2012

Expansion Dating: Calibrating Molecular Clocks in Marine Species from Expansions onto the Sunda Shelf following the Last Glacial Maximum

Eric D. Crandall
California State University, Monterey Bay, ecrandall@csumb.edu

Elizabeth J. Sbrocco

Timery S. DeBoer

Paul H. Barber

Kent E. Carpenter

Recommended Citation

Crandall, E.D., Sbrocco, E.J., DeBoer, T.S., Barber, P.H., and K.C. Carpenter. 2012. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum. Molecular Biology & Evolution 29(2): 707-719.

Date of Submission: 8/14/2011

Research Article

Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum.

Eric D. Crandall[1*], Elizabeth J. Sbrocco[2], Timery S. DeBoer[2], Paul H. Barber[3], Kent E. Carpenter[1]

1. Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, U.S.A.
2. Biology Department, Boston University, Boston, MA 02215, U.S.A.
3. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095, U.S.A.

*Corresponding author
Current Address: Southwest Fisheries Science Center, Santa Cruz, CA 95060, U.S.A.
Phone - +01-831-420-3959
Fax - +01-831-420-3980
email: eric.crandall@noaa.gov

Running Title: Expansion dating in Sunda Shelf marine species

Key words: Time-dependency, population expansion, mutation rate, molecular clock, marine invertebrates

Non-standard abbreviations: aDNA – Ancient DNA, BSP – Bayesian skyline plot, ESS – Effective sample size, LGM – Last Glacial Maximum

Abstract

The rate of change in DNA is an important parameter for understanding molecular evolution, and hence for inferences drawn from studies of phylogeography and phylogenetics. Most rate calibrations for mitochondrial coding regions in marine species have been made from divergence dating for fossils and vicariant events older than 1-2 million years, and are typically 0.5% - 2% per lineage per million years. Recently, calibrations made with ancient DNA from younger dates have yielded faster rates, suggesting that estimates of the molecular rate of change depend on the time of calibration, decaying from the instantaneous mutation rate to the phylogenetic substitution rate. Ancient DNA methods for recent calibrations are not available for most marine taxa so instead we use radiometric dates for sea-level rise onto the Sunda Shelf following the Last Glacial Maximum (starting ~18,000 years ago), which led to massive population expansions for marine species. Instead of divergence dating, we use a two-epoch coalescent model of logistic population growth preceded by a constant population size to infer a time in mutational units for the beginning of these expansion events. This model compares favorably to simpler coalescent models of constant population size, and exponential or logistic growth, and is far more precise than estimates from the mismatch distribution. Mean rates estimated with this method for mitochondrial coding genes in three invertebrate species are elevated in comparison to older calibration points (2.3% - 6.6% per lineage per million years), lending additional support to the hypothesis of calibration time-dependency for molecular rates.

Introduction

The observation that genetic distances between taxa are correlated with the amount of time since they diverged first gave rise to the idea that DNA may evolve at a relatively constant rate: the molecular clock (Zuckerkandl and Pauling 1965). Like radiometric dating, molecular clocks have allowed illumination of life's evolutionary history, but not without a good deal of scientific controversy (see reviews in Arbogast et al. 2002; Smith and Peterson 2002; Bromham and Penny 2003; Takahata 2007). The original observation of rate constancy led directly to the neutral theory of molecular evolution, which initially predicted that the rate of evolutionary change, k, would be equal to the neutral mutation rate μ, regardless of the effective population size (Kimura 1968). It wasn't long, however, before heterogeneity in amino acid substitution rates was observed (Kimura and Ohta 1971; Langley and Fitch 1974), and the neutral theory was soon amended to account for the effects of effective population size (N_e) and slightly deleterious selection ("nearly neutral theory"; Ohta 1972; Tachida 1991). Under the nearly neutral theory, the substitution rate scales inversely with N_e, so it is possible for a large number of alleles with neutral and slightly deleterious mutations to remain transient in a large population for relatively long periods of time before slowly being fixed or removed by genetic drift or purifying selection.

The event that is typically used for a molecular clock calibration is a divergence of known age between two genetically distinct lineages. However, as noted above, at the time of actual population divergence, the gene of interest was likely represented in the ancestral population by a number of divergent alleles. Thus, divergence in the gene of interest will predate the actual population divergence by an average of $2N_e$ generations

(Nei and Li 1979; Edwards and Beerli 2000). Because gene divergence is more easily measured than population divergence, the two measures are often conflated, leading to overestimates of population divergence times (in mutational units), and therefore of molecular substitution rates (Arbogast et al. 2002). This effect is magnified by larger effective population sizes, and more recent population divergence times (Peterson and Masel 2009). It is only recently that divergence-dating methods have been able to explicitly consider the effects of ancestral polymorphism under a coalescent model (Edwards and Beerli 2000; Nielsen and Wakeley 2001; Hickerson et al. 2003).

With increased availability of radiometrically-dated ancient DNA (aDNA) from sub-fossil tissues, it has become possible to calibrate molecular clocks using radiometric dates assigned to lineages that terminate at some known time in the past (Drummond et al. 2002; Ho et al. 2007a). In a coalescent framework, this method circumvents the problem of ancestral polymorphism by essentially sampling directly from the ancestral population (but see Navascues and Emerson 2009 for potential problems with this method). In 2005, Ho and co-workers used coalescent rate calibrations from contemporary and ancient DNA to show that there is an exponential relationship between the date used for calibration and the resultant rate, with high rates for recent calibration dates that decline to more familiar substitution rates for dates over 1-2 million years ago (the “lazy J”, Penny 2005), a result that supported previous observations from pedigree studies in humans (Parsons et al. 1997; Howell et al. 2003). In addition to avoiding, through coalescent models, the problem of ancestral polymorphism, their analysis suggests that this inverse relationship cannot be entirely explained by errors in calibration or sequencing, nor by saturation at fast-evolving sites (but see debate in Emerson 2007; Ho

et al. 2007b). Instead they suggest that this time-dependent relationship between apparent molecular rate and calibration time results from the prolonged action of purifying selection on slightly deleterious mutations.

This hypothesis of time-dependency for molecular rates may have far-reaching consequences for studies of genetic variation within populations and species, as substitution rates from relatively old calibration points are frequently used to estimate divergence time, as well as a number of population genetic parameters. For example, if time-dependency occurs, then the application of rates from old (> 2 mya) calibrations to analyses of recent timescales will result in the consistent overestimates of divergence times and effective population sizes, and consistent underestimates of the proportion of migrants in structured populations (Ho et al. 2008). However, further inquiry into this hypothesis of time-dependency is hindered by a lack of recent calibration points (Ho et al. 2007b), as aDNA samples are rare and pedigree studies can be prohibitively lengthy. In particular, we are not aware of any recent (< 1 mya) calibration points for marine fishes and invertebrates (but see Burridge et al. 2008 for an example from freshwater fish). Molecular clock calibration in these groups typically relies on vicariance due to the rise of the Isthmus of Panama, which led to the closure of the Panamanian Seaway 2.8 – 3.1 mya (Lessios 2008; Hellberg 2009), although calibrations from fossils and other geological events exist as well (e.g. Marko 2002; Wares 2002). Here, we develop a new coalescent-based method for the calibration of molecular clock rates that uses a population expansion rather than divergence as its calibrating event. We demonstrate the method for a recent, well-documented marine population expansion event, and use it to test whether time-dependency can be detected in marine invertebrates.

Extending from the Malay Peninsula to the eastern sides of Java and Borneo, the Sunda Shelf covers 1.8×10^6 km^2, making it the largest shelf area outside the polar regions (Hanebuth et al. 2000). With the onset of Pleistocene glaciation cycles about 3 mya, global sea levels have fluctuated with maximum amplitudes of up to 140 meters (Lambeck et al. 2002). Over the past ~120 ky sea level remained between 30 and 100 meters below present-day sea level, but dropped more than 120 meters during the Last Glacial Maximum (LGM; Chappell et al. 1996), leaving the Sunda Shelf completely exposed. Since the LGM, ^{14}C dating of corals and littoral debris has confirmed a rapid rise to current sea levels, with the fastest rate of sea level change occurring between 15 and 10 kya (Hanebuth et al. 2000). This sea level rise produced an unparalleled expansion in range (Figure 1; Voris 2000; Sathiamurthy and Voris 2006) for the numerous marine species that now inhabit the Sunda Shelf, and a genetic signature of the expansion can be detected in nearly every species sampled in the region (e.g. Chenoweth and Hughes 2003; Lind et al. 2007; Crandall et al. 2008a; Crandall et al. 2008b).

In the present study we analyze mitochondrial datasets from three invertebrate species sampled from the Sunda Shelf for the signature of range expansion using the traditional mismatch distribution (Rogers and Harpending 1992), as well as a novel two-epoch approach. For this new method, we first create a Bayesian Skyline Plot (Drummond et al. 2005), which portrays changes in N_e across multiple coalescent intervals. We use these results to inform Bayesian priors for a simpler two-epoch model of the coalescent (Shapiro et al. 2004). This model provides an estimate of the genealogical depth (in mutational units) at which the expansion occurred to produce a per-lineage rate of change (one-half of the commonly estimated divergence rate), together with associated error. By

explicitly and simultaneously incorporating into our model the demographic signal that is encoded in the genetic data we overcome any bias associated with ancestral polymorphism or fluctuating population sizes (Navascues and Emerson 2009). Our goals are to 1) develop a method for finding recent molecular calibrations in taxa that are not represented by aDNA or pedigree calibrations and 2) test the hypothesis of time-dependency of molecular rates (Ho et al. 2005) using this method.

