

Supplementary Materials for

The vaquita is not doomed to extinction by inbreeding depression

Jacqueline A. Robinson, Christopher C. Kyriazis, Sergio F. Nigenda-Morales, Annabel C. Beichman, Lorenzo Rojas-Bracho, Kelly M. Robertson, Michael C. Fontaine, Robert K. Wayne, Kirk E. Lohmueller, Barbara L. Taylor, and Phillip A. Morin

correspondence to: jacqueline.robinson@ucsf.edu, ckyriazis@g.ucla.edu,
lrojasbracho@gmail.com, barbara.taylor@noaa.gov, klohueller@g.ucla.edu,
phillip.morin@noaa.gov

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S6

Materials and Methods

Samples and sequencing

We obtained tissue samples of 19 vaquitas sampled from 1985-2017, spanning roughly three vaquita generations. The 12 samples from 1985-1993 included 10 adults and two fetuses, implying that this sample alone may span ~2 generations. Additionally, the animal captured, biopsied and released in 2017 was a juvenile, suggesting that this individual was born at least 30 years after the oldest individuals in our dataset. Other sampled individuals that were fetuses or calves are noted in table S1.

All samples except one (ID = z0186934) were obtained opportunistically from beach-cast or bycaught animals, and stored in salt-saturated 20% DMSO at -20°C. The single sample from a live individual was collected under veterinary care and all animal handling was conducted under permit no. SGPA/DGVS/07534/17. The skin biopsy sample was stored without preservative at -80°C. Samples were transferred internationally under MMPA and CITES permits. DNA was extracted using the PureGene or DNeasy Tissue kits (Qiagen), following the manufacturer's recommendations, or using a salt/ethanol precipitation (29). DNA quality was checked by agarose gel electrophoresis and quantified using a Qubit-v3 fluorometer. Genomic libraries were prepared using the Swift Biosciences Inc. Accel-NGS single-strand 1S DNA Library Kit or using a blunt-end ligation method with double indexing (30, 31). Libraries were sequenced with Illumina HiSeq-X to generate 150 bp paired end reads.

Variant calling and filtering

We aligned vaquita whole genome sequence reads to the previously assembled vaquita reference genome (mPhoSin1.pri, GCF_008692025.1) (12) as follows. First, sequence reads were trimmed to remove low quality bases with BBDUK v38.72 (<http://sourceforge.net/projects/bbmap>). We then used an adapted version of the Best Practices workflow for the Genome Analysis Toolkit v3 (GATK v3.8-1-0-gf15c1c3ef, (32)) to perform variant calling. Initial processing steps were carried out per-sample prior to joint genotype calling. First, we used Picard v2.20.3-0 (<http://broadinstitute.github.io/picard>) MarkIlluminaAdapters to flag any adapter sequence contamination, and then aligned reads to the vaquita reference genome with BWA MEM v0.7.17-r1188 (33). Duplicate reads were then flagged with Picard MarkDuplicates and indel realignment was performed with GATK RealignerTargetCreator and IndelRealigner. Next, we performed base quality score recalibration (BQSR) using a set of high confidence variant calls from within each individual for use as the set of "known variants" — a procedure referred to as "bootstrapping" for cases where no database of known variation exists. These variants were called using Bcftools v1.9 (34) mpileup/call/filter with the following requirements: reads mapped in proper pairs, minimum read mapping quality score of 20, minimum base quality score of 30, at least one forward and one reverse read supporting the alternate allele, and no more than one alternate allele. One round of recalibration was sufficient to equalize the reported and expected base quality score distributions.

Genotype calling was performed in each individual with GATK HaplotypeCaller to produce gVCF files using the options `-ERC BP_RESOLUTION -out_mode EMIT_ALL_SITES` to call genotypes at all sites in the reference. We then generated joint VCF files by passing all individual gVCF files to GATK GenotypeGVCFs using the options `allSites -stand_call_conf 0`. These settings produced VCF files with genotypes at all sites in the genome, including invariant sites. We applied filters and masks to exclude low-quality genotypes from subsequent analyses.

Variants that were not bi-allelic SNPs were excluded. Variants failing recommended GATK hard filtering criteria (QD < 4, FS > 60, MQ < 40, MQRankSum < -12.5, ReadPosRankSum < -8, SOR > 3) were then filtered out, as well as sites with excess missingness (>25%) or excess heterozygosity (>75% of individuals heterozygous). Genotypes were filtered on a per-individual basis to exclude those with insufficient or excess read depth (<1/3x or >2x mean read depth for that individual). Individual genotypes were further filtered on the basis of allele balance (percent of reads carrying the reference allele out of total reads covering a site): homozygotes with >10% of reads carrying a different allele and heterozygotes with allele balance <20% or >80% were filtered out. We masked repetitive sequences identified with RepeatMasker v.open-4.0.9 (35) or Tandem Repeats Finder v4.09 (36). Finally, we masked regions of extreme SNP density (>0.01), which likely reflect technical artifacts, such as errors in the underlying assembly. Only data from the autosomes were used in our analyses.

Estimation of recent vaquita abundance

To estimate the trajectory of vaquita abundance in recent decades (as in Fig. 1A), we fit an exponential model consistent with point estimates from recent surveys (37–39). Specifically, our aim was to match the trend given by three abundance estimates from complete surveys of the vaquita distribution: an estimate of 567 in 1997 (37), an estimate of 245 in 2008 (38), and an estimate of 59 in 2015 (39). Based on these estimates, we fitted a model assuming two rates of decline. We assumed a constant exponential rate of decline for the period prior to 2012 that fit the 1997 and 2008 abundance estimates. Projecting this growth rate ($\lambda = 0.92$) into the past resulted in reaching the population size estimated from genetic data (~5,000 vaquitas) by 1972. The second rate of decline ($\lambda = 0.67$) solved the exponential equation between the projected number in 2012 and the number in 2019 (~10). This simple model approximated the estimated abundances from recent surveys well (see Fig. 1A).

Genomic diversity, runs of homozygosity, and relatedness

We define heterozygosity as the proportion of all called genotypes within a single individual that are heterozygous. This value represents the average number of pairwise differences per site between homologous chromosomes within a single individual. As we generated VCF files that included invariant positions, the denominator includes all sites with called genotypes, including those that are invariant and match the reference. The distribution of genome-wide heterozygosity in vaquitas revealed two notable peaks with high diversity across many individuals: one at the end of Chromosome 2 in eight individuals, and one in the middle of Chromosome 18 in twelve individuals (fig. S1). The peak at the end of Chromosome 2 is a 150 kb region containing one pseudogene, five orthologs of human immunoglobulin genes, and six unnamed genes. Several genes in this region have multiple stop codons, suggesting they are non-functional or that the annotations contain errors. The immunoglobulin genes in this region are also highly paralogous, and an elevated proportion of sites in this region fail numerous quality filters. The peak on Chromosome 18 is a 300 kb region that contains three genes (*FAM71A*, *OGT*, *CEP78*) that, to our knowledge, have no clear significance for fitness. These peaks of heterozygosity may represent technical artifacts rather than genuine signals of high germline diversity.

We identified runs of homozygosity (ROH) ≥ 1 Mb with three methods: the *--homozyg* function in Plink v1.90b6.12 (40), the *--LROH* function in Vcftools v0.1.16, (41), and the *roh* function in Bcftools v1.10.2 (42). Plink and Vcftools were run with default parameters. VCF

files containing all genotyped sites were provided for Plink, whereas files containing just variant sites were provided for Vcftools. For Bcftools, we provided variants only, used option `-G 30`, and specified that the function should use allele frequencies from the 15 unrelated vaquitas in identifying ROH (see below). All three methods yielded similar total lengths of ROH (fig. S2). Bcftools produced the highest summed ROH lengths overall, and we chose to use those results for the main text as they represent the most conservative estimate of the maximum degree of inbreeding.

To calculate the proportion of the genome in ROH, we generated a pseudo-genome with 100% homozygous genotypes and included this when re-running Bcftools *roh*. Including the pseudo-genome had no effect on ROH estimated in the true genomes but provided a maximum possible ROH length for a fully homozygous genome. The summed ROH length from this pseudo-genome (2,237,397,028 bp, slightly less than the full autosomal genome length of 2,237,877,686 bp) was then used as the denominator for calculating F_{ROH} in the vaquitas. To estimate the coalescence time for the two copies of the chromosome within a ROH (when identical segments originated from a single chromosomal segment in a common ancestor), we used a model in which the length of ROH due to inbreeding decays exponentially through recombination each generation (43). Specifically, we calculated the number of generations to the common ancestor (g) as $g = 100/(2*L)$, where L is the mean length of ROH in megabases (Mb). Here, we assumed a generation time of 11.9 years (13), and, in the absence of data on recombination rates in cetaceans, a constant recombination rate of 1 cM/Mb.

We estimated genetic relatedness among vaquitas in order to exclude closely related individuals from population allele frequency-based analyses (***Demographic Inference*** and ***Inference of the distribution of selection coefficients***). To calculate relatedness, we used the “KING-homo” method (44) in SNPRelate v1.24.0 (45) to estimate pairwise genetic distances between individuals on the basis of identity-by-state (IBS). Prior to estimating genetic distance and relatedness metrics, SNPs were pruned for linkage disequilibrium (using options `method = “composite”` and `ld.threshold = 0.5`), resulting in 59,420 SNPs for analysis. As two pairs of individuals were known to be first-degree relatives (mother and fetus pairs; see table S1), we were able to easily identify two other pairs of first-degree relatives that shared a similar degree of IBS/relatedness (kinship ~ 0.25 ; z0004390/z0004394, z0183496/z0185383), and one individual (z0004382) with a more distant relationship (kinship = 0.079) to one of the mother-fetus pairs (z0004380/z0004393) (fig. S3). Among all other individuals, there was no evidence of close relatedness (mean kinship = 7.73×10^{-3}). For all analyses which required using unrelated individuals only, we kept the higher coverage individual of each relative-pair, excluding z0001663, z0004380, z0004393, z0004394, and z0185383.

Demographic inference

To reconstruct the demographic history of the vaquita based on the neutral site frequency spectrum (SFS), we selected SNPs in putatively neutrally evolving regions of the genome, constructed a folded SFS, and used two demographic inference methods. We identified putatively neutral regions as those located >10 kb from coding sequences which did not overlap with CpG islands (identified with `cpg_lh`, from https://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64). To exclude highly conserved unannotated regions under strong evolutionary constraint, which may be functional, we aligned the remaining regions against the zebra fish genome using BLAST (v2.7.1(46)) and removed any region which had a hit below a $1e-10$ threshold (indicating high sequence similarity).

To create the neutral SFS, we used a modified version of EasySFS (<https://github.com/isaacovercast/easySFS>), which implements $\partial a\partial i$'s (47) hypergeometric projection to account for missing genotypes at some sites. The maximum number of SNPs was recovered using a projection value of 26 haploid, or 13 diploid, individuals. The number of monomorphic sites passing the projection threshold in neutral regions were counted and incorporated into the 0-bin of the SFS.

We used the neutral SFS for demographic inference under a coalescent framework with fastsimcoal2 v2.6 (48) and a diffusion approximation method with $\partial a\partial i$ (47). The four models we tested were: *1-Epoch model*: no size change, infers the ancestral size (N_{anc}) of the population; *2-Epoch model*: one size change from N_{anc} to the current size (N_{cur}) occurring T generations ago; *3-Epoch model*: two size changes, from N_{anc} to a reduction (N_{bot}) T_{bot} generations ago, followed by a recovery to the current size (N_{cur}) T_{cur} generations ago; *4-Epoch model*: three size changes, N_{anc} to a reduction or expansion (N_{bot}) T_{bot} generations ago, followed by a recovery or reduction (N_{rec}), T_{rec} generations ago and a final change T_{cur} generations ago to the current size (N_{cur}) (fig. S4).

