



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

Journal:	<i>Molecular Biology and Evolution</i>
Manuscript ID:	MBE-11-0516
Manuscript Type:	Research Article
Date Submitted by the Author:	20-Jun-2011
Complete List of Authors:	Crandall, Eric; University of California, Santa Cruz, Institute of Marine Sciences; Southwest Fisheries Science Center Sbrocco, Elizabeth; Boston University, Biology DeBoer, Timery; Boston University, Biology Barber, Paul; University of California, Los Angeles, Ecology and Evolutionary Biology Carpenter, Kent; Old Dominion University, Biological Sciences
Key Words:	time-dependency, mutation rate, population expansion, molecular clock, marine invertebrates

SCHOLARONE™
Manuscripts

1

Date of Submission: 6/20/2011

2 Research Article

3 **Expansion dating: calibrating molecular clocks in marine species from expansions**
4 **onto the Sunda Shelf following the Last Glacial Maximum.**5 Eric D. Crandall^{1*}, Elizabeth J. Sbrocco², Timery S. DeBoer², Paul H. Barber³, Kent E.
6 Carpenter¹

7 1. Department of Biological Sciences, Old Dominion University, Norfolk, VA

8 23529, U.S.A.

9 2. Biology Department, Boston University, Boston, MA 02215, U.S.A.

10 3. Department of Ecology and Evolutionary Biology, University of California, Los
11 Angeles, Los Angeles, CA 90095, U.S.A.

12 *Corresponding author

13 Current Address: Southwest Fisheries Science Center, Santa Cruz, CA 95060, U.S.A.

14 Phone - +01-831-420-3959

15 Fax - +01-831-420-3980

16 email: eric.crandall@noaa.gov

17 Running Title: Expansion dating in Sunda Shelf marine species

18 Key words: Time-dependency, population expansion, mutation rate, molecular clock,
19 marine invertebrates20 Non-standard abbreviations: aDNA – Ancient DNA, BSP – Bayesian skyline plot, ESS –
21 Effective sample size, LGM – Last Glacial Maximum

22

1 **Abstract**

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

22

1 1 Introduction

2 The observation that genetic distances between taxa are correlated with the amount of
3 time since they diverged first gave rise to the idea that DNA may evolve at a relatively
4 constant rate: the molecular clock (Zuckerkandl and Pauling 1965). Like radiometric
5 dating, molecular clocks have allowed illumination of life's evolutionary history, but not
6 without a good deal of scientific controversy (see reviews in Arbogast et al. 2002; Smith
7 and Peterson 2002; Bromham and Penny 2003; Takahata 2007). The original observation
8 of rate constancy led directly to the neutral theory of molecular evolution, which initially
9 predicted that the rate of evolutionary change, k , would be equal to the neutral mutation
10 rate μ , regardless of the effective population size (Kimura 1968). It wasn't long, however,
11 before heterogeneity in amino acid substitution rates was observed (Kimura and Ohta
12 1971; Langley and Fitch 1974), and the neutral theory was soon amended to account for
13 the effects of effective population size (N_e) and slightly deleterious selection ("nearly
14 neutral theory"; Ohta 1972; Tachida 1991). Under the nearly neutral theory, the
15 substitution rate scales inversely with N_e , so it is possible for a large number of alleles
16 with neutral and slightly deleterious mutations to remain transient in a large population
17 for relatively long periods of time before slowly being fixed or removed by genetic drift
18 or purifying selection.

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

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

9 With increased availability of radiometrically-dated ancient DNA (aDNA) from sub-
10 fossil tissues, it has become possible to calibrate molecular clocks using radiometric dates
11 assigned to lineages that terminate at some known time in the past (Drummond et al.
12 2002; Ho et al. 2007a). In a coalescent framework, this method circumvents the problem
13 of ancestral polymorphism by essentially sampling directly from the ancestral population
14 (but see Navascues and Emerson 2009 for potential problems with this method). In 2005,
15 Ho and co-workers used coalescent rate calibrations from contemporary and ancient
16 DNA to show that there is an exponential relationship between the date used for
17 calibration and the resultant rate, with high rates for recent calibration dates that decline
18 to more familiar substitution rates for dates over 1-2 million years ago (the “lazy J”,
19 Penny 2005), a result that supported previous observations from pedigree studies in
20 humans (Parsons et al. 1997; Howell et al. 2003). In addition to avoiding the problem of
21 ancestral polymorphism through coalescent models, their analysis suggests that this
22 inverse relationship cannot be entirely explained by errors in calibration or sequencing,
23 nor by saturation at fast-evolving sites (but see debate in Emerson 2007; Ho et al. 2007b).

1
2 Instead they suggest that this time-dependent relationship between apparent molecular
3 rate and calibration time results from the prolonged action of purifying selection on
4 slightly deleterious mutations.
5
6

7 This hypothesis of time-dependency for molecular rates may have far-reaching
8 consequences for studies of genetic variation within populations and species, as
9 substitution rates from relatively old calibration points are frequently used to estimate
10 divergence time, as well as a number of population genetic parameters. For example, if
11 time-dependency occurs, then the application of rates from old (> 2 mya) calibrations to
12 analyses of recent timescales will result in the consistent overestimates of divergence
13 times and effective population sizes, and consistent underestimates of the proportion of
14 migrants in structured populations (Ho et al. 2008). However, further inquiry into this
15 hypothesis of time-dependency is hindered by a lack of recent calibration points (Ho et
16 al. 2007b), as aDNA samples are rare and pedigree studies can be prohibitively lengthy.
17 In particular, we are not aware of any recent (< 1 mya) calibration points for marine
18 fishes and invertebrates (but see Burridge et al. 2008 for an example from freshwater
19 fish). Molecular clock calibration in these groups typically relies on vicariance due to the
20 rise of the Isthmus of Panama, which led to the closure of the Panamanian Seaway 2.8 –
21 3.1 mya (Lessios 2008; Hellberg 2009), although calibrations from fossils and other
22 geological events exist as well (e.g. Marko 2002; Wares 2002). Here, we develop a new
23 coalescent-based method for the calibration of molecular clock rates that uses a
24 population expansion rather than divergence as its calibrating event. We demonstrate the
25 method using a recent, well-documented marine population expansion event, and use it to
26 test whether time-dependency can be detected in marine invertebrates.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Extending from the Malay Peninsula to the eastern sides of Java and Borneo, the Sunda Shelf covers $1.8 \times 10^6 \text{ km}^2$, 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 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), which 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

1
2
3 our model the demographic signal that is encoded in the genetic data we overcome any
4
5 bias associated with ancestral polymorphism or fluctuating population sizes (Navascues
6
7 and Emerson 2009). Our goals are to 1) develop a method for finding recent molecular
8
9 calibrations in taxa that are not represented by aDNA or pedigree calibrations and 2) test
10
11 the hypothesis of time-dependency of molecular rates (Ho et al. 2005) using this method.
12
13
14
15

16 Materials and Methods

17
18
19
20

21 Data Characterization

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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: Malacostraca), 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).

1 We used DNAsp 5.10 (Librado and Rozas 2009) to characterize these datasets for
2 standard population genetic statistics, including D^* and F_S , which identify departures
3 from neutrality due to an excess of recent mutations (Fu and Li 1993; Fu 1997), with
4 significance for these neutrality tests determined with 10,000 coalescent simulations. For
5 datasets where more than one locality was sampled, we also measured Φ_{ST} among
6 localities using the AMOVA algorithm implemented in Arlequin 3.11 (Excoffier et al.
7 2005), and only took calibrations from datasets for which $\Phi_{ST} = 0$. We ascertained the
8 best-fitting model of molecular evolution using the Akaike information criterion (AIC) in
9 ModelTest 3.7 (Posada and Crandall 1998) and PAUP* 4.0 (Swofford 2002). Finally, we
10 constructed statistical parsimony networks in TCS 1.21 (Clement et al. 2000), using a
11 95% connection limit.

12 Calibration Points and Demographic Analysis

13 Calibration of a molecular clock based on population expansion requires geological
14 dates for when the expansion began. We used two dates taken from studies of
15 radiocarbon dates for sediment cores and littoral organics (Geyh et al. 1979; Hesp et al.
16 1998; Hanebuth et al. 2000; summarized in Sathiamurthy and Voris 2006; Hanebuth et al.
17 2009). The first date, 19.6 kya, reflects the earliest time following the LGM that sea level
18 rise first became statistically distinguishable from the 2m tidal range, during a rise of
19 ~10m over 800 years (Hanebuth et al. 2009). However, sea level rise was gradual at first
20 (0.41m/100 years after this pulse) and did not result in the significant flooding of the
21 Sunda Shelf (Figure 1). We therefore used another calibration point, at 14.58 kya
22 (corresponding to the Bølling interstadial period), during which sea level rose at an
23 average of 5.33m/100 years and flooding of the Sunda Shelf began in earnest.

1
2 1 In addition to the information from the sources noted above, we estimated the area of
3 shallow water habitat (0-10m) that became available to these species with each 10m
4 increase in sea level using spatially gridded data from the ETOPO1 1 arc-minute Global
5 Relief Model (Amante and Eakins 2009). Using the spatial analyst toolkit in ArcMap
6 10.0, we quantified the number of cells between 10-meter isobaths and multiplied this
7 value by a cell size 3.16 km², which resulted from projecting the data to the ARC
8 coordinate system, zone 1.

9 8 The calibration also requires accurate estimates of the mutational depth in the
10 genealogy at which the transition to expansion growth began, together with an
11 assessment of the associated error. Here, we use two different methods to make this
12 estimate: the mismatch distribution (Rogers and Harpending 1992) and a two-epoch
13 coalescent model (Shapiro et al. 2004).

14 13 The mismatch distribution is the distribution of pairwise differences among
15 haplotypes, which Rogers and Harpending (1992) observed to be unimodal after
16 exponential population growth in a single deme. They described this with an analytical
17 model for which the modal value, $\tau = 2\mu t$ estimates the time of population expansion t in
18 terms of the mutation rate μ . Schneider and Excoffier (1999) amended this “sudden
19 expansion” method for a finite-sites model with rate heterogeneity, and fitted each
20 parameter using a least-squares approach. We analyzed our data under this model in
21 Arlequin 3.11 (Excoffier et al. 2005), and established confidence intervals for τ with
22 10,000 parametric bootstraps of the analytical model. Model fit was evaluated using the
23 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

1 the two calibration points mentioned above, and τ is scaled by the number of sites in the
2 dataset.

3 To get an overall image of the demographic history information that was available in
4 each genetic dataset, we analyzed them each under a Bayesian skyline model (Strimmer
5 and Pybus 2001; Drummond et al. 2005) implemented in BEAST 1.5.3 (Drummond and
6 Rambaut 2007). Under this model a large sample of possible coalescent genealogies are
7 broken into piecewise intervals and effective population size (N_e) is estimated at each
8 interval from the number of observed coalescent events. The resultant Bayesian skyline
9 plot (BSP) makes few *a priori* assumptions about the historical demographic trajectory of
10 the population, and can thus provide a framework for constructing more specific models.

11 For these analyses we used ten piecewise linear intervals and the molecular evolution
12 model from ModelTest. We used uniform, uninformative priors for the population size at
13 each interval, and a gamma prior for κ , the transition:transversion ratio. Each skyline
14 analysis was run, at minimum, three times for thirty million steps. We assessed
15 convergence with estimates of effective sample size (ESS) in Tracer and by comparing
16 the marginal posterior distributions for each parameter among runs. Our criterion was
17 ESS > 200 as indicated in the BEAST manual. Finally, we combined the logged
18 parameter values and trees from replicate runs using LogCombiner 1.4.9, and used Tracer
19 to create a Bayesian skyline plot for each dataset.

20 The BSP for each dataset indicated a period of constant population size, followed by
21 rapid growth that slowed as it reached the present. We therefore used a two-epoch
22 coalescent model (Shapiro et al. 2004), implemented in BEAST, that simulated either
23 two-parameter exponential growth (Θ_1 , and intrinsic growth rate, r), or three-parameter

1 logistic growth (Θ_1 , r , and time to reach $\frac{1}{2} \Theta_1$, t_{50}), preceded by one-parameter constant
2 growth (Θ_0), with a final parameter for the transition time between the two epochs
3 ($t_{\text{transition}}$). We used the same molecular evolution models that we had used for the skyline
4 plots, and $1/x$ priors on all parameters (Drummond et al. 2002) except for κ , for which we
5 used a gamma-distributed prior, and r , for which we used a simple uniform prior. We set
6 the upper and lower limits of the prior distribution for each parameter using the 95%
7 confidence intervals (CI) from the BSPs as guidelines. For example, upper limits for
8 $t_{\text{transition}}$ (the parameter of interest) were the upper limit of the 95% CI for T_{MRCA} from the
9 skyline model. The lower limit for $t_{\text{transition}}$ was set to 10^{-6} mutations per site, which
10 represents a prior assumption that rates will not be lower than 0.05% /my (about twenty
11 times slower than the 1% /my rate commonly used for mtDNA; Brown et al. 1979). We
12 calculated lineage mutation rates as $\mu = t_{\text{transition}} / c$.

13 To more rigorously test the hypothesis of constant population size followed by
14 logistic growth, we ran each dataset under a model of constant population size, two
15 models of population growth (exponential and logistic) without a stable period preceding
16 them, and a model of expansion growth (constant size followed by exponential growth)
17 that does not allow the time of expansion to vary. Where applicable, we used the same
18 priors as were used in two-epoch models. We then used Bayes factors to compare the
19 harmonic mean of the marginal likelihood from these models to those from the
20 exponential and logistic two-epoch models described above. Marginal likelihoods for
21 each model were calculated as the sum of the log-likelihoods for the genealogy and the
22 coalescent model at each recorded step (thus, the product of these two quantities), and the
23 harmonic mean for each was calculated in Tracer, using 1000 bootstrap replicates to

1 establish confidence intervals. We report calibrations from the model that received the
2 most support from this method following recommendations from Kass and Raftery
3 (1995).

4 **Results**

5 We found a more than ten-fold increase in the amount of shallow-water habitat (0-
6 10m) that became available to these species as a result of sea-level rise following the
7 LGM (Figure 1). Shallow-water habitat on the shelf went from 30,050 km² at the
8 lowstand to 358,680 km² in present sea level.

9 Summary statistics for each genetic dataset are given in Table 1. Fu's F_S was
10 negative, but not significantly so for the *Tridacna crocea* dataset. *Haptosquilla pulchella*
11 and *Protoreaster nodosus* had significantly negative values for F_S. The *H. pulchella*
12 dataset comprising multiple sampling localities had a significant Φ_{ST} value, reflecting
13 genetic structure between Pulau Seribu and the other sites. Φ_{ST} was zero once samples
14 from Pulau Seribu were removed from the dataset. There was no evidence of any
15 structure among pooled localities for *T. crocea*. Modeltest selected an HKY model of
16 molecular evolution for all datasets, with the addition of a parameter for invariant sites (I)
17 in *H. pulchella*. Statistical parsimony networks from TCS showed numerous star-like
18 polytomies that are diagnostic of population expansions (Supplemental Figure 1; Slatkin
19 and Hudson 1991)

20 Mismatch distributions were generally unimodal for *Protoreaster nodosus* CO1 and
21 *Haptosquilla pulchella* CO1, but bimodal for *Tridacna crocea* CO1 (Figure 2). The
22 bimodal distribution for *T. crocea* resulted in very large estimates of τ, resulting in

1 exceptionally high values for μ (Table 2). None of the datasets rejected a sudden
2 expansion model.

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

14 Mean values for the time of transition between constant population size and logistic
15 growth were generally a bit lower than the inflection point depicted in the skyline model
16 for each gene (Figure 3). However, the 95% confidence intervals for $t_{\text{transition}}$ generally
17 encompassed the period of growth detected by the skyline model, and fit within the very
18 large confidence intervals generated for the mismatch distributions (Table 2). For the
19 conservative calibration point at 19.60 kya, mean estimates of μ for *T. crocea* and *P.*
20 *nodosus* were very similar (2.30%/my and 2.61%/my respectively), while *H. pulchella*
21 had a mean rate that was over twice as fast (6.58%/my; Table 2). Posterior distributions
22 for $t_{\text{transition}}$, Θ_0 and Θ_1 were unimodal, while those for growth rate (r) and t_{50} were mostly
23 uninformative. With the exception of the *H. pulchella* samples from Pulau Seribu, which

1 showed little evidence of population growth, two-epoch models for individual sampling
2 localities showed strong concordance with the pooled-locality datasets in *T. crocea* and
3 *H. pulchella*, albeit with larger confidence intervals (results not shown).

4 **Discussion**

5 Traditional molecular clock calibrations have relied almost exclusively on vicariant
6 events or fossil calibrations, both of which have limited applicability in most marine taxa,
7 particularly for recent timescales. Our use of a well-documented, recent population
8 expansion instead of a divergence allows us to estimate a molecular rate of change at a
9 relatively recent timescale, and, in combination with an explicit coalescent model of
10 constant population size followed by expansion, allows us to account for polymorphisms
11 that predate the expansion (Peterson and Masel 2009). Point estimates from both
12 mismatch analysis and the two-epoch coalescent model resulted in lineage mutation rates
13 that are much higher than the 1%/my (= 2%/my divergence rate) commonly assumed for
14 many marine phylogeographic studies, although our confidence intervals often include
15 these lower values.

