Molecular Biology and Evolution, 2025, **42**, 1–13
https://doi.org/10.1093/molbev/msae266
Advance access publication 7 January 2025
Discoveries

MBE OXFORD

Direct Measurement of the Mutation Rate and Its Evolutionary Consequences in a Critically Endangered Mollusk

T. Brock Wooldridge ,*[,1] Sarah M. Ford ,[1] Holland C. Conwell ,[1] John Hyde ,[2] Kelley Harris ,[3,4] Beth Shapiro [1,5]

[1]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95060, USA
[2]Southwest Fisheries Science Center, La Jolla, CA 92037, USA
[3]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[4]Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA
[5]Colossal Biosciences, Austin, TX 95060, USA

***Corresponding author:** E-mail: tywooldr@ucsc.edu.
Associate editor: Deepa Agashe

Abstract

The rate at which mutations arise is a fundamental parameter of biology. Despite progress in measuring germline mutation rates across diverse taxa, such estimates are missing for much of Earth's biodiversity. Here, we present the first estimate of a germline mutation rate from the phylum Mollusca. We sequenced three pedigreed families of the white abalone *Haliotis sorenseni*, a long-lived, large-bodied, and critically endangered mollusk, and estimated a de novo mutation rate of 8.60×10^{-9} single nucleotide mutations per site per generation. This mutation rate is similar to rates measured in vertebrates with comparable generation times and longevity to abalone, and higher than mutation rates measured in faster-reproducing invertebrates. The spectrum of de novo mutations is also similar to that seen in vertebrate species, although an excess of rare C > A polymorphisms in wild individuals suggests that a modifier allele or environmental exposure may have once increased C > A mutation rates. We use our rate to infer baseline effective population sizes (N_e) across multiple Pacific abalone and find that abalone persisted over most of their evolutionary history as large and stable populations, in contrast to extreme fluctuations over recent history and small census sizes in the present day. We then use our mutation rate to infer the timing and pattern of evolution of the abalone genus *Haliotis*, which was previously unknown due to few fossil calibrations. Our findings are an important step toward understanding mutation rate evolution and they establish a key parameter for conservation and evolutionary genomics research in mollusks.

Keywords: mutation rate, mollusk, conservation, phylogenomics

Introduction

Mutations are the ultimate source from which variation arises. Although mutation is a fundamental feature of all life on Earth, the rate at which new mutations arise can vary considerably between and within species. Consequently, the mutation rate and the extent to which it is fine-tuned by natural selection have long held the interest of biologists (Sturtevant 1937; Lynch et al. 2016).

Mutations may occur in any cell type but only mutations that occur in an organism's germline contribute to subsequent generations and drive evolutionary innovation (Bergeron et al. 2023). Within multicellular eukaryotes, germline mutation rates (GMRs) vary by at least three orders of magnitude, and the etiology of this variation is not completely understood (Lynch 2010). Mutation rates per generation are generally highest in large-bodied, long-lived organisms with modest effective population sizes, and several mechanisms (which are not mutually exclusive) have been proposed to explain this trend. Long generation times could increase GMRs by allowing more time for mutations to accumulate in spermatocytes and oocytes prior to reproduction. Consequently, the length of time elapsed between puberty and reproduction has been proposed to explain observed mutation rate variation among vertebrates (Thomas et al. 2018). However, small effective population sizes could also explain this variation, since weakly deleterious alleles may drift to high frequencies in small populations, increasing observed mutation rates (Lynch 2010). Both generation time and effective population size are similarly correlated with mutation rate variation in empirical datasets, making it difficult to disentangle the relative etiological contributions of reproductive longevity and the drift-barrier effect (Wang and Obbard 2023).

Current understanding of the causes and extent of GMR variation is shaped by available estimates of GMRs. However, these data are not representative of Earth's biodiversity. Roughly 83% of animals with an estimated GMR are vertebrates (Wang and Obbard 2023), despite vertebrates representing only 4.6% of animal diversity (Bánki et al. 2024). Some animal phyla are entirely unrepresented among available data for GMRs, including Mollusca. Mollusks

Received: September 16, 2024. **Revised:** November 13, 2024. **Accepted:** December 17, 2024

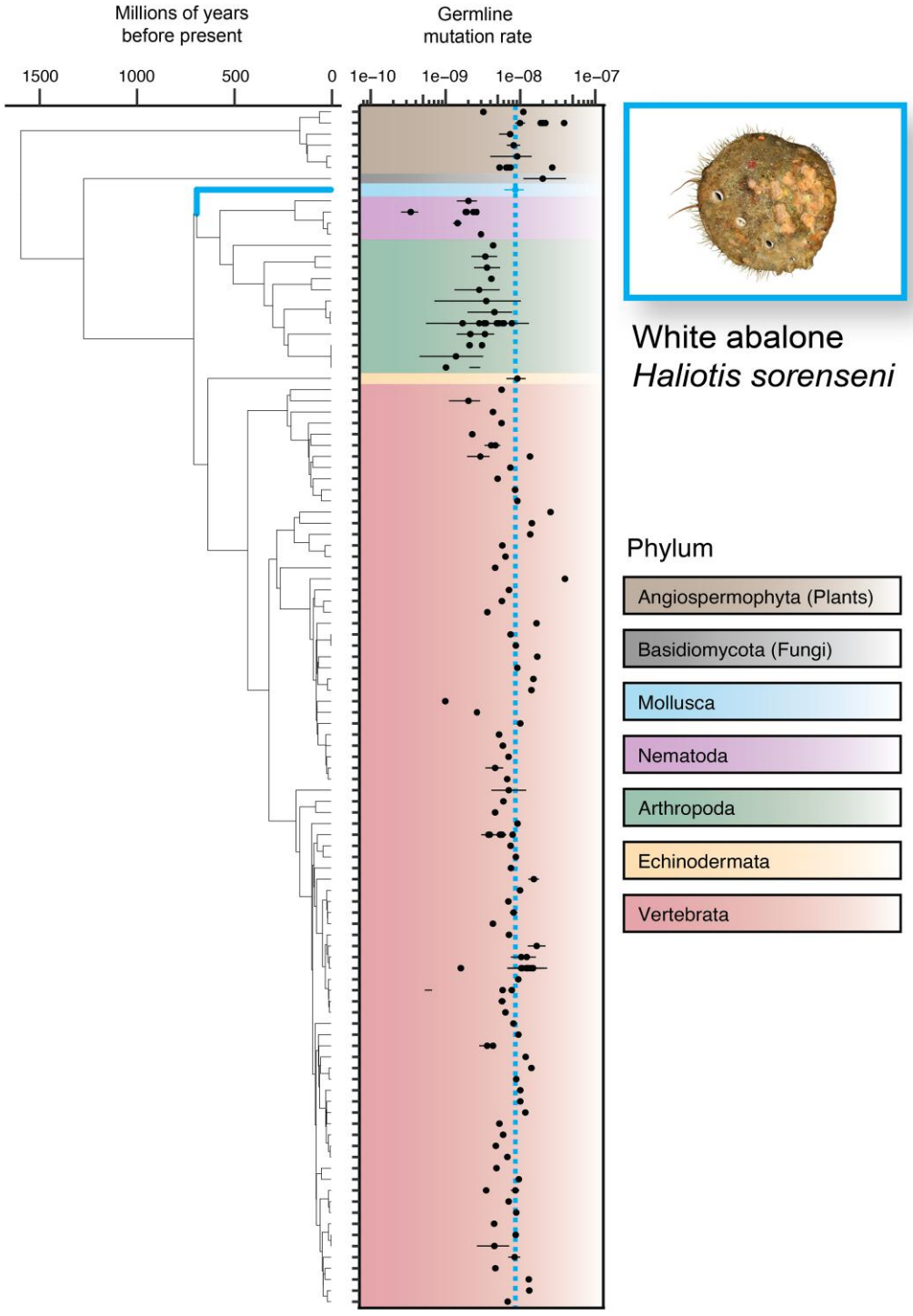

Fig. 1. Distribution of all SNV germline mutation rate estimates for multicellular eukaryotes, adapted from Wang and Obbard (2023). Multiple points per taxon represent multiple estimates of the GMR in that group. Time-scaled phylogeny to the left retrieved from *TimeTree* (Kumar et al. 2017).

encompass a broad diversity of form and function, spanning terrestrial species like the common garden snail to the deep ocean dwelling giant squid. Mollusca is also characterized by considerable variation among lineages in population size, gamete production, parental investment, and longevity (Ponder and Lindberg 2008). Therefore, mollusk diversity may help in understanding the evolution of variability among GMRs. The closest relatives of Mollusca with estimated GMRs share a common ancestor roughly 600 to 700 million years ago (Dohrmann and Wörheide 2017), a distance that represents more than 1,200 Ma of independent evolution (Fig. 1).

While the rate at which mollusks accumulate de novo mutations remains unknown, the high genetic diversity in many mollusk populations has led to speculation of a fast GMR (Hoeh et al. 1996; Launey and Hedgecock 2001; Zhang et al. 2012; Cutter, Jovelin, and Dey 2013). Previous work, for example, estimated a GMR for the Pacific oyster *Crassostrea gigas* that was 90 times faster than that of *Drosophila*, based

on the oyster's anomalously high deleterious mutation load (Plough, Shin, and Hedgecock 2016). This estimate was predicated on the theoretical relationship between the frequency of lethal mutations and mutation rates (Nei 1968), however, and was not a direct estimate of the GMR (Plough, Shin, and Hedgecock 2016). Although fast mutation rates can lead to high levels of genetic diversity, genetic diversity (π) can also be maintained in populations with large effective sizes (N_e) without requiring faster rates of mutation (μ) ($\pi = 4N_e\mu$; Nei and Tajima 1981). Many marine mollusks do have large census population sizes (N_c) but effective size may be much smaller due to processes like "sweepstakes reproduction", in which a handful of highly fecund individuals reproduce each generation (Hedrick 2005; Hedgecock and Pudovkin 2011). Therefore, uncertainty surrounding N_e complicates inference of a GMR based on population genetic diversity (Harrang et al. 2013).