In fastsimcoal2 (48), we estimated demographic parameters using 60 expectation conditional maximization (ECM) cycles, each with 500,000 simulated coalescent trees. The starting parameters were chosen from a uniform distribution with an enforced minimum value and flexible upper boundary. We compared the expected SFS under the fastsimcoal2 parameters to the empirical SFS and the multinomial log-likelihood was calculated. We performed 100 replicates of the inference to confirm that both parameters and log-likelihoods converged, and parameters with the maximum log-likelihood were chosen. All estimated effective population size parameter values were obtained as the number of haploids and converted to diploids, while time parameters were inferred as the number of generations before the present day. We calculated confidence intervals for the estimated parameters by generating 100 parametric bootstrap replicates of the SFS for the set of parameters in each model as follows. For each bootstrap replicate, we simulated 2,000,000 non-recombining segments of 1 kb (2 Gb total) in fastsimcoal2 under the parameters with the maximum likelihood for each model. Then, we estimated the parameters from these pseudo-observed data. We used the estimated values obtained from the data as the initial values for bootstrap parameter estimation. To determine which model explained our data better, we performed several tests. For each model we first evaluated the convergence for the estimated parameters and log-likelihood values in the five best replicates, then calculated the Akaike information criterion (AIC) for the replicate with the maximum log-likelihood and used these top replicates to carry out a likelihood ratio test (LRT) between nested models.

For demographic inference with $\partial a\partial i$ v2.1.1 (47), we permuted the starting parameter values for each of 100 runs for each model. The log-likelihood was calculated for the expected SFS using $\partial a\partial i$'s optimized parameter values and compared to the empirical SFS. We then estimated the best fit for each replicate and scaled by the total sequence length (L) and mutation rate (5.8×10^{-9} mutations/site/generation; see ***Vaquita mutation rate estimation***) to calculate the diploid ancestral size (N_{anc}) as $N_{anc} = \theta / (4\mu L)$. Parameter values for population size changes were scaled by N_{anc} , and time parameters were scaled by $2 * N_{anc}$, into units of diploids and number of generations, respectively. For each model, we selected the maximum likelihood estimate from the 100 runs. To determine the uncertainty in parameter estimates for the best run per model (i.e. with the maximum log-likelihood), we first performed a bootstrap analysis by partitioning the VCF file used to build the SFS into 1,000 independent fragments of 2 Mb (2 Gb total); then used

those fragments to calculate a Godambe information matrix to obtain the standard deviation and 95% confidence intervals of the best fitting parameters (49). The bootstrap and uncertainty analysis were done using *ada* functions *Misc.bootstraps_from_dd_chunks* and *dadi.Godambe.GIM_uncert*, respectively. As with *fastsimcoal2*, we evaluated the models based on convergence, AIC, and LRT.

Vaquita mutation rate estimation

Our inferences of the demographic history and distribution of selection coefficients in vaquitas, as well as our forward simulations, require that a per-site, per-generation mutation rate be specified. The mutation rate in vaquitas is unknown. Previously, Morin et al. (12) assumed a rate of 1.08×10^{-8} /site/generation, based on a previously estimated nuclear mutation rate of 9.10×10^{-10} /site/year in odontocetes (50), and a vaquita generation time of 11.9 years (13). However, this substitution rate estimate was based on a limited set of just four nuclear genes. We therefore used genome-wide divergence between the vaquita and two other porpoises (harbor porpoise, *Phocoena phocoena* (SRR8305658, (51)), and Indo-Pacific finless porpoise, *Neophocaena phocaenoides* (SRR940959, (52)) to estimate plausible alternative mutation rates.

Divergence (d_{xy}) is related to mutation rate (μ) through the equation $\mu \approx d_{xy}/2t$, where t is the time since two species diverged from a common ancestor. We calculated divergence as $d_{xy} = (h_1 + 2h_2) / (2g)$, where h_1 is the number of sites in which one or both members of the pair is heterozygous, h_2 is the number of sites in which individuals are homozygous for different alleles, and g is the number of sites in which both individuals had called genotypes. We estimated that d_{xy} between the vaquita and harbor porpoise is 4.25×10^{-3} /site, and d_{xy} between the vaquita and finless porpoise is 5.31×10^{-3} /site. Using a previous divergence time estimate of 5.42 million years (53) and a generation time of 11.9 years yields estimated mutation rates of 4.66- 5.83×10^{-9} /site/generation (fig. S12). These are potentially over-estimates, however, because they do not account for the level of diversity in the ancestral population prior to divergence (π_{anc}), which is a reasonable assumption only when $d_{xy} \gg \pi_{anc}$. Here, we calculate the mutation rate as $\mu = (d_{xy} - \pi_{anc})/2$, where we set π_{anc} to the genome-wide estimate of heterozygosity in the harbor or finless porpoise, as π_{anc} is unknown. Using this method, we estimate mutation rates of 3.00- 4.79×10^{-9} /site/generation (fig. S12). To include uncertainty in the divergence time estimate, we also calculated mutation rates using median estimated divergence times from TimeTree.org (54): 7.33 million years since the split between vaquitas and harbor porpoises, and 8.15 million years since the split between vaquitas and finless porpoises. With these divergence times, we obtain mutation rates of 3.45- 4.31×10^{-9} /site/generation without considering π_{anc} , and 2.22- 3.54×10^{-9} /site/generation when incorporating π_{anc} (fig. S12).

We used the minimum and maximum mutation rates estimated here (2.22×10^{-9} and 5.83×10^{-9} /site/generation, respectively) as well as the previously used rate of 1.08×10^{-8} /site/generation (12) for inferring the distribution of fitness effects and population viability simulations (see ***Inference of the distribution of selection coefficients*** and ***Simulation methods***). All findings presented in the main text used the rate of 5.83×10^{-9} /site/generation, which is intermediate between our estimated lower bound and the previously used rate. Our range of mutation rates spans the range of direct estimates from other mammals, including mice and humans (55, 56).

Annotation of putatively deleterious variants

We used the Sorting Intolerant From Tolerant program (SIFT (16)) to annotate synonymous and nonsynonymous mutations, and to identify nonsynonymous mutations that are putatively deleterious. SIFT designates mutations as putatively deleterious when they occur at highly conserved sites in alignments of homologous protein-coding sequences. We used SIFT4G (57) with the UniRef90 protein database (downloaded May 1, 2020) to develop custom SIFT databases for each of the species analyzed in our study, following the developer's recommendations. We used SnpEff v4.3t (build 2017-11-24 10:18) (58) to identify putative LOF mutations. Following (59), SnpEff identifies putative LOF mutations as those that introduce premature stop codons, disrupt splice sites, or induce frameshifts within the first 95% of a gene's coding sequence (LOF mutations in the last 5% of the sequence tend to be tolerated). Note that indels were excluded from our analyses and we therefore did not count frameshift mutations in LOF tallies. For sites with multiple annotations due to overlapping gene transcripts, we prioritized the most deleterious annotation (LOF > deleterious nonsynonymous > tolerated nonsynonymous > synonymous). Custom per-species databases for both SnpEff and SIFT were built using genome annotations (GTF files) provided by RefSeq.

Comparative analyses with cetaceans

We obtained whole genome sequence data for eleven other cetacean species from the NCBI Sequence Read Archive (SRA). Species were selected if they had: 1) an assembled reference genome in GenBank, 2) a genome annotation in RefSeq, and 3) short read Illumina resequencing data in the SRA. Reads from each species were aligned to the reference genome for that species (note: we used the Yangtze finless porpoise as the reference for the closely-related Indo-Pacific finless porpoise, which does not have its own genome assembly) and were processed using the same pipeline described above for the vaquita data. The only differences were that: 1) these samples were processed individually (no joint genotyping), and 2) slightly different filtering criteria were implemented, as follows. We excluded short scaffolds (<500 kb) and scaffolds with depth of coverage <75% or >150% of the genome-wide mean on a per-individual basis. Filtering and masking were the same as for the vaquita data with the following exceptions: we did not incorporate missingness or excess heterozygosity filters, or an excess SNP density mask, as those were implemented exclusively for joint VCF files containing multiple individuals. SIFT and SnpEff databases were generated for each reference genome and used to annotate variants (see *Annotation of putatively deleterious variants*).

For comparisons between the vaquita and other cetacean species (Fig. 2, fig. S8, S10), the value used for the vaquita was the average value calculated from all 20 individuals (see fig. S7 for plots showing values for all 20 vaquitas). All estimates of heterozygosity and heterozygosity-based analyses for comparisons between vaquitas and cetaceans used the same-species-reference dataset described above. In order to obtain counts of homozygous variants (fig. S8), which includes potentially fixed substitutions, we aligned reads from the vaquita and its closest relatives (porpoises, monodontids, and delphinids) to the reference genome of an outgroup, the blue whale (GCF_009873245.2_mBalMus1.pri.v3). We used an identical pipeline to the same-species pipeline described above to obtain filtered, annotated VCF files. More distantly related species were excluded from this analysis in order to mitigate biases from using a divergent outgroup reference genome.

We use several statistics to quantify and compare the burden of deleterious variation, or the "genetic load" (sometimes called the "mutation load"), between vaquitas and other cetacean species. Genetic load can arise from several distinct processes, leading to different types of load

(reviewed in (60)). One form of load is "drift load", or the increased frequency and fixation of weakly deleterious alleles as a result of genetic drift. Specifically, weakly deleterious alleles may accumulate because purifying selection is ineffective at removing mutations with selection coefficients $< 1/(2N_e)$. A second type of load is the "inbreeding load", or the load of recessive deleterious mutations that may reduce fitness by becoming homozygous through inbreeding. Unlike drift load, inbreeding load may be lower in small populations where recessive mutations are more frequently exposed to purifying selection as homozygotes, and higher in large populations where inbreeding seldom occurs and high heterozygosity masks recessive alleles from selection. In our study, we attempt to approximate both types of load using different statistics as described below.

To assess the effects of drift on deleterious variation, we calculated ratios of putatively deleterious variants (deleterious nonsynonymous mutations identified with SIFT and LOF mutations identified with SnpEff) to synonymous variants (Fig. 2A, B, fig. S8). With this approach, we use the number of synonymous variants as the "neutral" baseline to compare the relative burden of deleterious variants between species. This method accounts for varying evolutionary rates among species, and therefore corrects for branch length variation due to varying degrees of divergence between each focal species and the outgroup (in this case, the blue whale; see fig. S6), making it a valid approach for between-species comparisons. In particular, when comparing numbers of derived deleterious homozygotes between species, raw absolute counts are somewhat meaningless by themselves as each focal species has a different divergence time with respect to its outgroup species. A further advantage of our approach is that it inherently accounts for technical differences between samples that can influence absolute counts, such as varying sequence quality or genome coverage. Note, it has been suggested that the ratio of nonsynonymous to synonymous variants may not always be a good proxy for genetic load in non-equilibrium populations (61). However, normalization of putatively deleterious variation by putatively neutral variation is appropriate here because our objective is to compare patterns between divergent species, rather than between closely related populations of the same species. Moreover, the vast majority of homozygous derived variants that we identify are due to fixed substitutions that have accumulated over long branches with respect to the outgroup reference genome, and recent demography should therefore have little effect on this ratio (fig. S8). Because our analyses involve comparisons between divergent species, rather than between populations, we do not require a method that corrects for shared polymorphisms (e.g., R_{xy} (62)). For these reasons, normalization of the counts of nonsynonymous variants by the counts of synonymous variants is appropriate for assessing the effects of long-term N_e on the relative burden of deleterious variants per genome.

To quantify inbreeding load, we tabulated the absolute number of deleterious heterozygotes per genome, as this captures the deleterious variation that can reduce fitness if made homozygous through inbreeding (Fig. 2C, D, fig. S10). We normalized raw heterozygous genotype counts to correct for varying call rates across individuals due to technical differences in coverage, filtering, etc. Specifically, we divided the raw heterozygous genotype counts by the total number of genotype calls passing all filters within each individual, then multiplied this result by the mean total number of genotype calls across all individuals. For protein-coding heterozygotes (synonymous, tolerated nonsynonymous, deleterious nonsynonymous, LOF), we normalized by the total number of called genotypes in the exonic regions; for numbers of conserved noncoding heterozygotes, we normalized by the total number of calls in the conserved

noncoding regions (see *Identification of conserved noncoding regions in vaquita and cetacean genomes*).

