymorphisms in ancestral populations has, until recently, been overlooked as a potentially important explanation for variation in pairwise genetic distances among transisthmian taxa (Hickerson et al. 2003, 2006).

Gene divergence will almost always predate population divergence because of random sampling of ancestral polymorphisms by descendent populations; the degree to which gene divergence predates population divergence can be substantial when the effective population size ($N_e$) of the common ancestor is large relative to divergence time (Edwards and Beerli 2000). Hickerson et al. (2006) used an Approximate Bayesian Computational (ABC) method to test for simultaneous divergence of eight geminate echinoid species pairs while accounting for variance in coalescent times. This study provided an important first step in accounting for demographic sources of variation in transisthmian genetic distances; however, results from this study were limited as they were based on sampling of multiple alleles from a single mitochondrial locus. Population level sampling at a single nonrecombining locus can confirm if segregating alleles are reciprocally monophyletic and may reveal information about the $N_e$ of contemporary populations, but will not necessarily provide information on the coalescence time of ancestral alleles. Once reciprocal monophyly has been achieved, it is no longer possible to differentiate between variance due to divergence times and that resulting from ancestral coalescence (Wakeley and Hey 1997; Edwards and Beerli 2000). Theoretical and empirical evidence suggests that incorporating DNA sequence data from multiple, unlinked loci is the most effective way to improve the accuracy and precision of historical parameter estimates (Wakeley and Hey 1997; Edwards and Beerli 2000; Jennings and Edwards 2005). The expected variance in pairwise genetic distances from unlinked nuclear loci is directly proportional to ancestral $N_e$, thereby providing valuable information on coalescent processes in ancestral populations and reducing confidence intervals surrounding estimated divergence times. Several Maximum Likelihood and Bayesian methods capable of handling multilocus datasets have been developed which can simultaneously estimate divergence times and $N_e$ of ancestral and descendent populations while accounting for coalescent time in the common ancestor of two species.

Snapping shrimps (genus *Alpheus*) are among the most studied taxonomic groups affected by the closure of the Isthmus. Many factors contribute to making *Alpheus* the ideal study system for examining molecular divergence in transisthmian taxa. This genus contains more geminate species pairs than any other genus studied thus far, providing a naturally replicated experimental study group for testing evolutionary hypotheses. The taxonomic literature suggests that there are at least 20 transisthmian species pairs within *Alpheus* based on morphological characters and geographic distributions (Kim and Abele 1988). Second, *Alpheus* possesses well-characterized morphological traits that facilitate a rigorous analysis of the correlation between phenotypic and molecular divergences (Anker et al. 2007a,b, 2008a,b,c). Furthermore, past and ongoing research examining morphological and molecular systematics of *Alpheus* world-wide has enabled us to confidently assess transisthmian relationships, reducing the risk of misidentifying transisthmian pairs (Anker et al. 2008c; Williams et al. 2001).

Here we report results from a multiple locus Bayesian study testing for simultaneous divergence of eight transisthmian species pairs of *Alpheus*. To our knowledge, this is the first study to apply sequence data from multiple unlinked loci to address the question of simultaneous speciation across the Isthmus. Results from this study will be methodologically informative, highlighting the importance of taking a genome-wide approach to designing studies that rely on the calibration of a molecular clock. Furthermore, differentiating the subset of geminate species pairs that were most likely isolated at the time of the final closure of the Isthmus (approximately 3 mya) from those pairs that were isolated during earlier phases of the Isthmus' formation will enable us to accurately and confidently make full use of transisthmian species for measuring rates of evolution at many levels.

## Materials and Methods

### COLLECTIONS

Sequence data were obtained from multiple individuals belonging to 16 taxa (eight putative geminate species pairs) within the genus *Alpheus*. All specimens used in analysis were collected from the

Pacific and Caribbean coasts of Panama. Most samples were collected by hand, intertidally or subtidally, in crevices or under rocks (Pacific/Caribbean) (*A. cf. bouvieri*/*A. bouvieri*, *A. millsae*/*A. nuttingi*, *A. utriensis*/*A. cristulifrons*, *A. cf. malleator*/*A. malleator*, *A. panamensis*/*A. formosus*, and *A. umbo*/*A. schmitti*) whereas *A. saxidomus*/*A. simus* are unique among *Alpheus* species in that they occur in tunnels inside dead coral rubble. Finally, *A. colombiensis*/*A. estuariensis* were collected from mudflats in or near mangroves. These eight species pairs were selected based on strong evidence supporting their transisthmian geminate status (i.e., each is the other's closest relative from the other side of the Isthmus); this evidence includes geographical distribution, ecological and morphological similarities, and well-supported sister-status based on mitochondrial cytochrome oxidase I (COI) sequence (Kim and Abele 1988; Williams et al. 2001; Anker et al. 2007a, 2008b).

## LABORATORY METHODS

Between one and eight individuals were sequenced for each of seven nuclear partial coding genes as well as partial coding sequence from the mitochondrial gene COI (658 bp). The seven amplified nuclear loci were as follows: glucose phosphate dehydrogenase (GPH) (407 bp), tetrahydrofolate synthase (THS) (441 bp), alanyl-tRNA synthetase (ATS) (559 bp), elongation factor I α (EF1-α) (574 bp), elongation factor 2 (EF2) (704 bp), putative GTP-binding protein (PGBP) (726 bp), and glucose-6-phosphate isomerase (GPI) (388 bp). The size of each amplicon (indicated in parentheses) refers to the length of the PCR product in base pairs when amplified with internal primers, excluding primer sequences. In total, approximately 4457 bp of sequence across eight loci were generated for each individual.

Amplification of all nuclear and mitochondrial genes employed an RT-PCR strategy rather than direct amplification from genomic DNA. In the case of COI, this approach reduced the risk of amplification of nuclear pseudogenes, previously shown to be pervasive within the genus *Alpheus* (Williams and Knowlton 2001). With respect to nuclear loci, this method allowed us to accurately predict the size of the desired homologous amplicon across the range of species included in this study.

Total RNA was extracted using the SV Total RNA Isolation System (Promega, Madison, WI) following a modification of the manufacturers' protocol described by Regier and Schultz (1997). First-strand synthesis of cDNA was performed using MuLV reverse transcriptase (Applied Biosystems, Foster City, CA), RNase inhibitor (Applied Biosystems), and sequence-specific reverse primers. The resulting cDNA was then used as template in a polymerase chain reaction (PCR), which included sequence-specific forward primer and used thermocycler conditions described by Regier (1997). PCR product of the correct size was 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. If the initial PCR failed to generate sufficient product for sequencing, then a second PCR reaction was performed using internal primers and 2 μl of PCR product from the first amplification as template (see Table 1 for primer sequences).

Preliminary analyses in other systems as well as our results showed intraindividual variation at the selected nuclear loci to be low (Regier 2007). Furthermore, forward and reverse sequencing allowed unambiguous identification of heterozygous sites. Therefore, cloning of nuclear genes was not performed and PCR products were directly sequenced. An aliquot (2 μl) of gel-purified PCR product was quantified by electrophoresis on a 1% w/v agarose analytical gel and DNA concentration were determined by comparison of fluorescence with a standard DNA mass ladder. Cycle sequencing reactions were performed using 50–100 ng DNA and BigDye terminator version 3.1 (Applied Biosystems) following manufacturers' instructions for cycle sequencing. Reaction products were separated from unincorporated dye-terminators by centrifugation through Sephadex G-50 in a 96-well filter plate (Millipore, Billerica, MA). Products of sequencing reactions were analyzed on an ABI 3130*xl* Applied Biosystems automated capillary sequencer.

## MOLECULAR ANALYSIS

Sequence alignments were performed using the progressive alignment method as implemented by ClustalX and confirmed by eye. Heterozygous sites in nuclear genes 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 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 dataset was run a minimum of five 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. For each locus, unrooted phylogenies were estimated using the neighbor joining algorithm and the Kimura-2-Parameter (K2P) distance method as implemented by MEGA version 3.1 (Kumar et al. 2005). Branch support was estimated by bootstrap support (500 replicates) (Felsenstein 1985). MEGA was also used to estimate average K2P genetic distances between transisthmian species pairs for mitochondrial COI haplotypes.

The program IMa (Hey and Nielsen 2007) using the M-mode option was used to generate marginal distributions and multidimensional optima for population divergence times and

**Table 1.** List of forward (F) and reverse (R) primers, sources, and amplification strategies used to amplify COI, GPH,THS, ATS, EFI-α, EFII, GBP, and GPI.