Materials and Methods

Data Characterization

To ensure that any population expansions detected could be firmly attributed to sea-level rise on the Sunda Shelf following the last glacial maximum, we selected three species for which previous analyses revealed major clades that were largely confined to the Sunda Shelf or to central Indonesia. These species comprised the boring giant clam, *Tridacna crocea* (Mollusca: Bivalvia), whose 'black clade' is dominant from South Sumatra to Western Papua (DeBoer et al. 2008), the mantis shrimp *Haptosquilla pulchella* (Arthropoda: Malocostraca), whose 'white clade' is limited to the Sunda Shelf and the lesser Sunda islands (Barber et al. 2002b), and the chocolate-chip seastar *Protoreaster nodosus* (Echinodermata: Asteroidea), which has a relatively limited range stretching from Sri Lanka to New Caledonia that is dominated by the Sunda and Sahul shelves (Crandall et al. 2008b). All three species are found in shallow lagoonal waters no deeper than 10 meters. We sub-sampled mitochondrial Cytochrome-c Oxidase subunit 1 (CO1) sequence data from larger published datasets to only include localities on or very near the Sunda Shelf (Table 1, Figure 1H).

We used DNAsp 5.10 (Librado and Rozas 2009) to characterize these datasets for standard population genetic statistics, including D* and F_S, which identify departures from neutrality due to an excess of recent mutations (Fu and Li 1993; Fu 1997), with significance for these neutrality tests determined with 10,000 coalescent simulations. For datasets where more than one locality was sampled, we also measured Φ_{ST} among localities using the AMOVA algorithm implemented in Arlequin 3.11 (Excoffier et al. 2005), and only took calibrations from datasets for which $\Phi_{ST} = 0$. We ascertained the best-fitting model of molecular evolution using the Akaike information criterion (AIC) in ModelTest 3.7 (Posada and Crandall 1998) and PAUP* 4.0(Swofford 2002). Finally, we constructed statistical parsimony networks in TCS 1.21 (Clement et al. 2000), using a 95% connection limit.

Calibration Points and Demographic Analysis

Calibration of a molecular clock based on population expansion requires geological dates for when the expansion began. We used two dates taken from studies of radiocarbon dates for sediment cores and littoral organics (Geyh et al. 1979; Hesp et al. 1998; Hanebuth et al. 2000; summarized in Sathiamurthy and Voris 2006; Hanebuth et al. 2009). The first date, 19.6 kya, reflects the earliest time following the LGM that sea level rise first became statistically distinguishable from the 2m tidal range, during a rise of ~10m over 800 years (Hanebuth et al. 2009). However, sea level rise was gradual at first (0.41m/100 years after this pulse) and did not result in the significant flooding of the Sunda Shelf (Figure 1). We therefore used another calibration point, at 14.58 kya (corresponding to the Bølling interstadial period), during which sea level rose at an average of 5.33m/100 years and flooding of the Sunda Shelf began in earnest. Unless

stated otherwise, we report all rates in this paper as lineage mutation rates (= ½ divergence rate) in units of percent change per million years (%/my).

In addition to the information from the sources noted above, we estimated the area of shallow water habitat (0-10m) that became available to these species with each 10m increase in sea level using spatially gridded data from the ETOPO1 1 arc-minute Global Relief Model (Amante and Eakins 2009). Using the spatial analyst toolkit in ArcMap 10.0, we quantified the number of cells between 10-meter isobaths and multiplied this value by a cell size 3.16 km^2, which resulted from projecting the data to the ARC coordinate system, zone 1. These habitat area estimates are approximately correct at the scale of the ETOPO1 grid, but likely represent a slight underestimate.

The calibration also requires accurate estimates of the mutational depth in the genealogy at which the transition to expansion growth began, together with an assessment of the associated error. Here, we use two different methods to make this estimate: the mismatch distribution (Rogers and Harpending 1992) and a two-epoch coalescent model (Shapiro et al. 2004).

The mismatch distribution is the distribution of pairwise differences among haplotypes, which Rogers and Harpending (1992) observed to be unimodal after exponential population growth in a single deme. They described this with an analytical model for which the modal value, $\tau = 2\mu t$ estimates the time of population expansion t in terms of the mutation rate μ. Schneider and Excoffier (1999) amended this "sudden expansion" method for a finite-sites model with rate heterogeneity, and fitted each parameter using a least-squares approach. We analyzed our data under this model in

Arlequin 3.11 (Excoffier et al. 2005), and established confidence intervals for τ with 10,000 parametric bootstraps of the analytical model. Model fit was evaluated using the sum of squared deviations (SSD), with significance established with the same bootstrapped dataset. We calculated lineage mutation rates as $\mu = \tau / (2c)$, where c is one of the two calibration points mentioned above, and τ is divided by the number of sites in the dataset. We give results as % change per million years.

To get an overall image of the information on demographic history that was available in each genetic dataset, we analyzed them each under a Bayesian skyline model (Strimmer and Pybus 2001; Drummond et al. 2005) implemented in BEAST 1.5.3 (Drummond and Rambaut 2007). Under this model a large sample of possible coalescent genealogies are broken into piecewise intervals and effective population size (N_e) is estimated at each interval from the number of observed coalescent events. The resultant Bayesian skyline plot (BSP) makes few *a priori* assumptions about the historical demographic trajectory of the population, and can thus provide a framework for constructing more specific models. For these analyses we used ten piecewise linear intervals, a strict clock model, and the molecular evolution model from ModelTest. We used uniform, relatively uninformative priors for the population size at each interval, and a gamma prior for κ, the transition:transversion ratio. Each skyline analysis was run, at minimum, three times for thirty million steps. We assessed convergence with estimates of effective sample size (ESS) in Tracer and by comparing the marginal posterior distributions for each parameter among runs. Our criterion was ESS > 200 as indicated in the BEAST manual. Finally, we combined the logged parameter values and trees from

replicate runs using LogCombiner 1.4.9, and used Tracer to create a Bayesian skyline plot for each dataset.

The BSP for each dataset indicated a period of constant population size, followed by rapid growth that slowed as it reached the present. We therefore used a two-epoch coalescent model (Shapiro et al. 2004), implemented in BEAST, that simulated either two-parameter exponential growth (Θ_1, and intrinsic growth rate, r), or three-parameter logistic growth (Θ_1, r, and time to reach ½ Θ_1, t_{50}), preceded by one-parameter constant growth (Θ_0), with a final parameter for the transition time between the two epochs ($t_{transition}$). We used the same molecular evolution models that we had used for the skyline plots, and 1/x priors on all parameters (Drummond et al. 2002) except for κ, for which we used a gamma-distributed prior, and r, for which we used a simple uniform prior. We set the upper and lower limits of the prior distribution for each parameter using the 95% confidence intervals (CI) from the BSPs as guidelines. For example, upper limits for $t_{transition}$ (the parameter of interest) were the upper limit of the 95% CI for T_{MRCA} from the skyline model. The lower limits for $t_{transition}$ and t_{50} were set to 10^{-6} mutations per site, which represents a prior assumption that rates will not be lower than 0.05% /my (about twenty times slower than the 1% /my rate commonly used for mtDNA; Brown et al. 1979). We calculated lineage mutation rates as $\mu = t_{transition}/ c$, and give results in units of % change per million years.

To more rigorously test the hypothesis of constant population size followed by logistic growth, we ran each dataset under a model of constant population size, two models of population growth (exponential and logistic) without a stable period preceding

them, and a model of expansion growth (constant size followed by exponential growth) that does not allow the time of expansion to vary. Where applicable, we used the same priors as were used in two-epoch models. We then used Bayes factors to compare the harmonic mean of the marginal likelihood from these models to those from the exponential and logistic two-epoch models described above. Marginal likelihoods for each model were calculated as the sum of the log-likelihoods for the genealogy and the coalescent model at each recorded step (thus, the product of these two quantities), and the harmonic mean for each was calculated in Tracer, using 1000 bootstrap replicates to establish confidence intervals. We report calibrations from the model that received the most support from this method following recommendations from Kass and Raftery (1995).

Results

We found a more than ten-fold increase in the amount of shallow-water habitat (0-10m) that became available to these species as a result of sea-level rise following the Last Glacial Maximum (LGM; Figure 1). Shallow-water habitat on the shelf went from 30,050 km^2 at the lowstand to 358,680 km^2 in present sea level.

Summary statistics for each genetic dataset are given in Table 1. Fu's F_S was negative, but not significantly so for the *Tridacna crocea* dataset. *Haptosquilla pulchella* and *Protoreaster nodosus* had significantly negative values for F_S. The *H. pulchella* dataset comprising multiple sampling localities had a significant Φ_{ST} value, reflecting genetic structure between Pulau Seribu and the other sites. Φ_{ST} was zero once samples from Pulau Seribu were removed from the dataset. There was no evidence of any

structure among pooled localities for *T. crocea.* ModelTest selected an HKY model of molecular evolution for all datasets, with the addition of a parameter for invariant sites (I) in *H. pulchella*. Statistical parsimony networks from TCS showed numerous star-like polytomies that are diagnostic of population expansions (Supplemental Figure 1; Slatkin and Hudson 1991)

Mismatch distributions were generally unimodal for *Protoreaster nodosus* CO1 and *Haptosquilla pulchella* CO1, but bimodal for *Tridacna crocea* CO1 (Figure 2). The bimodal distribution for *T. crocea*, which is expected for a constant-sized population (Rogers and Harpending 1992), resulted in very large estimates of τ and exceptionally high values for μ (Table 2). None of the datasets rejected a sudden expansion model.