Phylogenetic correction for multi-species correlations

We used phylogenetic generalized least squares (PGLS) regression to account for phylogenetic non-independence among species when modelling the association between genome-wide heterozygosity and the burden of deleterious alleles. To generate a species phylogeny, we used alignments of vertebrate-specific single copy orthologous genes from each species. Briefly, we identified orthologs in each genome assembly with BUSCO v4.1.4 (63), generated an alignment for each ortholog with mafft v7.475 (64), and then inferred tree topology and branch lengths using maximum likelihood with IQ-TREE v2.1.2 (65). We included the cow genome (ARS-UCD1.2 (66)) as an outgroup, and excluded the Indo-Pacific finless porpoise as there is currently no genome assembly or annotation available for this species. Our species phylogeny therefore includes 11 cetaceans, including the vaquita, plus the cow. We used the vertebrata_odb10.2020-09-10 BUSCO dataset, which contains 3,354 orthologs from 67 species. Of these, 2,459 orthologs were found in all 12 genomes and used for inferring the species phylogeny. Alignments for each ortholog were trimmed with trimal v1.4.rev15 (67), and then concatenated into a supermatrix (containing 4,421,551 sites total) with catsequences (68). Phylogenies inferred with and without partitioning (69) yielded identical topologies with similar branch lengths, but the partitioned analysis yielded higher AIC and BIC scores. The best partitioning scheme and model, as identified with ModelFinder (70), was an edge-linked-proportional partition model of 167 partitions with separate substitution models and rates across sites. Consensus and maximum-likelihood tree topologies from partitioned and non-partitioned analyses were identical and consistent with the cetacean phylogeny inferred by (71) (fig. S6). In both partitioned and non-partitioned analyses, all branches received 100% support from 1000 bootstrap replicates (UFBoot2, (72)). We used the phylogeny inferred from the partitioned analysis for performing PGLS regression under a Brownian motion correlation structure in R v4.0.5 (73) with the ape v5.5 (74), geiger v2.0.7 (75), and nlme v3.1.152 (76) packages.

Identification of conserved noncoding regions in vaquita and cetacean genomes

Conserved noncoding elements are genomic sequences with high sequence similarity across species that may function in gene regulation (77). To identify conserved noncoding regions, we converted the coordinates of conserved noncoding elements in the mouse genome (mm10) provided by the UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgTables>) to their corresponding coordinates in the vaquita and other cetacean genome assemblies. Briefly, we extracted the coordinates of conserved elements ≥ 50 bp originally identified with PhastCons (78) from a 60-way alignment of placental mammal genomes, then followed the UCSC cross-species reciprocal best alignment protocol to obtain homologous regions in the cetacean genome assemblies (complete pipeline available at (27)). Converted coordinates that overlapped protein-coding sequences (annotated as “CDS”) or repetitive sequences were then discarded to retain only noncoding, non-repetitive conserved regions with a final total length of ~ 32 Mb.

Inference of the distribution of selection coefficients

We used the distributions of synonymous and nonsynonymous allele frequencies (fig. S13A) to infer the distribution of selection coefficients (s) of new nonsynonymous mutations (also known as the distribution of fitness effects) in vaquitas. Importantly, this distribution

characterizes the selective effects of new mutations, but the extent to which these mutations may segregate and potentially reach fixation is determined by the combined effects of natural selection and genetic drift (see *The impact of fixed mutations on vaquita recovery*). Despite being a central parameter in population genetics, this distribution has only been characterized for a few species, and little is known about how it differs between species (79–81). Thus, we provide an estimate of the distribution inferred for humans for comparison (80), given that this distribution has been studied most extensively using molecular data from humans.

We followed the methods of (79) and (80) to infer this distribution in vaquitas. First, site frequency spectra were generated from synonymous and nonsynonymous sites as annotated by SNPEff (58). Hypergeometric projection using easySFS (<https://github.com/isaacovercast/easySFS>), a wrapper for hypergeometric projection with $\partial a\partial i$ (v1.7.0 (47)), reduced the SFS down to 12 individuals to smooth over missing data and maximize the number of SNPs. Next, demographic parameters for the two-epoch model were inferred from the synonymous SFS using $\partial a\partial i$ to control for both demography and the effects of linked selection in coding regions. One hundred runs of $\partial a\partial i$ were carried out, and the parameters with the maximum multinomial log-likelihood were chosen. The optimal value of the population-scaled mutation rate for synonymous sites ($\theta_{synonymous}$) was inferred using $\partial a\partial i$ and then scaled up by a factor of 2.31 for $\theta_{nonsynonymous}$ sites, as in (79) and (80) (the value of 2.31 in those papers was chosen to reflect the higher number of nonsynonymous sites in the exome, and the higher mutation rate of synonymous sites due to their higher GC and CpG-content). The ancestral population size was calculated as:

$$N_a = \frac{\theta_{synonymous}}{4 * \mu * L_{synonymous}}$$

where μ is the mutation rate per site per generation, and L was estimated as:

$$L_{synonymous} = \left(\frac{1}{1 + 2.31} \right) * L_{cds}$$

$$L_{nonsynonymous} = \left(\frac{2.31}{1 + 2.31} \right) * L_{cds}$$

where L_{cds} is the total number of sites in coding regions (annotated as “CDS”) with at least 12 called genotypes (the projection value).

The distribution of selection coefficients of new mutations was then inferred from the nonsynonymous SFS using $\text{fit}\partial a\partial i$ (80), conditioning on the demographic parameters inferred from the synonymous SFS. A grid of expected site frequency spectra under the demographic model was created for a grid of 300 values of the population-scaled selection coefficient γ ($\gamma = s * 2N_a$). Following (80), we fit a gamma distribution for γ , with parameters for shape and scale. Shape and scale parameters were inferred from 25 runs of $\text{fit}\partial a\partial i$ by integrating over the grid of expected site frequency spectra and fitting model output to the empirical nonsynonymous SFS. Parameters with the highest Poisson log-likelihood were chosen. For plotting, the scale parameter was divided by $2N_a$ to no longer be scaled by the ancestral population size. For use in SLiM (19) simulations, the expected value of s ($E[s]$) was calculated by multiplying the shape and scale parameters together, making the sign negative, and multiplying by two since $\text{fit}\partial a\partial i$ computes fitness as $1+2s$ for homozygotes, whereas SLiM computes it as $1+s$:

$$E[s] = \frac{-2 * \text{shape} * \text{scale}}{2 * N_a}$$

This process was repeated for each of the three mutation rates used for demographic inference and forward simulations (2.2×10^{-9} , 5.8×10^{-9} , 1.08×10^{-8} mutations/site/generation).

Compared to humans (24), vaquitas have a higher fraction of new mutations that are neutral and nearly neutral ($0 < |s| \leq 10^{-5}$), a lower fraction of weakly to moderately deleterious mutations ($10^{-5} < |s| < 10^{-2}$), and a higher fraction of strongly deleterious mutations ($|s| \geq 10^{-2}$), though we caution that this latter category is the most difficult to estimate accurately (24, 83) and greatly depends on the mutation rate parameter (fig. S13B). To determine whether the vaquita distribution of selection coefficients (shape = 0.131, scale = 0.098) was an improvement over the distribution inferred for humans in (80), we generated an expected vaquita nonsynonymous SFS using parameters from the human distribution of fitness effects (shape = 0.186, scale = 0.0352), scaled using the estimated value of $2N_a$ for the vaquita described above and the value of $\theta_{\text{nonsynonymous}}$ in vaquitas. We then compared the fit of this expected SFS based on human parameters to the empirical vaquita nonsynonymous SFS and found that it had a lower log-likelihood (log-likelihood = -65.37) than the fit of the expected SFS under our inferred vaquita parameters (log-likelihood = -47.64), meaning that our inferred vaquita distribution of selection coefficients yields an expected SFS that is a better fit to the vaquita empirical data (fig. S13C). Thus, the vaquita distribution of selection coefficients is relatively similar to that of humans, but may result in slightly different selection dynamics. We carried out population viability simulations using both sets of vaquita and human parameters for comparison (see **Simulation methods**).

Simulation methods

Stochastic, individual-based simulations were conducted using the non-Wright-Fisher (nonWF) model in SLiM v3.3.2 (21). This model differs from the classical Wright-Fisher model in several key ways. First, generations in this model are overlapping (and can be thought of in this case as years), whereby the probability of an individual surviving from one year to the next is given by its absolute fitness. Absolute fitness for each individual ranges from 0 to 1 and is determined by its genetic composition. Thus, population size (N) in this model is not predetermined, but instead is an outcome of a stochastic process of reproduction and viability selection (Fig. 3A). To keep the simulated population from growing indefinitely, the nonWF model makes use of a carrying capacity (K), rescaling absolute fitness downward by the ratio of K/N when $N > K$. This rescaling can also optionally be used to model the effects of relaxed selection (e.g., due to reduced intraspecific competition) when $N < K$; however, by default we conservatively do not allow for this upwards rescaling of fitness (see **Fitness rescaling based on density dependence in simulations** for more details).

Each year in the nonWF model consists of separate reproduction and viability selection stages (Fig. 3A). At the start of each year, every reproductive-age female (>4 years) has a 75% probability of reproducing with a randomly-selected reproductive-age male (>4 years). This 75% probability was used to enforce a 1.5-year calving interval, given uncertainty in whether vaquitas calve annually or every two years (17). To examine the potential impacts of this uncertainty, we also ran simulations assuming calving intervals of one and two years. To model the effects of inbreeding depression on reproduction, we further scaled the probability of reproduction by the absolute fitness of the mother. For example, an adult female with absolute fitness 0.97 has a $0.75 * 0.97 = 0.73$ probability of mating each year assuming a 1.5-year calving interval. Each successful mating produces one offspring, with no cap on the number of reproductions per year for adult males. Following reproduction, viability selection occurs, in which each individual survives with a probability determined by its absolute fitness multiplied by any scaling factors for age or density dependence. In cases where mothers with newborn calves died, we also

enforced mortality for the calf, as newborn calves are dependent on their mothers for survival. To enforce age-related increases in mortality rates, fitness was rescaled by a factor of 0.6 for newborn individuals, 0.95 for yearlings, 0.9 for individuals aged 15-19, 0.2 for individuals aged 20-25, and 0.0 for individuals aged 25 (i.e., all individuals die after age 25). Note that these rescaling factors are multiplied by the absolute fitness of each individual, which also varies as a function of age (i.e., older individuals tend to have higher fitness given that they have survived many years of viability selection). These values were chosen such that the survivorship curves in this model reflect those of the closely-related harbor porpoise, as such information does not exist for the vaquita (fig. S11; (82)).

Our model assumes random mating and that individuals can find mates (no Allee effects), which is justified given the small size of the vaquita range. Virtually all vaquita detections are within a 12x24km area, which can be traversed within a day by a vaquita. Moreover, given the acoustic behavior of vaquitas, the remaining ~10 individuals are likely aware of each other. Finally, there is ample evidence for ongoing reproduction in the remaining vaquita individuals. Specifically, in 2019, three of the remaining ~10 individuals were calves, suggesting that reproductive effort is high.

Our aim in setting the genomic parameters was to model a vaquita exome with a combination of neutral and recessive deleterious mutations, given that these mutations are most likely responsible for inbreeding depression (3, 22). To do this, we modeled 18,173 genes, each of length 1,760 bp (the average length of coding sequence in a gene in the vaquita assembly, giving ~32 Mb of total coding sequence), with the number of genes on each chromosome taken from the vaquita genome assembly (12). We assumed no recombination within genes and a recombination rate of 1×10^{-3} between genes, reflecting the effective rate of crossing over in 100 kb noncoding regions between each gene, assuming a recombination rate of 1×10^{-8} per site per generation (83). These genes accumulated a combination of neutral and (partially) recessive deleterious mutations at our estimated rate of 5.8×10^{-9} per site per generation (see above), with the ratio of deleterious to neutral mutations set at 2.31:1 (79). Our model does not include heterozygote advantage, given that its impacts on inbreeding depression are likely to be small (3, 84). We also ran a subset of simulations with a lower estimated mutation rate of 2.2×10^{-9} and a higher rate of 1.08×10^{-8} , as estimated by (50). Importantly, we also rescaled the simulated demographic parameters according to the assumed mutation rate (i.e., $K = 5,200$ when assuming $\mu = 5.8 \times 10^{-9}$ was rescaled to $K = 5.8/2.2 * 5200 = 13,709$ when assuming $\mu = 2.2 \times 10^{-9}$). Finally, to examine sensitivity to the assumed deleterious mutation target size, we increased gene length by 50% to 2,640 bp, yielding a total sequence length of ~48 Mb. This analysis was intended to capture potential effects of both underestimating coding sequence length as well as not modeling functional noncoding variation, as we have no knowledge of how noncoding variation contributes to inbreeding depression and what the distribution of selection or dominance coefficients in these regions might be.