In contrast to inferences based on population genetic diversity, phylogenetic analyses of mollusks have suggested lower GMRs. Based on rates of nucleotide substitution between mollusk lineages calibrated by fossil ages, these estimates are on the order of 3×10^{-9} mutations/bp/generation (Allio et al. 2017; Li et al. 2021). Phylogenetic approaches to estimating mutation rates are, however, error-prone due to uncertainty in fossil calibration, generation time, mutation saturation over long timescales, and difficulty in identifying truly neutral sites for estimation (Scally and Durbin 2012; Wang and Obbard 2023). Because indirect estimates of mutation rates in mollusks vary to such an extent, a direct estimate based on pedigree sampling is needed.

Accurate estimates of GMRs have practical applications for conservation and evolutionary genomics research. Effective population size, which describes the size of an idealized Wright–Fisher population that would exhibit the magnitude of genetic drift and inbreeding as seen in a real-world wild population (Wright 1931), is used as a conservation metric and often requires a GMR to estimate. Minimum "healthy" N_e thresholds of 50 or 500 are often being used to guide wildlife management decisions (Jamieson and Allendorf 2012). Outgrowths of this have incorporated the census population size N_c, with high N_e/N_c ratios serving as an accurate indicator of high extinction risk (Palstra and Fraser 2012; Wilder et al. 2023). Knowing the germline mutation rate also has direct implications for evolutionary genomics and can help to date the origin of a clade (Besenbacher et al. 2019; Bergeron et al. 2021) or the age of a beneficial allele (Smith et al. 2018), particularly for groups where phylogenetic estimates of μ are absent or problematic.

To determine if the germline mutation rate of mollusks differs from that of other animals, we measured de novo mutation rates in three families of the white abalone, *Haliotis sorenseni* (supplementary fig. S1, Supplementary Material online). Individuals for this study were provided by the White Abalone Captive Breeding Program, which aims to restock wild populations of this critically endangered species with aquaculture-raised individuals. The Captive Breeding Program was initiated over 20 years ago using individuals collected from some of the last remaining southern California populations, and efforts since then have focused on breeding these individuals and their offspring. The white abalone is a gastropod mollusk typically found in 20 to 60 m of water along the coast of California and Mexico (Tutschulte 1976). Like many broadcast-spawning invertebrates, white abalone are highly fecund; one individual can release millions of eggs or sperm into the water column in a single spawning event, and several spawning events can occur in 1 year (Hobday et al. 2000). After fertilization, white abalone larvae will disperse for ~10 d before settling on the rocky substrate, after which point they will move very little, if at all, over the course of their adult lives (Leighton 1972; Lafferty et al. 2004). Growth in white abalone is slow, as sexual maturity in wild individuals occurs around 4 years of age with most individuals likely reproducing by 6 years (Tutschulte and Connell 1981). Individuals may live up to 30 years of age and grow to more than 20 cm (Hobday et al. 2000; Andrews et al. 2013). While much is still unknown about the biology of this species, white abalone represent a fascinating combination of traits typically associated with mollusks (e.g. broadcast spawning, metamorphosis) and those often associated with larger vertebrates (e.g. long-lived, late sexual maturity).

We then use our estimated GMR to resolve both recent and long-term questions in abalone conservation and evolutionary history. This species' high fecundity in a sea of dispersive currents was once thought to buffer white abalone—and organisms with similar life histories—against overexploitation (Jamieson 1993; Rogers-Bennett et al. 2016). However, slow growth and overharvesting of mature adults can reduce gamete abundance and concentration to the point of total recruitment failure (Stephens et al. 1999; Hobday et al. 2000). Despite an estimated historical population size of 360,000 in California, white abalone populations declined precipitously during the 20th century, due to a combination of intensive fishing, disease, and the aforementioned overly optimistic view of the species' ability to recover (Rogers-Bennett et al. 2002). Given the extent of decline, white abalone was the first marine invertebrate to be listed under the U.S. Federal Endangered Species Act. A better understanding of how current population sizes stack up against historical baselines (e.g. long-term N_e) will help guide management criteria. Additionally, the evolutionary history of white abalone and its relationship to other congeners, some of which are better studied or more robust to environmental stressors (Crosson and Friedman 2018), is largely unknown. Outlining the timescale of diversification in abalone can help set expectations for how much one species' traits might be true of another species in this understudied group.

Results

We sequenced three families consisting of nine offspring and five parents, three wild and two captive bred, to >50× coverage. In the offspring, we observe 107 unique de novo mutations (Table 1). Only 13 (12.1%) of these mutations were inherited by two or more offspring when they had parents in common (supplementary fig. S2, Supplementary Material online). After incorporating estimates of the false discovery rate (FDR) and false negative rate (FNR), 0.05 and 0.139, respectively, and the size of the callable genome (supplementary table S1, Supplementary Material online), we observe a median mutation rate of 7.99×10^{-9} mutations/bp/generation and a mean mutation rate of 8.60×10^{-9} (95% CI: 6.10 to 11.11×10^{-9}; supplementary fig. S3, Supplementary Material online). The mean rate is faster than most arthropod per-generation rates but falls within the distribution of per-generation rates estimated for most vertebrates, plants, and the one echinoderm (Fig. 1; Wang and Obbard 2023; Popovic et al. 2024). When our rate is plotted as a function of life history traits, including approximations of generation time (~6 years), age at sexual maturity (~4 years), and lifespan in the wild (~20 to 30 years), it remains consistent with vertebrate distributions

Table 1. Counts of observed mutations by family

Family	Num. offspring	Num. mutations	Num. multi-sib mutations	Mean rate ($\times 10^{-9}$)	Contribution (P/M/?)
#1	3	49	4	10.9	21/6/22[a]
#2	4	43	9	8.6	8/6/29
#3	2	17	0	5.1	5/6/6

Num. mutations = Count of unique mutations across the set of offspring. Families 2 and 3, which share a mother, share two of these mutations, such that the total number of unique mutations is 107, not 109. Num. multi-sib mutations = Count of unique mutations found in two or more offspring of a family. Mean rate = Mean of the individual-level rates for each family. Contribution = Mutations from the paternal line (P), maternal line (M), or unknown (?) as determined by read-backed phasing only.
[a]Significant parental bias; $P < 0.005$, chi-squared test.

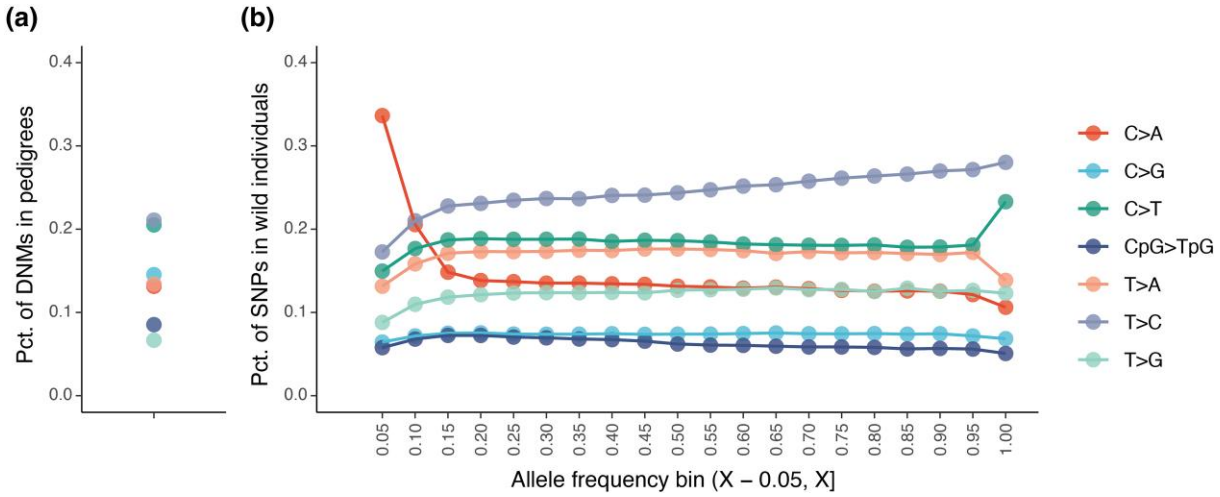

Fig. 2. a) Proportion of mutation types for all DNMs detected in pedigree sequencing. b) Proportion of single nucleotide polymorphisms (SNPs) in 11 wild-caught individuals as a function of minor allele frequency and polymorphism type.

(supplementary fig. S4a to c, Supplementary Material online; Hobday et al. 2000; Bergeron et al. 2023).

Using read-based haplotype phasing, we were able to attribute 29.2% of these mutations to either a mother or father. One of these families showed significant paternal bias (α), with the father contributing 21 mutations and the mother contributing only 6 (α = 3.5; Table 1). The two other families, which happened to share a mother, showed roughly equal contributions from both parents (α = 1.33, 0.83). As neither shell size, weight, nor age at time of reproduction is known for all parents, it is not possible to relate the age of each parent at spawning to the relative contribution of mutations. We know that two of the individuals born in captivity, the shared mother of Families 2 and 3 and the father of Family 1, were at least 19 years old at spawning. The three remaining parents were wild-caught as adults a few years prior to spawning, so their precise age is unknown.