Bayesian skyline models (Figure 3) for all datasets converged to the same posterior distribution, as demonstrated by effective sample size values >200 and agreement among multiple runs. Model runs for each different demographic scenario (constant size through two-epoch with logistic growth) also converged well for all datasets, with effective sample sizes generally much greater than 200, and agreement across multiple runs. The two-epoch model with logistical growth consistently had the strongest Bayes Factor support, beating out a similar model with exponential growth for all three species, as well as three simpler models of population growth (exponential, logistic, and expansion; Table 3). This model was also decisively better than a constant population size model for *H. pulchella* and *P. nodosus*, but was only weakly supported over a constant population size for *T. crocea*.

Mean values for the time of transition between constant population size and logistic growth were generally a bit lower than the inflection point depicted in the skyline model for each gene (Figure 3). However, the 95% confidence intervals for $t_{transition}$ generally encompassed the period of growth detected by the skyline model, and fit within the very large confidence intervals generated for the mismatch distributions (Table 2). For the conservative calibration point at 19.60 kya, mean estimates of μ for *T. crocea* and *P. nodosus* were very similar (2.30%/my and 2.61%/my respectively), while *H. pulchella* had a mean rate that was over twice as fast (6.58%/my; Table 2). Posterior distributions for $t_{transition}$, Θ_0 and Θ_1 were unimodal, while those for growth rate (r) and t_{50} were mostly uninformative. With the exception of the *H. pulchella* samples from Pulau Seribu, which showed little evidence of population growth, two-epoch models for individual sampling localities showed strong concordance with the pooled-locality datasets in *T. crocea* and *H. pulchella*, albeit with larger confidence intervals (results not shown).

Discussion

Traditional molecular clock calibrations have relied almost exclusively on vicariant events or fossil calibrations, both of which have limited applicability in most marine taxa, particularly for recent timescales. Our use of a well-documented, recent population expansion instead of a divergence allows us to estimate a molecular rate of change at a relatively recent timescale, and, in combination with an explicit coalescent model of constant population size followed by expansion, allows us to account for polymorphisms that predate the expansion (Peterson and Masel 2009). Point estimates from both mismatch analysis and the two-epoch coalescent model resulted in lineage mutation rates that are much higher than the 1%/my (= 2%/my divergence rate) commonly assumed for

many marine phylogeographic studies, although our confidence intervals often include these lower values.

Allowing for higher rates in future phylogeographic inference may often bring it into line with a region's recent palaeoclimate (see Ho et al. 2008). For example, divergence dates across the Maluku Sea between populations of *H. pulchella*, previously estimated to be about 470,000 years ago using a rate of 1.4%/my, (Barber et al. 2006), can now be estimated at about 100,000 years ago. This is much closer to the last time that sea levels approached current levels 120,000 years ago, before dropping again (Chappell et al. 1996). In the following discussion, we first compare estimates from the sudden expansion and two-epoch models. We then discuss potential sources of error in our calibration and compare our calibrated rates to rates from older calibrations.

Rate calibrations from the spatial expansion mismatch analyses were generally much higher than those from the two-epoch coalescent model (Table 2), although very wide confidence intervals on τ always included the two-epoch estimate. Of the two, we favor the two-epoch method, which takes advantage of the genealogical information in the data through Bayesian parameter estimation from an explicit coalescent model. In contrast, the mismatch analyses used here provide only an analytical approximation of substitution patterns expected from a spatial expansion, without considering any underlying genealogy, and so likely yield less precise estimations. As a case in point, the mismatch distributions for *T. crocea* were bimodal due to several star polytomies in the genealogies (supplementary figure 1). A bimodal distribution tends to increase the estimate of τ (relative to unimodal distributions), and therefore of the substitution rate. However, in a

time-reversed coalescent framework, these polytomies can be interpreted as a potentially simultaneous increase in the rate of coalescence.

Sources of Error

Our use of a two-epoch coalescent model (Shapiro et al. 2004) as a way to more accurately calibrate substitution rates is new, and uses an intra-specific process (population expansion) to make inferences for phylogeographic timescales, in contrast to inter-lineage divergence methods that have been used previously. As such, our calibration method is subject to some of the same sources of error and bias as divergence methods, while introducing new ones, and avoiding others (Arbogast et al. 2002). First and foremost, calibration points, as independent estimates of elapsed time, are central to any scheme for separating the effects of substitution rate and time, and are probably the largest sources of error in molecular clocks (Benton and Ayala 2003). Here, we used two possible points for the beginning of population expansion: one at 19.6 kya and the other at 14.6 kya, as a way to provide some assessment of the potential for calibration error. For the sake of discussion, we use rates from the date that is more conservative with respect to the time-dependency hypothesis (19.6 kya), but the 14.6 kya date may be more accurate, as this was the time that sea level rise first resulted in significant flooding of the Sunda Shelf (Figure 1). We are also making the assumption that coral reef communities colonized the Sunda Shelf as soon as new shallow marine habitat was available. This is justifiable, because coral reef communities have shown the potential for rapid, long-distance re-colonization of coastal waters, with species richness, coral cover and genetic diversities returning to their ambient levels less than 150 years after being obliterated by volcanic eruptions (Tomascik et al. 1996; Barber et al. 2002a; Starger et al. 2010).

Nevertheless, even the 14.6 kya date may be conservative, as the Sunda Shelf environment may not have been immediately appropriate for coral reef development following inundation of terrestrial habitats. Thus, as we learn more about re-colonization of the Sunda Shelf, it may become apparent that the above rates are actually conservative estimates and we may need to revise the estimated rates upward.

A related assumption is that populations of all three of these lagoonal species expanded onto the Sunda Shelf somewhat simultaneously. This is analogous to the assumption of simultaneity that has often been made in divergence dating across multiple taxa (but see Hickerson et al. 2006 et al. for a statistical test of simultaneous divergence), and rests on the idea that coral reef and their associated lagoonal communities would move concurrently onto the Sunda Shelf (Pandolfi and Jackson 2001; Tager et al. 2010). While no tests currently exist for simultaneity in expansion, the BSPs provide a sketch of demographic history for each species back to the most recent common mitochondrial ancestor, and have the potential to detect multiple demographic fluctuations (e.g. figure 4 in Crandall et al. 2008a), which are not evident for the present datasets. However, although the fact of recent marine population expansions onto the Sunda Shelf is not in doubt, the wide confidence intervals and weak support for a two-epoch model of logistical growth in *T. crocea* suggests that this species may have expanded later or less rapidly than the other two species. This could be the result of changes in other less important habitat factors such as sea surface temperature or primary production. More rigorous tests of this assumption of simultaneity await data from multiple loci (Heled and Drummond 2008).

By setting priors of the two-epoch model to confidence intervals for the Bayesian skyline plots, we essentially estimated expansion times that fit the BSPs (Figure 3). This approach offers a number of advantages. First, the resulting "fitted" two-epoch models were significantly better at explaining the data than simpler models of population growth (Table 3). Second, in using these models we have already accounted for the effects of population size change that are often neglected in divergence calibrations (Arbogast et al. 2002; Navascues and Emerson 2009). Third, by calibrating from an intra-specific process, we avoid the potential errors arising from phylogenetic divergence methodologies, such as problems with rooting, branching order and missing or extinct taxa (Smith and Peterson 2002). Finally, the 95% confidence intervals given in Table 2 should largely account for error due to Poisson-distributed mutations, rate variation across sites and the stochasticity of the coalescent.

However, it is important to note that our estimates will also be affected by any violations of the assumptions of the Bayesian skyline method (Drummond et al. 2005). One assumption is that the sampled population is unstructured, which is met for our datasets by zero or negative Φ_{ST} values in Table 1. Another assumption is that observed increases in the rate of coalescence are due to demographic changes, rather than positive or purifying selection, each of which can leave very similar population genetic patterns. We will address each of these alternate hypotheses in turn.

Because genes on the mitochondrial genome are strongly linked, an advantageous variant in one gene could potentially sweep to fixation through positive selection, bringing all variation on its particular genome with it (Maynard-Smith and Haigh 1974). As such, this genetic hitchhiking effect is nearly indistinguishable from a demographic

population expansion, since both result from growth in effective population size of a particular mitochondrial haplotype (Fu 1997; Bazin et al. 2006). However, if hitchhiking were occurring frequently in the mtDNA of these species, we might expect nucleotide diversities (Table 1) to be more similar to each other than they are (Bazin et al. 2006). Furthermore, we would not expect rapid increases in mtDNA effective size to occur at roughly the same genealogical depth in unlinked genetic regions, or across multiple species from the same geographic region. Yet we have observed population genetic patterns consistent with rapid population growth in multiple species, and across unlinked loci (E. Sbrocco, unpublished data from *A. ocellaris* mtDNA and scnDNA, Chenoweth and Hughes 2003; Lind et al. 2007; Crandall et al. 2008a; Crandall et al. 2008b). Population growth resulting from an expansion in demographic population size is therefore the most parsimonious explanation for the observed patterns.