Selection coefficients (s) for deleterious mutations were drawn from a gamma distribution estimated using the genomic data (see ***Inference of the distribution of selection coefficients***) with mean = -0.0257 and shape parameter = 0.1314. For sensitivity analyses varying the mutation rate, we used the parameters inferred under the respective mutation rate (mean = -0.00971 and shape = 0.1316 for $\mu = 2.2 \times 10^{-9}$ and mean = -0.0476 and shape = 0.1315 for $\mu = 1.08 \times 10^{-8}$; fig. S13). Note that this parameterization of the distribution of selection coefficients differs from that of *fitdadi*, as *fitdadi* assumes a homozygous fitness of $1+2s$ whereas SLiM assumes a homozygous fitness of $1+s$. In addition, to convert these per-generation

selection coefficients to per-year selection coefficients, we divided by the generation time (11.9 years), but only for more weakly deleterious mutations ($s > -0.01$; see **Converting selection coefficients from per generation to per year in simulations** for more details). For dominance coefficients (h), we assumed an inverse relationship between h and s (hs relationship) with $h = 0.0$ for very strongly deleterious mutations ($s < -0.1$), $h = 0.01$ for strongly deleterious mutations ($-0.1 \leq s < -0.01$), $h = 0.1$ for moderately deleterious mutations ($-0.01 \leq s < -0.001$), and $h = 0.4$ for weakly deleterious mutations ($s > -0.001$). To explore the sensitivity of our results to our assumed hs relationship (which is not estimated from data), we also ran simulations with an alternative hs relationship (hs -alt) of $h = 0.01$, $h = 0.05$, $h = 0.25$, and $h = 0.5$ for the same categories of mutations above, respectively, as well as with all mutations assumed to be fully recessive ($h = 0$) and fully additive ($h = 0.5$). To examine the sensitivity of our results to the assumed distribution of s , we also ran simulations using a distribution inferred for humans with mean -0.0131 and shape $= 0.186$ (80). In addition, given concerns that molecular-based estimates of this distribution may underestimate the presence of large-effect mutations (85), we ran simulations where we added a fraction of lethal mutations occurring at rates of either 1.0% or 5.0% of new nonsynonymous mutations. These percentages were chosen to yield a number of lethal mutations that is roughly consistent with estimates observed in humans of ~ 0.6 recessive lethals per diploid genome (86) (table S6). Finally, we also ran simulations under the selection and dominance parameters from Kardos et al. 2021 (85), which includes a distribution of selection coefficients with mean $= -0.05$, shape $= 0.5$, augmented with an additional 5% of mutations being recessive lethal, and a distribution of dominance coefficients following the relationship $h = 0.5 \cdot \exp(-13s)$ (presented in fig. S21 as the “Kardos” model). To maintain computational tractability, we approximated this relationship by partitioning this distribution as follows: $h = 0.48$ for $s \geq -0.01$, $h = 0.31$ for $-0.1 \leq s < -0.01$, $h = 0.07$ for $-0.4 \leq s < -0.1$, $h = 0.001$ for $-1.0 \leq s < -0.4$, and $h = 0.0$ for $s = -1.0$. For all simulations, mutations that became fixed in the simulated population were retained, such that fitness progressively declines with increasing numbers of fixed mutations.

We set the historical population size parameters of our simulations using our best-fit two-epoch demographic model (see **Demographic inference**). Specifically, this model consists of a historical population size of $N_1 = 4,485$ followed by a decline to $N_2 = 2,807$ for 25,727 years. To convert these effective population size estimates to carrying capacities (which here can be thought of as census sizes), we first estimated the ratio of N_e to K in our simulation model based on neutral heterozygosity at equilibrium under our default life-history parameters. From this estimate of $N_e/K = 0.54$, we obtained carrying capacities of $K_1 = 8,300$ ($4485/0.54$) and $K_2 = 5,200$ ($2807/0.54$). For all simulations, we ran burn-ins at K_1 for 50,000 years followed by K_2 for 26,000 years. This duration at K_1 was long enough for the inbreeding load to reach equilibrium, however it was not long enough for neutral heterozygosity to reach equilibrium (which takes $\sim 350,000$ years; fig. S17). Because neutral mutations do not impact fitness and therefore do not impact extinction dynamics, we used these shorter burn-ins to maintain computational feasibility. We also include a comparison of results when assuming burn-ins at K_1 of 100,000 and 350,000 years to demonstrate that this choice does not impact model results (fig. S18). Due to the high computational load of simulations where we increased the historical population size by a factor of 20, we only ran the second epoch of 26,000 years with $K_2 = 104,000$. Importantly, however, this burn-in duration was still long enough for the simulated inbreeding load to reach equilibrium (fig. S20).

Following these historical population dynamics, we initiated a decline by introducing stochastic bycatch mortality at a rate that matched observed patterns of decline (see *Estimation of recent vaquita abundance*) (fig. S19). To match this rate of decline in our simulations, we used a bycatch mortality probability of 0.15 per individual until the population declined below 250 individuals, after which we increased this mortality probability to 0.34, reflecting the more rapid pace of decline starting in 2011 due to the resurgence of the totoaba (*Totoaba macdonaldi*) fishery (see *The impact of gillnet fishing on vaquitas*). Bycatch mortalities were enforced by changing individual absolute fitness to 0.0 following reproduction, ensuring that those individuals died during the viability selection step. We maintained this rate of bycatch mortality of 0.34 until the population declined to a threshold of 5, 10, or 20 individuals or fewer, reflecting a plausible range of the current abundance (see (4) for discussion of census estimates). Recovery was initiated by reducing the bycatch mortality rate to 0.0, 0.034, or 0.068, a decline in mortality rates relative to 0.34 of 100%, 90%, and 80%, respectively. Simulations were run for 50 years following the start of recovery, or until the population went extinct. We ran 100 simulation replicates for all parameter combinations. We note that these recovery scenarios are greatly simplified as we do not expect a constant rate of bycatch mortality over the next 50 years, nor do we expect bycatch mortality rates to be immediately reduced to zero. However, we take this approach to maintain ease of interpretation of our results, with the general aim of demonstrating the influence of bycatch mortality rates in our model.

During our simulations, we kept track of several quantities relevant to the state of the population: mean absolute fitness (calculated multiplicatively across sites), mean inbreeding coefficient (here measured as F_{PED} tracing back to the grandparent generation, chosen to capture inbreeding due to mating among close relatives), mean inbreeding load, and the mean number of strongly deleterious alleles ($s < -0.01$), very strongly deleterious alleles ($s < -0.1$), and lethal alleles ($s < -0.5$) per individual. These quantities were estimated using a sample size of 60 individuals (or from the entire population if $N < 60$) every 1,000 years during the historical demographic trajectory, and every year following the initiation of the decline. In table S6, we report these quantities for each combination of simulation parameters following the conclusion of the burn-in. Simulation scripts are available at (28).

Converting selection coefficients from per generation to per year in simulations

The distribution of selection coefficients we estimate for vaquitas is inferred under a Wright-Fisher model, where selection occurs each generation, and selection coefficients are therefore scaled per generation. However, selection in the nonWF model occurs each year, meaning that selection coefficients should be scaled per year. To rescale these selection coefficients, we can simply divide them by the generation time (87). However, it is not clear that this rescaling should also apply to strongly deleterious mutations, which might express their effects in development or early age, and therefore do not operate over multiple years. Our solution was to rescale selection coefficients ($s \geq -0.01$) by the vaquita generation time (11.9; (13)), but not strongly deleterious mutations ($s < -0.01$). To test the sensitivity of our model to this assumption, we also ran simulations with no rescaling and with all mutations rescaled (fig. S21). We find negligible differences in extinction rates between the case with no scaling and the case with strongly deleterious mutations scaled, however we observe much lower extinction rates in the case where all mutations are scaled. Overall, we conclude that these considerations appear not to impact our model projections substantially, and, if anything, suggest that our projections may be conservative.

Fitness rescaling based on density dependence in simulations

To maintain a population near its carrying capacity (K), the SLiM nonWF model rescales fitness each generation by the ratio of K/N (i.e., if $N < K$, each individual's fitness is rescaled upward, and if $N > K$, fitness is rescaled downward). This rescaling is meant to model the effects of a relative abundance of resources and reduced intraspecific competition when $N < K$, and a relative scarcity of resources and increased intraspecific competition when $N > K$. By default, this rescaling occurs both in the case of $N < K$ and $N > K$, with no cap on the magnitude of the rescaling (though absolute fitness remains bounded by 0 and 1). Thus, in the case of our simulations, where a population decline occurs due to stochastic mortality that results in $N \ll K$ (i.e., N declines to < 20 while K stays fixed at 5,200), the default behavior of the model would be to rescale fitness upward, resulting in absolute fitness being rescaled to 1.0 for all individuals given the large ratio of K/N . As a consequence, viability selection would not occur, given that all individuals would have a 100% chance of survival with fitness of 1.0. To avoid this biologically unrealistic behavior, and ensure that the impacts of inbreeding depression were expressed in our model, we conservatively assumed that no upwards rescaling of fitness could occur (i.e., fitness rescaling was given by $\min(K/N, 1.0)$). We explored sensitivity to this assumption by also running simulations with upwards fitness rescaling capped at 5%. Here, we find much lower extinction rates, with only 10% of replicates going extinct compared to 27% when assuming no upwards rescaling of fitness (fig. S21). Thus, if inbreeding depression in vaquitas is at all buffered by the impacts of reduced intraspecific competition due to the very small vaquita population size, recovery potential may be further increased relative to our model projections.

Supplementary Text

The impact of gillnet fishing on vaquitas

Only three years after vaquitas were described as a new species in 1958, Norris & Prescott (88) reported the incidental mortality of vaquita in gillnets of totoaba, sharks, and rays. In 1975, the International Whaling Commission (IWC) Scientific Committee expressed concerns about incidental mortality in shark and totoaba fishery in the published report of the first meeting of the Small Cetaceans Subcommittee (89). At least 128 vaquita were killed in fishing gear between 1985 and early 1992, 65% of which were killed in totoaba fishery (90). D'Agrosa et al. (91) monitored fishing effort and incidental catch by El Golfo fishermen from January 1993 to January 1995 and using data collected by observers and fisherman interviews estimated the mortality of vaquita in 39 (95% CI 14-93) individuals from one fishing community and roughly double that for both communities adjacent to the vaquita's range. Once there was a full abundance estimate (37) and data on reproductive rate (92), it was clear that the gillnet mortality rate was unsustainable. The rate of decline between 1997 and 2008 was compatible with estimated rates of mortality only from gillnet entanglement (93). For a review of evidence of gillnet entanglement see (94). The recent catastrophic decline documented using vaquita acoustic data occurred when illegal fishing for totoaba resurged (4). Gulland et al. (18) examined nine vaquita carcasses from 2016-2018 when rapid decline was taking place and established the cause of death was bycatch in gillnets. In summary, the impact of gillnet fishing on the vaquita is well documented, with extensive evidence demonstrating that it has driven the decline of the species.

Simulating the ratio of heterozygosity at nonsynonymous to synonymous sites

Our empirical analysis of deleterious mutations in vaquitas demonstrate that vaquitas have the highest ratio of heterozygosity at nonsynonymous to synonymous sites of all surveyed cetaceans (Fig. 2), suggesting relaxed purifying selection. However, it is possible that this high ratio in vaquitas has been influenced by the recent bottleneck, rather than being a consequence of their small historical population size (61). To investigate this issue, we estimated this ratio during the bottleneck in 10 simulation replicates, using the same simulation framework described above. We observe only minimal change in this ratio during the bottleneck, with an average value of ~ 0.34 at the start of the bottleneck and ~ 0.35 at the end (fig. S9). Thus, these results confirm that the elevated ratio of deleterious alleles in vaquitas is largely due to relaxed selection as a result of historically low N_e , rather than more recent demographic processes.