The mutation spectrum of the detected de novo mutations is broadly similar to spectra previously reported in vertebrate species (Bergeron et al. 2023). Primarily, we observe a transition-to-transversion ratio of 1.06, and the percentage of C > T mutations occurring at CpG sites is 43.5%. The spectrum of DNMs in our pedigreed samples also broadly corresponds to the spectrum of common (MAF > 0.25) single nucleotide polymorphisms (SNPs) observed in wild individuals, although we do observe slightly higher rates of C > G and C > T mutations than their corresponding SNP frequencies might suggest (Fig. 2). Relative frequencies of different types of SNPs are generally constant regardless of allele frequency in the wild. However, we do observe one major exception to this pattern. C > A polymorphisms in wild individuals represent 33.6% of rare SNPs (MAF > 0.05), despite the observation that only 14.0% of DNMs were C > A (Fig. 2). This enrichment is absent for MAFs greater than 0.25.

Given our estimate of a germline mutation rate for *H. sorenseni*, we aimed to infer effective population size through time across *Haliotis* spp. and gather a sense of long-term baselines that could inform management of these species. To do this, we performed demographic inference and all available high coverage (>20×) whole-genome sequencing data for *Haliotis* spp. Demographic analysis of five species distributed throughout the Pacific Ocean points toward large and stable effective population sizes (N_e) over long timescales (Fig. 3a). The harmonic mean of N_e, summarized across roughly 1 million generations (1×10^4 to 1×10^6), is between 100,000 and 300,000 for all species. When we compare this value for our focal species, *H. sorenseni*, against the mean mutation rate, we see that values for white abalone are consistent with GMR~N_e relationships in vertebrates (supplementary fig. S4d, Supplementary Material online; Bergeron et al. 2023). *Haliotis sorenseni* does exhibit a gradual decline over the most recent 100,000 generations, rarely surpassing an N_e of 50,000. The blacklip abalone, *H. rubra*, exhibits a more abrupt decline, dropping by roughly ~90% in the past 10,000 generations. However, inferences of population size within this most recent time interval should be interpreted with caution when based on a single individual (Wilder et al. 2023). Therefore, the growth inferred for *H. cracherodii*, *H. rufescens*, and *H. laevigata* within this time interval suffers from similar limitations.

These historical estimates of N_e based on haplotype coalescence are supported by direct calculations of N_e from

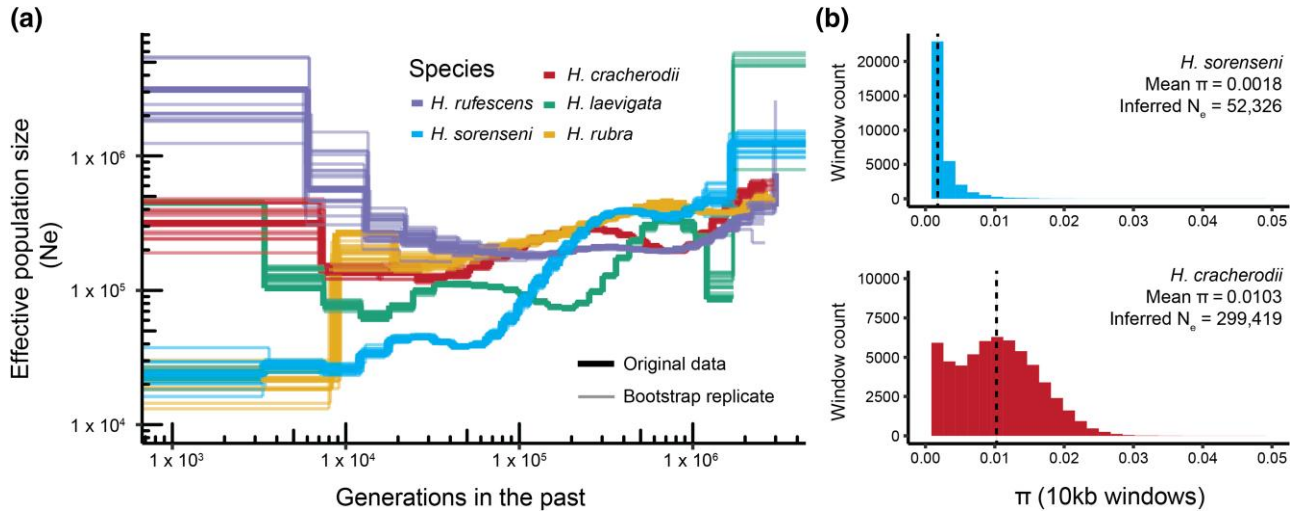

Fig. 3. a) Historical effective population size (N_e) for five abalone species as estimated by *MSMC2* (Schiffels and Wang 2020). Both original data and 20 bootstrap replicates per species are shown. b) Distribution of intraspecific nucleotide diversity (π) for species where multiple sequenced individuals were available. Effective population size is calculated from $\pi = 4N_e\mu$ using the mean observed π.

population-level sequence diversity (π). We called genotypes for 11 sequenced individuals from each of two species, *H. sorenseni* and *H. cracherodii*, which are the only abalone species with sufficient population-level sequencing data. Summarizing across genome-wide 10 kb windows, we observed intraspecific diversity (π) of 0.0018 and 0.0103 for *H. sorenseni* and *H. cracherodii*, respectively (Fig. 3b). By applying these values and our mean estimate of the mutation rate μ to the relationship $\pi = 4N_e\mu$ (Nei and Tajima 1981), we obtained N_e values of 52,236 and 299,419. These estimates, based on population-level nucleotide diversity (Hare et al. 2011; Nadachowska-Brzyska et al. 2022), closely match the historical values of N_e for both species (Fig. 3a).

We then applied our mutation rate to examine *Haliotis* evolution on deeper evolutionary timescales. To do this, we used genes present in the reference genomes for the above set of *Haliotis* species, the tropical abalone *Haliotis asinina* (for which a reference was also available), and the sea snail *Gibbula magus*, which served as an outgroup. We find a single highly supported topology based on the coding sequences of 2,525 coding genes (Fig. 4a). We then used a multispecies coalescent (MSC) approach to date this topology according to our germline mutation rate estimate (Tiley et al. 2020). Specifically, we used a subset of 150 clock-like genes, a mean and standard generation time of 6 and 2 years, respectively (Rogers-Bennett et al. 2004), and a mutation rate standard deviation of 3.26×10^{-9}. With this approach, we date the split between Western and Eastern Pacific abalone at 36.4 million years before present (95% CI: 33.6 to 39.1×10^6 Ma; Fig. 4). Within the Western Pacific clade, the lone tropical abalone representative in this analysis, *H. asinina*, splits off from *H. rubra* and *H. laevigata* early on at 19.9 Ma. In contrast, all three Eastern Pacific abalone share a relatively recent common ancestor at 4.3 Ma (95% CI: 3.9 to 4.7×10^6 Ma), with white (*H. sorenseni*) and red (*H. rufescens*) abalone diverging only 2.2 Ma (95% CI: 1.9 to 2.5×10^6 Ma).

Discussion

We estimated the germline mutation rate of white abalone to be 8.60×10^{-9} mutations/bp/generation. Even though *H. sorenseni* is diverged from vertebrates by over 1,200 Ma years of evolutionary history, the rate we estimate is not unlike rates previously estimated for vertebrates with similar generation time, age at sexual maturity, and effective population size (Fig. 1; supplementary fig. S4, Supplementary Material online; Bergeron et al. 2023). Our rate is also similar to the previously estimated GMR for the crown-of-thorns sea star, the only member of the phylum Echinodermata with a GMR estimate as well as the only other marine invertebrate and broadcast spawner for which such an analysis has been performed (9.13×10^{-9}; Popovic et al. 2024). Our direct estimate based on pedigreed samples is more technically reliable than previous indirect estimates of mollusk mutation rates, some of which were much higher (Hoeh et al. 1996; Plough et al. 2016) and some of which were much lower (Allio et al. 2017; Li et al. 2021) than ours. In sum, our findings suggest that GMRs in mollusks do not deviate from the established distributions and relationships observed in other branches of the tree of life.

We observed high variance in the degree of paternal bias (α) in the contribution of de novo mutations to offspring. Of the three families in our analysis, two show near equal contributions from both parents ($\alpha = 1.33; 0.83$), while the third family shows a clear paternal skew ($\alpha = 3.5$) (Table 1). A male mutation rate bias is not unexpected in GMR studies, and the magnitude of α is known to increase with longevity (Thomas et al. 2018). The larger number of germ cell divisions in males of many species is thought to drive a greater contribution of de novo mutations from fathers (Venn et al. 2014; Jónsson et al. 2017), although evidence of male bias independent of cell division number points to contributions from other processes, such as sex-specific differences in DNA damage and repair (de Manuel et al. 2022). Male and female white abalone appear to invest roughly the same amount of energy toward gonad development (Tutschulte and Connell 1981), but the number of sperm produced will far exceed the number of eggs produced for most abalone species (Babcock and Keesing 1999). Therefore, the greater number of germ cell divisions in male abalone could be driving the paternal bias we observe in one of our three families ($\alpha = 3.5$). However, it does not explain the lack of bias in the other two families, which exhibit values of α closer to species with similar male and female reproductive input, including the crown-of-thorns sea stars ($\alpha = 0.96$; Popovic et al. 2024) and several fish species ($\alpha = 0.8$;