| | Sequence 5′-3′ | Source | Amplification Strategy | GenBank Accession Numbers |
|---|---|---|---|---|
| **COI Primers** | | | | |
| LCOI (F) | 5′-TAW ACT TCD GGR TGH CCA AAR AAY CA-3′ | Modified from Folmer et al. 1994 | HCOI /LCOI | FJ013868–FJ013871 |
| HCOI (R) | 5′-GGT CAA CAA ATC ATA AAG ATA TTG G-3′ | Modified from Folmer et al. 1994 | Nested Reaction | FJ013874–FJ013948 |
| COI 815 (R) | 5′-TAT NCC TAR NGT TCC RAA KG-3′ | *Alpheus* sequence alignment | 1st amplification | EF092276 |
| | | | COI 815 /LCOI | EF092279–EF092283 |
| | | | 2nd amplification | EF532606–EF532613 |
| | | | HCOI /LCOI | |
| **GPH Primers** | | | | |
| 3007 79 (F) | 5′-TCA RGG TTA TAT TGA TTG CTG CAT-3′ | *Alpheus* sequence alignment | 3007 79(F) /3007 578(R) | FJ013479–FJ013529 |
| 3007 578 (R) | 5′-TTG CCA CGA ATG TTA RGA AG-3′ | *Alpheus* sequence alignment | Hemi-nested Reaction | |
| 3007 96 (F) | 5′-TGC TGC ATW AAG CTG AGT GG-3′ | *Alpheus* sequence alignment | 1st amplification | |
| | | | 3007 79(F) /3007 578(R) | |
| | | | 2nd amplification | |
| | | | 3007 96(F) /3007 578(R) | |
| **THS Primers** | | | | |
| 3017 554 (F) | 5′-GCC ATA ACT GCT GCC AAT AA-3′- | *Alpheus* sequence alignment | 3017 554(F) /3017 24(R) | FJ013841–FJ013867 |
| 3017 24 (R) | 5′-CAG ATT GAT GCC CGA ATG TT-3′ | *Alpheus* sequence alignment | Hemi-nested Reaction | |
| 3017 521 (F) | 5′-CAG ATT GAT GCC CGA ATG TT-3′ | *Alpheus* sequence alignment | 1st amplification | |
| | | | 3017 554(F) /3017 24(R) | |
| | | | 2nd amplification | |
| | | | 3017 521(F) /3017 24(R) | |
| **ATS Primers** | | | | |
| 3070 65F | 5′-GYT CRG AAA TTC ATT WYG ACC-3′ | *Alpheus* sequence alignment | 3070 65F /3070 728R | FJ013358–FJ013362, |
| 3070 728R | 5′-ARY CTC CAA GCC ACA TCA C-3′ | *Alpheus* sequence alignment | Nested Reaction | FJ003364–FJ013370, |
| M-3070–101F | 5′-ATG CTG CWC ATC TYG TGA AT-3′ | *Alpheus* sequence alignment | 1st amplification | FJ013374–FJ013400, |
| M-3070–705R | 5′-GAA TYT CCY TRC WAC CTC CT-3′ | *Alpheus* sequence alignment | 3070 65F /3070 728R | FJ013403–FJ013475 |
| | | | 2nd amplification | |
| | | | 3070 101F /3070 705R | |
| **EF1-α Primers** | | | | |
| EF1-α 419 (F) | 5′ ACA ACA TGC TGG AGA AGT CAG A-3′ | Williams et al. 2001 | EF1-α 419(F) /EFI-α 1240(R) | FJ013530–FJ013554 |
| EFI-α 1240 (R) | 5′-TTM ACG ATG CAR GAG TCM CC-3′ | Alignment between *Alpheus* sequence and GenBank cDNA sequence for crab[a] | Nested Reaction | FJ013556–FJ013566 |
| EFI- α 511 (F) | 5′-CCC TGG TGG AAG AAG AAG AA-3′ | *Alpheus* sequence alignment | 1st amplification | AF310786, AF310786, |
| EFI- α 1070 (R) | 5′-RCC RGT ACG CCT GTC AAT CT-3′ | *Alpheus* sequence alignment | EF1-α 419(F) /EFI-α | AF310789, AF310790, |
| | | | 1240(R) 2nd amplification | AF310792–AF310794, |
| | | | EF1-α 511(F) /EFI-α 1070(R) | AF310797–AF310798, |

Continued.

**Table 1.** Continued.

| | Sequence 5′-3′ | Source | Amplification Strategy | GenBank Accession Numbers |
|---|---|---|---|---|
| | | | | AF310807–AF310809, |
| | | | | AF310813–AF310814, |
| | | | | AF310820–AF310823, |
| | | | | AF310827, AF310828 |
| **EF2- Primers** | | | | |
| EF2–723(F) | 5′-MMA AGY TST GGG GTG ARA AC-3′ | Alignment between *Alpheus* sequence and GenBank cDNA sequence for crab[b] | EF2–723 F /EF2–1587 R | FJ013567–FJ013572 |
| EF2–1587 (R) | 5′-AYR ATG TGY TCT CCR GAY TC-3′ | Alignment between *Alpheus* sequence and GenBank cDNA sequence for crab[b] | Nested Reaction | FJ013674–FJ013578 |
| EF2–739 (F) | 5′-GAG RGC YTT CAA CAC CTA YA-3′ | *Alpheus* sequence alignment | 1st amplification | FJ013582–FJ013614 |
| EF2–1499 (R) | 5′-ART CGG AGG GGT TCT TGG-3′ | *Alpheus* sequence alignment | EF2–723(F)/EF2–1587(R) | FJ013617–FJ013664 |
| | | | 2nd amplification | |
| | | | EF2–739(F)/EF2–1499(R) | |
| **PGBP – Primers** | | | | |
| 42fin 1 (F) | 5′-GCN GAR AAY TTY CCI TTY TG-3′ | Regier 2007 | 42fin 1F /42fin 1R | FJ013666–FJ013671 |
| 42fin 2 (R) | 5′-ATD ATR AAN CCY TTY TCR AAR TC-3′ | Regier 2007 | Nested Reaction | FJ013675–FJ013710 |
| GBP 103 (F) | 5′-CTG GGC CTS CTG TGA AGW AG-3′ | *Alpheus* sequence alignment | 1st amplification | FJ013713–FJ013760 |
| GBP 781 (R) | 5′-TGC AAG YAA AGT CCC TGC ATT-3′ | *Alpheus* sequence alignment | 42fin 1F /42fin 1R | |
| | | | 2nd amplification | |
| | | | GBP 103(F) /GBP 781(R) | |
| **GPI – Primers** | | | | |
| DS1097 (F) | 5′-AAT CTA ATG GAA AGT AYG TAA C-3′ | Williams et al. 2001 | DS1097(F) /DS1574(R) | FJ013761–FJ013771 |
| DS1574 (R) | 5′-AGT AGC TCA ACA CCC CAC TGA TC-3′ | Williams et al. 2001 | Hemi-nested Reaction | FJ013775–FJ013797 |
| DS1523 (R) | 5′-TGG GTG AAA ATC TTG TGT TC-3′ | Williams et al. 2001 | 1st amplification | FJ013800–FJ013839 |
| | | | DS1097(F) /DS1574(R) | AF250552–AF250556, |
| | | | 2nd amplification | AF310730–AF310734, |
| | | | DS1097(F) /DS1523(R) | AF310736–AF310737, |
| | | | | AF310742–AF310743, |
| | | | | AF310748–AF310749, |
| | | | | AF310754–AF310755, |
| | | | | AF310761–AF310762, |
| | | | | AF310771–AF310772, |
| | | | | AF310773–AF310774, |

*Note*. GenBank Accession Nos.: [a]U90050; [b]AY305506.

theta values of ancestral and daughter populations. This program uses a maximum likelihood/Bayesian framework to fit a six-parameter model of divergence with migration to single or multilocus datasets from a pair of closely related species or populations. The six parameters estimated by this model include population mutation parameters for the ancestral population and two daughter populations ($\theta_A = 4N_A\mu$, $\theta_1 = 4N_1\mu$, and $\theta_2 = 4N_2\mu$, respectively), the scaled time parameter ($t = T\mu$), and directional migration rates scaled by the neutral mutation rate ($m_1 = m_1\mu$ and $m_2 = m_2\mu$), Here, $N_A$, $N_1$, and $N_2$ are the effective population sizes of the ancestral and daughter populations, respectively; $\mu$ is the geometric mean of the per gene mutation rate per generation; $T$ is the number of generations since population divergence; and $m_1$ and $m_2$ are the per generation migration rates from population one into population two and from population two into population one, respectively. Here we assume that no migration occurred after separation; this is accomplished by setting the upper limit of the prior distribution of $m_1$ and $m_2$ to zero. All resulting parameter estimates reflect the parameter value corresponding with the peak of the posterior distribution. Inheritance scalars were provided in the input file to account for the approximate fourfold smaller effective population size of mitochondrial versus nuclear alleles. Variation in mutation rates among loci can result in an upward bias in ancestral $N_e$ and underestimates of divergence times (Yang 2002). Because no prior information was available regarding locus-specific mutation rates, mutation rates were treated as parameters to be estimated from the data. IMa uses independent mutation rate scalars for each locus, which are drawn from a uniform log scale prior distribution and are free to vary under the constraint that their product is equal to one (Hey and Nielsen 2004).

The Hasegawa–Kishino–Yano (HKY) (Hasegawa et al. 1985) model of sequence evolution was applied to mitochondrial sequence datasets and the Infinite Sites (IS) model (Kimura 1969) of evolution was applied to all nuclear datasets. The scalar for $\theta$ maximum for the ancestral population and both daughter populations was set at 10; this sets the upper bound for the prior for $\theta_A$, $\theta_1$, and $\theta_2$ at 10 times the estimated value from the data. The Metropolis-Coupled Markov Chain Monte Carlo ($MC^3$) method option was selected to improve mixing. Preliminary analysis indicated that simple linear mixing with five chains and 10 swap attempts per step was sufficient to maintain adequate levels of chain swapping for all datasets.

As with any Markov Chain Monte Carlo method, it is critical to run the program for sufficient number of iterations in order for the sample distribution to reach convergence to the stationary distribution. To assess convergence, plot trend lines and the effective sample size (ESS) values were examined for each run. In addition, a minimum of three runs for each dataset were examined for consistency in parameter estimates.