The impact of fixed mutations on vaquita recovery

Much emphasis in conservation genetics has been placed on the role of fixed weakly deleterious mutations as drivers of extinction in small populations via a process termed “mutational meltdown” (14, 95, 96). In small populations, the accumulation of these fixed mutations can contribute to reduced fitness, resulting in an elevated “drift load”. This process may be relevant to vaquita recovery potential, given that weakly deleterious mutations are likely to have drifted to fixation in the small historical vaquita population.

To investigate the impact of fixed mutations in our SLiM nonWF simulation model, we tracked the quantity and total effect on fitness (drift load) of fixed mutations in 10 simulation replicates with 350,000 year burn-in durations (long enough for neutral mutations to reach equilibrium). At the end of the burn-in, we observe an average of 787 fixed deleterious mutations in each simulated population. However, the mean selection coefficient of these fixed mutations is -2.4×10^{-5} , resulting in a drift load of -0.19% (note that these selection coefficients are scaled

per year; fig. S15). Thus, although hundreds of weakly deleterious mutations become fixed, they have a relatively minor cumulative impact on fitness. This result helps explain why we do not see differences in extinction rates when running much shorter 50,000 year burn-ins (fig. S18), which only differ in the quantity of fixed mutations (7.8 deleterious mutations fixed on average, with a cumulative impact on absolute fitness of -2.25×10^{-5}). Note that our aim is not to match the number of fixed deleterious mutations in the sequenced vaquita genomes, but to explore the impacts of these mutations on extinction dynamics through simulations.

As another approach to investigate the impact of fixed mutations in our model, we ran simulations with a burn-in duration of 50,000 years where mutations were removed after fixation (i.e., their impact on fitness was not allowed to accumulate). Here, we observe negligible differences in extinction rates in this model compared to simulations where fixed mutations were retained (26% vs 27% extinction rates, respectively; fig. S16).

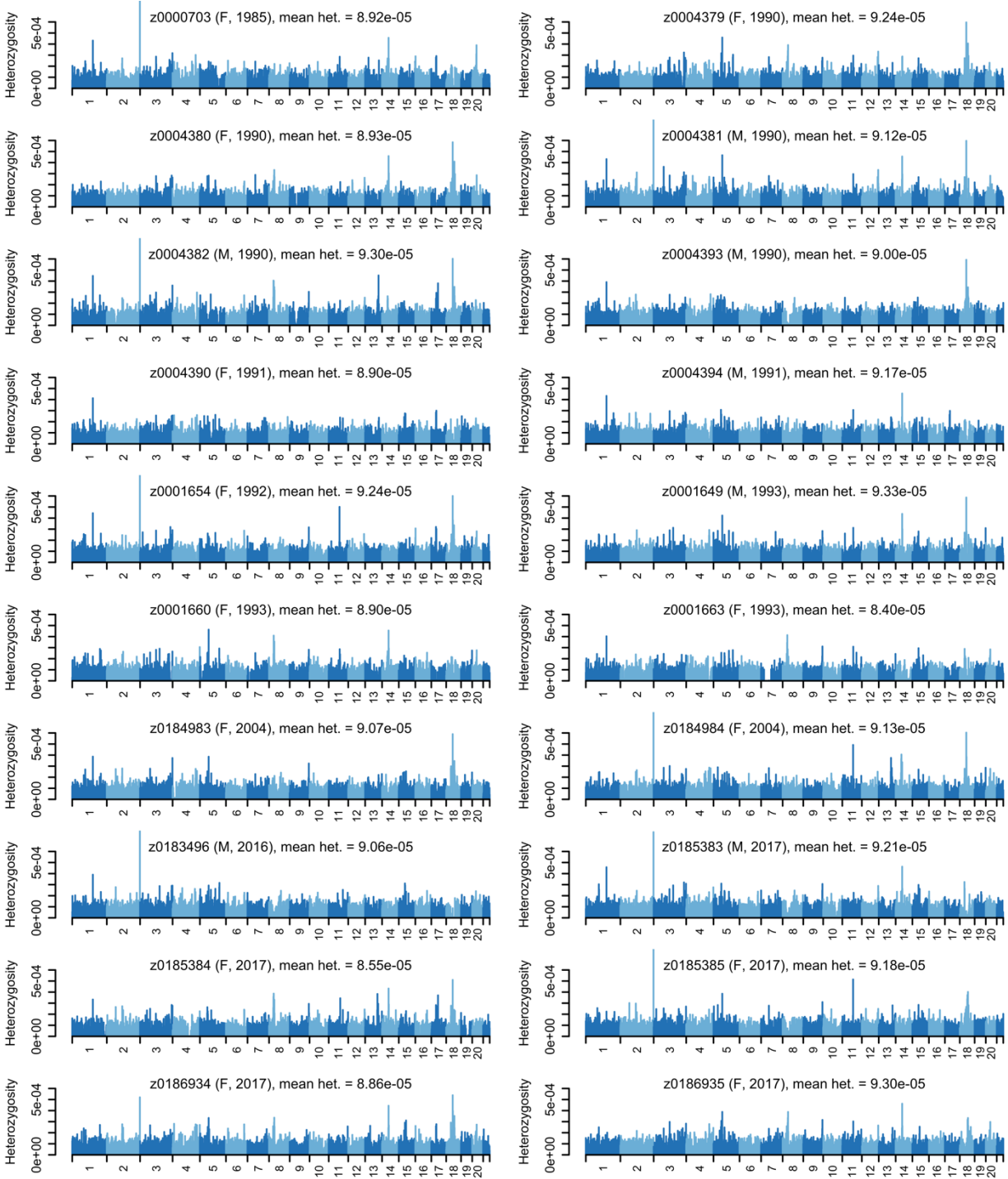


Fig. S1.

Genome-wide heterozygosity in vaquitas. Bar plots showing per-site heterozygosity in non-overlapping 1-Mb windows across the autosomes. Sex and sampling year are indicated next to sample names in parentheses. A peak of heterozygosity in the last window of Chromosome 2 extends beyond the plot in eight individuals to $\sim 1.16 \times 10^{-3}$. This peak contains a number of highly paralogous genes, and may be due to technical artifacts (see Materials and Methods).

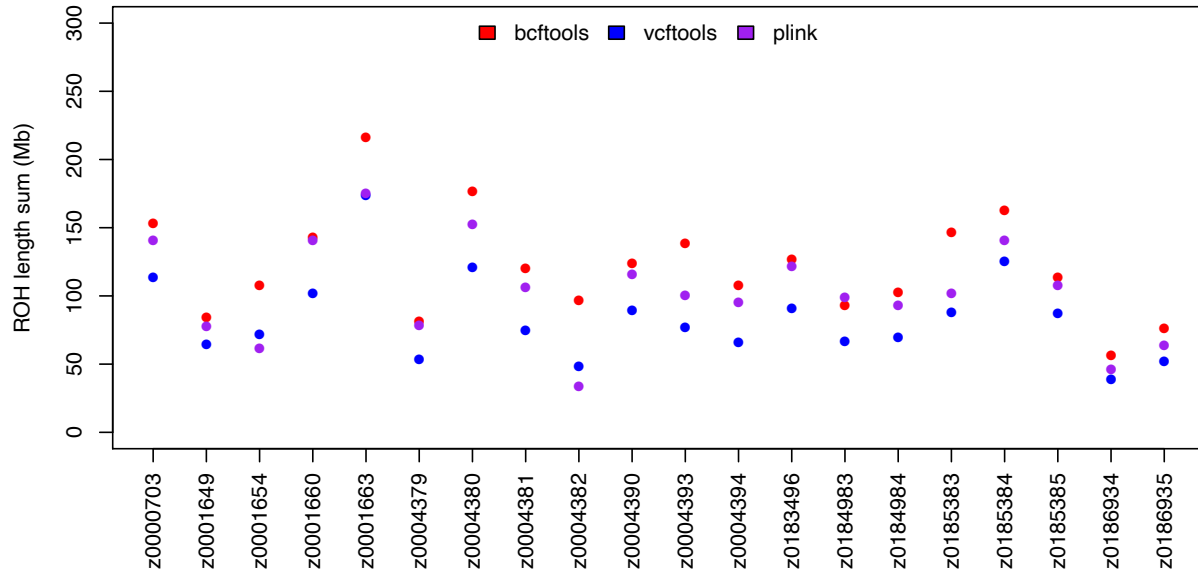


Fig. S2.

Comparison of ROH identification methods. Summed ROH lengths obtained from three different ROH calling methods were highly similar to one another. We provided variant sites only for running Bcftools (42) and Vcftools (41), and all sites (including invariant) for Plink (40). Plink and Vcftools were run with default parameters. Bcftools was run with option *-G 30* and allele frequencies estimated from 15 unrelated vaquitas (see Materials and Methods).

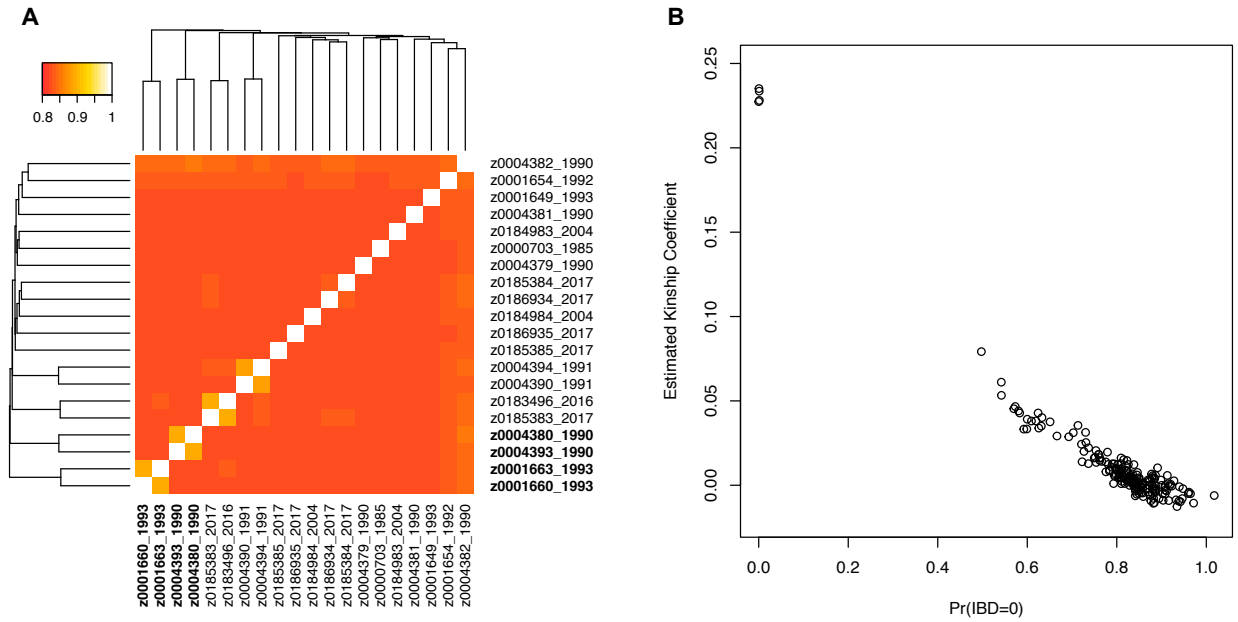


Fig. S3.

Genetic distance and relatedness between vaquitas. (A) Heatmap and hierarchical clustering based on identity by state (IBS) calculated from 59,420 SNPs pruned for linkage disequilibrium in SNPRelate (45). Sample year is indicated along with the sample name. Sample names in bold indicate known mother-fetus pairs. Two additional first-degree relationship pairs are evident: z0004390/z0004394 and z0183496/z0185383. Figure generated with the *heatmap.2* function in *gplots* (97). (B) Pairwise IBS values were used to estimate kinship coefficients with the KING-homo method (44) implemented in SNPRelate. X-axis values indicate the probability that two individuals share zero alleles identical by descent (IBD). Four kinship coefficients close to 0.25 (mean = 0.23) indicate four first-degree relationships, including the two known mother-fetus pairs. The next highest kinship coefficient is between samples z0004382 and z0004380 (kinship coefficient = 0.079), indicating a potential third-degree relationship.