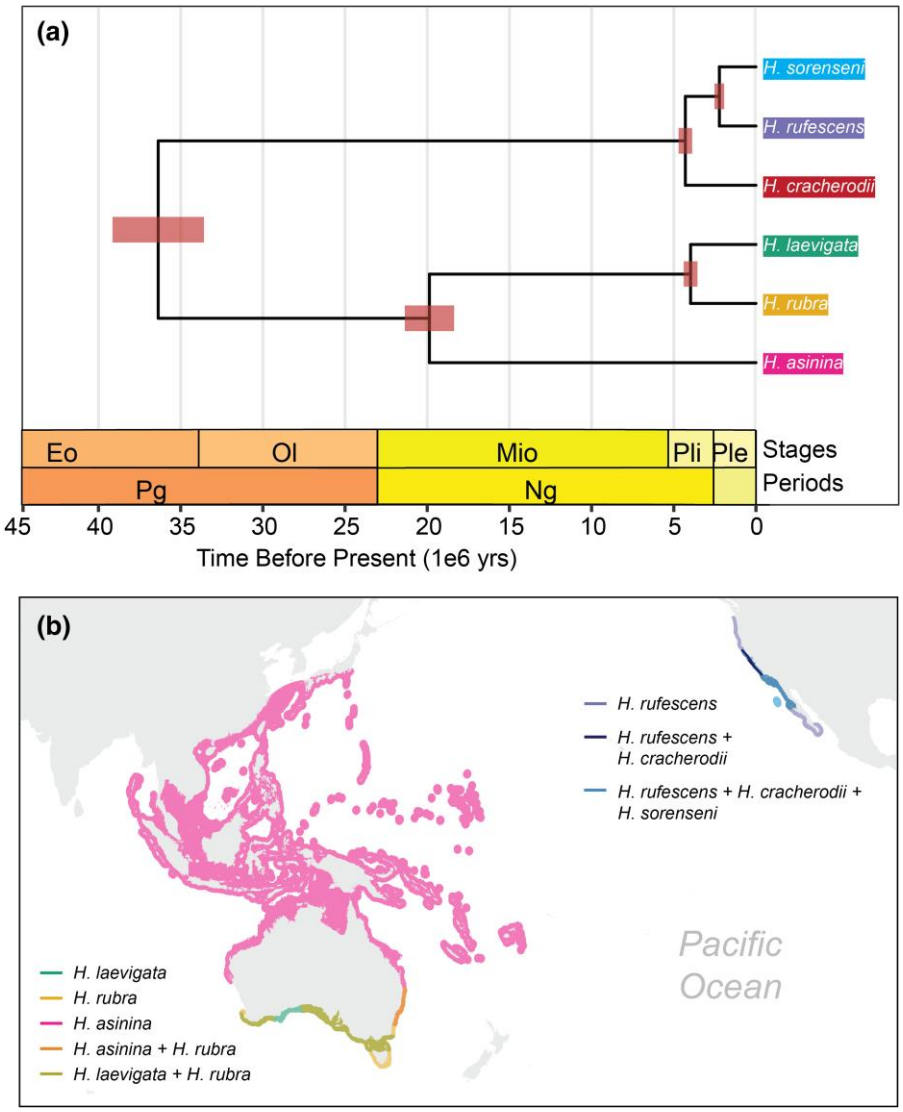

Fig. 4. a) Multispecies coalescent (MSC) phylogeny inferred with *BPP* (Flouri et al. 2018) and calibrated assuming μ.mean = 8.60×10^{-9}, μ.sd = 3.26×10^{-9}, g.mean = 6, and g.sd = 2, where μ is the mutation rate and g is the generation time. The tree is rooted with the outgroup *G. magus*, not shown. b) Distributions of the six abalone species presented in Fig. 3a. Range maps obtained from the IUCN Red List.

Bergeron et al. 2023). Because the extent of male bias scales with age, it is possible that age differences between males and females in two of the three families are modifying the parental contributions. However, we are unable to examine this relationship because we lack all parental ages at the time of spawning, and the wild origin of three of the five parents complicates age-size relationships measured in captivity (McCormick et al. 2016).

Reproductive life history affects not only the degree of paternal mutation bias but also the proportion of mutations that are shared among siblings. In our families, we see that 4% of mutations were transmitted to multiple offspring (Table 1; supplementary fig. S2, Supplementary Material online), which implies that at least 12% of the mutations we observe in the parents prior to primordial germ cell specification. Previous work has found that the sharing of germline mutations among siblings is most widespread in species with short generation times, and our results are consistent with that trend—the abalone generation time is intermediate between the generation times of mice and humans, and the proportion of mutations shared among siblings is similarly intermediate. In a previous study of mouse germline mutations, 23.9% were shared among siblings, while in humans the corresponding rate is just 4% (Lindsay et al. 2019). In the guppy *Poecilia reticulata*, which has a very short generation time of 3 months, the majority of de novo mutations are shared (Lin et al. 2023). Given the roughly 6-year generation time of wild white abalone, it is perhaps unsurprising that we observe a mutation rate closer to humans than that of mice or guppies.

The spectrum of de novo mutations showed a surprising relationship with standing variation in wild white abalone. C > A polymorphisms in wild individuals make up 33.6% of rare SNPs despite the observation that only 14.0% of DNMs were C > A (Fig. 2). This over-twofold enrichment comes in stark contrast to the expectation that the spectrum of rare polymorphisms should reflect the rate of input mutations. However, this signal is limited to rare variants. Among C > A polymorphisms that are more common in the wild (MAF > 0.25), C > As constitute only 13% to 14% of SNPs, equivalent to the proportion of C > A DNMs. Excluding C > As, the spectrum of common SNPs appears generally stable across all MAFs and broadly resembles the spectrum of DNMs. Some

differences, like the relatively low amount of C > G SNPs compared to C > G DNMs, might be explained by the effect of parental age on mutation rates (Jónsson et al. 2017). However, the absence of precise ages and limited sample size in our work prevents us from identifying such a relationship.

A "pulse" in the rate of one mutation type, like that we observe for C > A, has only previously been reported in humans (Harris and Pritchard 2017). The human pulse is clearly not active in modern populations, as evidenced by the mutation spectrum of rare variation, but the smaller sample size of wild *H. sorenseni* available for this study prevents us from determining whether the C > A mutation rate has declined from its peak in the wild. One possibility is that the C > A mutation rate became elevated in the wild due to an environmental change which did not affect the captive-born children of the trios sequenced in our study. Future work aimed at sequencing more trios and wild individuals will be needed to understand this result.

Our inference of effective population size showed large and stable abalone populations over evolutionary timescales, including our focal species *H. sorenseni*. Over roughly 1 million generations (10^3 to 10^6), N_e for all five species varied between 1×10^5 and 5×10^5 (Fig. 3). At face value, these absolute values of N_e are encouraging for the future of abalone conservation. Large N_e populations are thought to be less susceptible to genetic drift and inbreeding depression, and N_e values of 50 or 500 are often cited as desirable conservation thresholds (Jamieson and Allendorf 2012). N_e/N_c, another metric of population vulnerability (Palstra and Fraser 2012; Wilder et al. 2023), further indicates population stability when considering historical abalone N_c. For example, estimates of pre-collapse N_c for *H. sorenseni* and *H. cracherodii* in California are 360,000 and 3,500,000, respectively (Rogers-Bennett et al. 2002). When considering our long-term summary estimates of N_e (Fig. 3b), N_e/N_c ratios are 0.145 for *H. sorenseni* and 0.085 for *H. cracherodii*, close to the 0.1 metric typical of healthy wild populations (Frankham 1995; Palstra and Ruzzante 2008). However, natural populations of species that have high juvenile mortality and fecundity, such as abalone, often exhibit N_e/N_c ratios much lower than 0.1 (Hoban et al. 2020; Popovic et al. 2024). Uncertainties regarding life history, for example the influence of sweepstakes reproduction, make it difficult to generate clear expectations for N_e/N_c in healthy populations of abalone (Hedrick 2005). The historical census size is also a source of uncertainty for abalone, but fisheries landings data (Rogers-Bennett et al. 2002), written accounts (Vileisis 2020), and population genomic data (Wooldridge et al. 2024) all depict populations of *H. sorenseni* and *H. cracherodii* as large and continuous prior to collapse.

When we compare N_e to contemporary population sizes, we see the degree to which recent population collapses have rendered *H. sorenseni* and *H. cracherodii* vulnerable to extinction. Both *H. sorenseni* and *H. cracherodii* are thought to exist at less than 1% of their pre-collapse abundances (Rogers-Bennett et al. 2002), meaning extreme declines in N_c and inflation of N_e/N_c. For white abalone, contemporary populations number as few as 500 and no more than 5,000 (Stierhoff et al. 2012, 2014), resulting in N_e/N_c as high as 100, greater than that estimated for even the most threatened mammals (Wilder et al. 2023). Black abalone show greater variation in population size. They were completely extirpated in many southern California sites (VanBlaricom et al. 2009) and are showing incipient recovery at some locations ($N_c = 2{,}341$ at San Nicolas Island (Kenner 2021), while at the northern end of their range declines have been less severe (Neuman et al. 2010)). The range in N_e/N_c resulting from this variation further emphasizes the need for population-specific approaches to management. Even recovering sites like San Nicolas Island still fall in the "highly vulnerable" range ($N_e/N_c > 100$) while northern sites exhibiting minor decline are of less concern. While overreliance of these values is not recommended, especially when decades of census data sufficiently demonstrate a species' vulnerability, having some sense of N_e/N_c now enabled by our knowledge of the rate of input mutations can shed light on the magnitude of vulnerability and provide important guidance for species' recovery metrics (Robinson et al. 2022).