The process of purifying selection is conceptually more difficult to distinguish from demographic growth, since it is implicated in creating the time-dependency effect, by acting over longer periods of time than previously expected (Ho et al. 2005; Ho et al. 2007b). It is important to note that purifying selection is by no means limited to non-synonymous mutations. Through a comparison with pseudogenes, Ophir et al. (1999) found that an average of 75% of substitutions are non-neutral, roughly twice what is estimated by D_n/D_s ratios. Many synonymous changes are under very weak selective pressures, as arise from translational selection, or maintenance of nucleotide compositions in the face of biased mutational input (Montooth and Rand 2008). Thus, if purifying selection is occurring at the same timescales at which we are measuring population growth, it may be difficult to tease the two apart, because it can be difficult to

ascertain exactly how purifying selection would affect genetic variation in a reversed-time coalescent framework. Using a forward-time model, Peterson and Masel (2009) confirm that purifying selection on slightly deleterious variation leads to a period of elevation in the substitution rate at recent timescales, the length of which is increased by large N_e and a highly leptokurtic distribution of selection coefficients, as would be expected under nearly-neutral theory (Keightley and Eyre-Walker 2007).

Fortunately, purifying selection is expected to leave only a minor imprint on the shape of the genealogy, while having a similar effect as demographic growth or positive selection on the distribution of substitutions (shifting them towards the tips of the genealogy; Williamson and Orive 2002). If purifying selection were the only process occurring, we would expect to see it act more or less constantly throughout the history of the mitochondrial genome, following a simple model of exponential or logistic growth (e.g. Seger et al. 2010). Instead, as was initially indicated by the BSPs, two-epoch models of constant population size followed by a pulsed increase in the rate of coalescence are consistently favored by odds of more than 300:1 over these simpler growth models (Table 3).

The different topologies expected under purifying selection and demographic growth (Williamson and Orive 2002) may also be why Fu (1997) found that D* (which is based on distribution of substitutions) was more sensitive to purifying selection than F_S (which is based on haplotype frequencies), and vice versa. That Fu's F_S was significant in many more instances than D* (Table 1) also suggests that demographic change is more important than purifying selection in creating the observed patterns. Nevertheless, it will be important to account for sites under purifying selection when estimating rates from

population expansions, perhaps by including rate decay parameters in coalescent genealogy sampling schemes (O'Fallon 2010).

Rate Comparisons

Perhaps the most surprising result of calibrations from the two-epoch model is the disparity in substitution rate estimates. While *T. crocea* and *P. nodosus* have similar rates (mean values of 2.3% to 2.6% per million years), the substitution rates in *H. pulchella* appear twice as fast (6.6% per million years). It is possible that this disparity results from non-simultaneous dates of expansion, with *H. pulchella* experiencing an expansion event much earlier than the Sunda Shelf flooding, or *T. crocea* and *P. nodosus* expanding onto the Sunda Shelf about 10,000 years after *H. pulchella*. A simpler explanation for the rate differences among the three invertebrates can be found in their times to first reproduction, which we use as a conservative proxy for generation time. *T. crocea* and *P. nodosus* each take at least two years to reach reproductive maturity (Lucas 1988; Bos et al. 2008), whereas *H. pulchella* can reach reproductive maturity within one year (M.V. Erdmann, pers. comm.; Erdmann 1997). Thus, since *H. pulchella* replicates its germline more than twice as often as the two larger invertebrate species, the per-year rate disparities among the invertebrates can be easily reconciled in a per-generation context. A generation-time effect that is independent of body size and other correlates has recently been confirmed for invertebrates in general (Thomas et al. 2010).

Following correction for a generation-time effect, and removal of the calibration for *T. crocea* due to weak support for the two-epoch model, our mean estimates of CO1 lineage substitution rates ranged from 5.2% - 6.6% /million generations for two marine

invertebrate species. Per-generation lineage rates calibrated from the Isthmus of Panama that properly account for ancestral polymorphism are lower than this (1.4 – 3.2%, Hickerson et al. 2003; Hickerson et al. 2006), while fossil calibrated rates for marine invertebrates are still lower (0.5 – 1.2%, Marko 2002; Frey and Vermeij 2008; Malaquias and Reid 2009). Using least-squares regression, this decline in mean rates with calibration time (Figure 4) can be described as an exponential decay of rates over time (Ho et al. 2005):

$$\text{Rate}_{\text{Marine Invertebrate CO1}}(t) = 0.053e^{-0.40t} + 0.0065 \quad (1)$$

Under this relationship, marine invertebrate CO1 has an instantaneous mutation rate of 5.3% per million generations after lethal mutations have been removed, and declines to long-term "phylogenetic" rates of 0.65% per million generations with a much slower rate (given in the exponential term) than has been found in birds, mammals, or freshwater fish (Ho et al. 2005; Burridge et al. 2008).

The wide confidence intervals on a limited number of recent rate estimates suggest caution in interpreting this curve. Nevertheless, the slower rate of decay that we calculated makes sense in a nearly-neutral context, because with large census sizes and genetic neighborhoods, marine invertebrates can be expected to have larger long-term effective population sizes than other taxa (Palumbi 1994). The next slowest rate of decay can be found in birds, followed in rank order by mammals and freshwater fish, which agrees intuitively with what we know about effective population sizes in these taxa.

Conclusions

Although wide error bars on our rather conservative rate estimates preclude definitive statements, these results appear to show higher rates of molecular change for a recent calibration point, and thus they provide additional support for the hypothesis of time-dependency of molecular rates (Ho et al. 2005). This might provide insight into some persistent problems in marine phylogeography. If purifying selection is removing variation over longer timescales than previously assumed in the marine realm, it would help to explain the widespread pattern of shallow mitochondrial genealogies combined with deep genetic divergences between sister species (Grant and Bowen 1998), as well as the frequent departures from neutrality (Wares 2010) that are commonly observed in marine species.

Continued discovery of rate heterogeneity has nearly led to the abandonment of a global clock for any given region of the genome (Bromham and Penny 2003) in favor of local clocks that are particular to certain taxa (Yoder and Yang 2000) or relaxed clock models that allow the rate of evolution to vary across the phylogeny (Aris-Brosou and Yang 2002; Drummond et al. 2006). A wide variety of explanations for the observed rate variation have been offered, including the correlated effects of generation time (Laird et al. 1969), metabolic rate (Martin and Palumbi 1993; Gillooly et al. 2005), and DNA repair mechanisms (Li et al. 1996). Our results, together with other results showing time-dependent rates (Ho et al. 2005; Ho et al. 2007a; Saarma et al. 2007; Burridge et al. 2008; Subramanian et al. 2009) renew support for one of the original explanations for rate heterogeneity: the combined effect of weak negative selection and effective population size ($N_e\sigma_s$) proposed under the nearly neutral theory of molecular evolution (Ohta 1972). Specifically, these combined results suggest that the observed rate of molecular evolution

decays as a function of the calibration time, and that the rate of decay is a function of effective population size. With additional data from more species and more genes, we expect that our method of expansion dating will provide further insight into rates of molecular change at recent timescales.

Supplementary Material

Statistical parsimony networks for each species are available in Supplementary Figure 1.

Acknowledgements

Impetus for this project came from discussions in a population genetics graduate seminar, led by M. Sorenson, at Boston University in 2007. It was developed while EDC was a postdoctoral associate with Old Dominion University, and a visiting scholar at the Marine Sciences Institute of the University of the Philippines. This work is part of the Coral Triangle Partnerships for International Research and Education (CT-PIRE), funded by the National Science Foundation – OISE-0730256. Collection of DNA sequence data used in this study was supported by NSF grant OCE-0349177. We would like to acknowledge our Indonesian partners: Udayana University for hosting our project, as well as the Indonesian Institute of Sciences (LIPI) and the Indonesian Ministry of Science and Technology (RISTEK) for providing the necessary research permits. We thank N. Mahardika, H. Toha and Ambariyanto for their sponsorship and logistical support. Some BEAST analyses were carried out by using the online resources of the Computational Biology Service Unit of Cornell University. We thank C. Burridge, A. Drummond, M. Hellberg, and three anonymous reviewers for constructive comments that improved the manuscript. EDC thanks E. Anderson for statistical advice, and S. Gulamhussein for her

patience and support.

Literature Cited

Amante, C. and B. W. Eakins. 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp.

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu Rev Ecol Syst. **33**:707-740.

Aris-Brosou, S., and Z. H. Yang. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA Phylogeny. Syst Biol. **51**:703-714.

Barber, P. H., M. V. Erdmann, and S. R. Palumbi. 2006. Comparative phylogeography of three co-distributed stomatopods: origins and timing of regional lineage diversification in the coral triangle. Evolution **60**:1825-1839.

Barber, P. H., M. K. Moosa, and S. R. Palumbi. 2002a. Rapid recovery of genetic populations on Krakatau: diversity of stomatopod temporal and spatial scales of marine larval dispersal. Proc R Soc B **269**:1591-1597.

Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2002b. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. Mol Ecol. **11**:659-674.

Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science **312**:570-572.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science **300**:1698-1700.

Bos, A. R., G. S. Gumanao, J. C. E. Alipoyo, and L. T. Cardona. 2008. Population dynamics, reproduction and growth of the Indo-Pacific horned sea star, *Protoreaster nodosus* (Echinodermata: Asteroidea). Mar Biol. **156**:55-63.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nat Rev Gen. **4**:216-224.

Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. P Natl Acad Sci USA **76**:1967-1971.

Burridge, C. P., D. Craw, D. Fletcher, and J. M. Waters. 2008. Geological dates and molecular rates: Fish DNA sheds light on time dependency. Mol Biol Evol. **25**:624-633.

Chappell, J., A. Omura, T. Esat, M. McCulloch, J. Pandolfi, Y. Ota, and B. Pillans. 1996. Reconciliation of late Quaternary sea levels derived from coral terraces at Huon Peninsula with deep sea oxygen isotope records. Earth Planet Sc Lett. **141**:227-236.

Chenoweth, S. F., and J. M. Hughes. 2003. Oceanic interchange and nonequilibrium population structure in the estuarine dependent Indo-Pacific tasselfish, *Polynemus sheridani*. Mol Ecol. **12**:2387-2397.

Crandall, E. D., M. A. Frey, R. K. Grosberg, and P. H. Barber. 2008a. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. Mol Ecol. **17**:611-626.

Crandall, E. D., M. E. Jones, M. M. Muñoz, B. Akinronbi, M. V. Erdmann, and P. H. Barber. 2008b. Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. Mol Ecol. **17**:5276-5290.

DeBoer, T. S., M. D. Subia, Ambariyanto, M. V. Erdmann, K. Kovitvongsa, and P. H. Barber. 2008. Phylogeography and Limited Genetic Connectivity in the Endangered Boring Giant Clam across the Coral Triangle. Conserv Biol. **22**:1255-1266.

Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLOS Biol. **4**:e88.

Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. **161**:1307-1320.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. **7**:214.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. **22**:1185-1192.

Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution. **54**:1839-1854.

Emerson, B. C. 2007. Alarm Bells for the Molecular Clock? No Support for Ho et al.'s Model of Time-Dependent Molecular Rate Estimates. Syst Biol. **56**:337.

Erdmann, M. V. 1997. The ecology, distribution and bioindicator potential of Indonesian coral reef stomatopod communities. pp. 290. University of California, Berkeley.

Excoffier, L., L. G. Laval, and S. Schneider. 2005. Arlequin v.3.0: An integrated software package for population genetics data analysis. Evol Bioinform. **1**:47-50.

Frey, M. A., and G. J. Vermeij. 2008. Molecular phylogenies and historical biogeography of a circumtropical group of gastropods (Genus: *Nerita*): implications for regional diversity patterns in the marine tropics. Mol Phylogenet Evol. **48**:1067-1086.

Fu, Y.-X. 1997. Statistical tests of neutrality against population growth, hitchhiking and background selection. Genetics **147**:915-925.

Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693-809.

Geyh, M. A., H. R. Kudrass, and H. Streif. 1979. Sea-Level Changes during the Late Pleistocene and Holocene in the Strait of Malacca. Nature **278**:441-443.

Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. P Natl Acad Sci USA **102**:140-145.

Grant, W. S., and B. Bowen. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. J Hered. **89**:415-426.

Hanebuth, T., K. Stattegger, and P. M. Grootes. 2000. Rapid flooding of the Sunda Shelf: A late-glacial sea-level record. Science **288**:1033-1035.

Hanebuth, T. J. J., K. Stattegger, and A. Bojanowski. 2009. Termination of the Last Glacial Maximum sea-level lowstand: The Sunda-Shelf data revisited. Global Planet Change **66**:76-84.

Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from multiple loci. BMC Evol Biol. **8**:289.

Hellberg, M. 2009. Gene Flow and Isolation among Populations of Marine Animals. Annu Rev Ecol Syst. **40**:291-310.

Hesp, P. A., C. C. Hung, M. Hilton, C. L. Ming, and I. M. Turner. 1998. A first tentative Holocene sea-level curve for Singapore. J Coastal Res. **14**:308-314.

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. Lessios. 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution **60**:2435-2453.

Ho, S. Y., U. Saarma, R. Barnett, J. Haile, and B. Shapiro. 2008. The effect of inappropriate calibration: three case studies in molecular ecology. PLOS One. **3**:e1615.

Ho, S. Y. W., S. Kolokotronis, and R. G. Allaby. 2007a. Elevated substitution rates estimated from ancient DNA sequences. Biol Lett. **3**:702-705.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. **22**:1561-1568.

Ho, S. Y. W., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond. 2007b. Evidence for time dependency of molecular rate estimates. Syst Biol. **56**:515-522.

Howell, N., C. Smejkal, D. Mackey, P. Chinnery, D. Turnbull, and C. Herrnstadt. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. Am J Hum Genet. **72**:659-670.

Kass, R. E., and A. E. Raftery. 1995. Bayes Factors. J Am Stat Assoc. **90**:773-795.

Keightley, P. D., and A. Eyre-Walker. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics **177**:2251-2261.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature **217**:624-626.

Kimura, M., and T. Ohta. 1971. On the rate of molecular evolution. J Mol Evol. **1**:1-17.

Laird, C. D., B. L. McConaughy, and B. J. McCarthy. 1969. Rate of fixation of nucleotide substitutions in evolution. Nature **224**:149-154.

Lambeck, K., T. M. Esat, and E. K. Potter. 2002. Links between climate and sea levels for the past three million years. Nature **419**:199-206.

Langley, C. H., and W. M. Fitch. 1974. An examination of the constancy of the rate of molecular evolution. J Mol Evol. **3**:161-177.

Lessios, H. A. 2008. The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. Annu Rev Ecol Evol Syst. **39**:63-91.

Li, W. H., D. L. Ellsworth, J. Krushkal, B. H. J. Chang, and D. Hewett-Emmett. 1996. Rates of nucleotide substitution in primates and rodents and the generation time effect hypothesis. Mol Phylogenet Evol. **5**:182-187.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**:1451-1452.

Lind, C. E., B. S. Evans, J. J. U. Taylor, and D. R. Jerry. 2007. Population genetics of a marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago

shows differentiation and decreased diversity at range limits. Mol Ecol. **16**:5193-5203.

Lucas, J. S. 1988. Giant Clams: description, distribution, and life history. In: J. W. Copland, and J. S. Lucas, editors. Giant clams in Asia and the Pacific. ACIAR, Canberra, Australia. pp. 21-32

Malaquias, M., and D. Reid. 2009. Tethyan vicariance, relictualism and speciation: evidence from a global molecular phylogeny of the opisthobranch genus *Bulla*. J Biogeogr **36**:1760-1777.

Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. Mol Biol Evol. **19**:2005-2021.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. P Natl Acad Sci USA **90**:4087-4091.

Maynard-Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. Gen Res. **23**:23-35.

Montooth, K. L., and D. M. Rand. 2008. The spectrum of mitochondrial mutation differs across species. PLOS Biol. **6**:1634-1637.

Navascues, M., and B. C. Emerson. 2009. Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. Mol Ecol. **18**:4390-4397.

Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. P Natl Acad Sci USA **76**:5269-5273.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. Genetics **158**:885-896.

O'Fallon, B. D. 2010. A method to correct for the effects of purifying selection on genealogical inference. Mol Biol Evol. **27**:2406-2416.

Ohta, T. 1972. Population size and rate of evolution. J Mol Evol. **1**:305-314.

Ophir, R., T. Itoh, D. Graur, and T. Gojobori. 1999. A simple method for estimating the intensity of purifying selection in protein-coding genes. Mol Biol Evol. **16**:49-53.

Palumbi, S. R. 1994. Genetic Divergence, Reproductive Isolation, and Marine Speciation. Annu Rev Ecol Syst. **25**:547-572.

Pandolfi, J., and J. Jackson. 2001. Community structure of Pleistocene coral reefs of Curacao, Netherlands Antilles. Ecol Monogr. **71**:49-67.

Parsons, T., D. Muniec, K. Sullivan et al. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet. **15**:363-368.

Penny, D. 2005. Evolutionary biology - Relativity for molecular clocks. Nature **436**:183-184.

Peterson, G. I., and J. Masel. 2009. Quantitative Prediction of Molecular Clock and K-a/K-s at Short Timescales. Mol Biol Evol. **26**:2595-2603.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol. **9**:552-569.

Saarma, U., S. Y. W. Ho, O. G. Pybus et al. (19 co-authors) 2007. Mitogenetic structure of brown bears (*Ursus arctos* L.) in northeastern Europe and a new time frame for the formation of European brown bear lineages. Mol Ecol. **16**:401-413.

Sathiamurthy, E., and H. K. Voris. 2006. Maps of Holocene Sea Level Transgression and Submerged Lakes on the Sunda Shelf. The Natural History Journal of Chulalongkorn University **Supplement 2**:1-43.

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: Application to human mitochondrial DNA. Genetics **152**:1079-1089.

Seger, J., W. A. Smith, J. J. Perry, J. Hunn, Z. A. Kaliszewska, L. La Sala, L. Pozzi, V. J. Rowntree, and F. R. Adler. 2010. Gene Genealogies Strongly Distorted by Weakly Interfering Mutations in Constant Environments. Genetics **184**:529-545.

Shapiro, B., A. J. Drummond, A. Rambaut et al. 2004. Rise and fall of the Beringian steppe bison. Science **306**:1561-1565.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **123**:603-613.

Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades: Molecular clocks and the fossil record. Ann Rev Earth Planet Sci. **30**:65-88.

Starger, C., P. Barber, Ambariyanto, and A. Baker. 2010. The recovery of coral genetic diversity in the Sunda Strait following the 1883 eruption of Krakatau. Coral Reefs **29**:547-565.

Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol Biol Evol. **18**:2298-2305.

Subramanian, S., D. R. Denver, C. D. Millar, T. Heupink, A. Aschrafi, S. D. Emslie, C. Baroni, and D. M. Lambert. 2009. High mitogenomic evolutionary rates and time dependency. Trends Genet. **25**:482-486.

Swofford, D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony *and Other Methods. Sinauer Associates, Sunderland, MA.

Tachida, H. 1991. A study on a nearly neutral mutation model in finite populations. Genetics **128**:183-192.

Tager, D., J. M. Webster, D. C. Potts, W. Renema, J. C. Braga, and J. M. Pandolfi. 2010. Community dynamics of Pleistocene coral reefs during alternative climatic regimes. Ecology **91**:191-200.

Takahata, N. 2007. Molecular clock: an anti-neo-Darwinian legacy. Genetics **176**:1-6.

Thomas, J. A., J. J. Welch, R. Lanfear, and L. Bromham. 2010. A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates. Mol Biol Evol. **27**:1173-1180.

Tomascik, T., R. vanWoesik, and A. J. Mah. 1996. Rapid coral colonization of a recent lava flow following a volcanic eruption, Banda Islands, Indonesia. Coral Reefs **15**:169-175.

Voris, H. K. 2000. Special Paper 2:Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. J Biogeog. **27**:1153-1167.

Wares, J. P. 2010. Natural distributions of mitochondrial sequence diversity support new null hypotheses. Evolution **64**:1136-1142.

Wares, J. P. 2002. Community genetics in the Northwestern Atlantic intertidal. Mol Ecol. **11**:1131-1144.

Williamson, S., and M. E. Orive. 2002. The genealogy of a sequence subject to purifying selection at multiple sites. Mol Biol Evol. **19**:1376-1384.

Yoder, A. D., and Z. H. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol. **17**:1081-1090.

Zuckerkandl, E., and L. Pauling. 1965. Molecules as documents of evolutionary history. J Theor Biol. **8**:357-366.

Table 1. Summary statistics for the CO1 datasets used for calibration.

Taxon	Localities[1] included in dataset	n	# sites	S	h	π	F_S	D*	Φ_{ST}	Molec Evolut Mod
Tridacna crocea; 'Black Clade'	→K,P	49	485	22	0.810	0.01	-3.3	-1.35	-0.001	HK
	K	17	485	18	0.824	0.009	-0.93	-0.09	N/A	
	P	32	485	17	0.813	0.011	-0.93	0.13	N/A	
Haptosquilla pulchella; 'White Clade'	L,P,C	73	625	55	0.969	0.011	**-34.24**	**-1.99**	**0.052**	
	→L,C	59	625	52	0.964	0.009	**-29.95**	-1.76	-0.006	HKY
	L	43	625	43	0.965	0.009	**-18.71**	-0.96	N/A	
	C	16	625	28	0.983	0.008	**-7.91**	**-2.15**	N/A	
	P	14	625	30	0.956	0.015	-1.58	0.02	N/A	
Protoreaster nodosus	→K	38	803	15	0.694	0.002	**-10.36**	**-3.53**	N/A	HK

[1] Sampled localities are: Carita, West Java (C); Karimunjawa, East Java (K); Lovina, Bali (L); Pulau Seribu, Java (P). Arrows denote datasets on which final analyses were run, and for which calibrations are presented in Table 2.

Table 2. Mean parameter values and lineage mutation rates from the sudden expansion model for the mismatch distribution and the two-epoch model. All values are per site.

Taxon	Mismatch Distribution Parameters			Lineage Mutation Rate[a] (% per million years)		
	Θ_0	Θ_1	τ / 2b	μ 95% low	μ	μ 95% high
Tridacna crocea; 'Black Clade' - CO1	0	1.50×10^{-2}	1.20×10^{-2}	0.00%	**61.14%**[b]	529.26%
				0.00%	**82.18%**[c]	711.49%
Haptosquilla pulchella; 'White Clade' - CO1	4.00×10^{-3}	2.54×10^{-2}	3.65×10^{-3}	0.89%	**18.64%**[b]	73.52%
				1.20%	**25.06%**[c]	98.83%
Protoreaster nodosus - CO1	0	Inf	4.53×10^{-4}	0.80%	**2.31%**[b]	3.74%
				1.07%	**3.11%**[c]	5.03%
	Two-Epoch Model Parameters					
	Θ_0	Θ_1	$t_{transition}$			
Tridacna crocea; 'Black Clade' - CO1[d]	5.45×10^{-3}	7.47×10^{-2}	4.50×10^{-4}	0.05%	**2.30%**[b]	8.16%
				0.07%	**3.09%**[c]	10.97%
Haptosquilla pulchella; 'White Clade' - CO1	9.29×10^{-3}	2.37×10^{-1}	1.29×10^{-3}	2.16%	**6.58%**[b]	11.89%
				2.91%	**8.85%**[c]	15.98%
Protoreaster nodosus - CO1	9.35×10^{-4}	1.03×10^{-1}	5.11×10^{-4}	0.36%	**2.61%**[b]	5.51%
				0.48%	**3.50%**[c]	7.41%

[a] Lineage mutation rates calculated as $\mu = (\tau/b) / (2c)$ for the mismatch distribution, and $\mu = t_{transition}/ c$ for the two-epoch model, where b is the number of sites, c is the calibration point given below.

[b] Values for calibration point at 19.60 kya.

[c] Values for calibration point at 14.60 kya.

[d] A two-epoch model of logistical growth was only weakly supported over a model of constant population size for this species (Table 3).

Table 3. Bayes Factor tests comparing a model of constant population size to two different two-epoch models.

					2 Log$_e$ Bayes Factors				
					Exponential Growth	**Logistic Growth**	**Expansion Growth**	**Two-epoch (exponential growth)**	**Two-epoch (logistic growth)**
					# parameters				
Species	**Model[a,b]**	**Ln P (model)**	**SE**	**Constant**	**2**	**3**	**3**	**4**	**5**
Tridacna crocea – "Black Clade"	Exponential Growth	-713.89	0.48	-16.48	-	-6.60	-2.51	-16.17	-18.15
	Logistic Growth	-710.59	0.54	-9.88	6.60**	-	4.09*	-9.56	-11.54
	Expansion Growth	-712.63	0.50	-13.97	2.51*	-4.09	-	-13.65	-15.63
	Two-epoch (exp. growth)	-705.81	0.31	-0.32	16.17***	9.56**	13.65***	-	-1.98
	Two-epoch (log. growth)	-704.82	0.31	1.66	18.15***	11.54***	15.63***	1.98	-
Haptosquilla pulchella – "White Clade"	Constant	-1280.59	0.61	-	-3.81	-11.62	-16.71	-27.00	-37.36
	Exponential Growth	-1278.68	0.54	3.81*	-	-7.81	-12.89	-23.19	-33.55
	Logistic Growth	-1274.78	0.54	11.62***	7.81**	-	-5.08	-15.38	-25.74
	Expansion Growth	-1272.23	0.49	16.71***	12.89***	5.08*	-	-10.29	-20.65
	Two-epoch (exp. growth)	-1267.09	0.39	27.00***	23.19***	15.38***	10.29***	-	-10.36
	Two-epoch (log. growth)	-1261.91	0.57	37.36***	33.55***	25.74***	20.65***	10.36***	-
Protoreaster nodosus	Constant	-1130.94	0.34	-	-17.65	-9.80	-11.49	-29.32	-32.46
	Exponential Growth	-1122.11	0.38	17.65***	-	7.86**	6.16**	-11.66	-14.81
	Logistic Growth	-1126.04	0.40	9.80**	-7.86	-	-1.69	-19.52	-22.66
	Expansion Growth	-1125.20	0.35	11.49***	-6.16	1.69	-	-17.83	-20.97
	Two-epoch (exp. growth)	-1116.28	0.28	29.32***	11.66***	19.52***	17.83***	-	-3.15
	Two-epoch (log. growth)	-1114.71	0.45	32.46***	14.81***	22.66***	20.97***	3.15*	-

[a] Comparisons are row by column. [b] Models chosen for calibration are highlighted in bold.
* Positive support; ** Strong support; *** Very Strong Support

Figure Legends

Figure 1. Sea level curve for the Sunda Shelf based on radiocarbon dating of sediment cores, corals, and littoral detritus (Geyh et al. 1979; Hesp et al. 1998; Hanebuth et al. 2000; Hanebuth et al. 2009), and corresponding curves of newly submerged coastal habitat (0-10m) and total shelf area. Letters correspond to sea level maps below, and the stars highlight two calibration points at 19.60 kya and 14.58 kya. Red diamonds on map H denote collecting sites at Carita (C), Pulau Seribu (P), Karimunjawa (K), and Lovina (L). Data and maps for this figure were adapted with permission from Sathiamurthy and Voris (2006), © Field Museum of Natural History, Chicago, Illinois, USA, with slight modification for new data in Hanebuth et al. (2009), and a new analysis of shallow-water habitat presented herein.