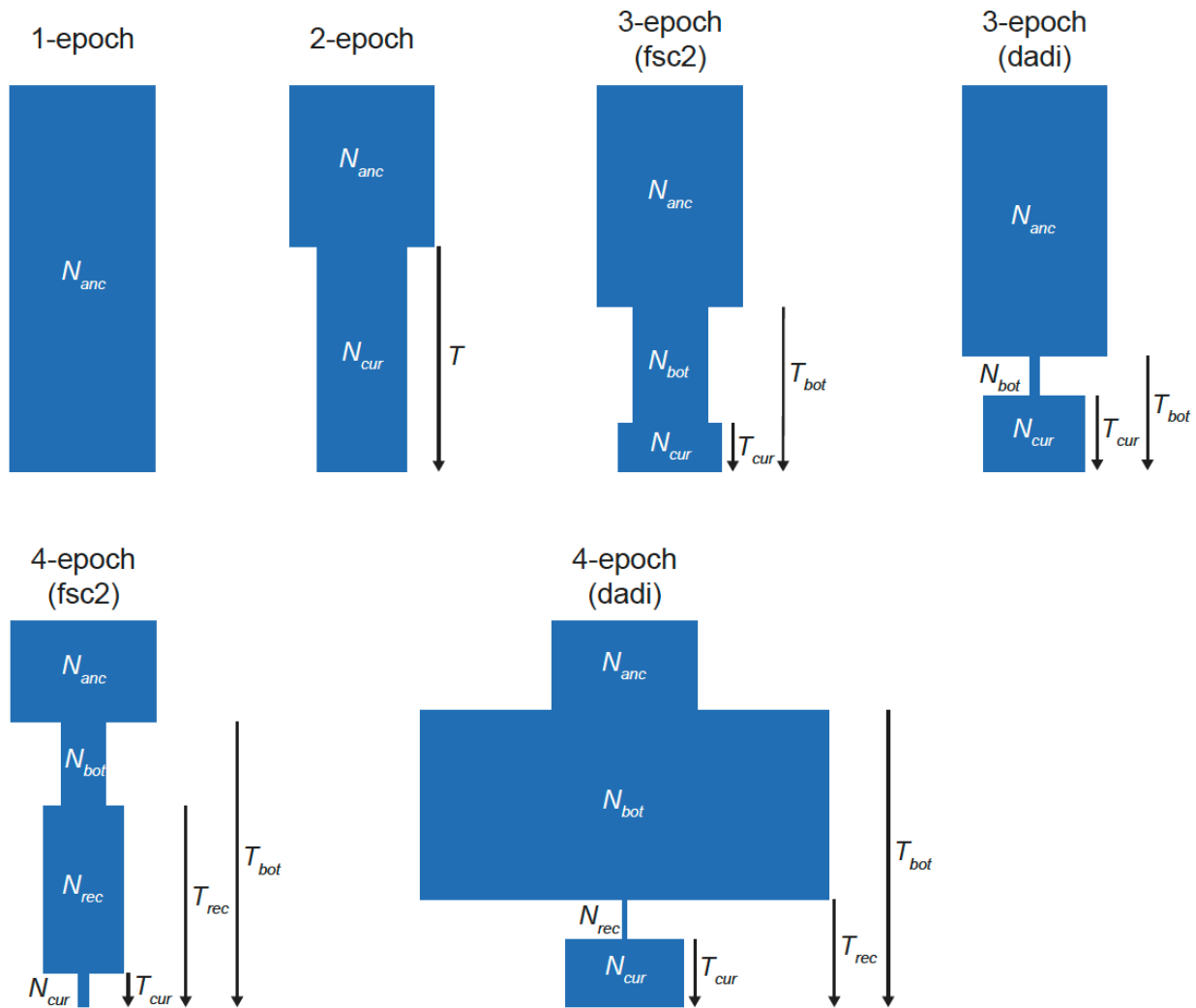


Fig. S4.

Graphical representation of the different demographic models tested. “1-epoch” represents a model with no population size change. “2-epoch” represents a model with one size change event. “3-epoch” represents a model with two size change events. “4-epoch” represents a model with three size change events. Population size parameters (N_{anc} , N_{bot} , N_{rec} , N_{cur}) represent the population size in diploid individuals. Time parameters (T , T_{bot} , T_{rec} , T_{cur}) represent the number of generations before the present day in which population size changes occurred (see Materials and Methods for the full description of the parameters). The size changes and time representations are not to scale, they are only approximations according to the inferred demographic scenario. See table S2 for detailed parameter and uncertainty estimates. For 1- and 2-epoch models, the representation applies for both $\partial a\partial i$ and fastsimcoal2 inference, while for the 3- and 4-epoch models the representation is different for both methods because the parameters estimated were substantially distinct.

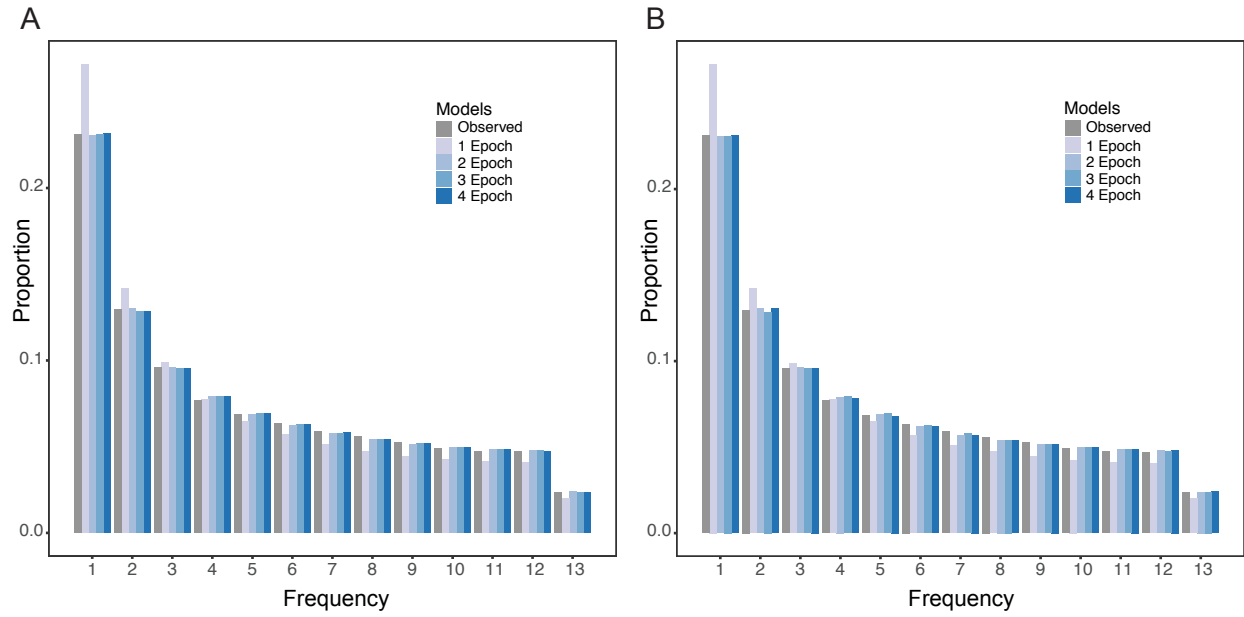


Fig. S5.

Fit of each demographic model with folded proportional site frequency spectrum (SFS).

Comparison of the SFS from each demographic model (1- to 4-epoch) obtained from *daDi* (A) and *fastsimcoal2* (B) inference with the SFS from the empirical data (denoted “Observed”).

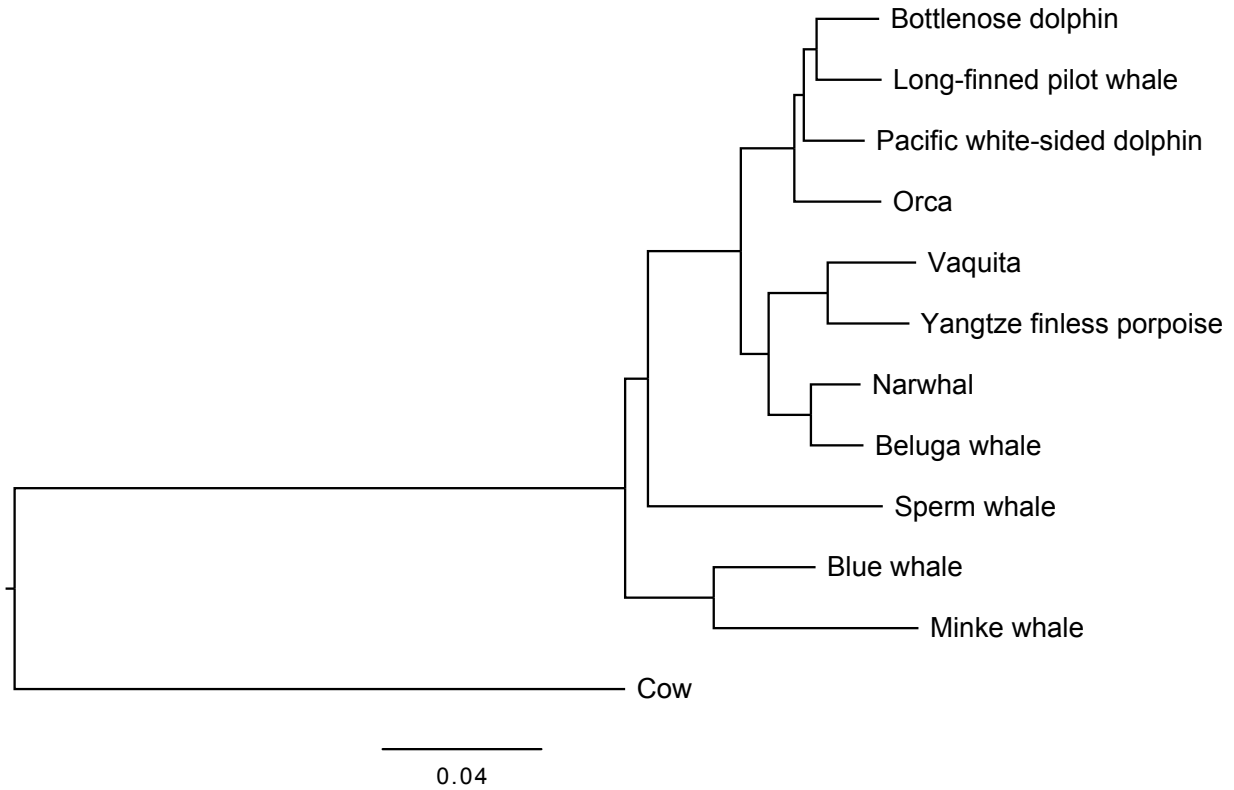


Fig. S6.

Cetacean phylogeny used in phylogenetic generalized least squares (PGLS) regression. The Indo-Pacific finless porpoise does not have an assembled genome and was therefore excluded. Maximum likelihood tree inferred with IQ-TREE (65) from 2,459 single copy orthologs present in all twelve species. All branches have 100% support from 1,000 bootstrap replicates (72). Scale bar represents genetic distance (average number of substitutions per site). See table S1 for cetacean genome information.