In addition to IMa, we also used an alternative Bayesian Markov chain Monte Carlo method, MCMCcoal 1.2 (Yang 2002; Rannala and Yang 2003), to estimate $t$ ($t = T\mu$), as well as $\theta_A$, $\theta_1$, and $\theta_2$ (as in IM, $\theta_A = 4N_A\mu$, $\theta_1 = 4N_1\mu$, and $\theta_2 = 4N_2\mu$, respectively). As in IMa, this analysis is improved by the incorporation of sequence data from multiple, independent loci and can accommodate multiple sequences per population. Both programs assume free recombination between loci, no recombination within loci, neutrality of existing variation, and a known species tree. However, MCMCcoal assumes no gene flow following initial population separation, whereas IMa can be used to estimate postdivergence migration rates, although we have assumed no migration following separation for these analyses so that both analyses will estimate the same set of parameters (this is also a more biologically reasonable assumption for the Isthmus). We employed the variable mutation rate option that estimates the relative per locus mutation rate from the data. This option assigns a prior distribution for mutation rates as a Dirichlet distributed variable with mean equal to one and a user-defined $\alpha$ (used to determine amount of rate variation among loci); $\alpha$ was set at two for all analyses (Yang 2002; Rannala and Yang 2003). As in IM, inheritance scalars were used in the input file to account for the approximate fourfold smaller effective population size of mitochondrial versus nuclear alleles. Parameter estimates presented in the results correspond to values at the peak of the posterior distribution.

MCMCcoal requires a user-specified gamma prior distribution for each of the parameters to be estimated; this can be strength or a limitation depending on the availability of pre-existing information. Hyperparameters $\alpha$ and $\beta$ are used to determine the mean ($\alpha/\beta$) and variance ($\alpha/\beta^2$) of the gamma prior. We specified identical "vague" priors for all eight species pairs to minimize the influence of the prior on the posterior distribution. Rannala and Yang (2003) recommend a minimum value of 1, because if $\alpha > 1$ the distribution will have a maximum greater than zero and it is likely that the true value for all parameters will have positive values. Therefore, a value of 1 was specified for each $\alpha$ prior, and $\beta$ priors were selected so that divergence time parameters had a mean value of 3.0 million years and ancestral effective population sizes of 50,000 with a per site mutation rate of $1 \times 10^{-9}$, similar to rates observed in other animal nuclear genomes (Moriyama and Gojobori 1992; Kumar and Subramanian 2002). Analyses were conducted using a burn-in value of 10,000 generations, a sample interval of 2, and a sample size of 1,000,000 generations. Each analysis was run at least three times with different random seeds to examine the consistency of parameter estimates and to assess convergence. Fine-tune parameters were optimized for each two-species dataset to maintain the acceptance proportions with the interval of (0.15, 0.7) as suggested by the authors.

Both IMa and MCMCcoal assume that existing sequence variation is neutral and that there is no recombination within loci. The Hudson–Kreitman–Aguade test (HKA test) was used to test for selection across multiple loci (Hudson et al. 1987) as performed by the software program HKA (Jody Hey, Rutgers University). This test is based on a prediction that follows from the Neutral Theory stating that loci exhibiting high rates of divergence between species will also contain high levels of variation (segregating sites) within species. To test for recombination within the seven nuclear loci we applied the "four-gamete test" (Hudson and Kaplan 1985) to all two-population datasets, as implemented by the software program DNAsp 4.10.9 (Rozas et al. 2003). This method estimates the minimum number of recombination events in the history of a sample of sequences. In population datasets in which recombination was detected, haplotypes were divided into nonrecombining segments. Only the largest segment from each locus was used in subsequent analysis, and flanking regions were discarded.

# Results

## DATA CHARACTERISTICS AND GENE TREE RECONSTRUCTIONS

Five of eight loci were successfully amplified in all eight species pairs; we were unable to obtain GPH and THS sequence from *A. colombiensis*/*A. estuariensis*, THS did not amplify in *A. cf malleator*/*A. malleator*, and EF2 did not amplify in *A. panamensis*/*A. formosus*. Sample sizes ranged from one to eight individuals per population and are summarized in Table 2. The "four-gamete test" (Hudson and Kaplan 1985) detected historical recombination events at all seven nuclear loci in at least one of the transisthmian species pairs. Recombination-free segment lengths for all datasets ranged from 229 bp to 657 bp (Table 2).

A total of 80 mitochondrial COI haplotypes were identified by direct sequencing of PCR products and 387 unique nuclear haplotypes were reconstructed from sequenced nuclear PCR products using PHASE (Stephens et al. 2001). The elimination of

**Table 2. Sample sizes for each locus. Italicized numbers at the top of each column indicate the total length (excluding primer sequence) of each PCR product (bp). Numbers in italics indicate the length of the largest nonrecombining fragment for each species pair as indicated by the four-gamete test.**

| Taxa | $COI_a$ | $GPH_b$ | $THS_b$ | $ATS_b$ | $EF1\alpha_b$ | $EF2_b$ | $PGBP_b$ | $GPI_b$ |
|---|---|---|---|---|---|---|---|---|
| | *648* | *406* | *458* | *559* | *574* | *704* | *726* | *388* |
| | *648* | — | — | *571* | *277* | *655* | *501* | *388* |
| *A. colombiensis* EP | 5 | — | — | 14 | 10 | 10 | 14 | 18 |
| *A. estuariensis* WA | 9 | — | — | 12 | 4 | 10 | 8 | 16 |
| | *648* | *406* | *458* | *545* | *391* | *655* | *657* | *371* |
| *A. cf. bouvieri* EP | 4 | 10 | 6 | 12 | 4 | 12 | 12 | 14 |
| *A. bouvieri* WA | 5 | 12 | 4 | 12 | 4 | 10 | 14 | 14 |
| | *648* | *406* | — | *230* | *391* | *655* | *500* | *388* |
| *A. cf. malleator* EP | 10 | 10 | — | 16 | 8 | 10 | 16 | 14 |
| *A. malleator* WA | 7 | 4 | — | 12 | 4 | 16 | 8 | 8 |
| | *648* | *399* | *458* | *364* | *286* | *655* | *657* | *388* |
| *A. millsae* EP | 7 | 8 | 6 | 14 | 16 | 14 | 14 | 16 |
| *A. nuttingi* WA | 6 | 8 | 4 | 16 | 10 | 8 | 10 | 12 |
| | *648* | *406* | *458* | *229* | *341* | — | *298* | *388* |
| *A. panamensis* EP | 7 | 6 | 2 | 14 | 6 | — | 10 | 10 |
| *A. formosus* WA | 4 | 6 | 2 | 16 | 4 | — | 12 | 14 |
| | *582* | *369* | *458* | *527* | *304* | *655* | *358* | *388* |
| *A. saxidomus* EP | 6 | 8 | 6 | 16 | 6 | 16 | 10 | 14 |
| *A. simus* WA | 3 | 2 | 4 | 14 | 8 | 12 | 8 | 12 |
| | *644* | *406* | *458* | *271* | *391* | *649* | *657* | *388* |
| *A. umbo* EP | 8 | 2 | 2 | 12 | 8 | 12 | 10 | 14 |
| *A. schmitti* WA | 2 | 6 | 4 | 12 | 8 | 10 | 10 | 14 |
| | *648* | *406* | *208* | *326* | *391* | *495* | *519* | *388* |
| *A. utriensis* EP | 6 | 10 | 14 | 14 | 8 | 12 | 12 | 8 |
| *A. cristulifrons* WA | 6 | 10 | 4 | 14 | 8 | 12 | 12 | 6 |

**Note: Subscript "a", Number of sequenced mitochondrial COI haplotypes for each taxa. Italicized numbers indicate the length in base pairs of the final sequence alignment for each species pair.**

**Subscript "b", Number of PHASE reconstructed haplotypes from genotype sequence data for each taxon.**

heterozygous genotypes that did not produce identical haplotype reconstructions across all five replicated PHASE runs resulted in the removal of only one genotype.

With one exception, alleles for geminate pairs formed paraphyletic or reciprocally monophyletic sister clades at all loci, supporting the putative sister species status of geminate taxa included in this study (Fig. 1). The exception was *A. utriensis*/*A. cristulifrons*, which formed well-supported sister clades in six of eight topologies suggesting sister taxa status according to the majority-rule criterion (Saitou and Nei 1986; Pamilo and Nei 1988). Neighbor-joining reconstruction of the mitochondrial COI phylogeny resulted in reciprocal monophyly of all populations. However, none of the species pairs were reciprocally monophyletic for all seven nuclear phylogenetic reconstructions. Paraphyletic and polyphyletic relationships among nuclear alleles, despite well-resolved mitochondrial clades, are a common transitory phase in phylogenetic reconstructions of recent speciation events; nuclear alleles typically require longer divergence times for reciprocal monophyly than mitochondrial haplotypes sampled from the same taxa due to their larger effective population size (Avise 2004).

Within population levels of variation varied greatly among taxa and among loci (Table 3); *A. bouvieri* had the most mutations of any population averaging 8.9 segregating sites per locus, whereas *A. umbo* had the least number of segregating sites with an average of 2.1 polymorphic sites per locus. The pair *A. utriensis*/*A. cristulifrons* had the most fixed difference of any species pair with an average of 25.5 bp differences averaged across all loci. The pair *A. cf. malleator*/*A. malleator* had the fewest fixed differences averaging 7.6 differences between them. Mitochondrial COI had higher levels of variation, both within populations and between species pairs, than nuclear loci, averaging 12.2 segregating sites per population and 60.3 differences between sister taxa versus 5.7 segregating sites and 3.7 fixed differences for nuclear loci. The average number of segregating sites for nuclear loci was 5.7 per population. No evidence of selection was detected for any of the species pairs using the HKA test; *P*-values ranged from 0.134 to 0.827 (Table 3).