Figure 2. Mismatch distributions for all three species calculated in Arlequin 3.11 (Excoffier et al. 2005). Histograms show the frequency of each pairwise difference in the sample, and dotted lines show the expected frequencies under a sudden expansion model and a spatial expansion model (Excoffier 2004). Sum of squared deviations (SSD) are given for each gene and model. Parameters of the spatial expansion model are given in Table 3.

Figure 3. Bayesian skyline plots (BSP) for all three species estimated in BEAST 1.5.3 (Drummond and Rambaut 2007). Plots are log-linear, and the x-axis is in mutational units. Dotted black lines depict the median value for Θ ($\frac{1}{2}N_e\mu$, righthand vertical axis), and thin lines depict 95% confidence intervals. Marginal posterior distributions for the time of transition to logistic growth from the two-epoch model are overlaid on the BSPs with gray shading (posterior density given on lefthand vertical axis). Thick lines depict the mean value for the time of transition ($t_{transition}$). Dark shaded areas depict areas beyond the region of 95% highest posterior density. The upper-right panel shows a schematic for the two-epoch model used to estimate the

time of population expansion, using BSP confidence intervals as priors. Black lines show parameters, while grey lines show prior boundaries. The lower limits for $t_{transition}$ and t_{50} were set to 10^{-6} mutations per site, which is not visible in this figure.

Figure 4. Lineage substitution rates (½ divergence rate) *per generation* for marine invertebrate CO1 plotted against their calibration date. Error bars represent 95% credibility intervals. For calibrations < 5 mya, we only plotted rates that account for ancestral polymorphism (Edwards and Beerli 2000) in a coalescent framework. For calibrations > 5 mya we corrected for ancestral polymorphism using net nucleotide divergence (Nei and Li 1979). Rates from present study: (a) *Haptosquilla pulchella*; (b) *Protoreaster nodosus*. (c) rate from simultaneous divergence of 7 echinoid species pairs at the Isthmus of Panama, (Hickerson et al. 2006). (d) average rate from two divergence times for 15 alpheid species pairs (Hickerson et al. 2003). Rates from Neritid fossil calibrations (Frey and Vermeij 2008): (e) *Nerita fulgurans* and *N. senegalensis*; (f) *N. scabricosta* and *N. peloranta* + *versicolor* (g) *Nerita exuvia* and *Nerita textilis* (j) *Nerita adanensis* + *Nerita planospira* and Nerita crown clade. (i) Fossil calibration for *Bulla striata* and *B. occidentalis* (Malaquias and Reid 2009). Rates from Arcid fossil calibrations (Marko 2002): (h) *Andara* and *Grandiarca*; (k) *Fugleria* and *Cucullaearca*

Figure 1.

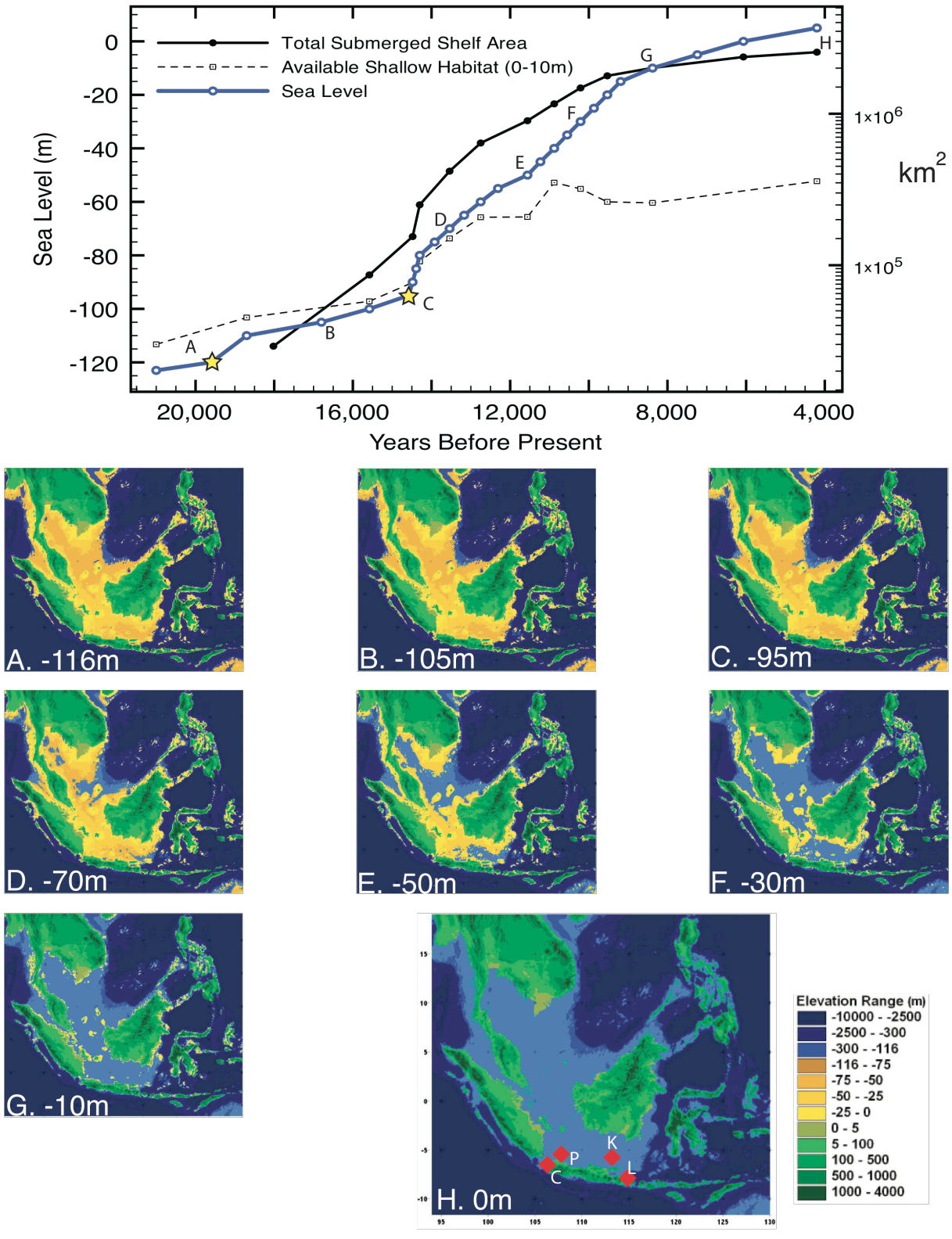

Total Submerged Shelf Area
Available Shallow Habitat (0-10m)
Sea Level
Sea Level (m)
0
-20
-40
-60
-80
-100
-120
1×10^6
1×10^5
km^2
A
B
C
D
E
F
G
H
20,000
16,000
12,000
8,000
4,000
Years Before Present
A. -116m
B. -105m
C. -95m
D. -70m
E. -50m
F. -30m
G. -10m
H. 0m
K
P
C
L
Elevation Range (m)
-10000 - -2500
-2500 - -300
-300 - -116
-116 - -75
-75 - -50
-50 - -25
-25 - 0
0 - 5
5 - 100
100 - 500
500 - 1000
1000 - 4000

Figure 2.