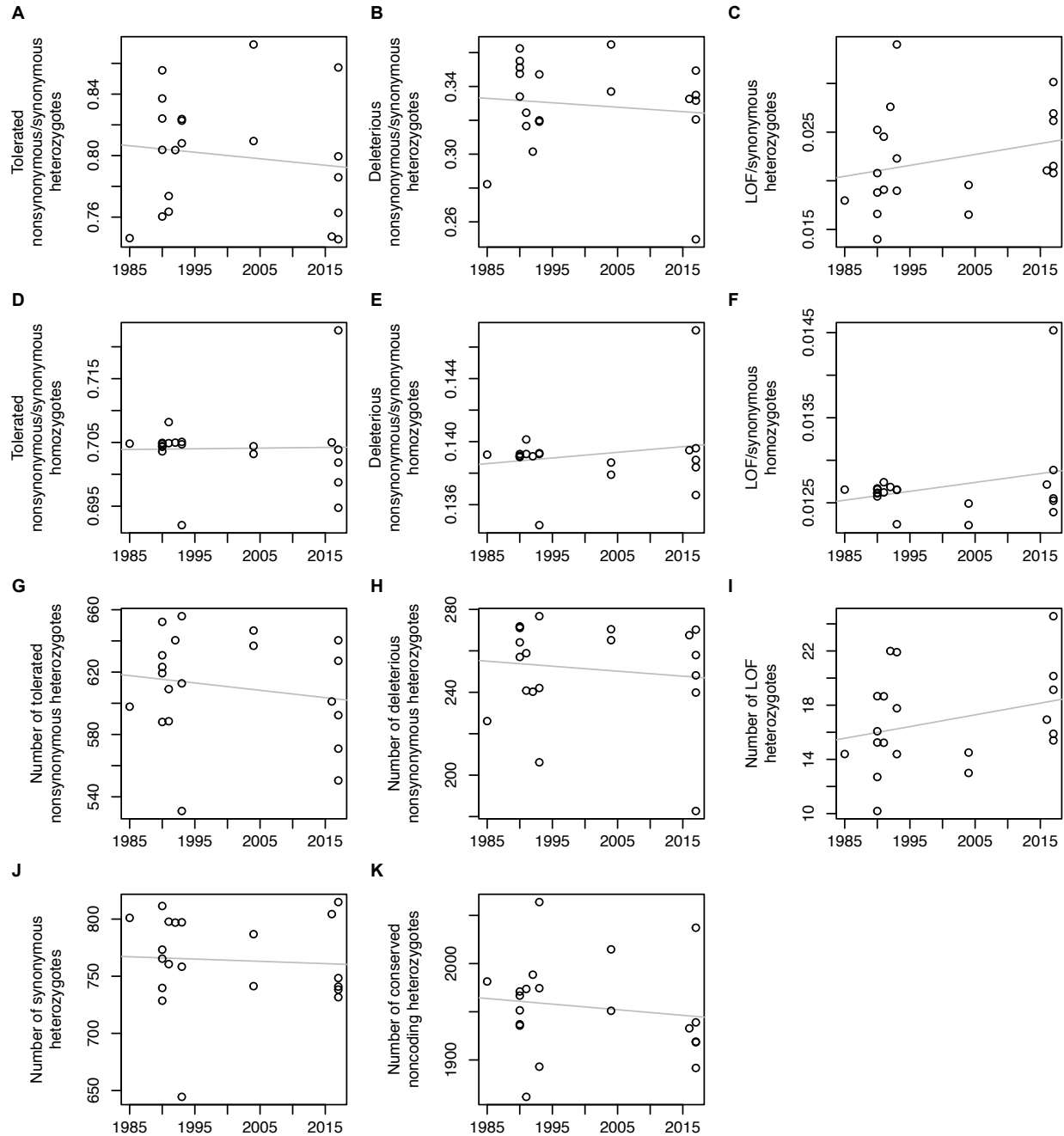


Fig. S7.

Ratios and counts of protein-coding and conserved noncoding variants in vaquitas. X-axis represents the sample year (table S1). Grey lines show linear regressions. There are no apparent trends with regard to sampling year, and none of the regressions are statistically significant ($p \gg 0.05$). (A-C) Ratios of tolerated nonsynonymous, deleterious nonsynonymous, and LOF heterozygotes to synonymous heterozygotes. (D-F) Ratios of tolerated nonsynonymous, deleterious nonsynonymous, and LOF homozygotes to synonymous homozygotes. (G-K) Numbers of tolerated nonsynonymous, deleterious nonsynonymous, LOF, synonymous, and conserved noncoding heterozygotes.

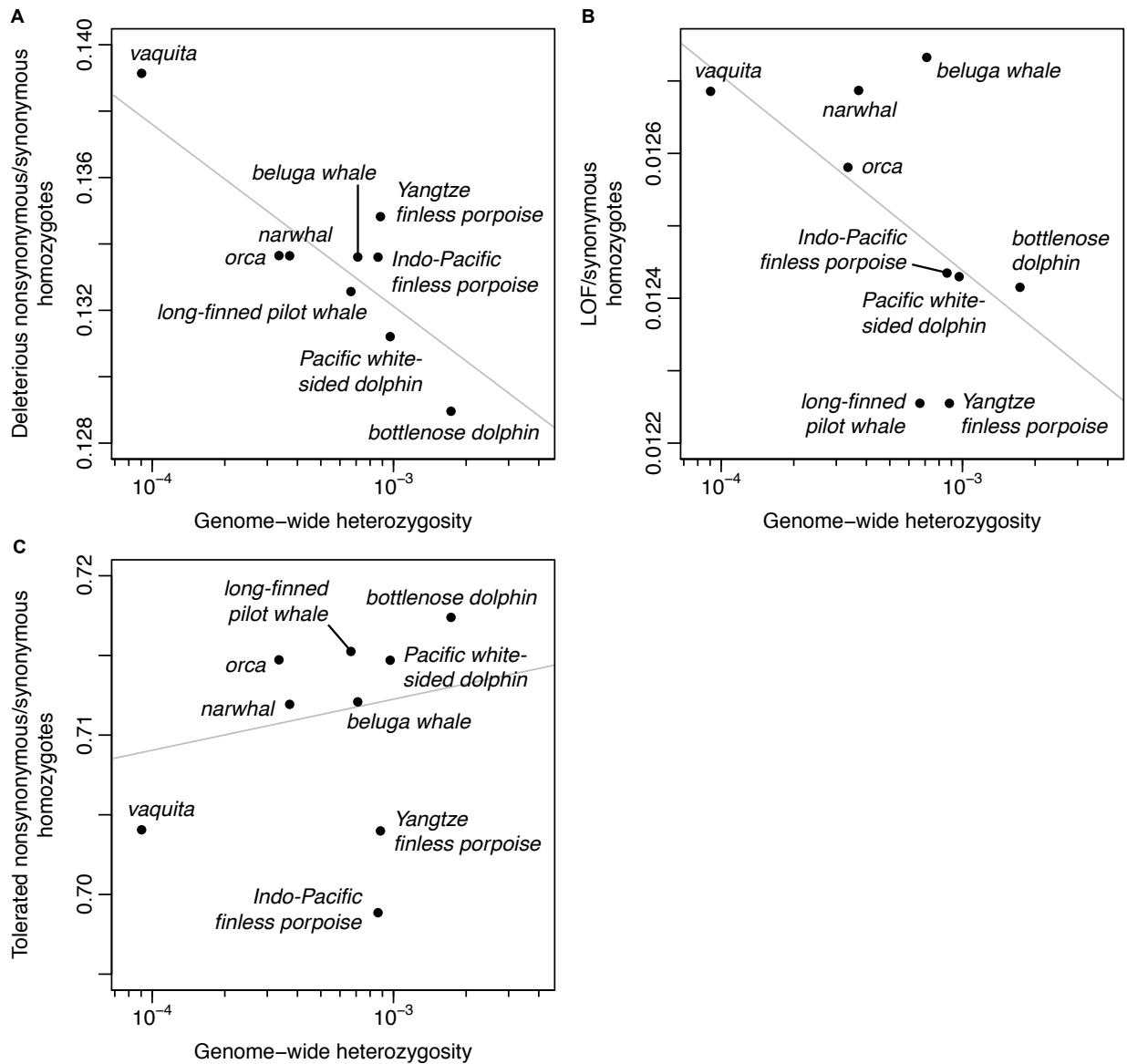


Fig. S8.

Genome-wide heterozygosity and the ratio of nonsynonymous and LOF homozygotes to synonymous homozygotes. (A) The relative proportion of deleterious nonsynonymous homozygotes is significantly negatively correlated with genome-wide heterozygosity (per bp, log-scaled) (phylogenetic generalized least squares regression (PGLS) $p = 2.95 \times 10^{-3}$). (B) A negative correlation appears to exist for LOF homozygotes also, but the relationship is not significant (PGLS $p = 0.124$). (C) The correlation for tolerated nonsynonymous variants appears slightly positive but is not significant (PGLS $p = 0.350$). Grey lines show phylogeny-corrected regressions (excluding the Indo-Pacific finless porpoise; see Materials and Methods). Variants were identified by alignment to an outgroup reference genome (blue whale). In order to mitigate the effects of reference bias, only the vaquita and species with approximately equivalent divergence from the reference are included (porpoises, monodontids, delphinids).

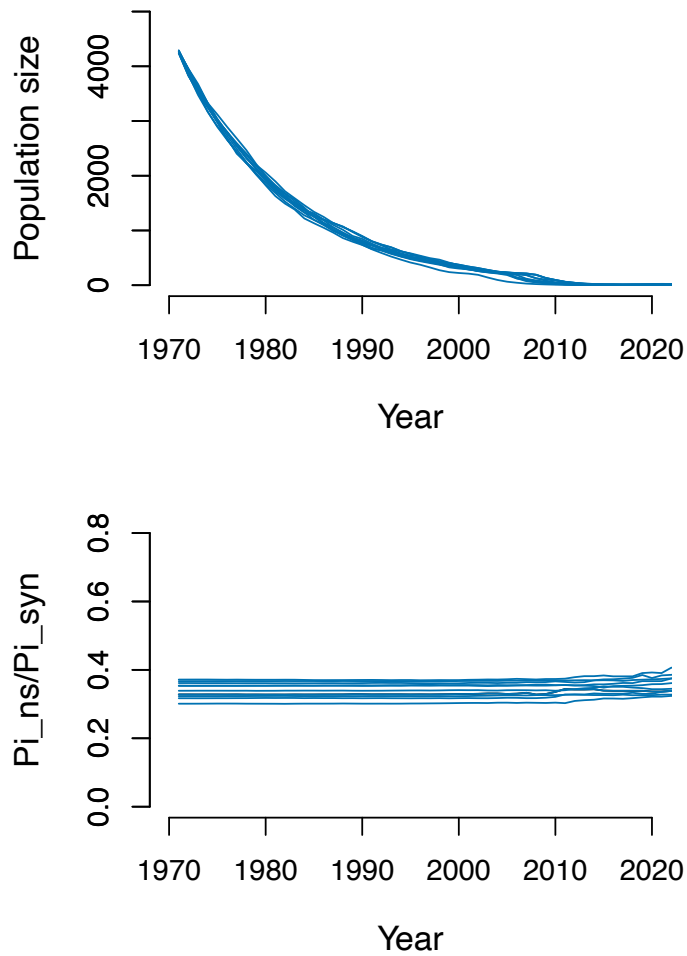


Fig. S9.

The ratio of nonsynonymous to synonymous heterozygosity is largely unaffected by a recent population bottleneck in simulations. (top) The change in population size in the simulations over time. (bottom) The ratio of nonsynonymous heterozygosity to synonymous heterozygosity, sampling individuals at different time-points in the simulations. Note the minimal increase in the ratio, which only starts to occur during the most recent portion of the bottleneck. Thus, the high ratio observed empirically for vaquitas (Fig. 2A, B) appears to be primarily due to an accumulation of nonsynonymous variants in the ancestral vaquita population and not driven by the recent bottleneck. Results are shown for 10 simulation replicates.