Our estimate of a germline mutation rate for *H. sorenseni* permitted the first time-calibrated phylogeny for the abalone genus *Haliotis* and resolved the timing of diversification of Pacific species. Despite their historical abundance, abalone are poorly represented in the fossil record (Geiger and Groves 1999). For the abalone fossils that do exist, morphological ambiguity in the fossilized shells makes it difficult to place these specimens in the context of present-day diversity (Geiger and Groves 1999). Therefore, a mutation rate-based approach to phylogenetic dating is particularly suited to this system, which has yet to see an attempt at divergence dating. Our inferred topology agrees with previously reported relationships (Gruenthal and Burton 2006; Streit et al. 2006; Masonbrink et al. 2019), and we identify a common ancestor for the analyzed species at 36.4 Ma, during the late Eocene (Fig. 4). These species represent the major lineages of Pacific abalone diversity, and the date is consistent with a late Cretaceous specimen found in California and an Eocene specimen from New Zealand (Geiger and Groves 1999; Estes et al. 2005). Furthermore, the estimated 4.3 Ma common ancestor of California abalone (Fig. 4) agrees with the handful of Pliocene (5.3 to 2.6 Ma) and much larger number of Pleistocene (2.6 Ma to 11.7 ka) fossils of *H. cracherodii* and *H. rufescens* from California (Geiger and Groves 1999). Similarly, single Pleistocene fossils of *H. laevigata* and *H. rubra* from Australia are consistent with the 4.0 Ma common ancestor of these species. While the agreement between our rate-based estimates and the limited fossil record are encouraging, it should of course be recognized that these figures are subject to change as a better understanding of life history (i.e. generation time) and mutation rate variation emerges (Tiley et al. 2020). Nevertheless, these results demonstrate the utility of such an approach for closely related clades with similar constraints on fossil information.

Deriving an estimate of a species' germline mutation rate drives at fundamental questions in biology but also contributes an essential resource for conservation and evolutionary genomics research. Here, we have added to the growing understanding of how GMRs vary, finding that a significant branch of Earth's biodiversity previously absent from the literature—in this case mollusks—exhibits mutation rate characteristics that match both empirical distributions and theoretical predictions.

Methods

Sample Collection and DNA Extraction

Samples for this study were derived from the White Abalone Captive Breeding Program at the Bodega Marine Lab of UC Davis, and offspring were sampled at the NOAA Southwest Fisheries Science Center in La Jolla, CA. White abalone at these

facilities are bred and reared in captivity under the National Marine Fisheries Service (NMFS) Endangered Species Act (ESA) Section 10(a)(1)(A) Research Permit 14344-3R. Controlled crosses were performed by pooling isolated gametes from each mother and father in vitro.

DNA extractions were prepared from epipodial tissue sampled from each individual. Samples from parents were obtained by excising a single epipodial tentacle from a live animal, while samples from offspring required both tentacle and epipodial fringe due to their small size. All samples came from fresh mortalities. Extractions were performed following the protocol of Gemmel and Akiyama (1996).

Sequencing Library Preparation

DNA extract concentration was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen) and fragment length was found with the Fragment Analyzer Genomic 50 kb DNA Kit (Agilent). Sequencing libraries were prepared following the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB) standard recommendations, using Y-Adapters rather than the NEBNext Adapters. All samples were diluted with 1× TE (10 mM Tris pH 8.0, 1 mM EDTA) to reach ≤100 ng inputs and incubated for 6 min during the enzymatic fragmentation step. Libraries were amplified for 7 to 8 cycles using dual unique indexes and were eluted in a final volume of 21 µL of 0.1× TE. DNA concentration was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen) and fragment length was determined using the Fragment Analyzer High Sensitivity NGS Kit (Agilent). Each library was then screened via low-coverage sequencing on an Illumina Nextseq 2000 (2 × 150 bp). Libraries were then sequenced on an Illumina NovaSeq X (2 × 150 bp) with a target depth of 50× genome-wide coverage at Duke University School of Medicine's Sequencing and Genomics Technologies Core Facility.

Alignment and Variant Calling

The read alignment, variant calling, and initial filtering steps were all directly informed by best practices established for shotgun resequencing data (Mirchandani et al. 2024). With our raw sequencing reads, we performed initial quality control and trimmed Illumina adapters by running *fastp* (v0.23.4) with default parameters on each lane *x* sample combination of paired-end reads (Chen et al. 2018). We then merged the post-*fastp* reads for each sample. We aligned these reads to the white abalone reference genome (https://abalone.dbgenome.org/) via *bwa-mem* with *–p* to indicate interleaved paired-end fastq input, *–M* to mark short split hits as secondary for compatibility with Picard, and *–a* to output alignments of unpaired reads (Li and Durbin 2009). Following mapping, we marked duplicate reads in two steps using *Sentieon*: (1) *driver –algo LocusCollector –fun score_info*, then (2) *driver –algo Dedup* providing the output of step (1) with *–score_info* (Kendig et al. 2019).

To begin variant calling, we used *Sentieon driver –algo Haplotyper –emit_mode gvcf* to create gvcf files for each individual sample. We then performed joint genotyping on this set of gvcfs using *Sentieon driver –algo GVCFtyper*, which produced cohort level variant sites across the white abalone genome. Finally, we filtered these variant sites using *GATK VariantFiltration*. We performed initial filtering on SNPs and INDELs independently, excluding SNPs with $QUAL < 30.0$, $QD < 2.0$, $FS > 60.0$, $MQ < 40.0$, $MQRankSum < -12.5$, $ReadPosRankSum < -8.0$ or $SOR > 3.0$ and excluding INDELs with $QUAL < 30.0$, $QD < 2.0$, $FS > 200.0$, $ReadPosRankSum < -20.0$, and $SOR > 10.0$ (Mirchandani et al. 2024).

We also called and filtered variants in parallel with *bcftools v1.13* (Danecek et al. 2021). First, we generated "pileup" files for the set of all samples by running *bcftools mpileup –annotate FORMAT/AD*, *FORMAT/ADF*, *FORMAT/ADR*, *FORMAT/DP*, *FORMAT/SP*, *INFO/AD*, *INFO/ADF*, *INFO/ADR –min-MQ 20 –min-BQ 20 –max-depth 500* on all input bam files. We then piped this output into *bcftools call -m –ploidy 2* to generate vcfs for the whole set of samples. Filtering was performed using the same criteria as stated above for the GATK variants, with the exception of all FisherStrand ("FS") filters, which were not available via the *bcftools* method.

Kinship Matrix

To validate our sample pedigrees, we estimated a kinship matrix with *plink2 v2.00a4.4LM –make-king square –allow-extra-chr* using the set of all genome-wide biallelic SNPs (Chang et al. 2015). We visualized the resulting kinship matrix in *R 4.3.3* and confirmed that families showed the expected degree of relatedness (supplementary fig. S1, Supplementary Material online).

Masking and Determining the Callable Genome

Accurate estimation of GMRs requires dividing the number of observed mutations by the proportion of the genome where such mutations could potentially be observed given (a) genome quality and (b) sequencing effort. We combined several masking approaches to determine this denominator.

We quantified mappability of the white abalone reference genome with *genmap v1.3.0* (Pockrandt et al. 2020). First, we indexed the genome with *genmap index*, then we determined the mappability of 150 length kmers with up to 2 mismatches using *genmap map -K 150 -E 2*. We then retained all regions where the 150 × 2 mappability score was less than 1.0 to create a "negative mappability mask", or a list of regions with poor mappability to exclude from downstream analyses.

In addition to the above mappability mask, we also generated masks based on sequencing depth for each individual. First, we generated base-pair level resolution depth files with *samtools depth -a* for each sample, and determined the mean, median, and standard deviation of sequencing depth based on the first (largest) chromosome in the white abalone reference genome. Given these parameters, we then identified regions for each individual where read depth was either less than 20 or greater than the mean read depth plus two standard deviations. Such regions, either too low to reliably call heterozygotes or outside the standard coverage distribution for each individual, were designated as regions to mask ("negative per-sample mask") from subsequent analyses.

Finally, for each family, we combined the above negative mappability mask and negative per-sample masks to create a conservative set of regions to exclude from mutation rate estimation. The genome remaining after this exclusion is referred to as the "callable genome", and averaged around 75% to 80% (supplementary table S1, Supplementary Material online).

Mutation Rate Estimation

The additional variant filtering and mutation rate estimation described below directly follows best practices established for such studies (Bergeron et al. 2022, 2023).

With variant calls from *GATK* and *bcftools* as well as callable regions of the genome, we then proceeded with mutation rate estimation. First, for each potential trio (two parents + one offspring), we selected variants within the callable regions that were heterozygous in the offspring and homozygous in the parents. We referred to these as candidate de novo mutations, and further filtered this set following the stringent criteria outlined in Bergeron et al. (2022). Specifically, we retained mutations with (a) genotype quality (GQ) greater than 60 in each member of the trio, (b) no reads containing the mutation present in either parent, (c) allelic balance > 0.30, and (d) no occurrence in other offspring except those sharing one of the parents.

We applied this same filtering pipeline to both GATK and bcftools variant calls and intersected those that appeared via both methods under the assumption that those appearing in only one approach were more likely to be spurious (Bergeron et al. 2022; Sendell-Price et al. 2023). Of the 2,817 de novo mutations detected via GATK and 1,023 detected via bcftools, 107 were shared. To understand why so few variants were shared by both *GATK* and *bcftools* we directly inspected the alignments at 20 randomly chosen variant positions, half of which passed GATK but not bcftools, and vice versa. Of those 20, only 2 (10%) had alignments that looked like genuine de novo mutations based on allelic balance and the absence of variant reads in either parent. For those that failed this test, the most common reasons were (1) reads present in one or both parents carrying the putative mutation, (2) low allelic balance (<30%) in the child carrying the putative mutation, or (3) no reads carrying the variant in the child. While this last reason is the most puzzling, it is not uncommon to see very low frequency variant calls with no obvious alignment support.

We used a similar approach to approximate the FDR. We manually inspected the read alignments of each trio for 20 of the 107 de novo mutations in IGV (Robinson 2017). Only one of the 20 mutations did not display convincing alignment-based evidence with low alternate allelic depth and map quality <40. Despite this result, the corresponding variant call appeared to have passed our strict filters because of read realignment during *GATK/bcftools* variant calling. Therefore, we set the FDR to 5%.