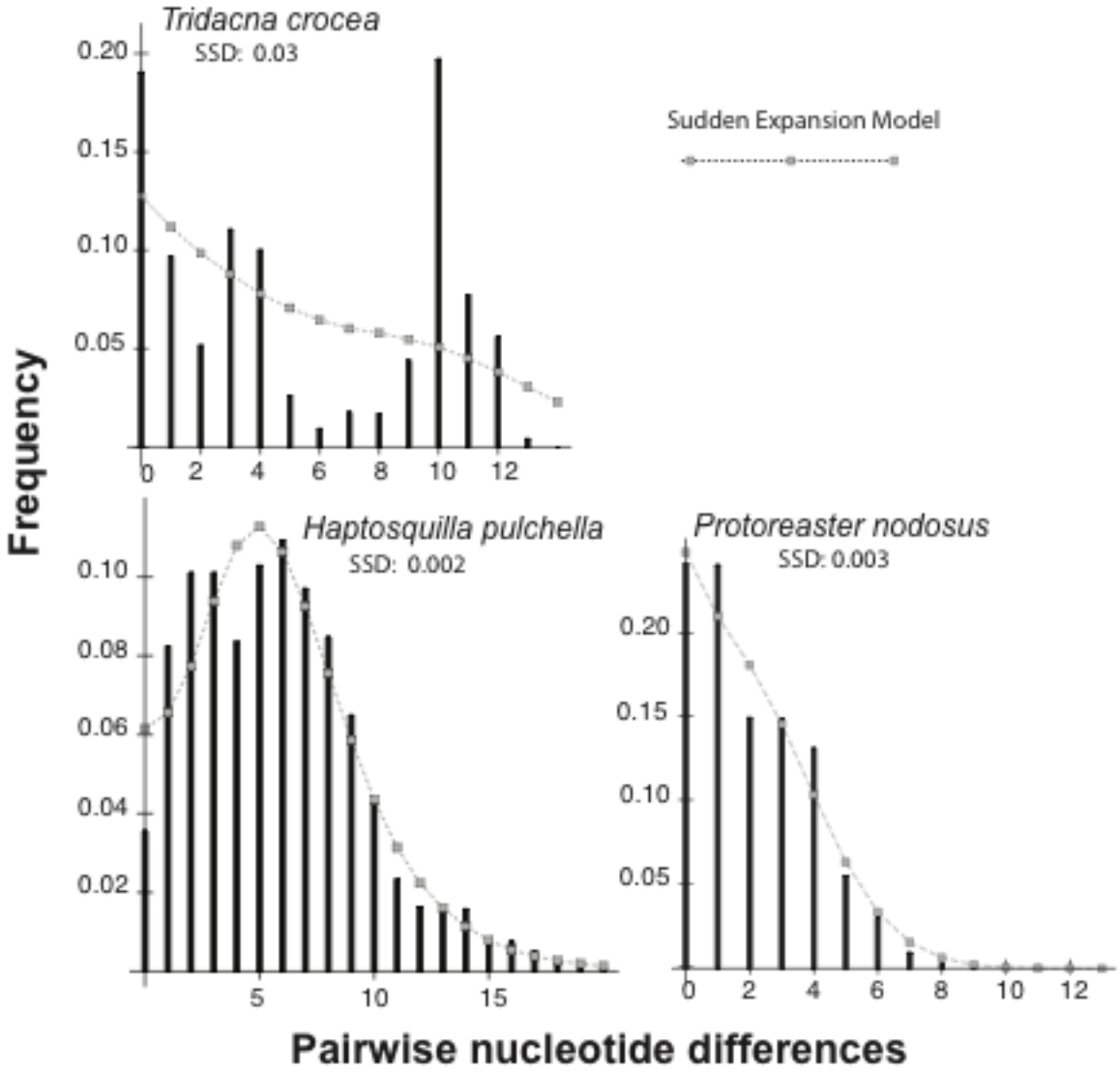

Tridacna crocea
SSD: 0.03
Sudden Expansion Model
Haptosquilla pulchella
SSD: 0.002
Protoreaster nodosus
SSD: 0.003
Frequency
Pairwise nucleotide differences

Figure 3.

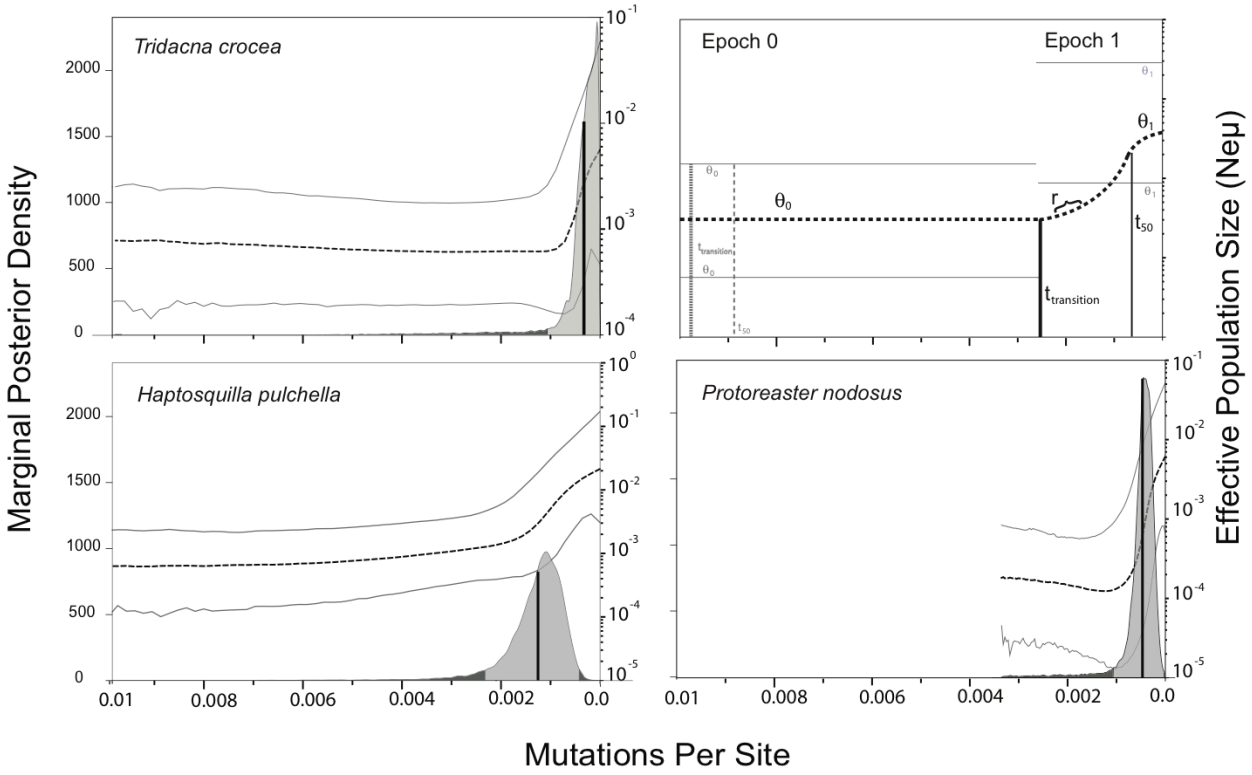
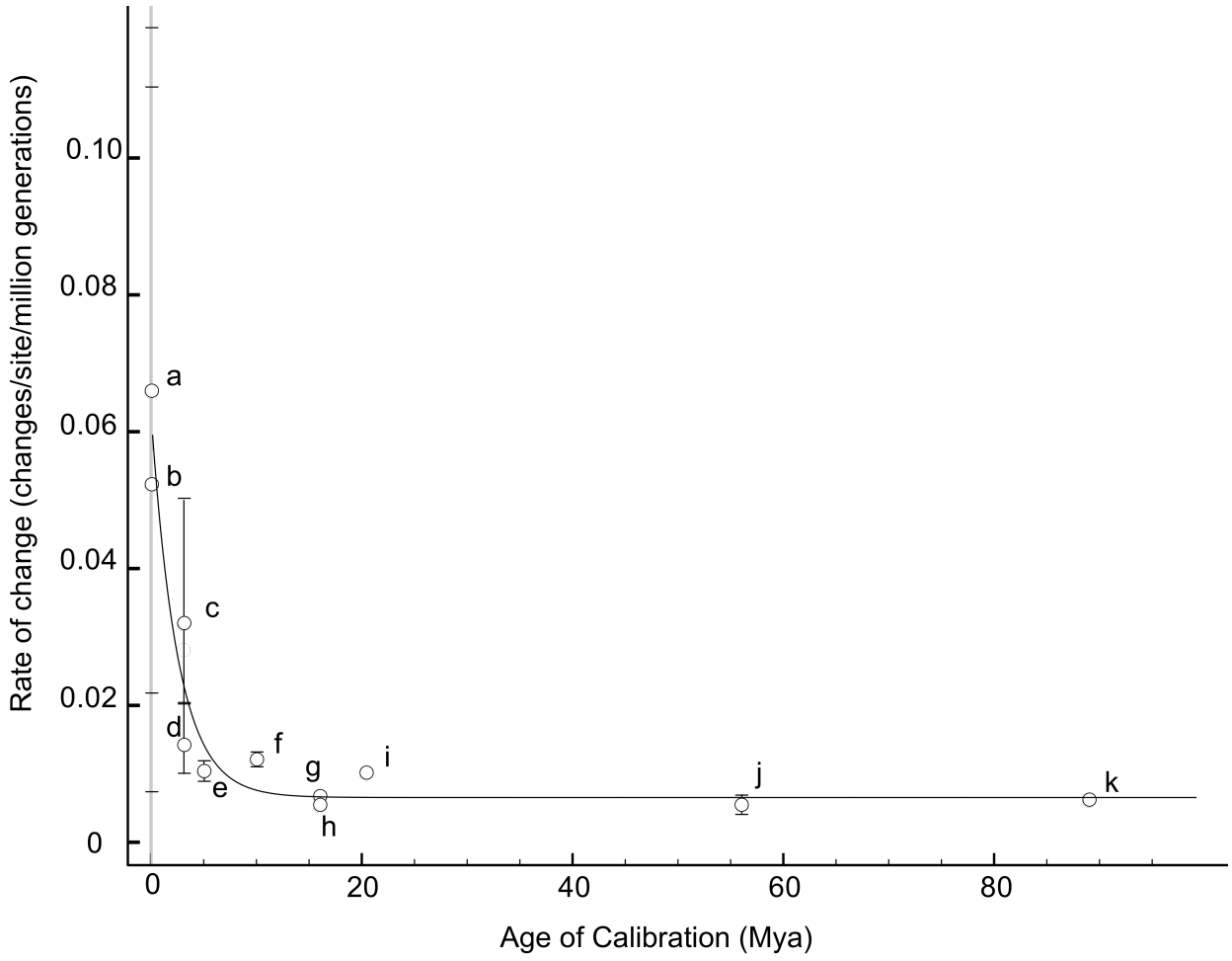

Tridacna crocea
Haptosquilla pulchella
Protoreaster nodosus
Epoch 0
Epoch 1
θ0
θ1
r
t50
ttransition
Marginal Posterior Density
Effective Population Size (Neμ)
Mutations Per Site
0.01
0.008
0.006
0.004
0.002
0.0

Figure 4.


Rate of change (changes/site/million generations)
0.10
0.08
0.06
0.04
0.02
0
a
b
c
d
e
f
g
h
i
j
k
0
20
40
60
80
Age of Calibration (Mya)