16 Allowing for higher rates in future phylogeographic inference may often bring it into
17 line with a region's recent palaeoclimate (see Ho et al. 2008). For example, divergence
18 dates across the Maluku Sea between populations of *H. pulchella*, previously estimated to
19 be about 470,000 years ago using a rate of 1.4%/my, (Barber et al. 2006), can now be
20 estimated at about 100,000 years ago. This is much closer to the last time that sea levels
21 approached current levels 120,000 years ago, before dropping again (Chappell et al.
22 1996). In the following discussion, we first compare estimates from the sudden expansion

1 and two-epoch models. We then discuss potential sources of error in our calibration and
2 compare our calibrated rates to rates from older calibrations.

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

16 Sources of Error

17 Our use of a two-epoch coalescent model (Shapiro et al. 2004) as a way to more
18 accurately calibrate substitution rates is new, and uses an intra-specific process
19 (population expansion) to make inferences for phylogeographic timescales, in contrast to
20 inter-lineage divergence methods that have been used previously. As such, our calibration
21 method is subject to some of the same sources of error and bias as divergence methods,
22 while introducing new ones, and avoiding others (Arbogast et al. 2002). First and

1 foremost, calibration points, as independent estimates of elapsed time, are central to any
2 scheme for separating the effects of substitution rate and time, and are probably the
3 largest sources of error in molecular clocks (Benton and Ayala 2003). Here, we used two
4 possible points for the beginning of population expansion: one at 19.6 kya and the other
5 at 14.6 kya, as a way to provide some assessment of the potential for calibration error.

6 For the sake of discussion, we use rates from the more conservative of the two dates (19.6
7 kya), but the 14.6 kya date may be more accurate, as this was the time that sea level rise
8 first resulted in significant flooding of the Sunda Shelf (Figure 1). We are also making
9 the assumption that coral reef communities colonized the Sunda Shelf as soon as new
10 shallow marine habitat was available. This is justifiable, because coral reef communities
11 have shown the potential for rapid, long-distance re-colonization of coastal waters, with
12 species richness, coral cover and genetic diversities returning to their ambient levels less
13 than 150 years after being obliterated by volcanic eruptions (Tomascik et al. 1996; Barber
14 et al. 2002a; Starger et al. 2010). Nevertheless, even the 14.6 kya date may be
15 conservative, as the Sunda Shelf environment may not have been immediately
16 appropriate for coral reef development following inundation of terrestrial habitats. Thus,
17 as we learn more about re-colonization of the Sunda Shelf, it may become apparent that
18 the above rates are actually conservative estimates and we may need to revise the
19 estimated rates upward.

20 A related assumption is that populations of all three of these lagoonal species
21 expanded onto the Sunda Shelf somewhat simultaneously. This is analogous to the
22 assumption of simultaneity that has often been made in divergence dating across multiple
23 taxa (but see Hickerson et al. 2006 et al. for a statistical test of simultaneous divergence),

1 and rests on the idea that coral reef and their associated lagoonal communities would
2 move concurrently onto the Sunda Shelf (Pandolfi and Jackson 2001; Tager et al. 2010).
3 While no tests currently exist for simultaneity in expansion, the BSPs provide a sketch of
4 demographic history for each species back to the most recent common mitochondrial
5 ancestor, and have the potential to detect multiple demographic fluctuations (e.g. figure 4
6 in Crandall et al. 2008a), which are not evident for the present datasets. However,
7 although the fact of recent marine population expansions onto the Sunda Shelf is not in
8 doubt, the wide confidence intervals and weak support for a two-epoch model of
9 logistical growth in *T. crocea* suggests that this species may have expanded later or less
10 rapidly than the other two species. More rigorous tests of this assumption await data from
11 multiple loci (Heled and Drummond 2008).

12 By setting priors of the two-epoch model to confidence intervals for the Bayesian
13 skyline plots, we essentially estimated expansion times that fit the BSPs (Figure 3). This
14 approach offers a number of advantages. First, the resulting “fitted” two-epoch models
15 were significantly better at explaining the data than simpler models of population growth
16 (Table 3). Second, in using these models we have already accounted for the effects of
17 population size change that are often neglected in divergence calibrations (Arbogast et al.
18 2002; Navascues and Emerson 2009). Third, by calibrating from an intra-specific
19 process, we avoid the potential errors arising from phylogenetic divergence
20 methodologies, such as problems with rooting, branching order and missing or extinct
21 taxa (Smith and Peterson 2002). Finally, the 95% confidence intervals given in Table 2
22 should largely account for error due to Poisson-distributed mutations and the stochasticity
23 of the coalescent.

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

8 Because genes on the mitochondrial genome are strongly linked, an advantageous
9 variant in one gene could potentially sweep to fixation through positive selection,
10 bringing all variation on its particular genome with it (Maynard-Smith and Haigh 1974).
11 As such, this genetic hitchhiking effect is nearly indistinguishable from a demographic
12 population expansion, since both result from growth in effective population size of a
13 particular mitochondrial haplotype (Fu 1997; Bazin et al. 2006). However, if hitchhiking
14 were occurring frequently in the mtDNA of these species, we might expect nucleotide
15 diversities (Table 1) to be more similar to each other than they are (Bazin et al. 2006).
16 Furthermore, we would not expect rapid increases in mtDNA effective size to occur at
17 roughly the same genealogical depth in unlinked genetic regions, or across multiple
18 species from the same geographic region. Yet we have observed population genetic
19 patterns consistent with rapid population growth in multiple species, and across unlinked
20 loci (E. Sbrocco, unpublished data from *A. ocellaris* mtDNA and scnDNA, Chenoweth
21 and Hughes 2003; Lind et al. 2007; Crandall et al. 2008a; Crandall et al. 2008b).
22 Population growth resulting from an expansion in demographic population size is
23 therefore the most parsimonious explanation for the observed patterns.

1 The process of purifying selection is conceptually more difficult to distinguish from
2 demographic growth, since it is implicated in creating the time-dependency effect, by
3 acting over longer periods of time than previously expected (Ho et al. 2005; Ho et al.
4 2007b). It is important to note that purifying selection is by no means limited to non-
5 synonymous mutations. Through a comparison with pseudogenes, Ophir et al. (1999)
6 found that an average of 75% of substitutions are non-neutral, roughly twice what is
7 estimated by D_n/D_s ratios. Many synonymous changes are under very weak selective
8 pressures, as arise from translational selection, or maintenance of nucleotide
9 compositions in the face of biased mutational input (Montooth and Rand 2008). Thus, if
10 purifying selection is occurring at the same timescales at which we are measuring
11 population growth, it may be difficult to tease the two apart, because it can be difficult to
12 ascertain exactly how purifying selection would affect genetic variation in a reversed-
13 time coalescent framework. Using a forward-time model, Peterson and Masel (2009)
14 confirm that purifying selection on slightly deleterious variation leads to a period of
15 elevation in the substitution rate at recent timescales, the length of which is increased by
16 large N_e and a highly leptokurtic distribution of selection coefficients, as would be
17 expected under nearly-neutral theory (Keightley and Eyre-Walker 2007).

18 Fortunately, purifying selection is expected to leave only a minor imprint on the
19 shape of the genealogy, while having a similar effect as demographic growth or positive
20 selection on the distribution of substitutions (shifting them towards the tips of the
21 genealogy; Williamson and Orive 2002). If purifying selection were the only process
22 occurring, we would expect to see it act more or less constantly throughout the history of
23 the mitochondrial genome, following a simple model of exponential or logistic growth

1 (e.g. Seger et al. 2010). Instead, as was initially indicated by the BSPs, two-epoch models
2 of constant population size followed by a pulsed increase in the rate of coalescence are
3 consistently favored by odds of more than 300:1 over these simpler growth models
4 (Table 3).

5 The different topologies expected under purifying selection and demographic growth
6 (Williamson and Orive 2002) may also be why Fu (1997) found that D* (which is based
7 on distribution of substitutions) was more sensitive to purifying selection than F_S (which
8 is based on haplotype frequencies), and vice versa. That Fu's F_S was significant in many
9 more instances than D* (Table 1) also suggests that demographic change is more
10 important than purifying selection in creating the observed patterns. Nevertheless, it will
11 be important to account for sites under purifying selection when estimating rates from
12 population expansions, perhaps by including rate decay parameters in coalescent
13 genealogy sampling schemes (O'Fallon 2010).

14 Rate Comparisons

15 Perhaps the most surprising result of calibrations from the two-epoch model is the
16 disparity in substitution rate estimates. While *T. crocea* and *P. nodosus* have similar rates
17 (mean values of 2.3% to 2.6% per million years), the substitution rates in *H. pulchella*
18 appear twice as fast (6.6% per million years). It is possible that this disparity results from
19 non-simultaneous dates of expansion, with *H. pulchella* experiencing an expansion event
20 much earlier than the Sunda Shelf flooding, or *T. crocea* and *P. nodosus* expanding onto
21 the Sunda Shelf about 10,000 years after *H. pulchella*. A simpler explanation for the rate
22 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 years after lethal mutations have been removed, and declines to long-term “phylogenetic” rates of 0.65% per million years with a much slower rate (given in

1 the exponential term) than has been found in birds, mammals, or freshwater fish (Ho et
2 al. 2005; Burridge et al. 2008).

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

10 Conclusions

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

21 Continued discovery of rate heterogeneity has nearly led to the abandonment of a
22 global clock for any given region of the genome (Bromham and Penny 2003) in favor of

1 local clocks that are particular to certain taxa (Yoder and Yang 2000) or relaxed clock
2 models that allow the rate of evolution to vary across the phylogeny (Aris-Brosou and
3 Yang 2002; Drummond et al. 2006). A wide variety of explanations for the observed rate
4 variation have been offered, including the correlated effects of generation time (Laird et
5 al. 1969), metabolic rate (Martin and Palumbi 1993; Gillooly et al. 2005), and DNA
6 repair mechanisms (Li et al. 1996). Our results, together with other results showing time-
7 dependent rates (Ho et al. 2005; Ho et al. 2007a; Saarma et al. 2007; Burridge et al. 2008;
8 Subramanian et al. 2009) renew support for one of the original explanations for rate
9 heterogeneity: the combined effect of weak negative selection and effective population
10 size ($N_e\sigma_s$) proposed under the nearly neutral theory of molecular evolution (Ohta 1972).
11 Specifically, these combined results suggest that the observed rate of molecular evolution
12 decays as a function of the calibration time, and that the rate of decay is a function of
13 effective population size. With additional data from more species and more genes, we
14 expect that our method of expansion dating will provide further insight into rates of
15 molecular change at recent timescales.

40 **Supplementary Material**

41
42
43 Statistical parsimony networks for each species are available in Supplementary Figure 1.
44

46 **Acknowledgements**

47
48 Impetus for this project came from discussions in a population genetics graduate seminar,
49 led by M. Sorenson, at Boston University in 2007. It was developed while EDC was a
50 postdoctoral associate with Old Dominion University, and a visiting scholar at the Marine
51 Sciences Institute of the University of the Philippines. This work is part of the Coral
52
53
54
55
56
57
58
59
60

1 Triangle Partnerships for International Research and Education (CT-PIRE), funded by the
2 National Science Foundation – OISE-0730256. Collection of DNA sequence data used in
3 this study was supported by NSF grant OCE-0349177. We would like to acknowledge
4 our Indonesian partners: Udayana University for hosting our project, as well as the
5 Indonesian Institute of Sciences (LIPI) and the Indonesian Ministry of Science and
6 Technology (RISTEK) for providing the necessary research permits. We thank N.
7 Mahardika, H. Toha and Ambaryanto for their sponsorship and logistical support. Some
8 BEAST analyses were carried out by using the online resources of the Computational
9 Biology Service Unit from Cornell University, which is partially funded by Microsoft
10 Corporation. EDC thanks E. Anderson for statistical advice, and S. Gulamhussein for her
11 patience and support.

1
2 1 **Literature Cited**
3
4
5
6 2 Amante, C. and B. W. Eakins. 2009.ETOPO1 1 Arc-Minute Global Relief Model:
7
8 3 Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS
9
10 4 NGDC-24, 19 pp.
11
12
13 5 Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002.
14
15 6 Estimating divergence times from molecular data on phylogenetic and population
16
17 7 genetic timescales. *Annu Rev Ecol Syst.* **33**:707-740.
18
19
20 8 Aris-Brosou, S., and Z. H. Yang. 2002. Effects of models of rate evolution on estimation
21
22 9 of divergence dates with special reference to the metazoan 18S ribosomal RNA
23
24
25 10 Phylogeny. *Syst Biol.* **51**:703-714.
26
27
28 11 Barber, P. H., M. V. Erdmann, and S. R. Palumbi. 2006. Comparative phylogeography of
29
30 12 three co-distributed stomatopods: origins and timing of regional lineage
31
32 13 diversification in the coral triangle. *Evolution* **60**:1825-1839.
33
34
35 14 Barber, P. H., M. K. Moosa, and S. R. Palumbi. 2002a. Rapid recovery of genetic
36
37 15 populations on Krakatau: diversity of stomatopod temporal and spatial scales of
38
39 16 marine larval dispersal. *Proc R Soc B* **269**:1591-1597.
40
41
42 17 Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2002b. Sharp genetic
43
44 18 breaks among populations of *Haplosquilla pulchella* (Stomatopoda) indicate
45
46 19 limits to larval transport: patterns, causes, and consequences. *Mol Ecol.* **11**:659-
47
48 20 674.
49
50
51 21 Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence
52
53 22 mitochondrial genetic diversity in animals. *Science* **312**:570-572.
54
55
56 23 Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. *Science* **300**:1698-1700.
57
58
59
60

- 1 1 Bos, A. R., G. S. Gumaniao, J. C. E. Alipoyo, and L. T. Cardona. 2008. Population
2 2 dynamics, reproduction and growth of the Indo-Pacific horned sea star,
3 3 *Protoreaster nodosus* (Echinodermata: Asteroidea). Mar Biol. **156**:55-63.
- 4 4 Bromham, L., and D. Penny. 2003. The modern molecular clock. Nat Rev Gen. **4**:216-
5 5 224.
- 6 6 Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid evolution of animal
7 7 mitochondrial DNA. P Natl Acad Sci USA **76**:1967-1971.
- 8 8 Burridge, C. P., D. Craw, D. Fletcher, and J. M. Waters. 2008. Geological dates and
9 9 molecular rates: Fish DNA sheds light on time dependency. Mol Biol Evol.
10 10 **25**:624-633.
- 11 11 Chappell, J., A. Omura, T. Esat, M. McCulloch, J. Pandolfi, Y. Ota, and B. Pillans. 1996.
12 12 Reconciliation of late Quaternary sea levels derived from coral terraces at Huon
13 13 Peninsula with deep sea oxygen isotope records. Earth Planet Sc Lett. **141**:227-
14 14 236.
- 15 15 Chenoweth, S. F., and J. M. Hughes. 2003. Oceanic interchange and nonequilibrium
16 16 population structure in the estuarine dependent Indo-Pacific tasselfish, *Polynemus*
17 17 *sheridani*. Mol Ecol. **12**:2387-2397.
- 18 18 Crandall, E. D., M. A. Frey, R. K. Grosberg, and P. H. Barber. 2008a. Contrasting
19 19 demographic history and phylogeographical patterns in two Indo-Pacific
20 20 gastropods. Mol Ecol. **17**:611-626.
- 21 21 Crandall, E. D., M. E. Jones, M. M. Muñoz, B. Akinronbi, M. V. Erdmann, and P. H.
22 22 Barber. 2008b. Comparative phylogeography of two seastars and their
23 23 ectosymbionts within the Coral Triangle. Mol Ecol. **17**:5276-5290.

- 1 1 DeBoer, T. S., M. D. Subia, Ambariyanto, M. V. Erdmann, K. Kovitvongsa, and P. H.
2 2 Barber. 2008. Phylogeography and Limited Genetic Connectivity in the
3 3 Endangered Boring Giant Clam across the Coral Triangle. *Conserv Biol.* **22**:1255-
4 4 1266.
- 5 5 Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics
6 6 and dating with confidence. *PLOS Biol.* **4**:e88.
- 7 7 Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating
8 8 mutation parameters, population history and genealogy simultaneously from
9 9 temporally spaced sequence data. *Genetics*. **161**:1307-1320.
- 10 10 Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by
11 11 sampling trees. *BMC Evol Biol.* **7**:214.
- 12 12 Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent
13 13 inference of past population dynamics from molecular sequences. *Mol Biol Evol.*
14 14 **22**:1185-1192.
- 15 15 Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population
16 16 divergence, and the variance in coalescence time in phylogeographic studies.
17 17 *Evolution*. **54**:1839-1854.
- 18 18 Emerson, B. C. 2007. Alarm Bells for the Molecular Clock? No Support for Ho et al.'s
19 19 Model of Time-Dependent Molecular Rate Estimates. *Syst Biol.* **56**:337.
- 20 20 Erdmann, M. V. 1997. The ecology, distribution and bioindicator potential of Indonesian
21 21 coral reef stomatopod communities. pp. 290. University of California, Berkeley.
- 22 22 Excoffier, L., L. G. Laval, and S. Schneider. 2005. Arlequin v.3.0: An integrated software
23 23 package for population genetics data analysis. *Evol Bioinform*. **1**:47-50.