## IM RESULTS

To test for convergence, three independent Metropolis-coupled Markov chains were performed for each of the two-population datasets. Posterior distributions for $\theta_A$, $\theta_1$, $\theta_2$, and $t$ were nearly identical across replicates, suggesting that the chains are converging to the correct stationary distribution given the model (Fig. 2). The posterior distribution of $t$ for *A. utriensis*/*A. cristulifrons* is flatter than for other species pairs, possibly resulting from the substantially older divergence time of this pair. Longer divergence times lead to reciprocal monophyly and fewer shared polymorphisms between taxa, a situation that results in a negative

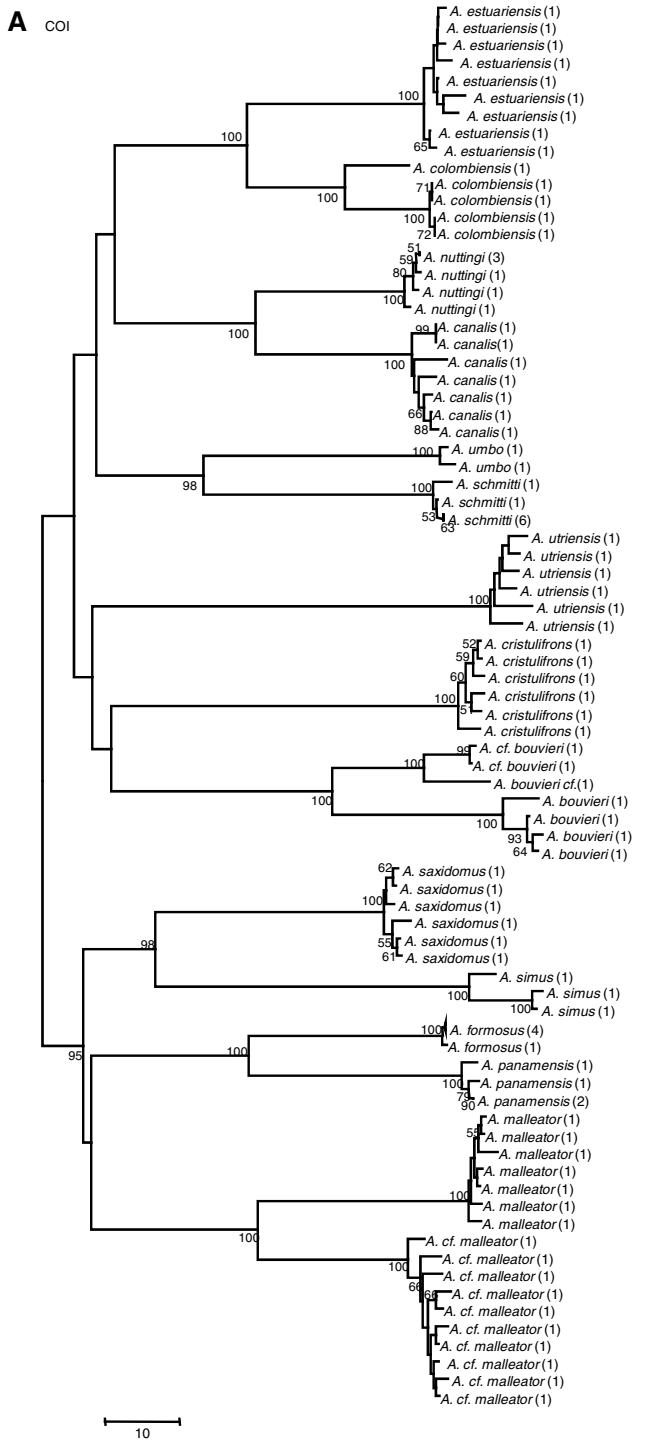


**Figure 1. Neighbor-joining phylograms of (A) mitochondrial cytochrome Oxidase I (COI) haplotypes and PHASE reconstructed haplotypes from (B) glucose phosphate dehydrogenase (GPH), (C) tetrahydrofolate synthase (THS), (D) alanyl-tRNA synthetase (ATS), (E) elongation factor I α (EF1-α), (F) elongation factor 2 (EF2), (G) putative GTP-binding protein (PGBP), and (H) glucose-6-phosphate isomerase (GPI) for *Alpheus* transisthmian species pairs. Numbers at nodes indicate percentage bootstrap support (1000 replicates). Numbers in parentheses indicate the number of copies of each haplotype.**

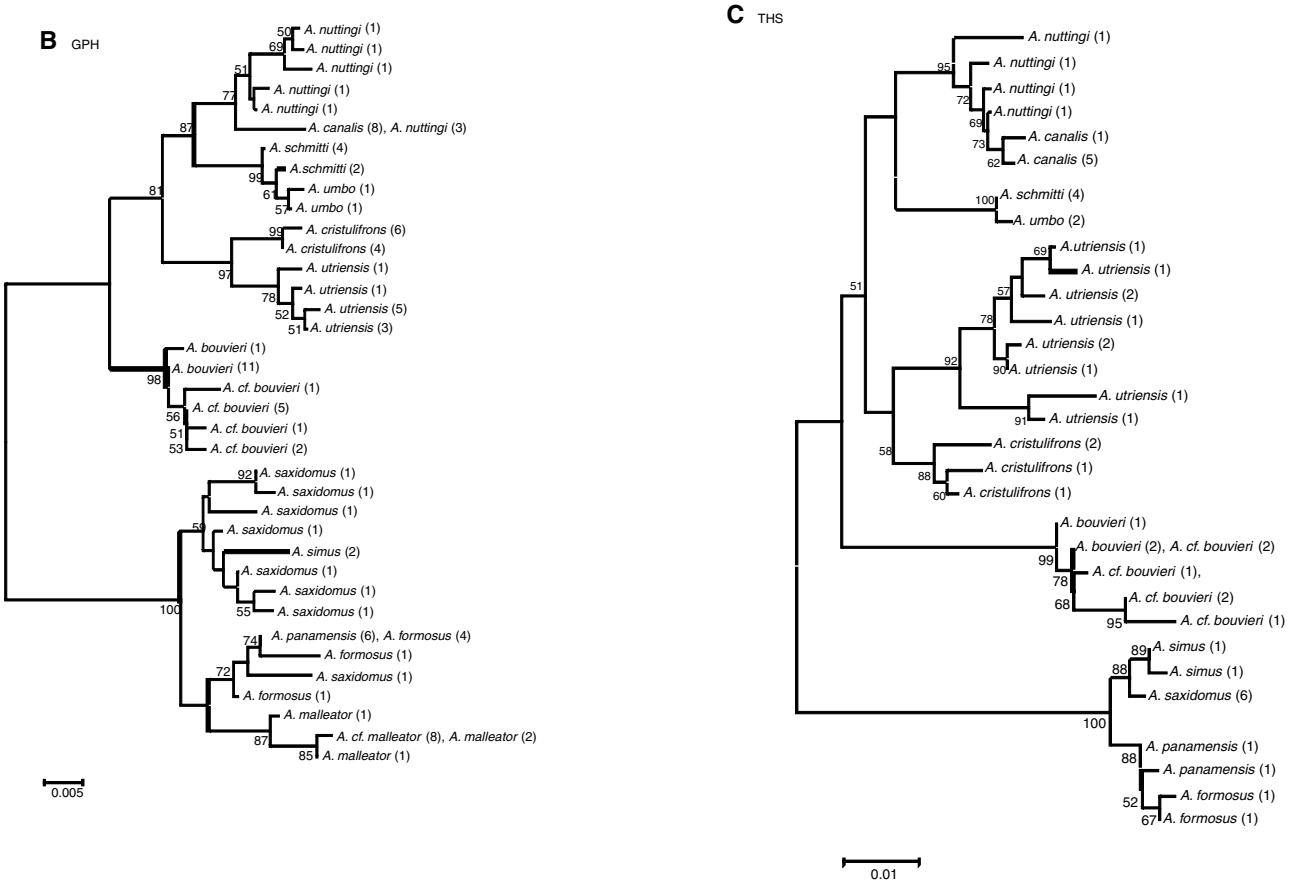


**Figure 1.** Continued.

correlation between estimates of $N_e$ and divergence time, thus increasing the variance in parameter estimates (Wakeley and Hey 1996).

Parameter estimates from IMa are scaled by the geometric mean of the per locus mutation rate. Estimates of $t$ ranged from 0.52 to 4.24. It has been suggested that mangrove geminate species pairs were likely the last species to be separated by the closing of the Isthmus (Knowlton and Weigt 1998). Consistent with this assumption, *A. colombiensis*/*A. estuarensis*, the only mangrove species pair included in this study, had the smallest estimated divergence time of any pairs examined ($t = 1.23$) (Fig. 2) in both Bayesian analyses. If we assume that *A. colombiensis*/*A. estuariensis* were isolated at the final closing of the Isthmus, approximately 3 mya and assuming annual generation times, the geometric mean of the per locus mutation rate across the loci included in this study is approximately $1.7 \times 10^{-7}$. To compare divergence time estimates with results from MCMCcoal analyses and previous estimates based on mitochondrial sequence data (Knowlton and Weigt 1998), this rate was used to obtain rough estimates of divergence times for all other species pairs, under the assumption that the average among locus mutation rate was the same across *Alpheus*. Estimated divergence time for *A. cf. malleator*/*A. malleator* was only slightly larger than that of the mangrove species and the 95% HPD for the two estimates overlapped greatly ($t = 0.72$ corresponding to a divergence time of 4.1 mya). The next three species pairs (going back in time)—*A. panamensis*/*A. formosus*, *A. millsae*/*A. nuttingi*, and *A. cf. bouvieri*/*A. bouvieri*—split at roughly the same time (6.0 mya, 7.1 mya, and 7.6 mya). *Alpheus umbo*/*A. schmitti* and *A. saxidomus*/*A. simus* show substantially older divergence times ($T = 12.7$ mya and 16.5 mya, respectively). Finally, the estimated divergence time for *A. utriensis*/*A. cristulifrons* predated all others by 8 million years ($T = 24.5$).