To formally calculate the FNR, we first set out to identify high quality variants that would have to be heterozygous in offspring (0/1) based on confident parent genotypes (e.g. 0/0 and 1/1). We then calculated what proportion of such variants, assumed to be true heterozygotes, would not pass the allelic balance, GQ, and depth filters listed above. This ratio was calculated on a per-child basis, and across all children exhibited a median of 0.139 and a mean of 0.181. Given that this approach may overestimate the FNR because of overconfidence in parent genotypes, we opted to implement the median value of 0.139 in our FNR correction.

Finally, our reported mutation rates were calculated as:

$$\mu = \frac{nbcandidateDNMs \times (1 - \mathrm{FDR})}{2 \times \mathrm{CG} \times (1 - \mathrm{FNR})}$$

where *nbcandidateDNMs* was the count of observed de novo mutations, and CG was the size, in base pairs, of the callable genome. We determined the confidence interval for the mutation rate based on the confidence interval reported by the *R* function *t.test* on the reported values of all nine offspring with estimates.

Parent-of-Origin Tracing

We used read phasing in order to determine which parents contributed de novo mutations. To do this, we applied *POOHA* (https://github.com/besenbacher/POOHA/) to each mother-father-offspring trio of bam alignments with the options *–min-parents-GQ 60 –min-child-GQ 60 –max-marker-distance 10000 –output_variants germline*. Due to insufficient haplotype information, we were only able to trace 33 of the 107 total mutations to a parent.

Mutation Rate Comparisons

To compare our estimated mutation rate to other published rates in multicellular eukaryotes (Fig. 1), we retrieved the set of estimates compiled by Wang and Obbard (2023). We also obtained the corresponding phylogeny for this list of species using *TimeTree* (Kumar et al. 2017), omitting *Amphilophus* and *Marasmius oreades* when they were not located in the database. All visualizations were executed in *R* with the packages *ggtree* (Yu et al. 2017) and *aplot* (Yu 2023).

Analysis of Polymorphisms in Wild-Caught *H. sorenseni* and *H. cracherodii*

To examine sequence diversity in wild populations, we first downloaded whole-genome shotgun data from NCBI's SRA database for *H. sorenseni* ($n = 11$) and *H. cracherodii* ($n = 11$), the only species which had multiple wild-caught individuals with such data (supplementary extended data table 1, Supplementary Material online). We then generated variant calls for these data following the pipeline detailed in "Alignment and variant calling". For both species, we used their respective reference genomes (supplementary extended data table 1, Supplementary Material online). At the "GVCFtyper" stage, in which cohort level VCFs are generated, we specified "–emit_mode ALL" in order to produce VCFs containing both invariant and variant sites. We filtered variant and invariant sites separately. For variant sites, we used the exact criteria specified for GATK filtering in "Alignment and variant calling". For invariant sites, we filtered based on site quality ("QUAL > 30") and the fraction of missing genotypes at a site ("F_MISSING < 0.25").

We then proceeded to analyze the variant + invariant site VCFs with *pixy* (Korunes and Samuk 2021), which estimates sequence diversity while accounting for the pitfalls in generating such estimates from heterogeneous data with high rates of missingness. We ran *pixy –stats pi –window_size 10000*, then examined the distribution of site missingness in 10 kb windows to determine filtering heuristics for downstream analysis. We reported values of π after retaining windows with more than 8,000 sites for *H. sorenseni* and more than 6,000 sites for *H. cracherodii*.

Demographic Inference

We reconstructed demographic histories of multiple *Haliotis* species with *MSMC2* (Schiffels and Wang 2020) using our new mutation rate. First, we downloaded whole-genome shotgun data from NCBI's SRA database for these additional *Haliotis* species, as well as their respective reference genomes (supplementary extended data table 1, Supplementary Material online). We then analyzed these data following the exact pipeline detailed above in "Alignment and variant calling". Following the production of filtered variant calls, we generated two data masks: (1) reference genome mappability masks

following the pipeline detailed above in "Masking and determining the callable genome" and (2) sequencing depth masks following *msmc-tools* recommendations. For the latter, we used the *msmc-tools* script *bamCaller.py* on the output of *samtools mpileup -q 20 -Q 20 -C 50 -u | bcftools call -c -V indels –ploidy 2*.

With our input variant calls, reference genome mask, and sample sequencing depth mask, we then generated the *MSMC2* input with the *msmc-tools* script *generate_multihetsep.py*. We created bootstrap replicates of this input data using the *multihetsep_bootstrap.py*, specifying *-n 20 -s 5000000 –chunks_per_chromosome 10 –nr_chromosomes 20* for all species except *H. laevigata*, which had a highly fragmented genome assembly. For *H. laevigata*, we specified *-n 20 -s 100000 –chunks_per_chromosome 10 –nr_chromosomes 1000* to generate bootstrap replicates with similar characteristics as the input data. Finally, we ran *MSMC2* on the original data and all bootstrap replicates with the time segment pattern specified as *-p 25*1 + 1*2 + 1*3*. All demographic histories were then visualized in *R* and scaled by the newly estimated mutation rate of 8.60×10^{-9} and generation time of 1.

Phylogenomic Inference

To supplement the *H. sorenseni* genome for phylogenomic analysis, we downloaded reference genome assemblies for five *Haliotis* species as well as the outgroup *G. magus* from NCBI's RefSeq Database (supplementary extended data table 1, Supplementary Material online). We subsequently annotated these genomes for BUSCO gene content using *compleasm* (v.02.6) with the *mollusca_odb10* database (Huang and Li 2023). Following this, we identified all genes that were present only as complete single copies in each of the seven taxa (six *Haliotis* + *G. magus*), and extracted the corresponding spliced CDS nucleotide sequences from each taxon using *gffread -x* (v.0.12.8) (Pertea and Pertea 2020). For each gene, we then aligned the seven sequences with *mafft* (v7.526) (Katoh and Standley 2013) and quality trimmed the resulting alignments with *trimal -gt 0.50 -cons 50* (v.1.4.rev15) (Capella-Gutiérrez et al. 2009). From this set of trimmed alignments, we selected only those greater than 900 bp in length, resulting in 2,525 total genes. We concatenated these genes into a single alignment with *seqkit concat* (v.0.16.1) (Shen et al. 2016) and inferred a phylogeny with *iqtree -bb 1000 -bnni -m MFP* (v.2.3.4) (Minh et al. 2020). We plotted the *G. magus*-rooted tree in *R* using the package *ggtree* (v.3.10.1) (Yu et al. 2017).

Having observed 100% bootstrap support at all nodes in this initial tree, we then proceeded with inference of species divergence times under this topology. Motivated by our estimate of the germline mutation rate and a lack of obvious fossil calibrations for *Haliotis*, we opted for a fossil-free approach under the MSC following principles outlined in Tiley et al. (2020). To do this, we first estimated the extent to which each of the 2,525 genes used for tree inference evolved in clock-like fashion along the seven lineages. Specifically, we estimated the parameter "rate.coefficientOfVariation" (CoV) for each gene alignment individually in *BEAST* (v.2.6.6) (Bouckaert et al. 2019). Following each BEAST analysis, we (a) filtered for genes which reached an effective sample size (ESS) greater than 200 for the posterior and CoV and (b) mean CoV < 0.50 and the upper and lower limits of the 95% HPD for CoV less than 1 and 0.1, respectively. This filtering resulted in 393 clock-like genes for further analysis.

With these genes and our inferred topology, we performed Bayesian estimation of divergence times under the MSC with *BPP* (Flouri et al. 2018). Specifically, we applied the A00 model to a partitioned alignment of 150 randomly selected genes from the original set of 393 genes and provided the fixed topology estimated by *IQTREE*.

We then recalibrated branch lengths in the inferred tree to real time in R using the *msc2time.r* function from the *R* package *bppr* (Campbell et al. 2021). To do this, we specified the mean mutation rate *u.mean* as 8.60×10^{-9} and the standard deviation *u.sd* as 3.26×10^{-9} based on our findings. For lack of a precise generation time for any abalone species, we used data from growth-reproduction curves in Eastern Pacific abalone like *H. sorenseni* and *H. rufescens* as a crude proxy. Specifically, we set the mean generation time *g.mean* as 6 and the standard deviation *g.sd* as 2, as most of these abalone start reproducing by at least 4 years of age and continue on into adulthood, with some evidence for reproductive senescence with age (Rogers-Bennett et al. 2004). After using these parameters to obtain a rate for the conversion of substitution rates to real time, we rescaled all nodes, edges, and 95% confidence intervals. Finally, we reran *BPP* to confirm convergence on these parameters across independent runs.

Species Range Maps

The world map was obtained from the *R* package *rnaturalearth v0.3.2* (Massicotte and South 2023) with the function *ne_countries* and transformed to a Pacific-centered Robinson projection with the function *st_transform(st_crs[+proj = robin + lon_0 = 0 +x_0 = 0 +y_0 = 0 +datum = WGS84 +units = m + pm = 180 +no_defs])* from the *R* package *sf v1.0.8* (Pebesma 2018). Species range shapefile were downloaded from the IUCN Red List of Species. We visualized the world map and range maps together with *ggplot2 v3.3.6*.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank the White Abalone Captive Breeding Program at the Bodega Marine Lab of UC Davis for their work in conserving this species.

Author Contributions

T.B.W. conceptualized the study with assistance from K.H. T.B.W. performed all bioinformatics analyses. S.F. and H.C. performed quality control of DNA extracts and generated all sequencing libraries. J.H. tracked down white abalone families and provided DNA extracts. B.S. provided funding for sequencing and analysis. T.B.W. wrote the manuscript with input and approval from all authors.

Funding

Brock Wooldridge was supported by the National Science Foundation Division of Ocean Sciences (NSF–OCE); No. 2307479).