- 1 Frey, M. A., and G. J. Vermeij. 2008. Molecular phylogenies and historical biogeography
2 of a circumtropical group of gastropods (Genus: *Nerita*): implications for regional
3 diversity patterns in the marine tropics. Mol Phylogen Evol. **48**:1067-1086.
- 4 Fu, Y.-X. 1997. Statistical tests of neutrality against population growth, hitchhiking and
5 background selection. Genetics **147**:915-925.
- 6 Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics
7 **133**:693-809.
- 8 Geyh, M. A., H. R. Kudrass, and H. Streif. 1979. Sea-Level Changes during the Late
9 Pleistocene and Holocene in the Strait of Malacca. Nature **278**:441-443.
- 10 Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA
11 evolution: Effects of body size and temperature on the molecular clock. P Natl
12 Acad Sci USA **102**:140-145.
- 13 Grant, W. S., and B. Bowen. 1998. Shallow population histories in deep evolutionary
14 lineages of marine fishes: Insights from sardines and anchovies and lessons for
15 conservation. J Hered. **89**:415-426.
- 16 Hanebuth, T., K. Stattegger, and P. M. Grootes. 2000. Rapid flooding of the Sunda Shelf:
17 A late-glacial sea-level record. Science **288**:1033-1035.
- 18 Hanebuth, T. J. J., K. Stattegger, and A. Bojanowski. 2009. Termination of the Last
19 Glacial Maximum sea-level lowstand: The Sunda-Shelf data revisited. Global
20 Planet Change **66**:76-84.
- 21 Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from
22 multiple loci. BMC Evol Biol. **8**:289.

- 1 Hellberg, M. 2009. Gene Flow and Isolation among Populations of Marine Animals.
- 2 Annu Rev Ecol Syst. **40**:291-310.
- 3 Hesp, P. A., C. C. Hung, M. Hilton, C. L. Ming, and I. M. Turner. 1998. A first tentative
- 4 Holocene sea-level curve for Singapore. J Coastal Res. **14**:308-314.
- 5 Hickerson, M. J., M. A. Gilchrist, and N. Takebayashi. 2003. Calibrating a molecular
- 6 clock from phylogeographic data: Moments and likelihood estimators. Evolution
- 7 **57**:2216-2225.
- 8 Hickerson, M. J., E. A. Stahl, and H. Lessios. 2006. Test for simultaneous divergence
- 9 using approximate Bayesian computation. Evolution **60**:2435-2453.
- 10 Ho, S. Y., U. Saarma, R. Barnett, J. Haile, and B. Shapiro. 2008. The effect of
- 11 inappropriate calibration: three case studies in molecular ecology. PLOS One.
- 12 **3**:e1615.
- 13 Ho, S. Y. W., S. Kolokotronis, and R. G. Allaby. 2007a. Elevated substitution rates
- 14 estimated from ancient DNA sequences. Biol Lett. **3**:702-705.
- 15 Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of
- 16 molecular rate estimates and systematic overestimation of recent divergence
- 17 times. Mol Biol Evol. **22**:1561-1568.
- 18 Ho, S. Y. W., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond. 2007b.
- 19 Evidence for time dependency of molecular rate estimates. Syst Biol. **56**:515-522.
- 20 Howell, N., C. Smejkal, D. Mackey, P. Chinnery, D. Turnbull, and C. Herrnstadt. 2003.
- 21 The pedigree rate of sequence divergence in the human mitochondrial genome:
- 22 There is a difference between phylogenetic and pedigree rates. Am J Hum Genet.
- 23 **72**:659-670.

- 1 Kass, R. E., and A. E. Raftery. 1995. Bayes Factors. *J Am Stat Assoc.* **90**:773-795.
- 2 Keightley, P. D., and A. Eyre-Walker. 2007. Joint inference of the distribution of fitness
3 effects of deleterious mutations and population demography based on nucleotide
4 polymorphism frequencies. *Genetics* **177**:2251-2261.
- 5 Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* **217**:624-626.
- 6 Kimura, M., and T. Ohta. 1971. On the rate of molecular evolution. *J Mol Evol.* **1**:1-17.
- 7 Laird, C. D., B. L. McConaughay, and B. J. McCarthy. 1969. Rate of fixation of
8 nucleotide substitutions in evolution. *Nature* **224**:149-154.
- 9 Lambeck, K., T. M. Esat, and E. K. Potter. 2002. Links between climate and sea levels
10 for the past three million years. *Nature* **419**:199-206.
- 11 Langley, C. H., and W. M. Fitch. 1974. An examination of the constancy of the rate of
12 molecular evolution. *J Mol Evol.* **3**:161-177.
- 13 Lessios, H. A. 2008. The Great American Schism: Divergence of Marine Organisms
14 After the Rise of the Central American Isthmus. *Annu Rev Ecol Evol Syst.* **39**:63-
15 91.
- 16 Li, W. H., D. L. Ellsworth, J. Krushkal, B. H. J. Chang, and D. Hewett-Emmett. 1996.
17 Rates of nucleotide substitution in primates and rodents and the generation time
18 effect hypothesis. *Mol Phylogenet Evol.* **5**:182-187.
- 19 Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of
20 DNA polymorphism data. *Bioinformatics* **25**:1451-1452.
- 21 Lind, C. E., B. S. Evans, J. J. U. Taylor, and D. R. Jerry. 2007. Population genetics of a
22 marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago

- 1 1 shows differentiation and decreased diversity at range limits. Mol Ecol. **16**:5193-
2 2 5203.
- 3 3 Lucas, J. S. 1988. Giant Clams: description, distribution, and life history. In: J. W.
4 4 Copland, and J. S. Lucas, editors. Giant clams in Asia and the Pacific. ACIAR,
5 5 Canberra, Australia. pp. 21-32
- 6 6 Malaquias, M., and D. Reid. 2009. Tethyan vicariance, relictualism and speciation:
7 7 evidence from a global molecular phylogeny of the opisthobranch genus *Bulla*. J
8 8 Biogeogr **36**:1760-1777.
- 9 9 Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of
10 10 geminate species pairs separated by the Isthmus of Panama. Mol Biol Evol.
11 11 **19**:2005-2021.
- 12 12 Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and
13 13 the molecular clock. P Natl Acad Sci USA **90**:4087-4091.
- 14 14 Maynard-Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. Gen
15 15 Res. **23**:23-35.
- 16 16 Montooth, K. L., and D. M. Rand. 2008. The spectrum of mitochondrial mutation differs
17 17 across species. PLOS Biol. **6**:1634-1637.
- 18 18 Navascues, M., and B. C. Emerson. 2009. Elevated substitution rate estimates from
19 19 ancient DNA: model violation and bias of Bayesian methods. Mol Ecol. **18**:4390-
20 20 4397.
- 21 21 Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms
22 22 of restriction endonucleases. P Natl Acad Sci USA **76**:5269-5273.

- 1 Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov
2 Chain Monte Carlo approach. *Genetics* **158**:885-896.
- 3 O'Fallon, B. D. 2010. A method to correct for the effects of purifying selection on
4 genealogical inference. *Mol Biol Evol*. **27**:2406-2416.
- 5 Ohta, T. 1972. Population size and rate of evolution. *J Mol Evol*. **1**:305-314.
- 6 Ophir, R., T. Itoh, D. Graur, and T. Gojobori. 1999. A simple method for estimating the
7 intensity of purifying selection in protein-coding genes. *Mol Biol Evol*. **16**:49-53.
- 8 Palumbi, S. R. 1994. Genetic Divergence, Reproductive Isolation, and Marine Speciation.
9 *Annu Rev Ecol Syst*. **25**:547-572.
- 10 Pandolfi, J., and J. Jackson. 2001. Community structure of Pleistocene coral reefs of
11 Curacao, Netherlands Antilles. *Ecol Monogr*. **71**:49-67.
- 12 Parsons, T., D. Muniec, K. Sullivan et al. 1997. A high observed substitution rate in the
13 human mitochondrial DNA control region. *Nat Genet*. **15**:363-368.
- 14 Penny, D. 2005. Evolutionary biology - Relativity for molecular clocks. *Nature* **436**:183-
15 184.
- 16 Peterson, G. I., and J. Masel. 2009. Quantitative Prediction of Molecular Clock and K-
17 a/K-s at Short Timescales. *Mol Biol Evol*. **26**:2595-2603.
- 18 Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution.
19 *Bioinformatics* **14**:817-818.
- 20 Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the
21 distribution of pairwise genetic differences. *Mol Biol Evol*. **9**:552-569.

- 1
2
3 1 Saarma, U., S. Y. W. Ho, O. G. Pybus et al. (19 co-authors) 2007. Mitogenetic structure
4 2 of brown bears (*Ursus arctos* L.) in northeastern Europe and a new time frame for
5 3 the formation of European brown bear lineages. *Mol Ecol.* **16**:401-413.
6
7 4 Sathiamurthy, E., and H. K. Voris. 2006. Maps of Holocene Sea Level Transgression and
8 5 Submerged Lakes on the Sunda Shelf. *The Natural History Journal of*
9 6 *Chulalongkorn University Supplement* **2**:1-43.
10
11 7 Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from
12 8 the distribution of pairwise differences when the mutation rates vary among sites:
13 9 Application to human mitochondrial DNA. *Genetics* **152**:1079-1089.
14
15 10 Seger, J., W. A. Smith, J. J. Perry, J. Hunn, Z. A. Kaliszewska, L. La Sala, L. Pozzi, V. J.
16 11 Rountree, and F. R. Adler. 2010. Gene Genealogies Strongly Distorted by
17 12 Weakly Interfering Mutations in Constant Environments. *Genetics* **184**:529-545.
18
19 13 Shapiro, B., A. J. Drummond, A. Rambaut et al. 2004. Rise and fall of the Beringian
20 14 steppe bison. *Science* **306**:1561-1565.
21
22 15 Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA
23 16 sequences in stable and exponentially growing populations. *Genetics* **123**:603-
24 17 613.
25
26 18 Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades:
27 19 Molecular clocks and the fossil record. *Ann Rev Earth Planet Sci.* **30**:65-88.
28
29 20 Starger, C., P. Barber, Ambariyanto, and A. Baker. 2010. The recovery of coral genetic
30 21 diversity in the Sunda Strait following the 1883 eruption of Krakatau. *Coral Reefs*
31 22 **29**:547-565.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1 Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA
2 sequences using the generalized skyline plot. *Mol Biol Evol.* **18**:2298-2305.
- 3 Subramanian, S., D. R. Denver, C. D. Millar, T. Heupink, A. Aschrafi, S. D. Emslie, C.
4 Baroni, and D. M. Lambert. 2009. High mitogenomic evolutionary rates and time
5 dependency. *Trends Genet.* **25**:482-486.
- 6 Swofford, D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony *and Other
7 Methods. Sinauer Associates, Sunderland, MA.
- 8 Tachida, H. 1991. A study on a nearly neutral mutation model in finite populations.
9 *Genetics* **128**:183-192.
- 10 Tager, D., J. M. Webster, D. C. Potts, W. Renema, J. C. Braga, and J. M. Pandolfi. 2010.
11 Community dynamics of Pleistocene coral reefs during alternative climatic
12 regimes. *Ecology* **91**:191-200.
- 13 Takahata, N. 2007. Molecular clock: an anti-neo-Darwinian legacy. *Genetics* **176**:1-6.
- 14 Thomas, J. A., J. J. Welch, R. Lanfear, and L. Bromham. 2010. A Generation Time
15 Effect on the Rate of Molecular Evolution in Invertebrates. *Mol Biol Evol.*
16 **27**:1173-1180.
- 17 Tomascik, T., R. vanWoesik, and A. J. Mah. 1996. Rapid coral colonization of a recent
18 lava flow following a volcanic eruption, Banda Islands, Indonesia. *Coral Reefs*
19 **15**:169-175.
- 20 Voris, H. K. 2000. Special Paper 2:Maps of Pleistocene sea levels in Southeast Asia:
21 Shorelines, river systems and time durations. *J Biogeog.* **27**:1153-1167.
- 22 Wares, J. P. 2010. Natural distributions of mitochondrial sequence diversity support new
23 null hypotheses. *Evolution* **64**:1136-1142.

- 1
2
3 1 Wares, J. P. 2002. Community genetics in the Northwestern Atlantic intertidal. *Mol Ecol.*
4 2 **11**:1131-1144.
5
6
7 3 Williamson, S., and M. E. Orive. 2002. The genealogy of a sequence subject to purifying
8
9 4 selection at multiple sites. *Mol Biol Evol.* **19**:1376-1384.
10
11
12 5 Yoder, A. D., and Z. H. Yang. 2000. Estimation of primate speciation dates using local
13
14 6 molecular clocks. *Mol Biol Evol.* **17**:1081-1090.
15
16
17 7 Zuckerkandl, E., and L. Pauling. 1965. Molecules as documents of evolutionary history. *J*
18
19 8 *Theor Biol.* **8**:357-366.
20
21
22 9
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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}	Molecular Evolution Model
<i>Tridacna</i> <i>crocea</i> ; 'Black Clade'	→K,P	49	485	22	0.810	0.01	-3.3	-1.35	-0.001	HKY
	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	
<i>Haptosquilla</i> <i>pulchella</i> ; 'White Clade'	L,P,C	73	625	55	0.969	0.011	-34.24	-1.99	0.052	HKY+I
	→L,C	59	625	52	0.964	0.009	-29.95	-1.76	-0.006	
	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	
<i>Protereaster</i> <i>nodosus</i>	→K	38	803	15	0.694	0.002	-10.36	-3.53	N/A	HKY

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	$\tau / 2$	μ 95% low	μ	μ 95% high
<i>Tridacna crocea</i> ; '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%
<i>Haptosquilla pulchella</i> ; '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%
<i>Protoreaster nodosus</i> - CO1	8.12×10^{-6}	3.72×10^{-3}	1.99×10^{-3}	1.40%	10.17%^b	17.76%
				1.88%	13.67%^c	23.87%
	Two-Epoch Model Parameters					
	Θ_0	Θ_1	$t_{\text{transition}}$			
<i>Tridacna crocea</i> ; '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%
<i>Haptosquilla pulchella</i> ; '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%
<i>Protoreaster nodosus</i> - 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 $\mu = t_{\text{transition}} / c$ 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.

Species	Model ^{a,b}	Ln P (model)	SE	Constant	2 Log _e Bayes Factors					
					# parameters					
					Exponential Growth	Logistic Growth	Expansion Growth	Two-epoch (exponential growth)	Two-epoch (logistic growth)	
<i>Tridacna crocea</i> - "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	-	
<i>Haptocholla pulchella</i> - "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	
<i>Protoreaster nodosus</i>	Two-epoch (log. growth)	-1261.91	0.57	37.36***	33.55***	25.74***	20.65***	10.36***	-	
	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	
<i>Two-epoch (exp. growth)</i>	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*	-	

^aComparisons are row by column. ^bModels chosen for calibration are highlighted in bold.

* Positive support; ** Strong support; *** Very Strong Support

1
2
3 Figure Legends
4
5

6 Figure 1. Sea level curve for the Sunda Shelf based on radiocarbon dating of sediment cores,
7 corals, and littoral detritus (Geyh et al. 1979; Hesp et al. 1998; Hanebuth et al. 2000; Hanebuth et
8 al. 2009), and corresponding curves of newly submerged coastal habitat (0-10m) and total shelf
9 area. Letters correspond to sea level maps below, and the stars highlight two calibration points at
10 19.60 kya and 14.58 kya. Red diamonds on map H denote collecting sites at Carita (C), Pulau
11 Seribu (P), Karimunjawa (K), and Lovina (L). Data and maps for this figure were adapted with
12 permission from Sathiamurthy and Voris (2006), © Field Museum of Natural History, Chicago,
13 Illinois, USA, with slight modification for new data in Hanebuth et al. (2009), and a new analysis
14 of shallow-water habitat presented herein.
15
16

17 Figure 2. Mismatch distributions for all three species calculated in Arlequin 3.11 (Excoffier et al.
18 2005). Histograms show the frequency of each pairwise difference in the sample, and dotted
19 lines show the expected frequencies under a sudden expansion model and a spatial expansion
20 model (Excoffier 2004). Sum of squared deviations (SSD) are given for each gene and model;
21 starred values deviated significantly from that model. Parameters of the spatial expansion model
22 are given in Table 3.
23
24

25 Figure 3. Bayesian skyline plots (BSP) for all three species estimated in BEAST 1.5.3
26 (Drummond and Rambaut 2007). Plots are log-linear, and the x-axis is in mutational units.
27 Dotted black lines depict the median value for Θ ($\frac{1}{2}N_e\mu$, righthand vertical axis), and thin lines
28 depict 95% confidence intervals. Marginal posterior distributions for the time of transition to
29 logistic growth from the two-epoch model are overlaid on the BSPs with gray shading (posterior
30 density given on lefthand vertical axis). Thick lines depict the mean value for the time of
31 transition ($t_{transition}$). Dark shaded areas depict areas beyond the region of 95% highest posterior
32 density.
33
34

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.