No consistent trends regarding relative population size estimates between Eastern Pacific ($\theta_{EP}$) and Western Atlantic ($\theta_{WA}$) populations were observed (Table 4A). The 95% HPDs for $\theta_{EP}$ and $\theta_{WA}$ overlapped substantially for all species pairs. Ratios of ancestral to daughter effective population sizes ($(\theta_{EP} + \theta_{WA})/\theta_A$) were computed to detect change in population size since speciation. Results suggested population expansion in the daughter populations relative to the ancestral populations for five of eight pairs, whereas *A. colombiensis*/*A. estuariensis*, *A. umbo*/*A. schmitti*, and *A. utriensis*/*A. cristulifrons* showed patterns of population decline; values ranged from 0.68 to 3.39.

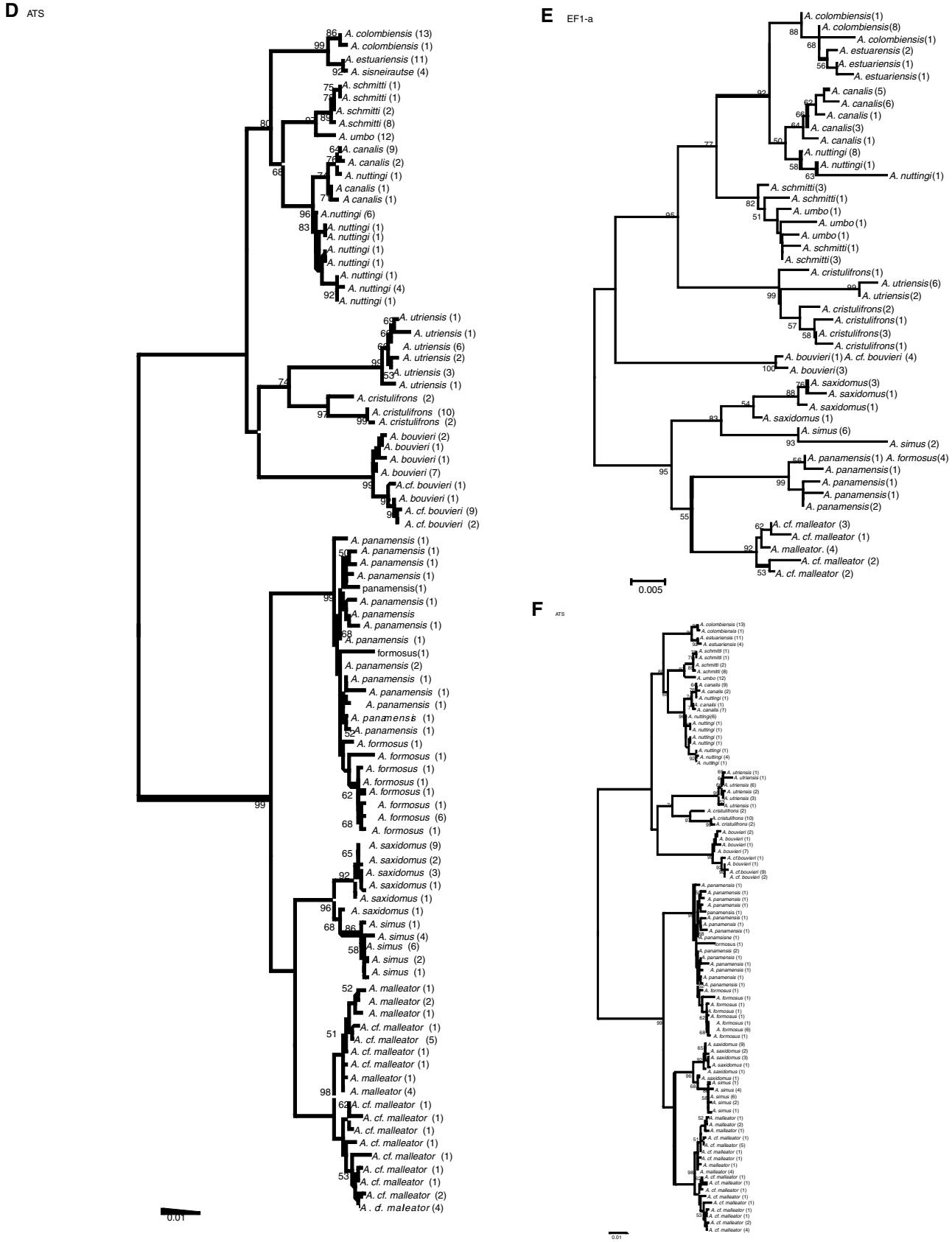


**Figure 1.** Continued.

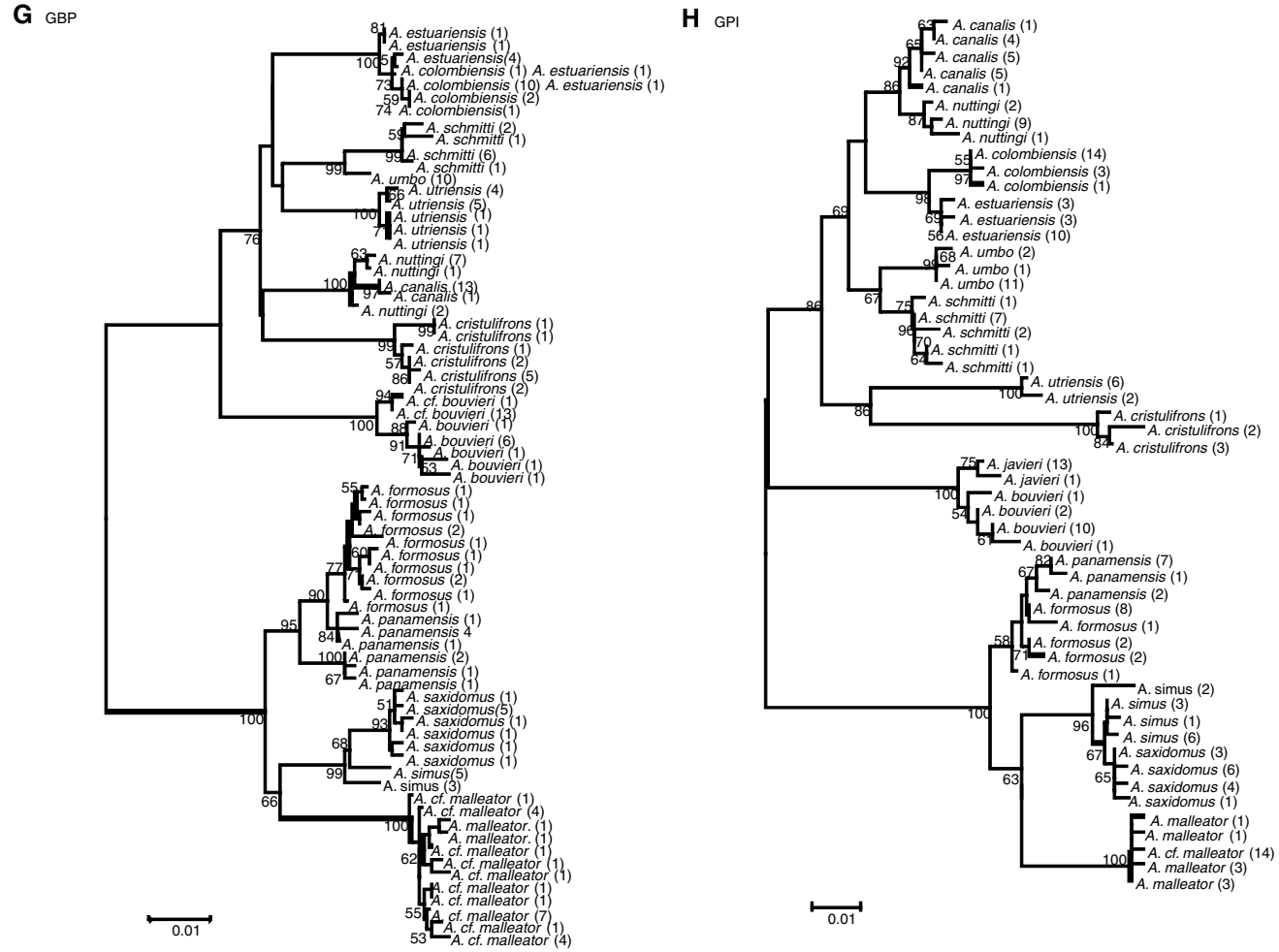


**Figure 1.** Continued.

### MCMCcoal RESULTS

As in IMa, multiple runs of MCMCcoal were nearly identical across replicates, suggesting that there were a sufficient number of simulations to achieve convergence. Consistent with our previous findings, *A. colombiensis*/*A. estuariensis* had the smallest divergence time estimate ($t = 0.0008$). Following the above reasoning, we used this value for $t$ and a divergence time of 3 million years to obtain a rough estimate of the average per site mutation rate across loci ($\mu = 2.3 \times 10^{-10}$). Using this calibration, the MCMCcoal estimated divergence times across species pairs were much more spread out than times calculated from IMa, spanning more than 70 million years (compared to less than 30 million years) (see Discussion). The peak posterior probability estimate for the divergence time of *A. cf. malleator*/*A. malleator*, *A. panamensis*/*A. formosus*, *A. millsae*/*A. nuttingi*, and *A. saxidomus*/*A. simus* were very similar for the two analyses (4.9 mya, 7.9 mya, 9.4 mya, and 17.7 mya, respectively). However, estimated divergence times for *A. cf. bouvieri*/*A. bouvieri*, *A. schmitti*/*A. umbo*, and *A. utriensis*/ *A. cristulifrons* were much larger than those obtained by IMa (17.5 mya, 26.2 mya, and 74.1 mya, respectively).