Data Availability

All novel sequence data generated by this study will be available on NCBI's SRA database under PRJNA1186246 upon publication of this manuscript. Code used to perform these analyses will be made available on the lead author's github at https://github.com/twooldridge/.

References

Allio R, Donega S, Galtier N, Nabholz B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol*. 2017:34(11):2762–2772. https://doi.org/10.1093/molbev/msx197.

Andrews AH, Leaf RT, Rogers-Bennett L, Neuman M, Hawk H, Cailliet GM. Bomb radiocarbon dating of the endangered white abalone (*Haliotis sorenseni*): investigations of age, growth and lifespan. *Mar Freshw Res*. 2013:64(11):1029–1039. https://doi.org/10.1071/MF13007.

Babcock R, Keesing J. Fertilization biology of the abalone *Haliotis laevigata*: laboratory and field studies. *Can J Fish Aquat Sci*. 1999:56(9):1668–1678. https://doi.org/10.1139/f99-106.

Bánki O, Roskov Y, Döring M, Ower G, Hernández Robles DR, Plata Corredor CA, Stjernegaard Jeppesen T, Örn A, Vandepitte L, Hobern D, *et al. Catalogue of life (version 2024-07-18)*. Amsterdam (NL): Catalogue of Life; 2024.

Bergeron LA, Besenbacher S, Bakker J, Zheng J, Li P, Pacheco G, Sinding MS, Kamilari M, Gilbert MTP, Schierup MH, *et al*. The germline mutational process in rhesus macaque and its implications for phylogenetic dating. *GigaScience*. 2021:10(5):giab029. https://doi.org/10.1093/gigascience/giab029.

Bergeron LA, Besenbacher S, Turner T, Versoza CJ, Wang RJ, Price AL, Armstrong E, Riera M, Carlson J, Chen HY, *et al*. The mutationathon highlights the importance of reaching standardization in estimates of pedigree-based germline mutation rates. *eLife*. 2022:11(January):e73577. https://doi.org/10.7554/eLife.73577.

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St Leger J, Shao C, *et al*. Evolution of the germline mutation rate across vertebrates. *Nature*. 2023:615(7951):285–291. https://doi.org/10.1038/s41586-023-05752-y.

Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol*. 2019:3(2):286–292. https://doi.org/10.1038/s41559-018-0778-x.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, *et al*. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2019:15(4):e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Campbell CR, Tiley GP, Poelstra JW, Hunnicutt KE, Larsen PA, Lee HJ, Thorne JL, Dos Reis M, Yoder AD. Pedigree-based and phylogenetic methods support surprising patterns of mutation rate and spectrum in the gray mouse lemur. *Heredity (Edinb)*. 2021:127(2):233–244. https://doi.org/10.1038/s41437-021-00446-5.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009:25(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015:4(1):7. https://doi.org/10.1186/s13742-015-0047-8.

Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Crosson LM, Friedman CS. Withering syndrome susceptibility of northeastern Pacific abalones: a complex relationship with phylogeny and thermal experience. *J Invertebr Pathol*. 2018:151(January):91–101. https://doi.org/10.1016/j.jip.2017.11.005.

Cutter AD, Jovelin R, Dey A. Molecular hyperdiversity and evolution in very large populations. *Mol Ecol*. 2013:22(8):2074–2095. https://doi.org/10.1111/mec.12281.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al*. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021:10(2):giab008. https://doi.org/10.1093/gigascience/giab008

de Manuel M, Wu FL, Przeworski M. A paternal bias in germline mutation is widespread in amniotes and can arise independently of cell division numbers. *eLife*. 2022:11(August):e80008. https://doi.org/10.7554/eLife.80008.

Dohrmann M, Wörheide G. Dating early animal evolution using phylogenomic data. *Sci Rep*. 2017:7(1):3599. https://doi.org/10.1038/s41598-017-03791-w.

Estes JA, Lindberg DR, Wray C. Evolution of large body size in abalones (*Haliotis*): patterns and implications. *Paleobiology*. 2005:31(4):591–606. https://doi.org/10.1666/04059.1.

Flouri T, Jiao X, Rannala B, Yang Z. Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol Biol Evol*. 2018:35(10):2585–2593. https://doi.org/10.1093/molbev/msy147.

Frankham R. Effective population size/adult population size ratios in wildlife: a review. *Genet Res (Camb)*. 1995:66(2):95–107. https://doi.org/10.1017/S0016672300034455.

Geiger DL, Groves LT. Review of fossil abalone (Gastropoda: Vetigastropoda: Haliotidae) with comparison to recent species. *J Paleontol*. 1999:73(5):872–885. https://doi.org/10.1017/S0022336000040713.

Gemmell NJ, Akiyama S. An efficient method for the extraction of DNA from vertebrate tissues. *Trends Genet*. 1996:9(12):338–339. https://doi.org/10.1016/s0168-9525(96)80005-9.

Gruenthal KM, Burton RS. Genetic diversity and species identification in the endangered white abalone (*Haliotis sorenseni*). *Conserv Genet*. 2006:6(6):929–939. https://doi.org/10.1007/s10592-005-9079-4.

Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Ruegg K, Palstra F. Understanding and estimating effective population size for practical application in marine species management. *Conserv Biol*. 2011:25(3):438–449. https://doi.org/10.1111/j.1523-1739.2010.01637.x.

Harrang E, Lapègue S, Morga B, Bierne N. A high load of non-neutral amino-acid polymorphisms explains high protein diversity despite moderate effective population size in a marine bivalve with sweepstakes reproduction. *G3 (Bethesda)*. 2013:3(2):333–341. https://doi.org/10.1534/g3.112.005181.

Harris K, Pritchard JK. Rapid evolution of the human mutation spectrum. *eLife*. 2017:6:e24284. https://doi.org/10.7554/eLife.24284.

Hedgecock D, Pudovkin AI. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull Mar Sci*. 2011:87(4):971–1002. https://doi.org/10.5343/bms.2010.1051.

Hedrick P. Large variance in reproductive success and the N_e/N ratio. *Evolution*. 2005:59(7):1596–1599. https://doi.org/10.1111/j.0014-3820.2005.tb01809.x.

Hoban S, Bruford MW, da Silva JM, Funk WC, Frankham R, Gill MJ, Grueber CE, Heuertz M, Hunter ME, Kershaw F, *et al*. Genetic diversity targets and indicators in the CBD post-2020 global biodiversity framework must be improved. *Biol Conserv*. 2020:248:108654. https://doi.org/10.1016/j.biocon.2020.108654.

Hobday AJ, Tegner MJ, Haaker PL. Over-exploitation of a broadcast spawning marine invertebrate: decline of the white abalone. *Rev Fish Biol Fish*. 2000:10(4):493–514. https://doi.org/10.1023/A:1012274101311.

Hoeh WR, Stewart DT, Sutherland BW, Zouros E. Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). *Mol Biol Evol*. 1996:13(2):418–421. https://doi.org/10.1093/oxfordjournals.molbev.a025600.

Huang N, Li H. Compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics*. 2023:39(10):btad595. https://doi.org/10.1093/bioinformatics/btad595.

Jamieson G. Marine invertebrate conservation: evaluation of fisheries over-exploitation concerns. *Integr Comp Biol*. 1993:33(6):551–567. https://doi.org/10.1093/icb/33.6.551.

Jamieson IG, Allendorf FW. How does the 50/500 rule apply to MVPs? *Trends Ecol Evol*. 2012:27(10):578–584. https://doi.org/10.1016/j.tree.2012.07.001.

Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, *et al*. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*. 2017:549(7673):519–522. https://doi.org/10.1038/nature24018.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kendig KI, Baheti S, Bockol MA, Drucker TM, Hart SN, Heldenbrand JR, Hernaez M, Hudson ME, Kalmbach MT, Klee EW, *et al*. Sentieon DNASeq variant calling workflow demonstrates strong computational performance and accuracy. *Front Genet*. 2019:10(August):736. https://doi.org/10.3389/fgene.2019.00736.

Kenner MC. *Black abalone surveys at Naval Base Ventura County, San Nicolas Island, California—2020, annual report*. Reston (VA): US Geological Survey; 2021.

Korunes KL, Samuk K. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour*. 2021:21(4):1359–1368. https://doi.org/10.1111/1755-0998.13326.

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017:34(7):1812–1819. https://doi.org/10.1093/molbev/msx116.

Lafferty KD, Behrens MD, Davis GE, Haaker PL, Kushner DJ, Richards DV, Taniguchi IK, Tegner MJ. Habitat of endangered white abalone, *Haliotis sorenseni*. *Biol Conserv*. 2004:116(2):191–194. https://doi.org/10.1016/S0006-3207(03)00189-7.

Launey S, Hedgecock D. High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics*. 2001:159(1):255–265. https://doi.org/10.1093/genetics/159.1.255.

Leighton DL. Laboratory observations on the early growth of the abalone, *Haliotis sorenseni*, and the effect of temperature on larval development and settling success. *Fish Bull*. 1972:70(2):373–381.

Li A, Dai H, Guo X, Zhang Z, Zhang K, Wang C, Wang X, Wang W, Chen H, Li X, *et al*. Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity. *Commun Biol*. 2021:4(1):1287. https://doi.org/10.1038/s42003-021-02823-6.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009:25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lin Y, Darolti I, van der Bijl W, Morris J, Mank JE. Extensive variation in germline de novo mutations in *Poecilia reticulata*. *Genome Res*. 2023:33(8):1317–1324. https://doi.org/10.1101/gr.277936.123.

Lindsay SJ, Rahbari R, Kaplanis J, Keane T, Hurles ME. Similarities and differences in patterns of germline mutation between mice and humans. *Nat Commun*. 2019:10(1):4053. https://doi.org/10.1038/s41467-019-12023-w.