Figure 4. Lineage substitution rates ($\frac{1}{2}$ 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 *Cucullaeearca*

Figure 1.

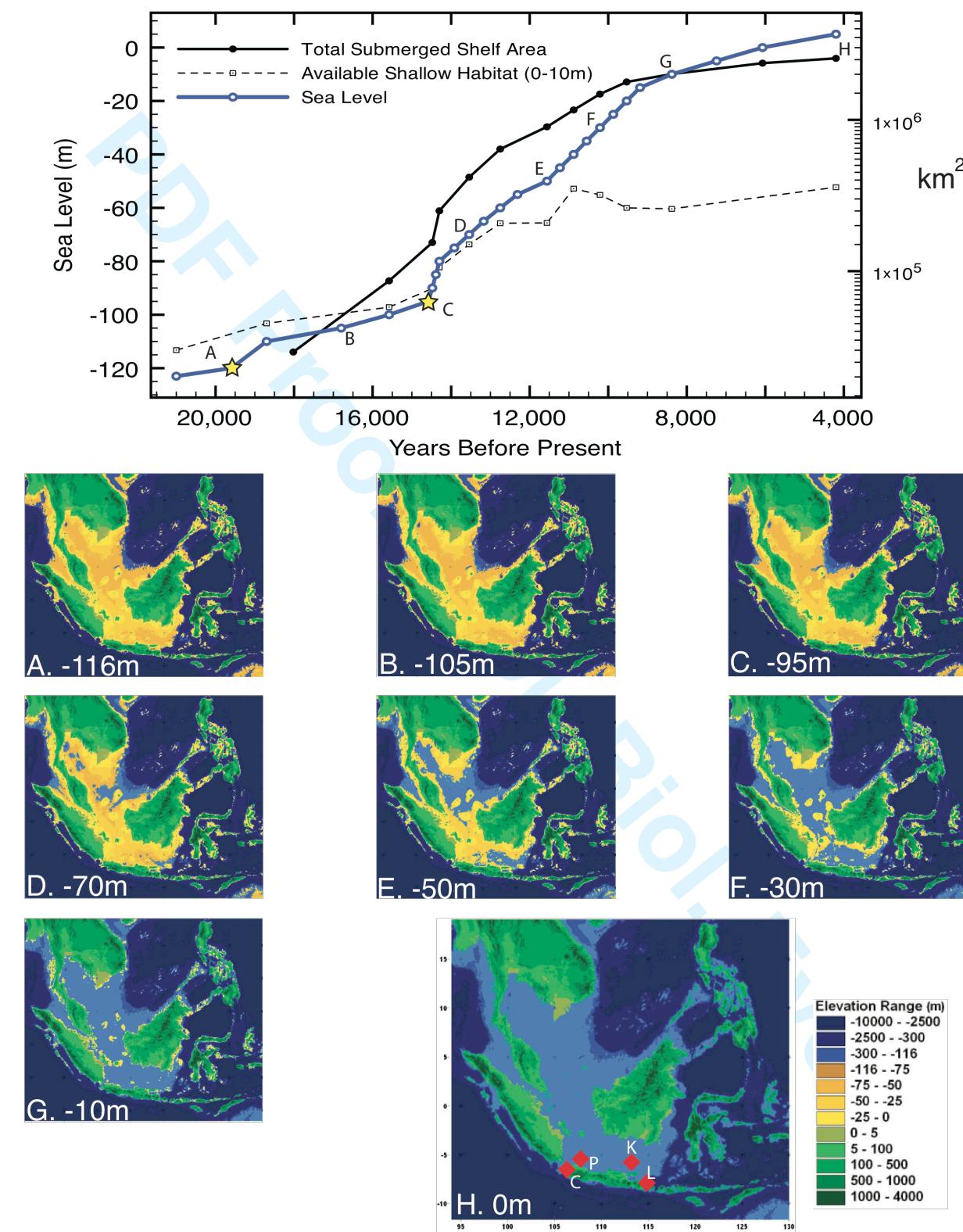


Figure 2.

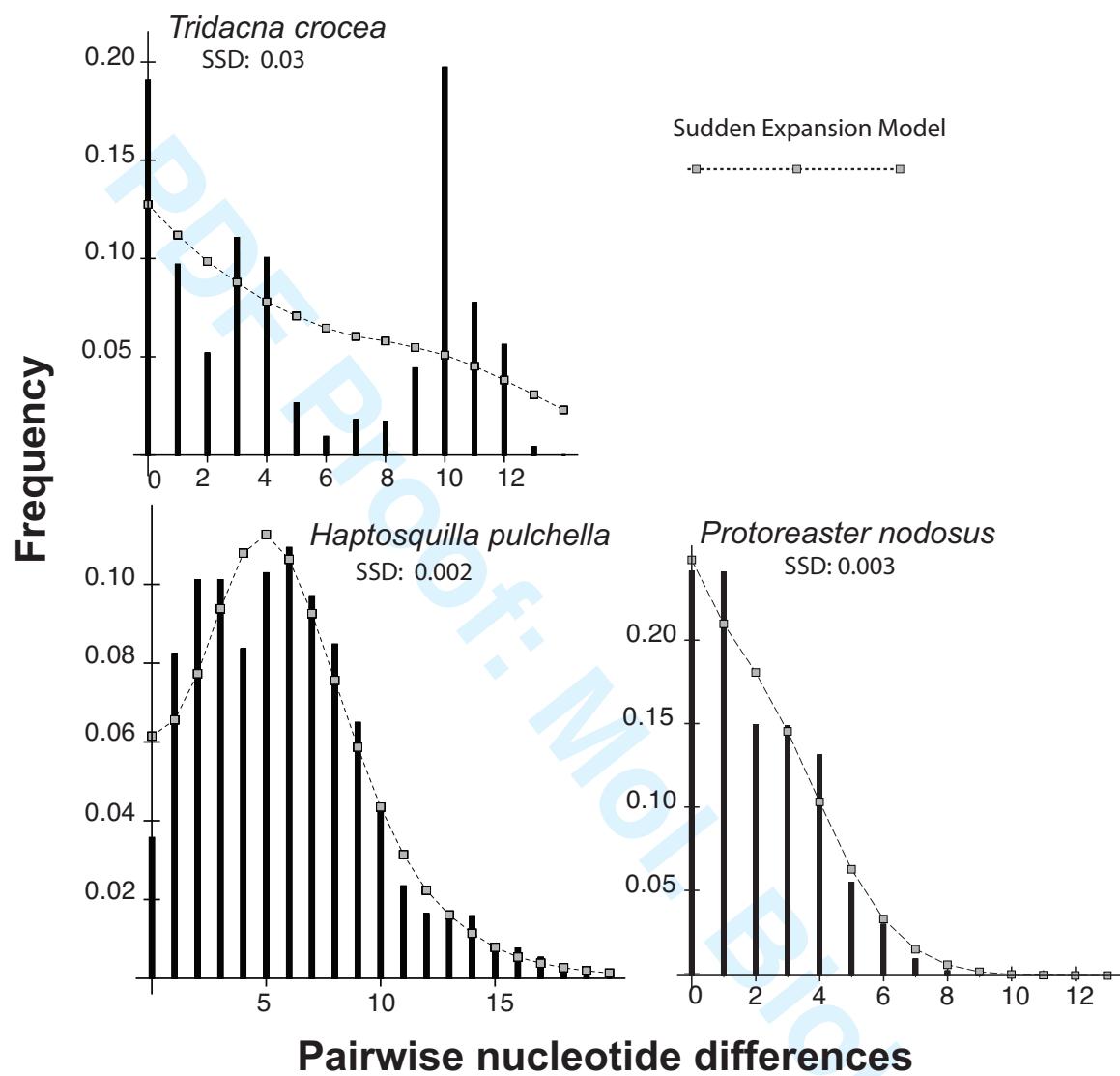


Figure 3.

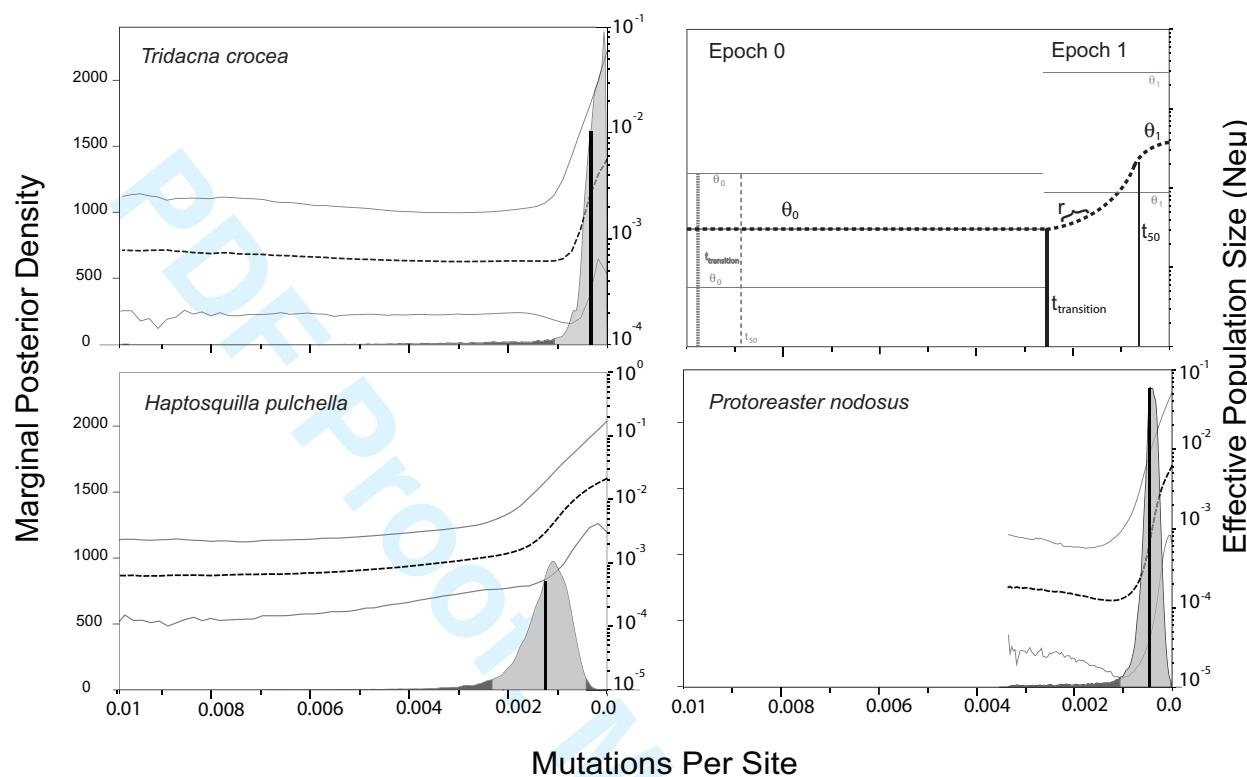
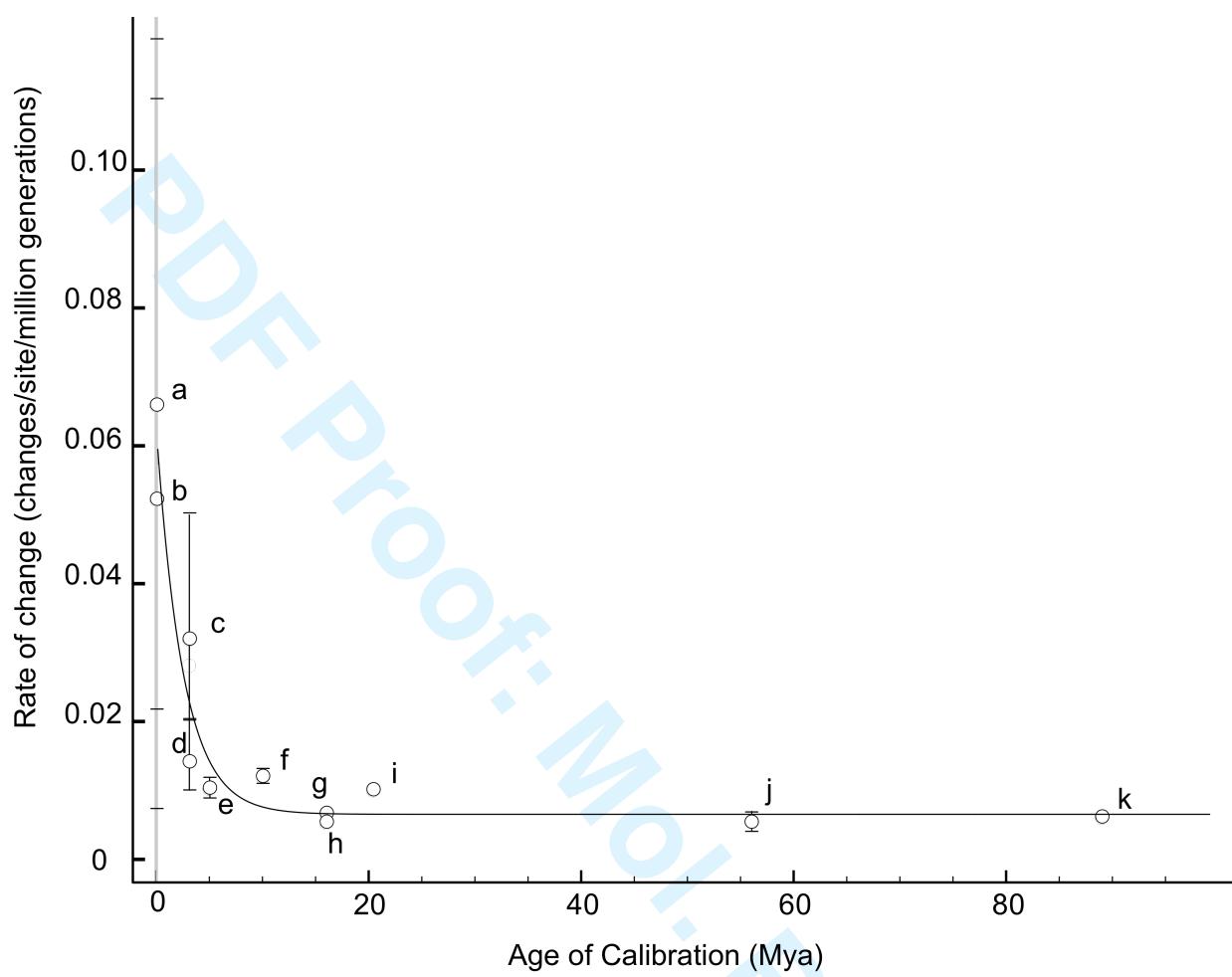


Figure 4.



Eric Crandall 6/20/11 7:37 PM
Style Definition: Comment Reference
Eric Crandall 6/20/11 7:37 PM
Style Definition: Emphasis
Eric Crandall 6/20/11 7:37 PM
Style Definition: Hyperlink
Eric Crandall 6/20/11 7:37 PM
Deleted: 2/3

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60Date of Submission: 6/20/20111 |
2 Research Article3 **Expansion dating: calibrating molecular clocks in marine species from expansions**
4 onto the Sunda Shelf following the Last Glacial Maximum.5 Eric D. Crandall^{1*}, Elizabeth J. Sbrocco², Timery S. DeBoer², Paul H. Barber³, Kent E.
6 Carpenter¹7 1. Department of Biological Sciences, Old Dominion University, Norfolk, VA
8 23529, U.S.A.

9 2. Biology Department, Boston University, Boston, MA 02215, U.S.A.

10 3. Department of Ecology and Evolutionary Biology, University of California, Los
11 Angeles, Los Angeles, CA 90095, U.S.A.

12 *Corresponding author

13 Current Address: Southwest Fisheries Science Center, Santa Cruz, CA 95060, U.S.A.

14 Phone - +01-831-420-3959

15 Fax - +01-831-420-3980

16 email: eric.crandall@noaa.gov

17 Running Title: Expansion dating in Sunda Shelf marine species

18 Key words: Time-dependency, population expansion, mutation rate, molecular clock,
19 marine invertebrates20 Non-standard abbreviations: aDNA – Ancient DNA, BSP – Bayesian skyline plot, ESS –
21 Effective sample size, LGM – Last Glacial Maximum

1
2 1 **Abstract**
3
4
5

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

Eric Crandall 6/20/11 7:37 PM
Deleted: the
Eric Crandall 6/20/11 7:37 PM
Deleted: , a method that compares favorably to the mismatch distribution in predicting the genetic diversity of our datasets.