MCMCcoal estimates of $N_e$ indicated very little difference in the relative population sizes of Eastern Pacific and Western Atlantic species pairs (Table 3B). The ratio of $(\theta_{EP} + \theta_{WA})/\theta_{Ancestral}$ was generally similar to what we found for the IMa analysis with two exceptions; relative ancestral to daughter $N_e$ suggested increasing populations for both *A. saxidomus*/*A. simus* and *A. utriensis*/*A. cristulifrons* in MCMC analyses.

## Discussion

The well-dated completion of the Isthmus of Panama provides unparalleled opportunities for the study of virtually all aspects of evolutionary change. Transisthmian species pairs are ideal models for studying rates of morphological divergence, adaptation of life-history traits to different environments, and accumulation of reproductive barriers leading to speciation. However, use of this natural laboratory to its full potential requires a confident assessment of when individual transisthmian species pairs ceased to exchange genetic material. Assignment of absolute divergence times from molecular sequence data is not trivial, however, as

**Table 3.** Number of segregating sites within Eastern Pacific/Western Atlantic populations and fixed differences between populations in parentheses. *P*-values indicate results of HKA test for selection at multiple loci for each species pair.

| Taxa | COI | GPH | THS | ATS | EFI- α | EFII | GBP | GPI | *P*-value HKA |
|---|---|---|---|---|---|---|---|---|---|
| *A. cf. bouvieri/A. bouvieri* | 1/17 (37) | — | 7/8 (0) | 3/9 (0) | 0/1 (0) | 1/4 (1) | 1/9 (5) | 2/5 (2) | 0.635 |
| *A. canalis/A. nuttingi* | 15/6 (45) | 0/9 (0) | 2/8 (10) | 4/9 (0) | 3/5 (2) | 3/1 (0) | 1/3 (3) | 5/4 (2) | 0.827 |
| *A. colombiensis/A. estuariensis* | 21/18 (42) | — | — | 1/2 (4) | 3/2 (0) | 4/3 (0) | 2/4 (0) | 2/2 (4) | 0.459 |
| *A. cf. malleator/A. malleator* | 22/12 (49) | 1/3 (0) | — | 15/7 (0) | 4/0 (1) | 7/2 (1) | 2/9 (1) | 0/3 (1) | 0.405 |
| *A. panamensis/A. formosus* | 2/8 (60) | 0/4 (0) | 1/1 (1) | 20/18 (0) | 4/0 (0) | 4/6 (2) | 15/10 (2) | 3/5 (1) | 0.168 |
| *A. saxidomus/A. simus* | 7/14 (74) | 14/0 (4) | 5/6 (1) | 9/5 (3) | 5/5(6) | 7/2 (2) | 9/7 (4) | 1/6 (1) | 0.364 |
| *A. umbo/A. schmitti* | 17/3 (69) | 1/1 (3) | 0/0 (1) | 3/0 (6) | 4/4 (0) | 2/7 (6) | 6/06 (9) | 5/2 (7) | 0.411 |
| *A. utriensis/A. cristulifrons* | 21/11 (106) | 4/1 (5) | 7/20 (5) | 9/12 (21) | 1/7 (3) | 6/1 (14) | 8/8 (22) | 2/6 (28) | 0.134 |

measures of genetic distances between members of a species pair are confounded by stochastic demographic factors such as ancestral population sizes and population structure. Multilocus coalescent-based approaches can decouple the variance in genetic distance estimates, which results from random sampling of ancestral lineages and variance due to differences in divergence times. Here we employed two different Bayesian Markov chain Monte Carlo methods to test the null hypotheses of simultaneous isolation among eight geminate species pairs in the genus *Alpheus*. Results from these analyses indicate that both population demographic processes as well as staggered divergence times have contributed to the variance in pairwise genetic distances observed in mitochondrial data, and emphasize the benefits of employing coalescent-based methods in studies of comparative phylogeography.

Coalescent methods such as IMa and MCMCcoal, which account for gene divergence within the common ancestor, are most critically applied when populations have recently split; when divergence times are small, genetic polymorphisms that were present in the common ancestor can account for a significant proportion of the total amount of observed genetic divergence at any single locus. Failure to account for ancestral coalescent times may explain discrepancies observed in pairwise genetic distance estimates from COI and our Bayesian estimates of *t*. For the five most recently diverged species pairs, transisthmian genetic distances from COI and Bayesian parameter estimates appear uncorrelated and both IMa and MCMC confidence intervals are widely overlapping (Fig. 3), suggesting that much of the variance in COI distances may be accounted for by ancestral lineage sorting. However, when we include more anciently diverged species, *A. umbo*/*A. schmitti*, *A. saxidomus*/*A. simus*, and *A. utriensis*/*A. cristulifrons*, positive correlation between distance estimates emerges; as divergence times increase, the overestimation of population divergence times resulting from ancestral lineage sorting is less problematic.

These results have important implications for selecting appropriate geminate species pairs for investigating evolutionary processes. Both IMa and MCMCcoal results indicate that population divergence for the four most recently diverged geminate pairs—*A. colombiensis*/*A. estuariensis*, *A. cf. malleator*/*A. malleator*, *A. panamensis*/*A. formosus*, and *A. millsae*/*A. nuttingi*—occurred within a small time frame, and confidence intervals for these parameters were widely overlapping. It is likely that isolation in these taxa was contemporaneous with the final stages of the formation of the Isthmus. Given these results, these four geminate species pairs are particularly promising candidate study systems for investigators interested in examining patterns of differentiation across the Isthmus. Although simultaneous divergences between *A. simus*/*A. saxidomus* and *A. cf. bouvieri*/*A. bouvieri* cannot be rejected based on 95% HPD (Highest Posterior Density) intervals, inconsistent results from MCMCcoal and IMa make this conclusion less reliable.

Although parameter estimates for recently diverged taxa were similar for the two analysis methods, inconsistencies in IMa and MCMCcoal parameter estimates and flattened posterior probability densities were observed in those species pairs with older divergence times, suggesting that coalescent approaches may not be optimal for divergence time estimates in anciently diverged taxa. As time since population divergence increases, there is an increased probability that two populations will be reciprocally monophyletic at a given locus; on average, will occur after $4N_e$ generations (Wakeley and Hey 1997; Edwards and Beerli 2000). Once a locus has achieved reciprocal monophyly, shared ancestral polymorphisms will no longer be segregating within daughter populations; these shared polymorphisms are a valuable source of information regarding $N_e$ of the common ancestor. Multilocus methods provide information regarding ancestral $N_e$, as variance in coalescence times among multiple unlinked loci is expected to increase with increasing ancestral population size. However, in the absence of shared polymorphisms,

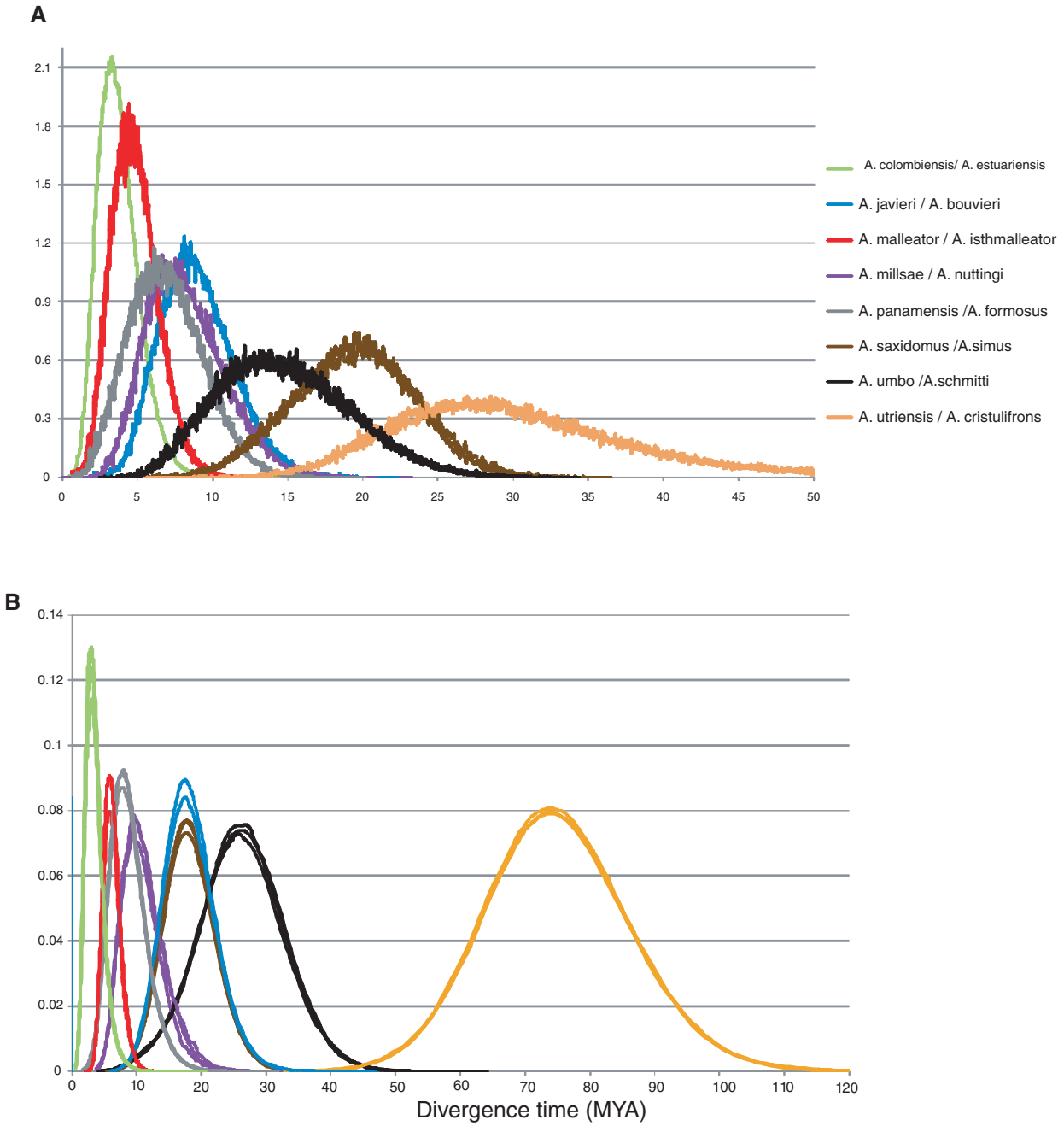


**Figure 2.** **Posterior distributions of divergence time parameter estimates for eight transisthmian species pairs generated using (A) IM software and (B) MCMCcoal1.2 software. *x*-axis is in million years. Three overlapping replicates are shown for each species pair analysis. Species pairs are specified by color.**

many more loci may be needed to obtain reliable parameter estimates.