Lynch M. Evolution of the mutation rate. *Trends Genet*. 2010:26(8):345–352. https://doi.org/10.1016/j.tig.2010.05.003.

Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet*. 2016:17(11):704–714. https://doi.org/10.1038/nrg.2016.104.

Masonbrink RE, Purcell CM, Boles SE, Whitehead A, Hyde JR, Seetharam AS, Severin AJ. An annotated genome for *Haliotis rufescens* (red abalone) and resequenced green, pink, pinto, black, and white abalone species. *Genome Biol Evol*. 2019:11(2):431–438. https://doi.org/10.1093/gbe/evz006.

Massicotte P, South A. *Rnaturalearth: world map data from natural earth*; 2023.

Mccormick TB, Navas G, Buckley LM, Biggs C. Effect of temperature, diet, light, and cultivation density on growth and survival of larval and juvenile white abalone *Haliotis sorenseni* (Bartsch, 1940). *J Shellfish Res*. 2016:35(4):981–992. https://doi.org/10.2983/035.035.0421.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mirchandani CD, Shultz AJ, Thomas GWC, Smith SJ, Baylis M, Arnold B, Corbett-Detig R, Enbody E, Sackton TB. A fast, reproducible, high-throughput variant calling workflow for population genomics. *Mol Biol Evol*. 2024:41(1):msad270. https://doi.org/10.1093/molbev/msad270.

Nadachowska-Brzyska K, Konczal M, Babik W. Navigating the temporal continuum of effective population size. *Methods Ecol Evol*. 2022:13(1):22–41. https://doi.org/10.1111/2041-210X.13740.

Nei M. The frequency distribution of lethal chromosomes in finite populations. *Proc Natl Acad Sci U S A*. 1968:60(2):517–524. https://doi.org/10.1073/pnas.60.2.517.

Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. *Genetics*. 1981:97(1):145–163. https://doi.org/10.1093/genetics/97.1.145.

Neuman M, Tissot B, VanBlaricom G. Overall status and threats assessment of black abalone (*Haliotis cracherodii* Leach, 1814) populations in California. *J Shellfish Res*. 2010:29(3):577–586. https://doi.org/10.2983/035.029.0305.

Palstra FP, Fraser DJ. Effective/census population size ratio estimation: a compendium and appraisal. *Ecol Evol*. 2012:2(9):2357–2365. https://doi.org/10.1002/ece3.329.

Palstra FP, Ruzzante DE. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol*. 2008:17(15):3428–3447. https://doi.org/10.1111/j.1365-294X.2008.03842.x.

Pebesma E. Simple features for R: standardized support for spatial vector data. *R J*. 2018:10(1):439. https://doi.org/10.32614/RJ-2018-009.

Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Res*. 2020:9:ISCB Comm J-304. https://doi.org/10.12688/f1000research.23297.1.

Plough LV, Shin G, Hedgecock D. Genetic inviability is a major driver of type III survivorship in experimental families of a highly fecund marine bivalve. *Mol Ecol*. 2016:25(4):895–910. https://doi.org/10.1111/mec.13524.

Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K. GenMap: ultra-fast computation of genome mappability. *Bioinformatics*. 2020:36(12):3687–3692. https://doi.org/10.1093/bioinformatics/btaa222.

Ponder W, Lindberg DR. *Phylogeny and evolution of the mollusca*. Berkeley (CA): University of California Press; 2008.

Popovic I, Bergeron LA, Bozec Y-M, Waldvogel A-M, Howitt SM, Damjanovic K, Patel F, Cabrera MG, Wörheide G, Uthicke S, *et al*. High germline mutation rates, but not extreme population outbreaks, influence genetic diversity in a keystone coral predator. *PLoS Genet*. 2024:20(2):e1011129. https://doi.org/10.1371/journal.pgen.1011129.

Robinson JA, Kyriazis CC, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, Robertson KM, Fontaine MC, Wayne RK, Lohmueller KE, Taylor BL, *et al*. The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science*. 2022:376(6593):635–639. https://doi.org/10.1126/science.abm1742.

Robinson P. Integrative genomics viewer (IGV): visualizing alignments and variants. In: *Computational exome and genome analysis*. Chapman and Hall/CRC; 2017. p. 233–245.

Rogers-Bennett L, Aquilino KM, Catton CA, Kawana SK, Walker BJ, Ashlock LW, Marshman BC, Moore JD, Taniguchi IK, Gilardi KV, *et al*. Implementing a restoration program for the endangered white abalone (*Haliotis sorenseni*) in California. *J Shellfish Res*. 2016:35(3):611–618. https://doi.org/10.2983/035.035.0306.

Rogers-Bennett L, Dondanville RF, Kashiwada J. Size specific fecundity of red abalone (*Haliotis rufescens*): evidence for reproductive senescence? *J Shellfish Res*. 2004:23(2):553–560.

Rogers-Bennett L, Haaker PL, Huff TO, Dayton PK. Estimating baseline abundances of abalone in California for restoration. *Calif Coop Ocean Fish Investig Rep*. 2002:4:97–111.

Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*. 2012:13(10):745–753. https://doi.org/10.1038/nrg3295.

Schiffels S, Wang K. MSMC and MSMC2: the multiple sequentially Markovian coalescent. *Methods Mol Biol*. 2020:2090:147–166. https://doi.org/10.1007/978-1-0716-0199-0_7.

Sendell-Price AT, Tulenko FJ, Pettersson M, Kang D, Montandon M, Winkler S, Kulb K, Naylor GP, Phillippy A, Fedrigo O, *et al*. Low mutation rate in epaulette sharks is consistent with a slow rate of evolution in sharks. *Nat Commun*. 2023:14(1):6628. https://doi.org/10.1038/s41467-023-42238-x.

Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*. 2016:11(10):e0163962. https://doi.org/10.1371/journal.pone.0163962.

Smith J, Coop G, Stephens M, Novembre J. Estimating time to the common ancestor for a beneficial allele. *Mol Biol Evol*. 2018:35(4):1003–1017. https://doi.org/10.1093/molbev/msy006.

Stephens PA, Sutherland WJ, Freckleton RP. What is the allee effect? *Oikos*. 1999:87(1):185–190. https://doi.org/10.2307/3547011.

Stierhoff KL, Mau SA, Murfin DW, Neumann M. *White abalone at San Clemente Island: population estimates and management recommendations*. USA: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center; 2014.

Stierhoff KL, Neuman M, Butler JL. On the road to extinction? Population declines of the endangered white abalone, *Haliotis sorenseni*. *Biol Conserv*. 2012:152(August):46–52. https://doi.org/10.1016/j.biocon.2012.03.013.

Streit K, Geiger DL, Lieb B. Molecular phylogeny and the geographic origin of Haliotidae traced by haemocyanin sequences. *J Molluscan Stud*. 2006:72(1):105–110. https://doi.org/10.1093/mollus/eyi048.

Sturtevant AH. Essays on evolution. I. On the effects of selection on mutation rate. *Q Rev Biol*. 1937:12(4):464–467. https://doi.org/10.1086/394543.

Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DST, Murali SC, Williams LE, Doddapaneni H, Muzny DM, *et al*. Reproductive longevity predicts mutation rates in primates. *Curr Biol*. 2018:28(19):3193–3197.e5. https://doi.org/10.1016/j.cub.2018.08.050.

Tiley GP, Poelstra JW, Dos Reis M, Yang Z, Yoder AD. Molecular clocks without rocks: new solutions for old problems. *Trends Genet*. 2020:36(11):845–856. https://doi.org/10.1016/j.tig.2020.06.002.

Tutschulte T, Connell JH. Reproductive biology of three species of abalones (*Haliotis*) in southern California. *Veliger*. 1981:23(3):195–206.

Tutschulte TC. The comparative ecology of three sympatric abalones [doctoral dissertation, University]. ProQuest; 1976.

VanBlaricom G, Neuman M, Butler JL, De Vogelaere A, Gustafson RG, Mobley C, Richards D, Rumsey S, Louise Taylor B. *Status review report for black abalone*. Long Beach (CA): National Marine Fisheries Service; 2009.

Venn O, Turner I, Mathieson I, de Groot N, Bontrop R, McVean G. Strong male bias drives germline mutation in chimpanzees. *Science*. 2014:344(6189):1272–1275. https://doi.org/10.1126/science.344.6189.1272.

Vileisis A. *Abalone: the remarkable history and uncertain future of California's iconic shellfish*. Corvallis (OR): Oregon State University Press; 2020.

Wang Y, Obbard DJ. Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. *Evol Lett*. 2023:7(4):216–226. https://doi.org/10.1093/evlett/qrad027.

Wilder AP, Supple MA, Subramanian A, Mudide A, Swofford R, Serres-Armero A, Steiner C, Koepfli KP, Genereux DP, Karlsson EK, *et al*. The contribution of historical processes to contemporary extinction risk in placental mammals. *Science*. 2023:380(6643):eabn5856. https://doi.org/10.1126/science.abn5856.

Wooldridge B, Orland C, Enbody E, Escalona M, Mirchandani C, Corbett-Detig R, Kapp JD, Fletcher N, Cox-Ammann K, Raimondi P, *et al*. Limited genomic signatures of population collapse in the critically endangered black abalone (*Haliotis cracherodii*). *Mol Ecol*. 2024:e17362. https://doi.org/10.1111/mec.17362.

Wright S. Evolution in Mendelian populations. *Genetics*. 1931:16(2):97–159. https://doi.org/10.1093/genetics/16.2.97.

Yu G. *Aplot: decorate a 'Ggplot' with associated information*; 2023.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017:8(1):28–36. https://doi.org/10.1111/2041-210X.12628.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, *et al*. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*. 2012:490(7418):49–54. https://doi.org/10.1038/nature11413.