22 Mean rates estimated with this method for mitochondrial coding genes in three
23 invertebrate species are elevated in comparison to older calibration points (2.3% - 6.6%
24 per lineage per million years), lending additional support to the hypothesis of calibration
25 time-dependency for molecular rates.
26
27
28
29
30
31
32
33
34

1
2 1 Introduction
3
4

5 2 The observation that genetic distances between taxa are correlated with the amount of
6 time since they diverged first gave rise to the idea that DNA may evolve at a relatively
7 constant rate: the molecular clock (Zuckerkandl and Pauling 1965). Like radiometric
8 dating, molecular clocks have allowed illumination of life's evolutionary history, but not
9 without a good deal of scientific controversy (see reviews in Arbogast et al. 2002; Smith
10 and Peterson 2002; Bromham and Penny 2003; Takahata 2007). The original observation
11 of rate constancy led directly to the neutral theory of molecular evolution, which initially
12 predicted that the rate of evolutionary change, k , would be equal to the neutral mutation
13 rate μ , regardless of the effective population size (Kimura 1968). It wasn't long, however,
14 before heterogeneity in amino acid substitution rates was observed (Kimura and Ohta
15 1971; Langley and Fitch 1974), and the neutral theory was soon amended to account for
16 the effects of effective population size (N_e) and slightly deleterious selection ("nearly
17 neutral theory"; Ohta 1972; Tachida 1991). Under the nearly neutral theory, the
18 substitution rate scales inversely with N_e , so it is possible for a large number of alleles
19 with neutral and slightly deleterious mutations to remain transient in a large population
20 for relatively long periods of time before slowly being fixed or removed by genetic drift
21 or purifying selection.

Eric Crandall 6/20/11 7:37 PM
Deleted: for neutral mutations

22 19 The event that is typically used for a molecular clock calibration is a divergence of
23 known age between two genetically distinct lineages. However, as noted above, at the
24 time of actual population divergence, the gene of interest was likely represented in the
25 ancestral population by a number of divergent alleles. Thus, divergence in the gene of
26 interest will predate the actual population divergence by an average of $2N_e$ generations

Eric Crandall 6/20/11 7:37 PM
Deleted: population

1
2 1 (Nei and Li 1979; Edwards and Beerli 2000). Because gene divergence is more easily
3 measured than population divergence, the two measures are often conflated, leading to
4 overestimates of population divergence times, and therefore of molecular substitution
5 rates (Arbogast et al. 2002). This effect is magnified by larger effective population sizes,
6 and more recent population divergence times (Peterson and Masel 2009). It is only
7 recently that divergence-dating methods have been able to explicitly consider the effects
8 of ancestral polymorphism under a coalescent model (Edwards and Beerli 2000; Nielsen
9 and Wakeley 2001; Hickerson et al. 2003).

17
18 9 With increased availability of radiometrically-dated ancient DNA (aDNA) from sub-
19 fossil tissues, it has become possible to calibrate molecular clocks using radiometric dates
20

22 11 assigned to lineages that terminate at some known time in the past (Drummond et al.
23 2002; Ho et al. 2007a). In a coalescent framework, this method circumvents the problem
24 of ancestral polymorphism by essentially sampling directly from the ancestral population
25
26

Eric Crandall 6/20/11 7:37 PM
Deleted: nodes

27 13 (but see Navascues and Emerson 2009 for potential problems with this method). In 2005,
28

29 15 Ho, and co-workers used coalescent rate calibrations from contemporary and ancient
30 DNA to show that there is an exponential relationship between the date used for
31 calibration and the resultant rate, with high rates for recent calibration dates that decline
32 to more familiar substitution rates for dates over 1-2 million years ago (the “lazy J”,
33 Penny 2005), a result that supported previous observations from pedigree studies in
34 humans (Parsons et al. 1997; Howell et al. 2003). In addition to avoiding the problem of
35

Eric Crandall 6/20/11 7:37 PM
Deleted: . et al.
Eric Crandall 6/20/11 7:37 PM
Deleted: a

36 17 ancestral polymorphism through coalescent models, their analysis suggests that this
37 inverse relationship cannot be entirely explained by errors in calibration or sequencing,
38 nor by saturation at fast-evolving sites (but see debate in Emerson 2007; Ho et al. 2007b).
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 1 Instead they suggest that this time-dependent relationship between apparent molecular
3 2 rate and calibration time results from the prolonged action of purifying selection on
4 3 slightly deleterious mutations.

Eric Crandall 6/20/11 7:37 PM
Deleted: slower than expected

5 4 This hypothesis of time-dependency for molecular rates may have far-reaching
6 5 consequences for studies of genetic variation within populations and species, as
7 6 substitution rates from relatively old calibration points are frequently used to estimate
8 7 divergence time, as well as a number of population genetic parameters. For example, if
9 8 time-dependency occurs, then the application of rates from old (> 2 mya) calibrations to
10 9 analyses of recent timescales will result in the consistent overestimates of divergence
11 10 times and effective population sizes, and consistent underestimates of the proportion of
12 11 migrants in structured populations (Ho et al. 2008). However, further inquiry into this
13 12 hypothesis of time-dependency is hindered by a lack of recent calibration points (Ho et
14 13 al. 2007b), as aDNA samples are rare and pedigree studies can be prohibitively lengthy.
15 14 In particular, we are not aware of any recent (< 1 mya) calibration points for marine
16 15 fishes and invertebrates (but see Burridge et al. 2008 for an example from freshwater
17 16 fish). Molecular clock calibration in these groups typically relies on vicariance due to the
18 17 rise of the Isthmus of Panama, which led to the closure of the Panamanian Seaway 2.8 –
19 18 3.1 mya (Lessios 2008; Hellberg 2009), although calibrations from fossils and other
20 19 geological events exist as well (e.g. Marko 2002; Wares 2002). Here, we develop a new
21 20 coalescent-based method for the calibration of molecular clock rates that uses a
22 21 population expansion rather than divergence as its calibrating event. We demonstrate the
23 22 method using a recent, well-documented marine population expansion event, and use it to
24 23 test whether time-dependency can be detected in marine invertebrates.

Eric Crandall 6/20/11 7:37 PM
Deleted: are

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

Eric Crandall 6/20/11 7:37 PM

Deleted: 100

Eric Crandall 6/20/11 7:37 PM

Deleted: to

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

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

6 Materials and Methods

7 Data Characterization

8 To ensure that any population expansions detected could be firmly attributed to sea-
9 level rise on the Sunda Shelf following the last glacial maximum, we selected three
10 species for which previous analyses revealed major clades that were largely confined to

11 the Sunda Shelf or to central Indonesia. These species comprised the boring giant clam,
12 *Tridacna crocea* (Mollusca: Bivalvia), whose ‘black clade’ is dominant from South
13 Sumatra to Western Papua (DeBoer et al. 2008), the mantis shrimp *Haptosquilla*
14 *pulchella* (Arthropoda: Malacostraca), whose ‘white clade’ is limited to the Sunda Shelf
15 and the lesser Sunda islands (Barber et al. 2002b), and the chocolate-chip seastar
16 *Protoreaster nodosus* (Echinodermata: Asteroidea), which has a relatively limited range
17 stretching from Sri Lanka to New Caledonia that is dominated by the Sunda and Sahul
18 shelves (Crandall et al. 2008b). All three species are found in shallow lagoonal waters no
19 deeper than 10 meters. We sub-sampled mitochondrial Cytochrome-c Oxidase subunit 1
20 (CO1) sequence data from larger published datasets to only include localities on or very
21 near the Sunda Shelf (Table 1, Figure 1H).

Eric Crandall 6/20/11 7:37 PM

Deleted: We sub-sampled mitochondrial Cytochrome-c Oxidase subunit 1 (CO1) sequence data from larger published datasets (Barber et al. 2002b; Crandall et al. 2008b; DeBoer et al. 2008) of the boring giant clam *Tridacna crocea* (Mollusca: Bivalvia), the mantis shrimp *Haptosquilla pulchella* (Arthropoda: Malacostraca) and the chocolate-chip seastar *Protoreaster nodosus* (Echinodermata: Asteroidea),

Eric Crandall 6/20/11 7:37 PM

Deleted:

1 We used DNAsp 5.10 (Librado and Rozas 2009) to characterize these datasets for
2 standard population genetic statistics, including D* and F_S, which identify departures
3 from neutrality due to an excess of recent mutations (Fu and Li 1993; Fu 1997), with
4 significance for these neutrality tests determined with 10,000 coalescent simulations. For
5 datasets where more than one locality was sampled, we also measured Φ_{ST} among
6 localities using the AMOVA algorithm implemented in Arlequin 3.11 (Excoffier et al.
7 2005), and only took calibrations from datasets for which Φ_{ST} = 0. We ascertained the
8 best-fitting model of molecular evolution using the Akaike information criterion (AIC) in
9 ModelTest 3.7 (Posada and Crandall 1998) and PAUP* 4.0 (Swofford 2002). Finally, we
10 constructed statistical parsimony networks in TCS 1.21 (Clement et al. 2000), using a
11 95% connection limit.

12 Calibration Points and Demographic Analysis

13 Calibration of a molecular clock based on population expansion requires geological
14 dates for when the expansion began. We used two dates taken from studies of
15 radiocarbon dates for sediment cores and littoral organics (Geyh et al. 1979; Hesp et al.
16 1998; Hanebuth et al. 2000; summarized in Sathiamurthy and Voris 2006; Hanebuth et al.
17 2009). The first date, 19.6 kya, reflects the earliest time following the LGM that sea level
18 rise first became statistically distinguishable from the 2m tidal range, during a rise of
19 ~10m over 800 years (Hanebuth et al. 2009). However, sea level rise was gradual at first
20 (0.41m/100 years after this pulse) and did not result in the significant flooding of the
21 Sunda Shelf (Figure 1). We therefore used another calibration point, at 14.58 kya
22 (corresponding to the Bølling interstadial period), during which sea level rose at an
23 average of 5.33m/100 years and flooding of the Sunda Shelf began in earnest.

1 1 In addition to the information from the sources noted above, we estimated the area of
2 2 shallow water habitat (0-10m) that became available to these species with each 10m
3 3 increase in sea level using spatially gridded data from the ETOPO1 1 arc-minute Global
4 4 Relief Model (Amante and Eakins 2009). Using the spatial analyst toolkit in ArcMap
5 5 10.0, we quantified the number of cells between 10-meter isobaths and multiplied this
6 6 value by a cell size 3.16 km², which resulted from projecting the data to the ARC
7 7 coordinate system, zone 1.

16 8 The calibration also requires accurate estimates of the mutational depth in the
17 9 genealogy at which the transition to expansion growth began, together with an
18 10 assessment of the associated error. Here, we use two different methods to make this
19 11 estimate: the mismatch distribution (Rogers and Harpending 1992) and a two-epoch
20 12 coalescent model (Shapiro et al. 2004).

26 13 The mismatch distribution is the distribution of pairwise differences among
27 14 haplotypes, which Rogers and Harpending (1992) observed to be unimodal after
28 15 exponential population growth in a single deme. They described this with an analytical
29 16 model for which the modal value, $\tau = 2\mu t$ estimates the time of population expansion t in
30 17 terms of the mutation rate μ . Schneider and Excoffier (1999) amended this “sudden
31 18 expansion” method for a finite-sites model with rate heterogeneity, and fitted each
32 19 parameter using a least-squares approach. We analyzed our data under this model in
33 20 Arlequin 3.11 (Excoffier et al. 2005), and established confidence intervals for τ with
34 21 10,000 parametric bootstraps of the analytical model. Model fit was evaluated using the
35 22 sum of squared deviations (SSD), with significance established with the same
36 23 bootstrapped dataset. We calculated lineage mutation rates as $\mu = \tau / 2c$, where c is one of

1 1 the two calibration points mentioned above, and τ is scaled by the number of sites in the
2 2 dataset.

3 3 To get an overall image of the demographic history information that was available in
4 4 each genetic dataset, we analyzed them each under a Bayesian skyline model (Strimmer
5 5 and Pybus 2001; Drummond et al. 2005) implemented in BEAST 1.5.3 (Drummond and
6 6 Rambaut 2007). Under this model a large sample of possible coalescent genealogies are
7 7 broken into piecewise intervals and effective population size (N_e) is estimated at each
8 8 interval from the number of observed coalescent events. The resultant Bayesian skyline
9 9 plot (BSP) makes few *a priori* assumptions about the historical demographic trajectory of
10 10 the population, and can thus provide a framework for constructing more specific models.

11 11 For these analyses we used ten piecewise linear intervals and the molecular evolution
12 12 model from ModelTest. We used uniform, uninformative priors for the population size at
13 13 each interval, and a gamma prior for κ , the transition:transversion ratio. Each skyline
14 14 analysis was run, at minimum, three times for thirty million steps. We assessed
15 15 convergence with estimates of effective sample size (ESS) in Tracer and by comparing
16 16 the marginal posterior distributions for each parameter among runs. Our criterion was
17 17 ESS > 200 as indicated in the BEAST manual. Finally, we combined the logged
18 18 parameter values and trees from replicate runs using LogCombiner 1.4.9, and used Tracer
19 19 to create a Bayesian skyline plot for each dataset.

20 20 The BSP for each dataset indicated a period of constant population size, followed by
21 21 rapid growth that slowed as it reached the present. We therefore used a two-epoch
22 22 coalescent model (Shapiro et al. 2004), implemented in BEAST, that simulated either
23 23 two-parameter exponential growth (Θ_1 , and intrinsic growth rate, r), or three-parameter

1 1 logistic growth (Θ_1 , r, and time to reach $\frac{1}{2} \Theta_1$, t_{50}), preceded by one-parameter constant
2 2 growth (Θ_0), with a final parameter for the transition time between the two epochs
3 3 ($t_{\text{transition}}$). We used the same molecular evolution models that we had used for the skyline
4 4 plots, and $1/x$ priors on all parameters (Drummond et al. 2002) except for κ , for which we
5 5 used a gamma-distributed prior, and r, for which we used a simple uniform prior. We set
6 6 the upper and lower limits of the prior distribution for each parameter using the 95%
7 7 confidence intervals (CI) from the BSPs as guidelines. For example, upper limits for
8 8 $t_{\text{transition}}$ (the parameter of interest) were the upper limit of the 95% CI for T_{MRCA} from the
9 9 skyline model. The lower limit for $t_{\text{transition}}$ was set to 10^{-6} mutations per site, which
10 10 represents a prior assumption that rates will not be lower than 0.05% /my (about twenty
11 11 times slower than the 1% /my rate commonly used for mtDNA; Brown et al. 1979). We
12 12 calculated lineage mutation rates as $\mu = t_{\text{transition}} / c$.

Eric Crandall 6/20/11 7:37 PM

Deleted: uniform

Eric Crandall 6/20/11 7:37 PM

Deleted:

13 To more rigorously test the hypothesis of constant population size followed by
14 logistic growth, we ran each dataset under a model of constant population size, two
15 models of population growth (exponential and logistic) without a stable period preceding
16 them, and a model of expansion growth (constant size followed by exponential growth)
17 that does not allow the time of expansion to vary. Where applicable, we used the same
18 priors as were used in two-epoch models. We then used Bayes factors to compare the
19 harmonic mean of the marginal likelihood from these models to those from the
20 exponential and logistic two-epoch models described above. Marginal likelihoods for
21 each model were calculated as the sum of the log-likelihoods for the genealogy and the
22 coalescent model at each recorded step (thus, the product of these two quantities), and the
23 harmonic mean for each was calculated in Tracer, using 1000 bootstrap replicates to

Eric Crandall 6/20/11 7:37 PM

Deleted: and used Bayes Factors to compare the harmonic mean of the marginal likelihood from this

1 establish confidence intervals. We report calibrations from the model that received the
2 most support from this method following recommendations from Kass and Raftery
3 (1995).

4 **Results**

5 We found a more than ten-fold increase in the amount of shallow-water habitat (0-
6 10m) that became available to these species as a result of sea-level rise following the
7 LGM (Figure 1). Shallow-water habitat on the shelf went from 30,050 km² at the
8 lowstand to 358,680 km² in present sea level.

9 Summary statistics for each genetic dataset are given in Table 1. Fu's F_S was
10 negative, but not significantly so for the *Tridacna crocea* dataset. *Haptosquilla pulchella*
11 and *Protoreaster nodosus* had significantly negative values for F_S. The *H. pulchella*
12 dataset comprising multiple sampling localities had a significant Φ_{ST} value, reflecting
13 genetic structure between Pulau Seribu and the other sites. Φ_{ST} was zero once samples
14 from Pulau Seribu were removed from the dataset. There was no evidence of any
15 structure among pooled localities for *T. crocea*. Modeltest selected an HKY model of
16 molecular evolution for all datasets, with the addition of a parameter for invariant sites (I)
17 in *H. pulchella*. Statistical parsimony networks from TCS showed numerous star-like
18 polytomies that are diagnostic of population expansions (Supplemental Figure 1; Slatkin
19 and Hudson 1991)

20 Mismatch distributions were generally unimodal for *Protoreaster nodosus* CO1 and
21 *Haptosquilla pulchella* CO1, but bimodal for *Tridacna crocea* CO1 (Figure 2). The
22 bimodal distribution for *T. crocea* resulted in very large estimates of τ, resulting in

1 1 exceptionally high values for μ (Table 2). None of the datasets rejected a sudden
2 2 expansion model.
3
4 3 Bayesian skyline models (Figure 3) for all datasets converged to the same posterior
5 4 distribution, as demonstrated by effective sample size values >200 and agreement among
6 5 multiple runs. Model runs for each different demographic scenario (constant size through
7 6 two-epoch with logistic growth) also converged well for all datasets, with effective
8 7 sample sizes generally much greater than 200, and agreement across multiple runs. The
9 8 two-epoch model with logistical growth consistently had the strongest Bayes Factor
10 9 support, beating out a similar model with exponential growth for all three species, as well
11 10 as three simpler models of population growth (exponential, logistic, and expansion); Table
12 11 3). This model was also decisively better than a constant population size model for *H.*
13 12 *pulchella* and *P. nodosus*, but was only weakly supported over a constant population size
14 13 for *T. crocea*.