IMa and MCMCcoal differ in their definitions of μ and the methods they use to estimate locus-specific mutation rates from the data; such differences may have contributed to the conflict in parameter estimates for the two methods. Parameter estimates for both programs are scaled by the mutation rate μ, however, MCMCcoal defines μ as the arithmetic mean of the per site mutation rate and assigns it a Dirichlet distributed prior distribution to estimate locus-specific rates from the data, whereas IMa defines μ as the geometric mean of the per locus mutation rate, assigning it a uniform log scale prior distribution for estimating rates.

Despite large confidence intervals surrounding divergence time parameters estimates for *A. utriensis*/*A. cristulifrons* and *A. umbo*/*A. schmitti*, multiple lines of evidence indicate that

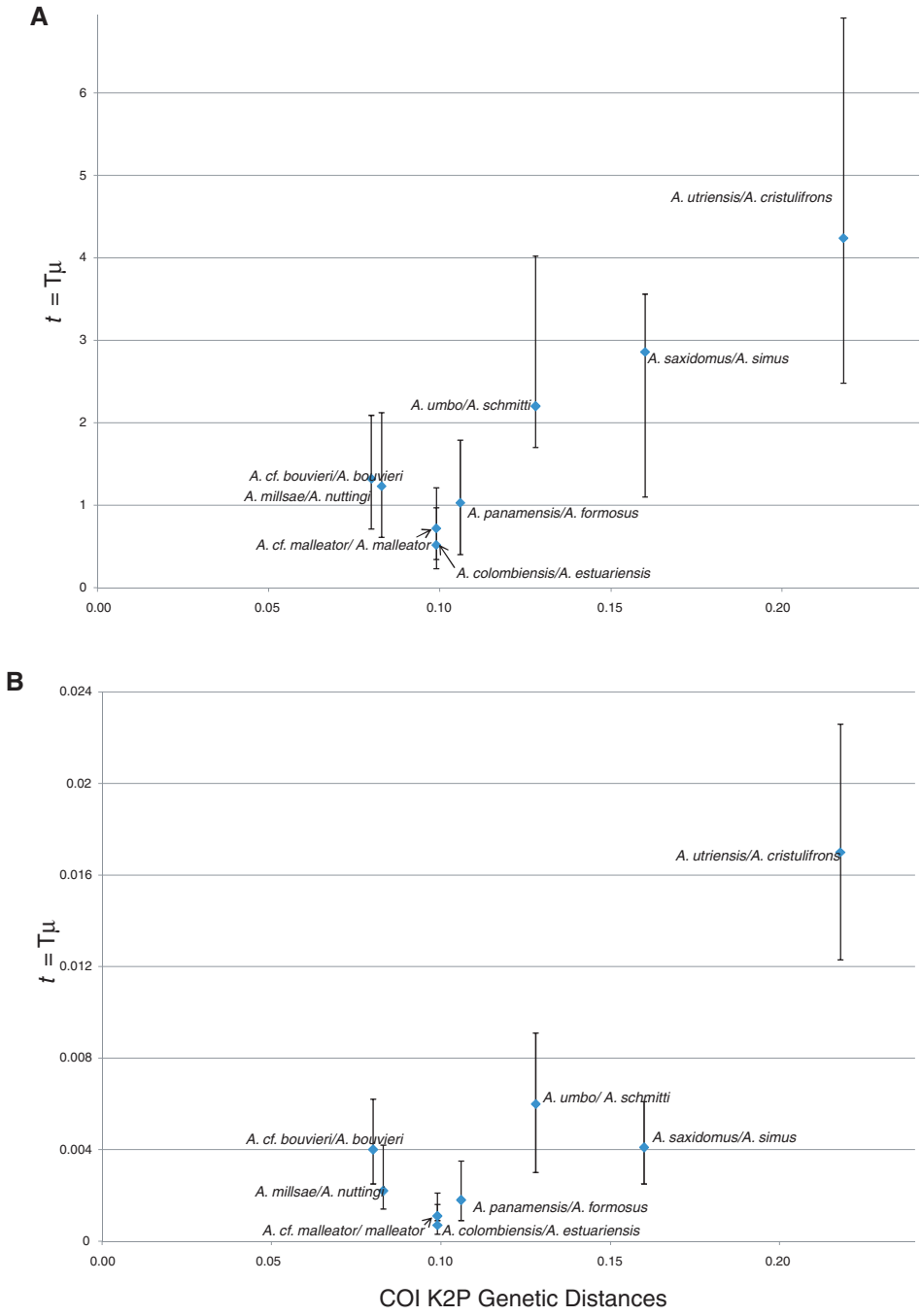


**Figure 3.** **Relationship between COI Kimura-2-Parameter genetic distances and *t* estimated from (A) IM and (B) MCMCcoal.**

speciation in these groups occurred much earlier than 3 mya. Such staggering of speciation events is not entirely unexpected, as the gradual nature of the formation of the Isthmus undoubtedly had stronger affects on some groups than others. Evolutionary changes dating to the late Miocene have been documented in some corals (Collins et al. 1996), Caribbean mollusks (Jackson et al. 1993), and Pacific species of the gastropod Strombina (Jung, P. 1989). By the late Miocene and early Pliocene, the Tropical American Seaway had shallowed significantly, affecting both ocean circulation and the physical environment; specifically, the Caribbean experienced an increase in temperature and salinity and carbonate content relative to waters in the eastern Pacific. Therefore, even though absolute physical barriers to dispersal were not in place until nearly 3 mya, it is likely that divergence of taxa with narrow physiological tolerances and limited larval vagility took place much earlier. It is also possible that earlier speciation events unrelated to the Isthmus formation and subsequent extinction could result in pseudo-sister taxa, and may explain larger

**Table 4.** (A) IM and (B) MCMCcoal derived peak posterior distribution parameter estimates and 95% Highest Posterior Densities (HPD) of divergence times (*T*), and effective population sizes for EP, WA, and ancestral populations ($\theta_1$, $\theta_2$, and $\theta_A$, respectively). All parameters are scaled by the neutral mutation rate where $\mu$ is the geometric mean of the per locus rate of mutation in IM analyses and is calculated as the arithmetic mean of the per site mutation rate in MCMCcoal analyses. Descendent/Ancestor ratios are used to indicate long-term population growth or decline and are calculated as $(\theta_1+\theta_2)/\theta_A$.

(A)

| Species Pair | *T* (95% HPD) | $\theta_1$ (95% HPD) | $\theta_2$ (95% HPD) | $\theta_A$ (95% HPD) | Descendent/ Ancestor |
|---|---|---|---|---|---|
| *A. colombiensis/A. estuariensis* | 0.52 (0.23–0.97) | 1.31 (0.56–2.57) | 1.15 (0.50–2.19) | 3.63 (1.48–7.53) | 0.68 |
| *A. cf. bouvieri/A. bouvieri* | 1.32 (0.71–2.09) | 2.24 (1.35–4.89) | 1.17 (0.70–2.16) | 1.01 (0.11–3.77) | 3.39 |
| *A. cf. malleator/A. malleator* | 0.72 (0.34–1.21) | 2.83 (1.50–4.94) | 1.58 (0.74–2.89) | 2.08 (0.61–4.30) | 2.13 |
| *A. millsae/A. nuttingi* | 1.23 (0.61–2.12) | 1.10 (0.65–1.99) | 2.00 (1.25–3.70) | 2.19 (0.39–6.10) | 1.41 |
| *A. panamensis/A. formosus* | 1.03 (0.40–1.79) | 1.70 (0.79–3.10) | 2.81 (1.16–5.98) | 2.27 (0.24–6.25) | 1.99 |
| *A. umbo/A. schmitti* | 2.20 (1.10–3.56) | 1.48 (0.77–2.52) | 1.01 (0.52–1.72) | 3.32 (0.16–10.08) | 0.75 |
| *A. saxidomus/A. simus* | 2.86 (1.70–4.02) | 2.65 (1.62–4.07) | 1.90 (1.06–3.12) | 2.17 (0.07–7.46) | 2.10 |
| *A. utriensis/A. cristulifrons* | 4.24 (2.48–6.91) | 3.27 (2.11–4.82) | 2.30 (1.44–3.50) | 12.14 (2.64–28.04) | 0.46 |