Eric Crandall 6/20/11 7:37 PM

Deleted: Constant and

Eric Crandall 6/20/11 7:37 PM

Deleted: models

Eric Crandall 6/20/11 7:37 PM

Deleted: strong

Eric Crandall 6/20/11 7:37 PM

Deleted: (

14 Mean values for the time of transition between constant population size and logistic
15 growth were generally a bit lower than the inflection point depicted in the skyline model
16 for each gene (Figure 3). However, the 95% confidence intervals for $t_{\text{transition}}$ generally
17 encompassed the period of growth detected by the skyline model, and fit within the very
18 large confidence intervals generated for the mismatch distributions (Table 2). For the
19 conservative calibration point at 19.60 kya, mean estimates of μ for *T. crocea* and *P.*
20 *nodosus* were very similar (2.30%/my and 2.61%/my respectively), while *H. pulchella*
21 had a mean rate that was over twice as fast (6.58%/my; Table 2). Posterior distributions
22 for $t_{\text{transition}}$, Θ_0 and Θ_1 were unimodal, while those for growth rate (r) and t_{50} were mostly
23 uninformative. With the exception of the *H. pulchella* samples from Pulau Seribu, which

1 1 showed little evidence of population growth, two-epoch models for individual sampling
2 2 localities showed strong concordance with the pooled-locality datasets in *T. crocea* and
3 3 *H. pulchella*, albeit with larger confidence intervals (results not shown).

4 4 **Discussion**

5 5 Traditional molecular clock calibrations have relied almost exclusively on vicariant
6 6 events or fossil calibrations, both of which have limited applicability in most marine taxa,
7 7 particularly for recent timescales. Our use of a well-documented, recent population
8 8 expansion instead of a divergence allows us to estimate a molecular rate of change at a
9 9 relatively recent timescale, and, in combination with an explicit coalescent model of
10 10 | constant population size followed by expansion, allows us to account for polymorphisms
11 11 that predate the expansion (Peterson and Masel 2009). Point estimates from both
12 12 mismatch analysis and the two-epoch coalescent model resulted in lineage mutation rates
13 13 that are much higher than the 1%/my (= 2%/my divergence rate) commonly assumed for
14 14 many marine phylogeographic studies, although our confidence intervals often include
15 15 these lower values.

16 16 Allowing for higher rates in future phylogeographic inference may often bring it into
17 17 line with a region's recent palaeoclimate (see Ho et al. 2008). For example, divergence
18 18 dates across the Maluku Sea between populations of *H. pulchella*, previously estimated to
19 19 be about 470,000 years ago using a rate of 1.4%/my, (Barber et al. 2006), can now be
20 20 estimated at about 100,000 years ago. This is much closer to the last time that sea levels
21 21 | approached current levels 120,000 years ago, before dropping again (Chappell et al.
22 22 1996). In the following discussion, we first compare estimates from the sudden expansion

Eric Crandall 6/20/11 7:37 PM
Deleted: about 110,000 years ago

1 and two-epoch models. We then discuss potential sources of error in our calibration and
2 compare our calibrated rates to rates from older calibrations.
3

4 Rate calibrations from the spatial expansion mismatch analyses were generally much
5 higher than those from the two-epoch coalescent model (Table 2), although very wide
6 confidence intervals on τ always included the two-epoch estimate. Of the two, we favor
7 the two-epoch method, which takes advantage of the genealogical information in the data
8 through Bayesian parameter estimation from an explicit coalescent model. In contrast, the
9 mismatch analyses used here provide only an analytical approximation of substitution
10 patterns expected from a spatial expansion, without considering any underlying
11 genealogy, and so likely yield less precise estimations. As a case in point, the mismatch
12 distributions for *T. crocea* were bimodal due to several star polytomies in the genealogies,
13 ([supplementary figure 1](#)). A bimodal distribution tends to increase the estimate of τ
14 (relative to unimodal distributions), and therefore of the substitution rate. However, in a
15 time-reversed coalescent framework, these polytomies can be interpreted as a potentially
16 simultaneous increase in the rate of coalescence.

Eric Crandall 6/20/11 7:37 PM
Deleted: accurate

Eric Crandall 6/20/11 7:37 PM
Deleted: .

32 Sources of Error

33 Our use of a two-epoch coalescent model (Shapiro et al. 2004) as a way to more
34 accurately calibrate substitution rates is new, and uses an intra-specific process
35 (population expansion) to make inferences for phylogeographic timescales, in contrast to
36 inter-lineage divergence methods that have been used previously. As such, our calibration
37 method is subject to some of the same sources of error and bias as divergence methods,
38 while introducing new ones, and avoiding others (Arbogast et al. 2002). First and

Eric Crandall 6/20/11 7:37 PM
Deleted: species

1 foremost, calibration points, as independent estimates of elapsed time, are central to any
2 scheme for separating the effects of substitution rate and time, and are probably the
3 largest sources of error in molecular clocks (Benton and Ayala 2003). Here, we used two
4 possible points for the beginning of population expansion: one at 19.6 kya and the other
5 at 14.6 kya, as a way to provide some assessment of the potential for calibration error.
6 For the sake of discussion, we use rates from the more conservative of the two dates (19.6
7 kya), but the 14.6 kya date may be more accurate, as this was the time that sea level rise
8 first resulted in significant flooding of the Sunda Shelf (Figure 1). We are also making
9 the assumption that coral reef communities colonized the Sunda Shelf as soon as new
10 shallow marine habitat was available. This is justifiable, because coral reef communities
11 have shown the potential for rapid, long-distance re-colonization of coastal waters, with
12 species richness, coral cover and genetic diversities returning to their ambient levels less
13 than 150 years after being obliterated by volcanic eruptions (Tomascik et al. 1996; Barber
14 et al. 2002a; Starger et al. 2010). Nevertheless, even the 14.6 kya date may be
15 conservative, as the Sunda Shelf environment may not have been immediately
16 appropriate for coral reef development following inundation of terrestrial habitats. Thus,
17 as we learn more about re-colonization of the Sunda Shelf, it may become apparent that
18 the above rates are actually conservative estimates and we may need to revise the
19 estimated rates upward.

20 A related assumption is that populations of all three of these lagoonal species
21 expanded onto the Sunda Shelf somewhat simultaneously. This is analogous to the
22 assumption of simultaneity that has often been made in divergence dating across multiple
23 taxa (but see Hickerson et al. 2006 et al. for a statistical test of simultaneous divergence),

1 and rests on the idea that coral reef and their associated lagoonal communities would
2 move concurrently onto the Sunda Shelf (Pandolfi and Jackson 2001; Tager et al. 2010).
3 While no tests currently exist for simultaneity in expansion, the BSPs provide a sketch of
4 demographic history for each species back to the most recent common mitochondrial
5 ancestor, and have the potential to detect multiple demographic fluctuations (e.g. figure 4
6 in Crandall et al. 2008a), which are not evident for the present datasets. However,
7 although the fact of recent marine population expansions onto the Sunda Shelf is not in
8 doubt, the wide confidence intervals and weak support for a two-epoch model of
9 logistical growth in *T. crocea* suggests that this species may have expanded later or less
10 rapidly than the other two species. More rigorous tests of this assumption await data from
11 multiple loci (Heled and Drummond 2008).

12 By setting priors of the two-epoch model to confidence intervals for the Bayesian
13 skyline plots, we essentially estimated expansion times that fit the BSPs (Figure 3). This
14 approach offers a number of advantages. First, the resulting “fitted” two-epoch models
15 were significantly better at explaining the data than simpler models of population growth
16 (Table 3). Second, in using these models we have already accounted for the effects of
17 population size change that are often neglected in divergence calibrations (Arbogast et al.

Eric Crandall 6/20/11 7:37 PM
Deleted: Additional calibration error can arise more directly from our two-epoch coalescent estimates of the genealogical depth of the expansion.

Eric Crandall 6/20/11 7:37 PM
Deleted: In so doing

18 2002; Navascues and Emerson 2009). Third, by calibrating from an intra-specific
19 process, we avoid the potential errors arising from phylogenetic divergence
20 methodologies, such as problems with rooting, branching order and missing or extinct
21 taxa (Smith and Peterson 2002). Finally, the 95% confidence intervals given in Table 2
22 should largely account for error due to Poisson-distributed mutations and the stochasticity
23 of the coalescent.

Eric Crandall 6/20/11 7:37 PM
Deleted: Moreover

Eric Crandall 6/20/11 7:37 PM
Deleted: The

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

8 Because genes on the mitochondrial genome are strongly linked, an advantageous
9 variant in one gene could potentially sweep to fixation through positive selection,
10 bringing all variation on its particular genome with it (Maynard-Smith and Haigh 1974).

Eric Crandall 6/20/11 7:37 PM
Deleted: linked haplotypes

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

1 1 The process of purifying selection is conceptually more difficult to distinguish from
2 demographic growth, since it is implicated in creating the time-dependency effect, by
3 acting over longer periods of time than previously expected (Ho et al. 2005; Ho et al.
4 2007b). It is important to note that purifying selection is by no means limited to non-
5 synonymous mutations. Through a comparison with pseudogenes, Ophir et al. (1999)
6 found that an average of 75% of substitutions are non-neutral, roughly twice what is
7 estimated by D_n/D_s ratios. Many synonymous changes are under very weak selective
8 pressures, as arise from translational selection, or maintenance of nucleotide
9 compositions in the face of biased mutational input (Montooth and Rand 2008). Thus, if
10 purifying selection is occurring at the same timescales at which we are measuring
11 population growth, it may be difficult to tease the two apart, because it can be difficult to
12 ascertain exactly how purifying selection would affect genetic variation in a reversed-
13 time coalescent framework. Using a forward-time model, Peterson and Masel (2009)
14 confirm that purifying selection on slightly deleterious variation leads to a period of
15 elevation in the substitution rate at recent timescales, the length of which is increased by
16 large N_e and a highly leptokurtic distribution of selection coefficients, as would be
17 expected under nearly-neutral theory (Keightley and Eyre-Walker 2007).

Eric Crandall 6/20/11 7:37 PM
Deleted: things

18 Fortunately, purifying selection is expected to leave only a minor imprint on the
19 shape of the genealogy, while having a similar effect as demographic growth or positive
20 selection on the distribution of substitutions (shifting them towards the tips of the
21 genealogy; Williamson and Orive 2002). If purifying selection were the only process
22 occurring, we would expect to see it act more or less constantly throughout the history of
23 the mitochondrial genome, following a simple model of exponential or logistic growth

Eric Crandall 6/20/11 7:37 PM
Deleted: If purifying selection were the only process occurring, we would not expect the star-shaped polytomies that we see in each genealogy (Supplemental Figure 1), indicating a sudden increase in the rate of coalescence, and inferred to be population growth in a Bayesian skyline context (Figure 3).

1 1 (e.g. Seger et al. 2010). Instead, as was initially indicated by the BSPs, two-epoch models
2 2 of constant population size followed by a pulsed increase in the rate of coalescence are
3 3 consistently favored by odds of more than 300:1 over these simpler growth models
4 4 (Table 3).

5 5 The different topologies expected under purifying selection and demographic growth
6 6 (Williamson and Orive 2002) may also be why Fu (1997) found that D* (which is based
7 7 on distribution of substitutions) was more sensitive to purifying selection than F_S (which
8 8 is based on haplotype frequencies), and vice versa. That Fu's F_S was significant in many
9 9 more instances than D* (Table 1) also suggests that demographic change is more
10 10 important than purifying selection in creating the observed patterns. Nevertheless, it will
11 11 be important to account for sites under purifying selection when estimating rates from
12 12 population expansions, perhaps by including rate decay parameters in coalescent
13 13 genealogy sampling schemes (O'Fallon 2010).

14 Rate Comparisons

15 Perhaps the most surprising result of calibrations from the two-epoch model is the
16 disparity in substitution rate estimates. While *T. crocea* and *P. nodosus* have similar rates
17 (mean values of 2.3% to 2.6% per million years), the substitution rates in *H. pulchella*
18 appear ~~twice as fast~~ (6.6% per million years). It is possible that this disparity results from
19 non-simultaneous dates of expansion, with *H. pulchella* experiencing an expansion event
20 much earlier than the Sunda Shelf flooding, or *T. crocea* and *P. nodosus* expanding onto
21 the Sunda Shelf about 10,000 years after *H. pulchella*. A simpler explanation for the rate
22 differences among the three invertebrates can be found in their times to first reproduction,

Eric Crandall 6/20/11 7:37 PM
Deleted: to be moving over

Eric Crandall 6/20/11 7:37 PM
Deleted: did.

1 which we use as a conservative proxy for generation time. *T. crocea* and *P. nodosus* each
2 take at least two years to reach reproductive maturity (Lucas 1988; Bos et al. 2008),
3 whereas *H. pulchella* can reach reproductive maturity within one year (M.V. Erdmann,
4 pers. comm.; Erdmann 1997). Thus, since *H. pulchella* replicates its germline more than
5 twice as often as the two larger invertebrate species, the per-year rate disparities among
6 the invertebrates can be easily reconciled in a per-generation context. A generation-time
7 effect that is independent of body size and other correlates has recently been confirmed
8 for invertebrates in general (Thomas et al. 2010).

18 Following correction for a generation-time effect, and removal of the calibration for
19 *T. crocea* due to weak support for the two-epoch model, our mean estimates of CO1
20 lineage substitution rates ranged from 5.2% - 6.6% /million generations for two marine
21 invertebrate species. Per-generation lineage rates calibrated from the Isthmus of Panama
22 that properly account for ancestral polymorphism are lower than this (1.4 – 3.2%,
23 Hickerson et al. 2003; Hickerson et al. 2006), while fossil calibrated rates for marine
24 invertebrates are still lower (0.5 – 1.2%, Marko 2002; Frey and Vermeij 2008; Malaquias
25 and Reid 2009). Using least-squares regression, this decline in mean rates with
26 calibration time (Figure 4) can be described as an exponential decay of rates over time
27 (Ho et al. 2005):

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

38 Under this relationship, marine invertebrate CO1 has an instantaneous mutation rate of
39 40 41 42 43 44 45

42 43 44 45

Eric Crandall 6/20/11 7:37 PM
Deleted: 2.8

Eric Crandall 6/20/11 7:37 PM
Deleted: 054e
Eric Crandall 6/20/11 7:37 PM
Deleted: 30r

Eric Crandall 6/20/11 7:37 PM
Deleted: 4

1 the exponential term) than has been found in birds, mammals, or freshwater fish (Ho et
2 al. 2005; Burridge et al. 2008).

Eric Crandall 6/20/11 7:37 PM
Deleted: While the wide confidence intervals
on recent rate estimates

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

Eric Crandall 6/20/11 7:37 PM
Deleted:

10 Conclusions

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

21 Continued discovery of rate heterogeneity has nearly led to the abandonment of a
22 global clock for any given region of the genome (Bromham and Penny 2003) in favor of

1 local clocks that are particular to certain taxa (Yoder and Yang 2000) or relaxed clock
2 models that allow the rate of evolution to vary across the phylogeny (Aris-Brosou and
3 Yang 2002; Drummond et al. 2006). A wide variety of explanations for the observed rate
4 variation have been offered, including the correlated effects of generation time (Laird et
5 al. 1969), metabolic rate (Martin and Palumbi 1993; Gillooly et al. 2005), and DNA
6 repair mechanisms (Li et al. 1996). Our results, together with other results showing time-
7 dependent rates (Ho et al. 2005; Ho et al. 2007a; Saarma et al. 2007; Burridge et al. 2008;
8 Subramanian et al. 2009) renew support for one of the original explanations for rate
9 heterogeneity: the combined effect of weak negative selection and effective population
10 size ($N_e\sigma_s$) proposed under the nearly neutral theory of molecular evolution (Ohta 1972).
11 Specifically, these combined results suggest that the observed rate of molecular evolution
12 decays as a function of the calibration time, and that the rate of decay is a function of
13 effective population size. With additional data from more species and more genes, we
14 expect that our method of expansion dating will provide further insight into rates of
15 molecular change at recent timescales.

31 16 **Supplementary Material**

32 17 Statistical parsimony networks for each species are available in Supplementary Figure 1.
33

34 18 **Acknowledgements**

35 19 Impetus for this project came from discussions in a population genetics graduate seminar,
36 20 led by M. Sorenson, at Boston University in 2007. It was developed while EDC was a
37 21 postdoctoral associate with Old Dominion University, and a visiting scholar at the Marine
38 22 Sciences Institute of the University of the Philippines. This work is part of the Coral
39

1 1 Triangle Partnerships for International Research and Education (CT-PIRE), funded by the
2 2 National Science Foundation – OISE-0730256. Collection of DNA sequence data used in
3 3 this study was supported by NSF grant OCE-0349177. We would like to acknowledge
4 4 our Indonesian partners: Udayana University for hosting our project, as well as the
5 5 Indonesian Institute of Sciences (LIPI) and the Indonesian Ministry of Science and
6 6 Technology (RISTEK) for providing the necessary research permits. We thank N.
7 7 Mahardika, H. Toha and Ambariyanto for their sponsorship and logistical support. Some
8 8 BEAST analyses were carried out by using the online resources of the Computational
9 9 Biology Service Unit from Cornell University, which is partially funded by Microsoft
10 10 Corporation. EDC thanks E. Anderson for statistical advice, and S. Gulamhussein for her
11 11 patience and support.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 1 Literature Cited

2 2 Amante, C. and B. W. Eakins. 2009.ETOPO1 1 Arc-Minute Global Relief Model:
3 3 Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS
4 4 NGDC-24, 19 pp.

Eric Crandall 6/20/11 7:37 PM
Deleted: -

- 5 5 Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002.
6 6 Estimating divergence times from molecular data on phylogenetic and population
7 7 genetic timescales. *Annu Rev Ecol Syst.* **33**:707-740.
8 8 Aris-Brosou, S., and Z. H. Yang. 2002. Effects of models of rate evolution on estimation
9 9 of divergence dates with special reference to the metazoan 18S ribosomal RNA
10 10 Phylogeny. *Syst Biol.* **51**:703-714.
11 11 Barber, P. H., M. V. Erdmann, and S. R. Palumbi. 2006. Comparative phylogeography of
12 12 three co-distributed stomatopods: origins and timing of regional lineage
13 13 diversification in the coral triangle. *Evolution* **60**:1825-1839.
14 14 Barber, P. H., M. K. Moosa, and S. R. Palumbi. 2002a. Rapid recovery of genetic
15 15 populations on Krakatau: diversity of stomatopod temporal and spatial scales of
16 16 marine larval dispersal. *Proc R Soc B* **269**:1591-1597.
17 17 Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2002b. Sharp genetic
18 18 breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate
19 19 limits to larval transport: patterns, causes, and consequences. *Mol Ecol.* **11**:659-
20 20 674.
21 21 Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence
22 22 mitochondrial genetic diversity in animals. *Science* **312**:570-572.
23 23 Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. *Science* **300**:1698-1700.

- 1 1 Bos, A. R., G. S. Gumanao, J. C. E. Alipoyo, and L. T. Cardona. 2008. Population
2 2 dynamics, reproduction and growth of the Indo-Pacific horned sea star,
3 3 *Protoreaster nodosus* (Echinodermata: Asteroidea). Mar Biol. **156**:55-63.
- 4 4 Bromham, L., and D. Penny. 2003. The modern molecular clock. Nat Rev Gen. **4**:216-
5 5 224.
- 6 6 Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid evolution of animal
7 7 mitochondrial DNA. P Natl Acad Sci USA **76**:1967-1971.
- 8 8 Burridge, C. P., D. Craw, D. Fletcher, and J. M. Waters. 2008. Geological dates and
9 9 molecular rates: Fish DNA sheds light on time dependency. Mol Biol Evol.
10 10 **25**:624-633.
- 11 11 Chappell, J., A. Omura, T. Esat, M. McCulloch, J. Pandolfi, Y. Ota, and B. Pillans. 1996.
12 12 Reconciliation of late Quaternary sea levels derived from coral terraces at Huon
13 13 Peninsula with deep sea oxygen isotope records. Earth Planet Sc Lett. **141**:227-
14 14 236.
- 15 15 Chenoweth, S. F., and J. M. Hughes. 2003. Oceanic interchange and nonequilibrium
16 16 population structure in the estuarine dependent Indo-Pacific tasselfish, *Polynemus*
17 17 *sheridani*. Mol Ecol. **12**:2387-2397.
- 18 18 Crandall, E. D., M. A. Frey, R. K. Grosberg, and P. H. Barber. 2008a. Contrasting
19 19 demographic history and phylogeographical patterns in two Indo-Pacific
20 20 gastropods. Mol Ecol. **17**:611-626.
- 21 21 Crandall, E. D., M. E. Jones, M. M. Muñoz, B. Akinronbi, M. V. Erdmann, and P. H.
22 22 Barber. 2008b. Comparative phylogeography of two seastars and their
23 23 ectosymbionts within the Coral Triangle. Mol Ecol. **17**:5276-5290.

- 1
2 1 DeBoer, T. S., M. D. Subia, Ambariyanto, M. V. Erdmann, K. Kovitvongsa, and P. H.
3
4 2 Barber. 2008. Phylogeography and Limited Genetic Connectivity in the
5
6 3 Endangered Boring Giant Clam across the Coral Triangle. *Conserv Biol.* **22**:1255-
7
8 4 1266.
9
10 5 Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics
11 6 and dating with confidence. *PLOS Biol.* **4**:e88.
12
13 7 Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating
14 8 mutation parameters, population history and genealogy simultaneously from
15
16 9 temporally spaced sequence data. *Genetics*. **161**:1307-1320.
17
18 10 Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by
19
20 11 sampling trees. *BMC Evol Biol.* **7**:214.
21
22 12 Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent
23
24 13 inference of past population dynamics from molecular sequences. *Mol Biol Evol.*
25
26 14 **22**:1185-1192.
27
28
29 15 Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population
30
31 16 divergence, and the variance in coalescence time in phylogeographic studies.
32
33 17 *Evolution*. **54**:1839-1854.
34
35 18 Emerson, B. C. 2007. Alarm Bells for the Molecular Clock? No Support for Ho et al.'s
36
37 19 Model of Time-Dependent Molecular Rate Estimates. *Syst Biol.* **56**:337.
38
39 20 Erdmann, M. V. 1997. The ecology, distribution and bioindicator potential of Indonesian
40
41 21 coral reef stomatopod communities. pp. 290. University of California, Berkeley.
42
43 22 Excoffier, L., L. G. Laval, and S. Schneider. 2005. Arlequin v.3.0: An integrated software
44
45 23 package for population genetics data analysis. *Evol Bioinform*. **1**:47-50.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1 1 Frey, M. A., and G. J. Vermeij. 2008. Molecular phylogenies and historical biogeography
2 2 of a circumtropical group of gastropods (Genus: *Nerita*): implications for regional
3 3 diversity patterns in the marine tropics. *Mol Phylogenet Evol.* **48**:1067-1086.
4 4 Fu, Y.-X. 1997. Statistical tests of neutrality against population growth, hitchhiking and
5 5 background selection. *Genetics* **147**:915-925.
6 6 Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics*
7 7 **133**:693-809.
8 8 Geyh, M. A., H. R. Kudrass, and H. Streif. 1979. Sea-Level Changes during the Late
9 9 Pleistocene and Holocene in the Strait of Malacca. *Nature* **278**:441-443.
10 10 Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA
11 11 evolution: Effects of body size and temperature on the molecular clock. *P Natl
12 12 Acad Sci USA* **102**:140-145.
13 13 Grant, W. S., and B. Bowen. 1998. Shallow population histories in deep evolutionary
14 14 lineages of marine fishes: Insights from sardines and anchovies and lessons for
15 15 conservation. *J Hered.* **89**:415-426.
16 16 Hanebuth, T., K. Stattegger, and P. M. Grootes. 2000. Rapid flooding of the Sunda Shelf:
17 17 A late-glacial sea-level record. *Science* **288**:1033-1035.
18 18 Hanebuth, T. J. J., K. Stattegger, and A. Bojanowski. 2009. Termination of the Last
19 19 Glacial Maximum sea-level lowstand: The Sunda-Shelf data revisited. *Global
20 20 Planet Change* **66**:76-84.
21 21 Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from
22 22 multiple loci. *BMC Evol Biol.* **8**:289.

- 1
2 1 Hellberg, M. 2009. Gene Flow and Isolation among Populations of Marine Animals.
3
4 2 Annu Rev Ecol Syst. **40**:291-310.
5
6 3 Hesp, P. A., C. C. Hung, M. Hilton, C. L. Ming, and I. M. Turner. 1998. A first tentative
7 4 Holocene sea-level curve for Singapore. J Coastal Res. **14**:308-314.
8
9 5 Hickerson, M. J., M. A. Gilchrist, and N. Takebayashi. 2003. Calibrating a molecular
10 6 clock from phylogeographic data: Moments and likelihood estimators. Evolution
11 7 **57**:2216-2225.
12
13 8 Hickerson, M. J., E. A. Stahl, and H. Lessios. 2006. Test for simultaneous divergence
14 9 using approximate Bayesian computation. Evolution **60**:2435-2453.
15
16 10 Ho, S. Y., U. Saarma, R. Barnett, J. Haile, and B. Shapiro. 2008. The effect of
17 11 inappropriate calibration: three case studies in molecular ecology. PLOS One.
18 12 **3**:e1615.
19
20 13 Ho, S. Y. W., S. Kolokotronis, and R. G. Allaby. 2007a. Elevated substitution rates
21 14 estimated from ancient DNA sequences. Biol Lett. **3**:702-705.
22
23 15 Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of
24 16 molecular rate estimates and systematic overestimation of recent divergence
25 17 times. Mol Biol Evol. **22**:1561-1568.
26
27 18 Ho, S. Y. W., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond. 2007b.
28 19 Evidence for time dependency of molecular rate estimates. Syst Biol. **56**:515-522.
29
30 20 Howell, N., C. Smejkal, D. Mackey, P. Chinnery, D. Turnbull, and C. Herrnstadt. 2003.
31 21 The pedigree rate of sequence divergence in the human mitochondrial genome:
32 22 There is a difference between phylogenetic and pedigree rates. Am J Hum Genet.
33 23 **72**:659-670.

- 1 1 Kass, R. E., and A. E. Raftery. 1995. Bayes Factors. J Am Stat Assoc. **90**:773-795.
- 2 2 Keightley, P. D., and A. Eyre-Walker. 2007. Joint inference of the distribution of fitness
- 3 3 effects of deleterious mutations and population demography based on nucleotide
- 4 4 polymorphism frequencies. Genetics **177**:2251-2261.
- 5 5 Kimura, M. 1968. Evolutionary rate at the molecular level. Nature **217**:624-626.
- 6 6 Kimura, M., and T. Ohta. 1971. On the rate of molecular evolution. J Mol Evol. **1**:1-17.
- 7 7 Laird, C. D., B. L. McConaughay, and B. J. McCarthy. 1969. Rate of fixation of
- 8 8 nucleotide substitutions in evolution. Nature **224**:149-154.
- 9 9 Lambeck, K., T. M. Esat, and E. K. Potter. 2002. Links between climate and sea levels
- 10 10 for the past three million years. Nature **419**:199-206.
- 11 11 Langley, C. H., and W. M. Fitch. 1974. An examination of the constancy of the rate of
- 12 12 molecular evolution. J Mol Evol. **3**:161-177.
- 13 13 Lessios, H. A. 2008. The Great American Schism: Divergence of Marine Organisms
- 14 14 After the Rise of the Central American Isthmus. Annu Rev Ecol Evol Syst. **39**:63-
- 15 15 91.
- 16 16 Li, W. H., D. L. Ellsworth, J. Krushkal, B. H. J. Chang, and D. Hewett-Emmett. 1996.
- 17 17 Rates of nucleotide substitution in primates and rodents and the generation time
- 18 18 effect hypothesis. Mol Phylogenetic Evol. **5**:182-187.
- 19 19 Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of
- 20 20 DNA polymorphism data. Bioinformatics **25**:1451-1452.
- 21 21 Lind, C. E., B. S. Evans, J. J. U. Taylor, and D. R. Jerry. 2007. Population genetics of a
- 22 22 marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago

Eric Crandall 6/20/11 7:37 PM
Formatted: Font:Italic

- 1 1 shows differentiation and decreased diversity at range limits. Mol Ecol. **16**:5193-
2 2 5203.
- 3 3 Lucas, J. S. 1988. Giant Clams: description, distribution, and life history. In: J. W.
4 4 Copland, and J. S. Lucas, editors. Giant clams in Asia and the Pacific. ACIAR,
5 5 Canberra, Australia. pp. 21-32
- 6 6 Malaquias, M., and D. Reid. 2009. Tethyan vicariance, relictualism and speciation:
7 7 evidence from a global molecular phylogeny of the opisthobranch genus *Bulla*. J
8 8 Biogeogr **36**:1760-1777.
- 9 9 Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of
10 10 geminate species pairs separated by the Isthmus of Panama. Mol Biol Evol.
11 11 **19**:2005-2021.
- 12 12 Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and
13 13 the molecular clock. P Natl Acad Sci USA **90**:4087-4091.
- 14 14 Maynard-Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. Gen
15 15 Res. **23**:23-35.
- 16 16 Montooth, K. L., and D. M. Rand. 2008. The spectrum of mitochondrial mutation differs
17 17 across species. PLOS Biol. **6**:1634-1637.
- 18 18 Navascues, M., and B. C. Emerson. 2009. Elevated substitution rate estimates from
19 19 ancient DNA: model violation and bias of Bayesian methods. Mol Ecol. **18**:4390-
20 20 4397.
- 21 21 Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms
22 22 of restriction endonucleases. P Natl Acad Sci USA **76**:5269-5273.

- 1
2 1 Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov
3
4 2 Chain Monte Carlo approach. *Genetics* **158**:885-896.
5
6 3 O'Fallon, B. D. 2010. A method to correct for the effects of purifying selection on
7
8 4 genealogical inference. *Mol Biol Evol.* **27**:2406-2416.
9
10 5 Ohta, T. 1972. Population size and rate of evolution. *J Mol Evol.* **1**:305-314.
11
12 6 Ophir, R., T. Itoh, D. Graur, and T. Gojobori. 1999. A simple method for estimating the
13
14 7 intensity of purifying selection in protein-coding genes. *Mol Biol Evol.* **16**:49-53.
15
16 8 Palumbi, S. R. 1994. Genetic Divergence, Reproductive Isolation, and Marine Speciation.
17
18 9 *Annu Rev Ecol Syst.* **25**:547-572.
19
20 10 Pandolfi, J., and J. Jackson. 2001. Community structure of Pleistocene coral reefs of
21
22 Curacao, Netherlands Antilles. *Ecol Monogr.* **71**:49-67.
23
24 12 Parsons, T., D. Muniec, K. Sullivan et al. 1997. A high observed substitution rate in the
25
26 13 human mitochondrial DNA control region. *Nat Genet.* **15**:363-368.
27
28 14 Penny, D. 2005. Evolutionary biology - Relativity for molecular clocks. *Nature* **436**:183-
29
30 15 184.
31
32 16 Peterson, G. I., and J. Masel. 2009. Quantitative Prediction of Molecular Clock and K-
33
34 17 a/K-s at Short Timescales. *Mol Biol Evol.* **26**:2595-2603.
35
36 18 Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution.
37
38 19 *Bioinformatics* **14**:817-818.
39
40 20 Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the
41
42 21 distribution of pairwise genetic differences. *Mol Biol Evol.* **9**:552-569.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2 1 Saarma, U., S. Y. W. Ho, O. G. Pybus et al. (19 co-authors) 2007. Mitogenetic structure
3
4 2 of brown bears (*Ursus arctos* L.) in northeastern Europe and a new time frame for
5
6 3 the formation of European brown bear lineages. *Mol Ecol.* **16**:401-413.
7
8 4 Sathiamurthy, E., and H. K. Voris. 2006. Maps of Holocene Sea Level Transgression and
9
10 5 Submerged Lakes on the Sunda Shelf. *The Natural History Journal of*
11
12 6 **Chulalongkorn University Supplement** **2**:1-43.
13
14 7 Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from
15
16 8 the distribution of pairwise differences when the mutation rates vary among sites:
17
18 9 Application to human mitochondrial DNA. *Genetics* **152**:1079-1089.
19
20 10 [Seger, J., W. A. Smith, J. J. Perry, J. Hunn, Z. A. Kaliszewska, L. La Sala, L. Pozzi, V. J.](#)
21
22 11 [Rowntree, and F. R. Adler. 2010. Gene Genealogies Strongly Distorted by](#)
23
24 12 [Weakly Interfering Mutations in Constant Environments. Genetics](#) **184**:529-545.
25
26 13 Shapiro, B., A. J. Drummond, A. Rambaut et al. 2004. Rise and fall of the Beringian
27
28 14 steppe bison. *Science* **306**:1561-1565.
29
30 15 Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA
31
32 16 sequences in stable and exponentially growing populations. *Genetics* **123**:603-
33
34 17 613.
35
36 18 Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades:
37
38 19 Molecular clocks and the fossil record. *Ann Rev Earth Planet Sci.* **30**:65-88.
39
40 20 Starger, C., P. Barber, Ambariyanto, and A. Baker. 2010. The recovery of coral genetic
41
42 21 diversity in the Sunda Strait following the 1883 eruption of Krakatau. *Coral Reefs*
43
44 22 **29**:547-565.