(B)

| Species pair | T (95% HPD) | $\theta_1$ (95% HPD) | $\theta_2$ (95% HPD) | $\theta_A$ (95% HPD) | Descendent/ Ancestor |
|---|---|---|---|---|---|
| *A. colombiensis/A. estuariensis* | 0.0007 (0.0003–0.0016) | 0.0008 (0.0006–0.0027) | 0.0011 (0.0008–0.0029) | 0.0031 (0.0019–0.0046) | 0.71 |
| *A. cf. bouvieri/A. bouvieri* | 0.004 (0.0025–0.0062) | 0.0028 (0.0018–0.0047) | 0.0016 (0.001–0.0029) | 0.0016 (0.0008–0.0036) | 2.75 |
| *A. cf. malleator/A. malleator* | 0.0011 (0.0009–0.0021) | 0.0024 (0.0015–0.0041) | 0.0013 (0.0008–0.0025) | 0.0015 (0.0009–0.0028) | 2.47 |
| *A. millsae/A. nuttingi* | 0.0022 (0.0014–0.0042) | 0.0021 (0.0014–0.0033) | 0.0032 (0.0022–0.005) | 0.0038 (0.0024–0.0063) | 1.39 |
| *A. panamensis/A. formosus* | 0.0018 (0.0009–0.0035) | 0.0015 (0.0009–0.0029) | 0.0023 (0.0014–0.0045) | 0.0024 (0.0013–0.0046) | 1.58 |
| *A. saxidomus/A. simus* | 0.0041 (0.0025–0.0061) | 0.0029 (0.002–0.0047) | 0.0017 (0.0011–0.0031) | 0.0028 (0.0017–0.0053) | 1.64 |
| *A. umbo/A. schmitti* | 0.006 (0.003–0.0091) | 0.002 (0.0012–0.0036) | 0.0012 (0.0007–0.0023) | 0.0023 (0.0013–0.0048) | 1.39 |
| *A. utriensis/A. cristulifrons* | 0.017 (0.0123–0.0226) | 0.0031 (0.0021–0.0049) | 0.0029 (0.0019–0.0047) | 0.0025 (0.0012–0.0053) | 2.40 |

than expected trans-Isthmian genetic distance values observed in certain putative geminate pairs. This scenario may explain the very large distances obtained for *A. utriensis*/*A. cristulifrons*. Indeed, recent analysis of eastern Atlantic populations suggests that the eastern Pacific *A. utriensis* is somewhat closer to the eastern Atlantic member of the clade (*A. xanthocarpus*) than it is to the western Atlantic *A. cristulifrons* (Anker et al. 2008c).

## Conclusion

This study incorporates recent theoretical and analytical work emphasizing the importance of more holistic genome-wide approaches to addressing questions of recent evolutionary events. Our results support the expectation that the incorporation of sequence data from multiple nuclear loci and coalescent-based analyses can greatly enhance parameter estimates. Only recently have such multilocus approaches to questions relating to comparative phylogeography become feasible. Reduced costs and increased efficiency associated with DNA sequencing, as well as the availability of universal nuclear primers (Regier 2007), have made collection of nuclear sequence data accessible in nonmodel taxa. Analytically demanding complex evolutionary models are now testable using methods that incorporate stochastic simulations and Bayesian principles. Markov-chain Monte Carlo methods have reduced much of the computational expense related to Bayesian approaches and recent developments in approximate Bayesian methods show promise in reducing loads even further, allowing for the analysis of more complex, parameter-rich models (Beaumont et al. 2002; Hickerson et al. 2007). Undoubtedly, the future of phylogeography will continue to experience a shift away from single mitochondrial datasets toward genome-wide approaches and rigorous statistical tests of realistic evolutionary models that can encompass the complexity of life.

### ACKNOWLEDGMENTS

We thank C. Cunningham and B. Ball for help in developing molecular techniques and markers. We are also appreciative to E. Gómez and J. Jara for their assistance with laboratory and field work. Analysis benefited greatly from suggestions by M. Hickerson, B. Rannala, and Z. Yang. We also thank P. Hedrick and two anonymous reviewers for their suggestions.

### LITERATURE CITED

Anker, A., C. Hurt, and N. Knowlton. 2007a. Revision of the *Alpheus nuttingi* (Schmitt) species complex, with description of a new species from the tropical eastern Pacific (Crustacea: Decapoda: Alpheidae). Zootaxa 1577:41–60.

———. 2007b. Three transisthmian snapping shrimps (Crustacea: Decapoda: Alpheidae: *Alpheus*) associated with innkeeper worms (Echiura: Thalassematidae). Zootaxa 1626:1–23.

———. 2008a. Revision of the *Alpheus websteri* Kingsley, 1880 species complex (Crustacea: Decapoda: Alpheidae), with revalidation of *A. arenensis* (Chace, 1937). Zootaxa 1694:51–68.

———. 2008b. Revision of the *Alpheus formosus* Gibbes 1850 species complex, with redescription of *A. formosus* and description of a new species from the tropical western Atlantic (Crustacea: Decapoda: Alpheidae). Zootaxa 1707:1–22.

——— 2008c. Revision of the *Alpheus cristulifrons* Rathbun species complex, with description of a new species from the tropical eastern Atlantic (Crustacea: Decapoda: Alpheidae). J Mar Biol Ass UK 88:543–562..

Avise, J. 2004. Molecular markers, natural history, and evolution. 2nd ed. Sinauer Associates Inc,. Sunderland, MA

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Bermingham, E., and H. A. Lessios. 1993. Rate variation of protein and mtDNA evolution as revealed by sea urchins separated by the Isthmus of Panama. PNAS 90:2434–2738.

Bermingham, E., S. S. McCafferty, and A. P. Martin. 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. Pp. 113–128, *in* T. D. Kocher and C. A. Stepien, eds. Molecular systematics of fishes. Academic Press. San Diego, CA.

Coates, A. G., and J. A. Obando. 1996. The geological evolution of the Central American Isthmus. Pp. 21–56, *in* J. B. C. Jackson, A. F. Budd, and A. G. Coates, eds. Evolution and environment in tropical America. Univ. of Chicago Press, Chicago, IL.

Coates, A. G., J. G. C. Jackson, L. S. Collins, T. M. Cronin, H. Dowsett, L. M. Bybell, P. Jung, and J. A. Obando. 1992. Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. Geol. Soc. Am. Bull. 104:814–828.

Collins T. M. 1996. Molecular comparisons of transisthmian species pairs: rates and patterns of evolution. Pp. 303–334, *in* J. B. C. Jackson, A. F. Budd and A. G. Coates, eds. Evolution and environment in tropical America. Univ. of Chicago Press, Chicago, IL.

Collins, L. S., A. F. Budd, and A. G. Coates. 1996. Earliest evolution associated with closure of the Tropical American Seaway. Proc. Natl. Acad. Sci. USA 93:6069–6072.

Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and variance in coalescent time in phylogeographic studies. Evolution 54:1839–1854.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3:294–299.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Hey, J. and R. Nielsen. 2004. MuLTilocus method for estimating population sizes, migration rates, and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis.* Genetics 167:747–760.

———. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA 104:2785–2790.

Hickerson, M. J., M. A. Gilchrist, and N. Takebayashi 2003. Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. Evolution 57:2216–2225.

Hickerson, M. J., E. A. Stahl, and H. A. Lessios 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60:2435–2453.

Hickerson, M. J., E. A. Stahl, and N. Takebayashi. 2007. MsBayes: pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. BMC Bioinform. 2007:268–274.

Hudson, R. R., and N. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164.

Hudson, R. R., M. Kreitman, and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Jackson, J. B. C., P. Jung, A. G. Coates, and L. S. Collins 1993. Diversity and extinction of tropical American mollusks and emergence of the Isthmus of Panama. Science 260:1624–1626.

Jordan, D. S. 1908. The law of the geminate species. Am. Nat. 42:73–80.

Jung, P. 1989. Revision of the Strombina-group (Gastropoda: Columbellidae), fossil and living: distribution, biostratigraphy, and systematics. Schweiz. Paleontol. Abh. 111:1–298.

Keigwin, L. D. 1982. Isotopic paleoceanography of the Caribbean and east Pacific: role of Panama uplift in Late Neogene time. Science 217:350–353.

Kimura, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61:893–903.

Knowlton, N., and L. A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. Proc. R. Soc. Lond. B 265:2257–2263.

Knowlton, N., L. A. Weigt, L. A. Solorzano, D. K. Mills, and E. Bermingham. 1993. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. Science 260:1629–1632.

Kim, W., and L. G. Abele. 1988. The snapping shrimp genus *Alpheus* from the eastern Pacific (Decapoda: Caridea: Alpheidae). Smithson. Contrib. Zool. 454:1–119.

Kumar, S., and S. Subramanian. 2002. Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. USA 99:803–808.

Kumar, S., T. Koichiro, and M. Nei. 2005. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Bioinformatics 10:198–191.

Lessios, H. A. 1979. Use of Panamanian sea urchins to test the molecular clock. Nature 280:599–601.

Moriyama, E. N., and T. Gojobori. 1992. Rates of synonymous substitution and base composition of nuclear genes in Drosophila. Genetics 130:855–864.

Pamilo P. and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.

Regier, J. C. 2007. Protocols, Concepts, and Reagents for preparing DNA sequencing templates. Version 8/30/05. www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf.

Regier, J. C., and R. Schultz. 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of Crustaceans and a new hypothesis for the origin of Hexapods. Mol. Biol. Evol. 14:902–913.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 10:2496–2497.

Saitou, N., and M. Nei. 1986. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. J. Mol. Evol. 24:189–204.

Stephens, M., N. Smith, and P. Donnelly 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Vawter, A. T., R. H. Rosenblatt, and G. C. Gorman. 1980. Genetic divergence among fishes of eastern Pacific and Caribbean: Support for the molecular clock. Evolution 34:705–711.

Wakeley, J., and J. Hey. 1997. Estimating ancestral population parameters. Genetics 145:847–855.

Williams, S. T., and N. Knowlton. 2001. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. Mol. Biol. Evol. 18:1484–1493.

Williams, S. T., N. Knowlton, L. A. Weigt, and J. A. Jara. 2001. Evidence of three major clades within the snapping shrimp genus *Alpheus* inferred from nuclear and mitochondrial gene sequence data. Mol. Phylogenet. Evol. 20:375–389.

Yang, Z. 2002. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. Genetics 162:1811–1823.

Associate Editor: D. Posada