Eric Crandall 6/20/11 7:37 PM
Formatted: Font:Italic

- 1 1 Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA
2 2 sequences using the generalized skyline plot. *Mol Biol Evol.* **18**:2298-2305.
3 3 Subramanian, S., D. R. Denver, C. D. Millar, T. Heupink, A. Aschrafi, S. D. Emslie, C.
4 4 Baroni, and D. M. Lambert. 2009. High mitogenomic evolutionary rates and time
5 5 dependency. *Trends Genet.* **25**:482-486.
6 6 Swofford, D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony *and Other
7 7 Methods. Sinauer Associates, Sunderland, MA.
8 8 Tachida, H. 1991. A study on a nearly neutral mutation model in finite populations.
9 9 *Genetics* **128**:183-192.
10 10 Tager, D., J. M. Webster, D. C. Potts, W. Renema, J. C. Braga, and J. M. Pandolfi. 2010.
11 11 Community dynamics of Pleistocene coral reefs during alternative climatic
12 12 regimes. *Ecology* **91**:191-200.
13 13 Takahata, N. 2007. Molecular clock: an anti-neo-Darwinian legacy. *Genetics* **176**:1-6.
14 14 Thomas, J. A., J. J. Welch, R. Lanfear, and L. Bromham. 2010. A Generation Time
15 15 Effect on the Rate of Molecular Evolution in Invertebrates. *Mol Biol Evol.*
16 16 **27**:1173-1180.
17 17 Tomascik, T., R. vanWoesik, and A. J. Mah. 1996. Rapid coral colonization of a recent
18 18 lava flow following a volcanic eruption, Banda Islands, Indonesia. *Coral Reefs*
19 19 **15**:169-175.
20 20 Voris, H. K. 2000. Special Paper 2:Maps of Pleistocene sea levels in Southeast Asia:
21 21 Shorelines, river systems and time durations. *J Biogeog.* **27**:1153-1167.
22 22 Wares, J. P. 2010. Natural distributions of mitochondrial sequence diversity support new
23 23 null hypotheses. *Evolution* **64**:1136-1142.

- 1
2 1 Wares, J. P. 2002. Community genetics in the Northwestern Atlantic intertidal. *Mol Ecol*.
3
4 2 **11**:1131-1144.
5
6 3 Williamson, S., and M. E. Orive. 2002. The genealogy of a sequence subject to purifying
7
8 4 selection at multiple sites. *Mol Biol Evol*. **19**:1376-1384.
9
10 5 Yoder, A. D., and Z. H. Yang. 2000. Estimation of primate speciation dates using local
11
12 6 molecular clocks. *Mol Biol Evol*. **17**:1081-1090.
13
14 7 Zuckerkandl, E., and L. Pauling. 1965. Molecules as documents of evolutionary history. *J*
15
16 8 Theor Biol. **8**:357-366.
17
18 9

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

Taxon	Localities ¹ included in dataset	n	# sites	S	h	π	Fs	D*	Φ_{ST}	Molecular Evolution Model
<i>Tridacna</i> <i>crocea</i> ; 'Black Clade'	→K,P	49	485	22	0.810	0.01	-3.3	-1.35	-0.001	HKY
	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	
<i>Haptosquilla</i> <i>pulchella</i> ; '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+I
	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	
<i>Protoreaster</i> <i>nodosus</i>	→K	38	803	15	0.694	0.002	-10.36	-3.53	N/A	HKY

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](#).

Eric Crandall 6/20/11 7:37 PM

Deleted: the main text

1
2 Table 2. Mean parameter values and lineage mutation rates from the sudden expansion model for
3 the mismatch distribution and the two-epoch model. All values are per site.
4
5

Taxon	Mismatch Distribution Parameters			Lineage Mutation Rate ^a (% per million years)		
	Θ_0	Θ_1	$\tau / 2$	μ 95% low	μ	μ 95% high
<i>Tridacna crocea</i> ; '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%
<i>Haptosquilla pulchella</i> ; '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%
<i>Protoreaster nodosus</i> - CO1	8.12×10^{-6}	3.72×10^{-3}	1.99×10^{-3}	1.40%	10.17%^b	17.76%
				1.88%	13.67%^c	23.87%
Two-Epoch Model Parameters						
	Θ_0	Θ_1	$t_{\text{transition}}$			
<i>Tridacna crocea</i> ; '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%
<i>Haptosquilla pulchella</i> ; '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%
<i>Protoreaster nodosus</i> - 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%

34 ^a Lineage mutation rates calculated as $\mu = (\tau/b) / 2c$ for the mismatch distribution, and $\mu =$
35 $t_{\text{transition}}/c$ for the two-epoch model, where b is the number of sites, c is the calibration point given
36 below.

Eric Crandall 6/20/11 7:37 PM

Deleted: τ

37 ^b Values for calibration point at 19.60 kya.

38 ^c Values for calibration point at 14.60 kya.

39 ^d A two-epoch model of logistical growth was only weakly supported over a model of constant
40 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.

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

Eric Crandall 6/20/11 7:37 PM

[...]

Eric Crandall 6/20/11 7:37 PM

Deleted: -

* Substantial

Eric Crandall 6/20/11 7:37 PM

Formatted: Indent: Left: 0", First line: 0", Line spacing: single

Eric Crandall 6/20/11 7:37 PM

Deleted: -

Eric Crandall 6/20/11 7:37 PM

Deleted: Decisive

Eric Crandall 6/20/11 7:37 PM

Deleted: -

1
2
3 Page 38: [1] Deleted

4 Eric Crandall

5 6/20/11 7:37 PM

6 Log₁₀ Bayes Factors

7 Species	8 Model ^{a,b}	9 Ln P (model)	10 SE	11 Constant	12 Two-epoch (exp. growth)	13 Two-epoch (log. growth)
10 <i>Tridacna crocea</i> 11 'Black Clade' - CO1	Constant	-705.65	0.44	-	0.07	-0.36
	Two-epoch (exp. growth)	-705.81	0.30	-0.069	-	-0.43
	Two-epoch (logistic growth)	-704.82	0.31	0.36	0.43	-
14 <i>Haplosquilla</i> 15 <i>pulchella</i> 16 'White Clade' - CO1	Constant	-1280.59	0.57	-	-5.86	-8.11
	Two-epoch (exp. growth)	-1267.09	0.39	5.86***	-	-2.25
	Two-epoch (logistic growth)	-1261.91	0.54	8.11***	2.25***	-
20 <i>Protoreaster</i> 21 <i>nodosus</i>	Constant	-1130.94	0.33	-	-6.37	-7.05
	Two-epoch (exponential growth)	-1116.28	0.28	6.37***	-	-0.68
	Two-epoch (logistic growth)	-1114.71	0.45	7.05***	0.68*	-

1
2 Figure Legends
3

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 habitat presented here.

Eric Crandall 6/20/11 7:37 PM
Formatted: Font:10 pt

Eric Crandall 6/20/11 7:37 PM

Deleted: a

Eric Crandall 6/20/11 7:37 PM

Deleted: curve

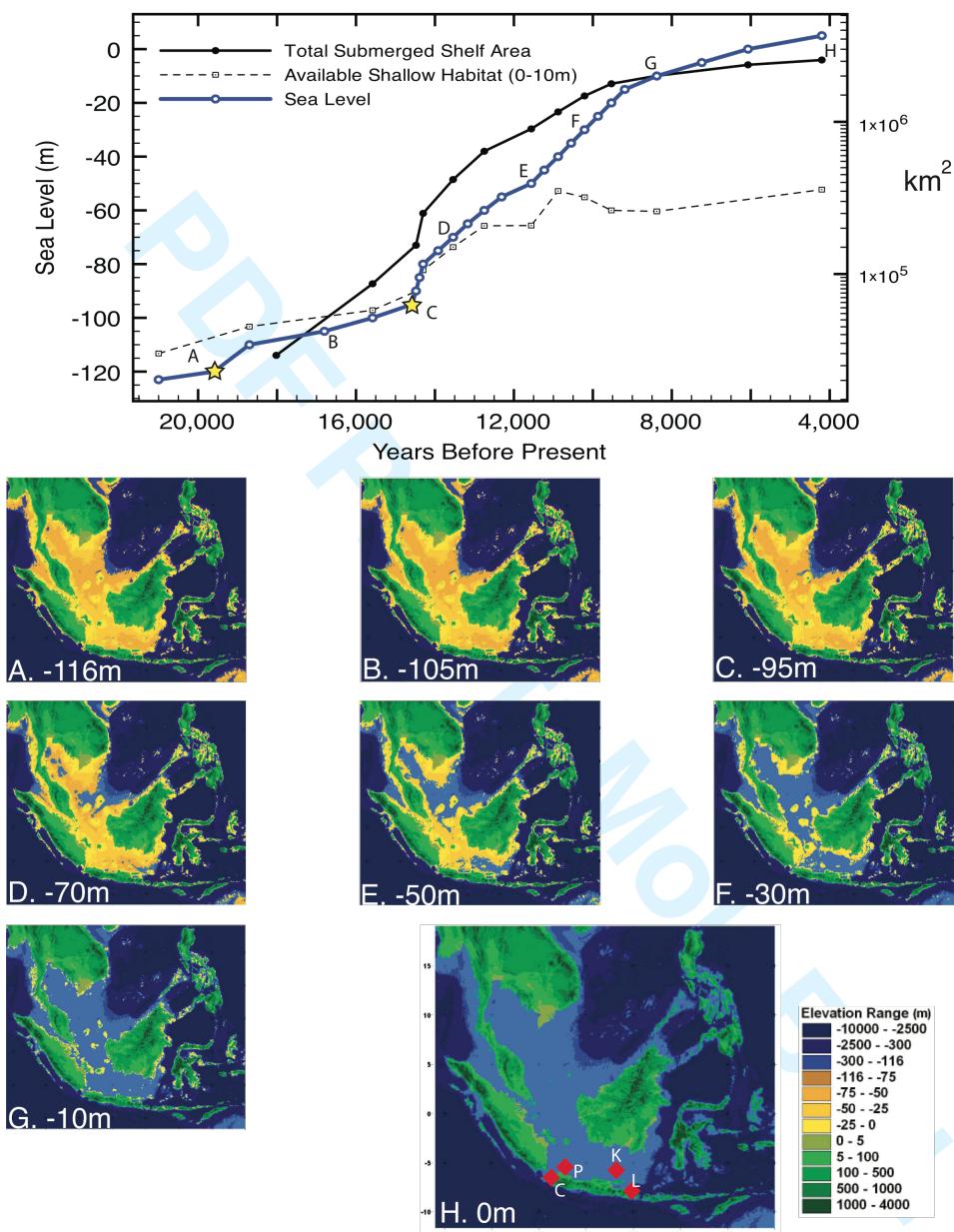
Figure 2. Mismatch distributions for all three species calculated in Arlequin 3.11 (Excoffier et al. 2005). Histograms show the frequency of each pairwise difference in the sample, and dotted lines show the expected frequencies under a sudden expansion model and a spatial expansion model (Excoffier 2004). Sum of squared deviations (SSD) are given for each gene and model; starred values deviated significantly from that 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 ($1/2N_e\mu$, 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

1 density. The upper-right panel shows a schematic for the two-epoch model used to estimate the
2 time of population expansion, using BSP confidence intervals as priors. Black lines show
3 parameters, while grey lines show prior boundaries.
4
5

6
7 Figure 4. Lineage substitution rates ($\frac{1}{2}$ divergence rate) per generation for marine
8 invertebrate CO1 plotted against their calibration date. Error bars represent 95%
9 credibility intervals. For calibrations < 5 mya, we only plotted rates that account for
10 ancestral polymorphism (Edwards and Beerli 2000) in a coalescent framework. For
11 calibrations > 5 mya we corrected for ancestral polymorphism using net nucleotide
12 divergence (Nei and Li 1979). Rates from present study: (a) *Haptosquilla pulchella*; (b)
13
14 *Protoreaster nodosus*. (c) rate from simultaneous divergence of 7 echinoid species pairs at
15 the Isthmus of Panama, (Hickerson et al. 2006). (d) average rate from two divergence times
16 for 15 alpheid species pairs (Hickerson et al. 2003). Rates from Neritid fossil calibrations
17 (Frey and Vermeij 2008): (e) *Nerita fulgurans* and *N. senegalensis*; (f) *N. scabrida* and *N.*
18
19 *peloranta* + *versicolor* (g) *Nerita exuvia* and *Nerita textilis* (j) *Nerita adanensis* + *Nerita*
20
21 *planospira* and Nerita crown clade. (i) Fossil calibration for *Bulla striata* and *B. occidentalis*
22 (Malaquias and Reid 2009). Rates from Arcid fossil calibrations (Marko 2002): (h) *Andara*
23
24 and *Grandiarca*; (k) *Fugleria* and *Cucullaeearca*
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1.



Eric Crandall 6/20/11 7:37 PM

Deleted:

Unknown

Formatted: Font:Times New Roman, 10 pt

Figure 2.

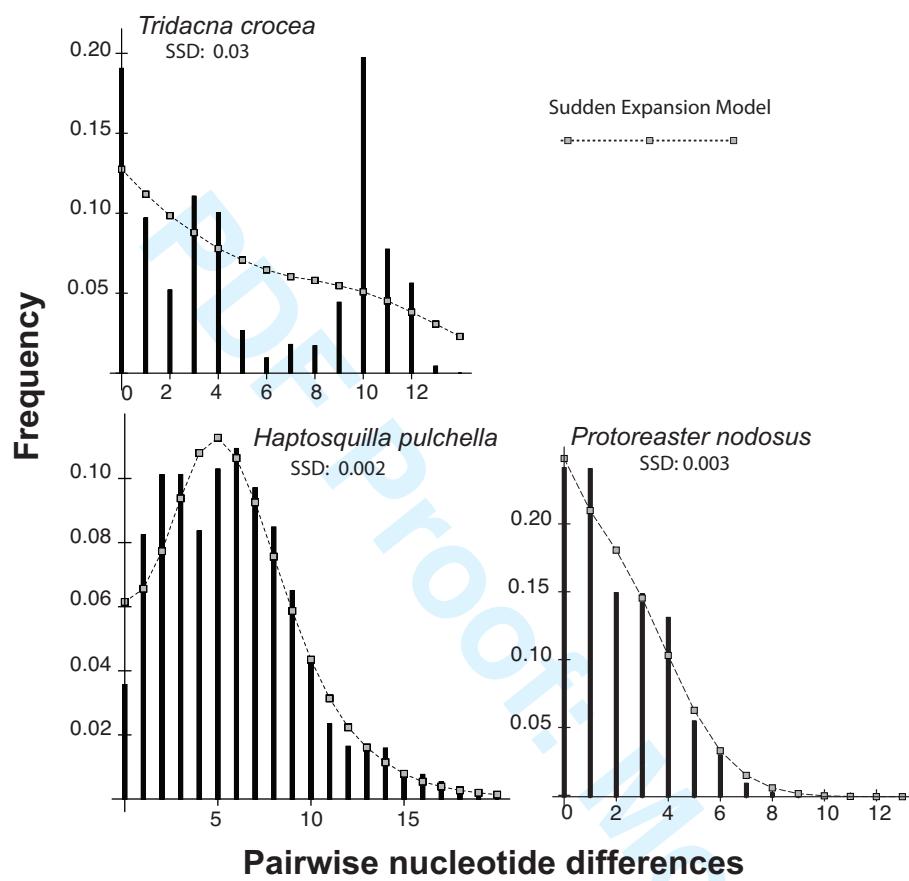


Figure 3.

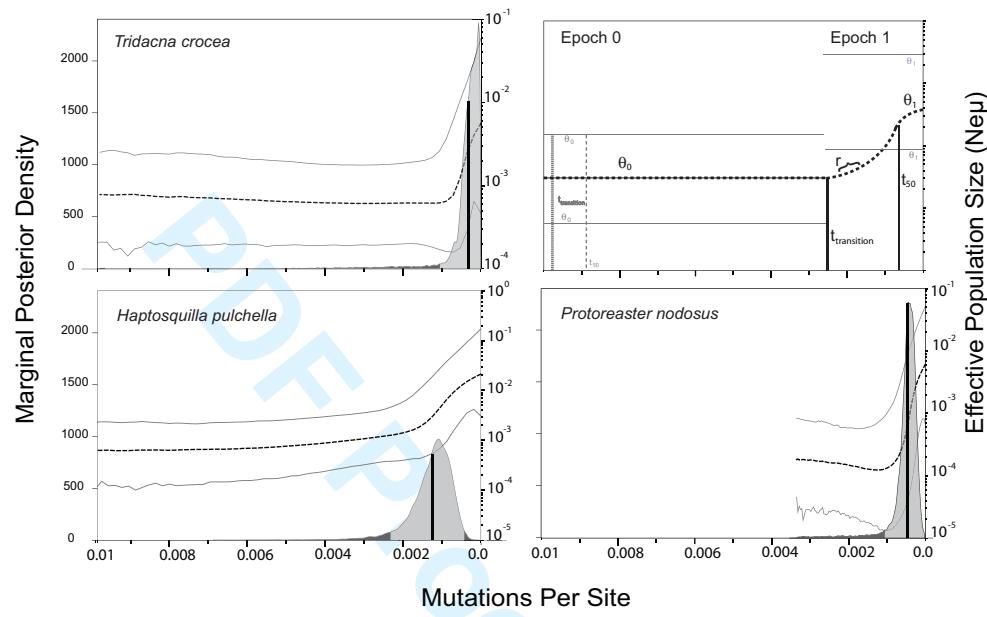


Figure 4.