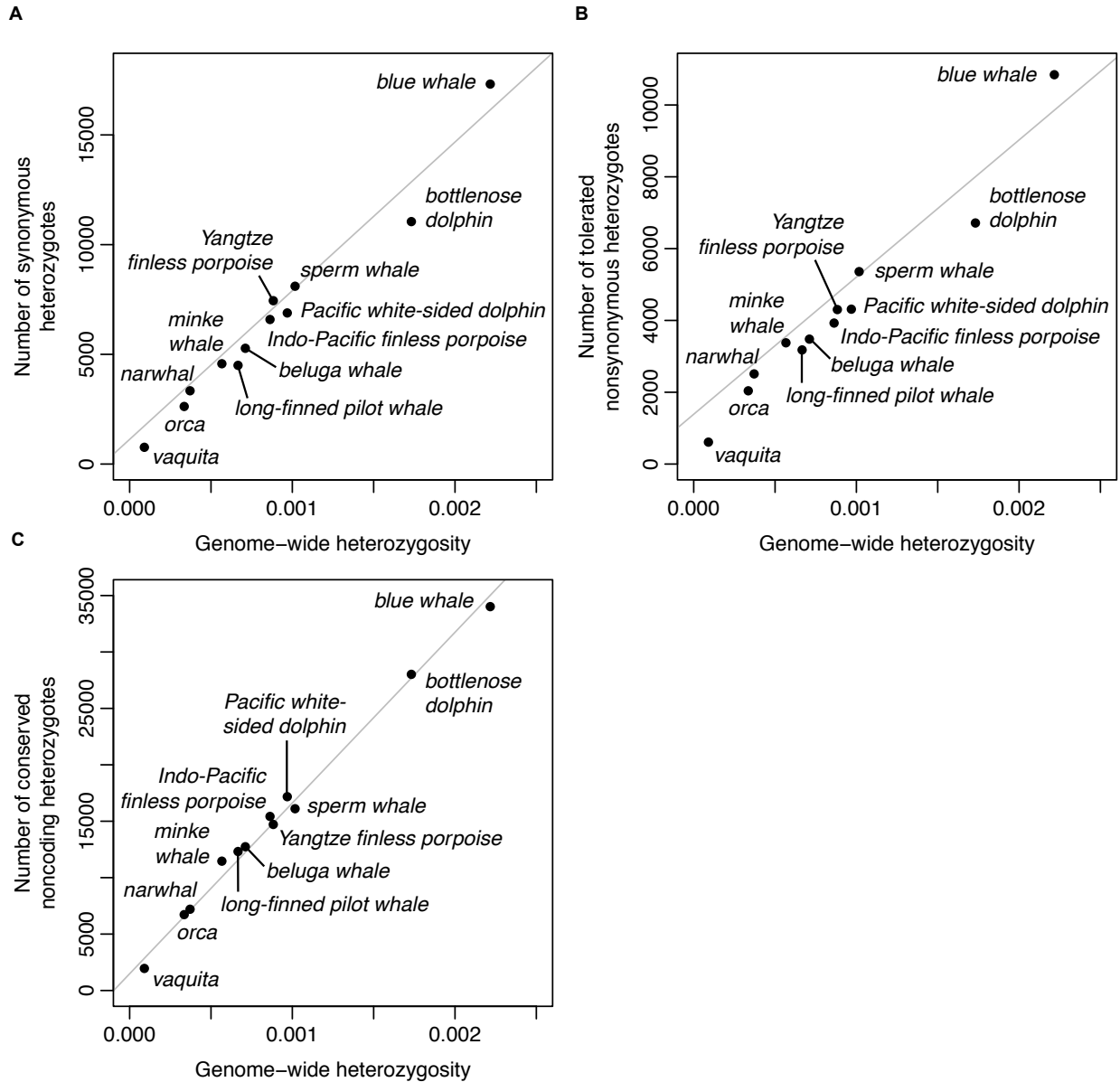


Fig. S10.

Genome-wide heterozygosity and the number of heterozygotes in protein-coding and conserved noncoding regions. The numbers of heterozygous synonymous (A), tolerated missense (B), and conserved noncoding (C) variants are significantly positively correlated with genome-wide heterozygosity (per bp). Grey lines show phylogeny-corrected regressions (excluding the Indo-Pacific finless porpoise; see Materials and Methods). All relationships are significant (PGLS regression; synonymous $p = 2.12 \times 10^{-8}$, tolerated nonsynonymous $p = 9.14 \times 10^{-8}$, conserved noncoding $p = 1.13 \times 10^{-11}$).

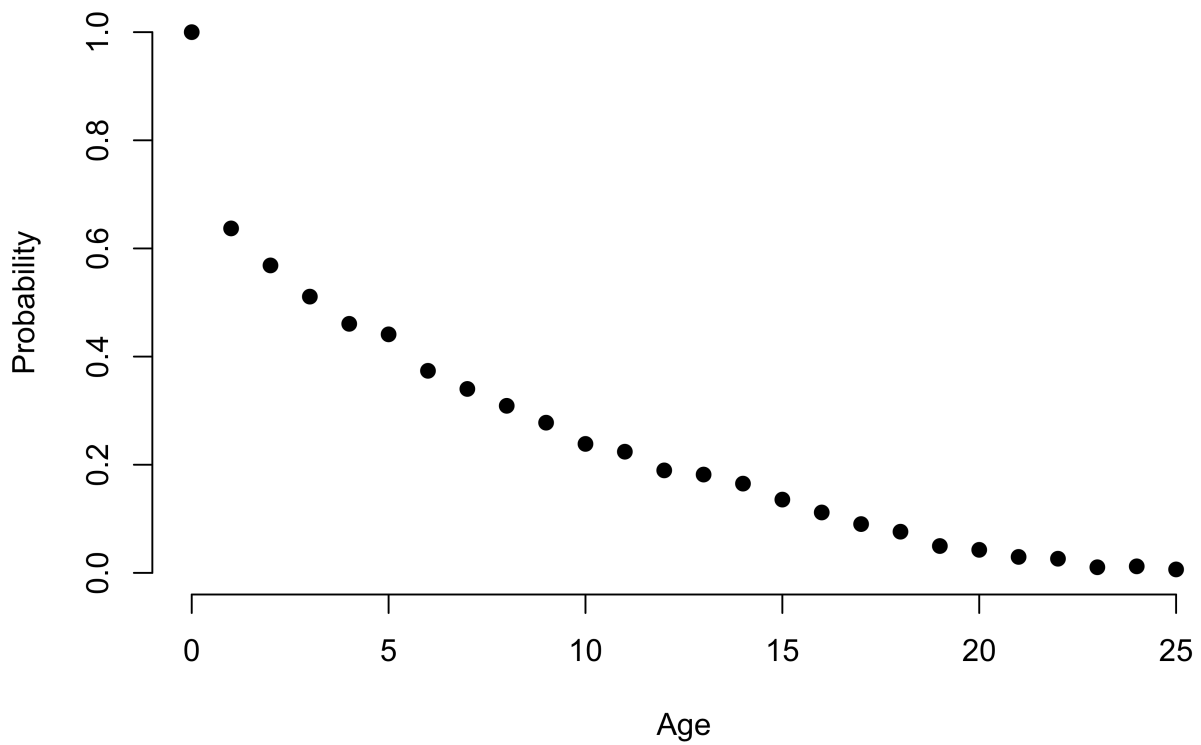


Fig. S11. Survivorship curves for the simulation model. Age-specific fitness rescaling was employed such that vaquita survivorship closely matched empirical estimates of survivorship from the closely-related harbor porpoise. See Fig. 2A in (82) for comparison.

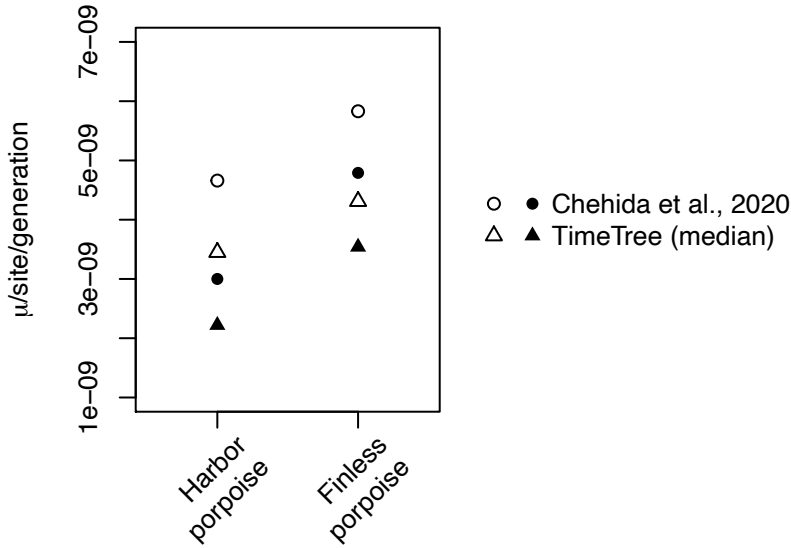


Fig. S12.

Range of estimated mutation rates in vaquitas based on divergence from other porpoise species. We used divergence time (t) estimates from a recent study of complete mitochondrial genomes in porpoises (53), as well as median estimated divergence times from TimeTree.org to estimate mutation rates (μ) from genome-wide divergence (d_{xy}). Open symbols show the estimates for $\mu = d_{xy}/2t$, and filled symbols show the estimates for $\mu = (d_{xy} - \pi)/2t$, where π is genome-wide heterozygosity in the harbor or finless porpoise genome.

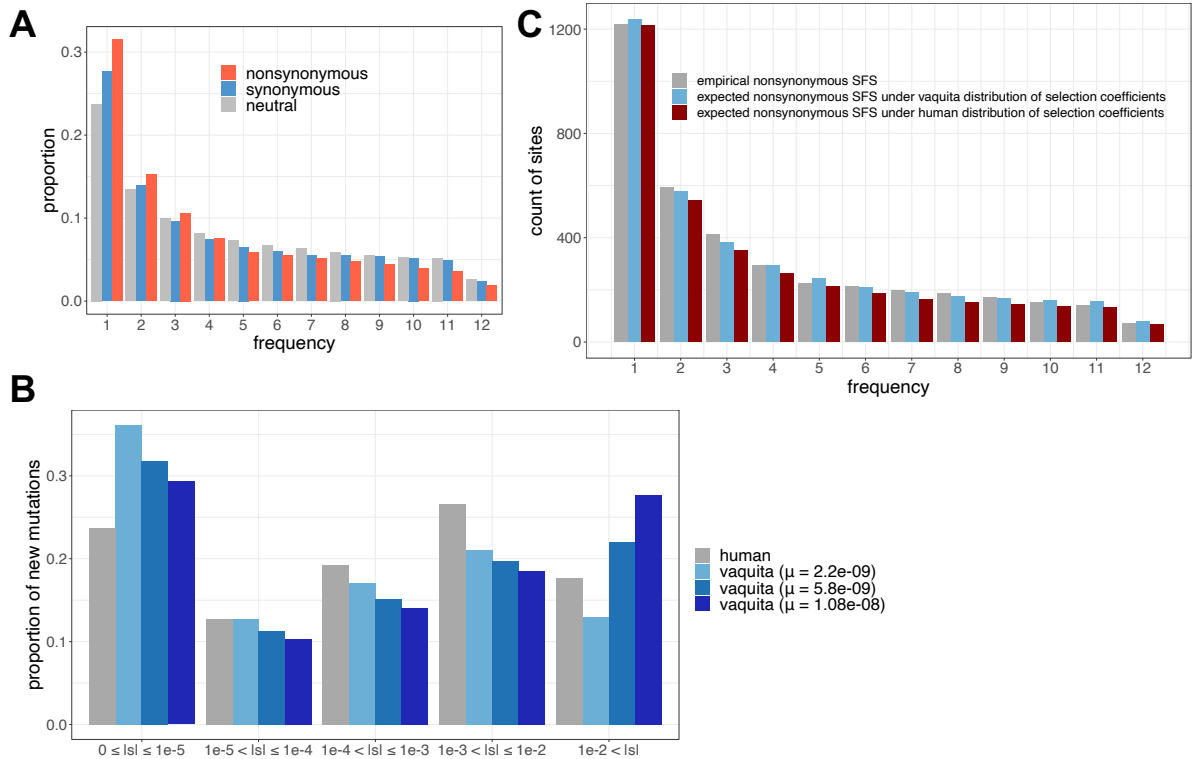


Fig. S13.

Inference of the distribution of selection coefficients. (A) The folded and projected site frequency spectra from coding regions (synonymous and nonsynonymous sites) show a skew toward low-frequency variants relative to the SFS from neutral sites far from genes. The skew is largest for nonsynonymous variants, consistent with the action of purifying selection removing deleterious amino-acid changing alleles. The skew for synonymous variants likely reflects background selection. (B) The distribution of selection coefficients of new mutations (also known as the distribution of fitness effects) for different vaquita mutation rates. The strongly-deleterious category ($10^{-2} < |s|$) is most affected by the assumed mutation rate. A distribution of selection coefficients inferred for humans (80) is shown in gray for comparison. (C) Comparison of the empirical vaquita nonsynonymous SFS (gray), the expected nonsynonymous SFS under the vaquita distribution of selection coefficients (blue), and the expected nonsynonymous SFS under the human distribution of selection coefficients (80) and rescaled by vaquita ancestral size and vaquita $\theta_{nonsynonymous}$ (red).

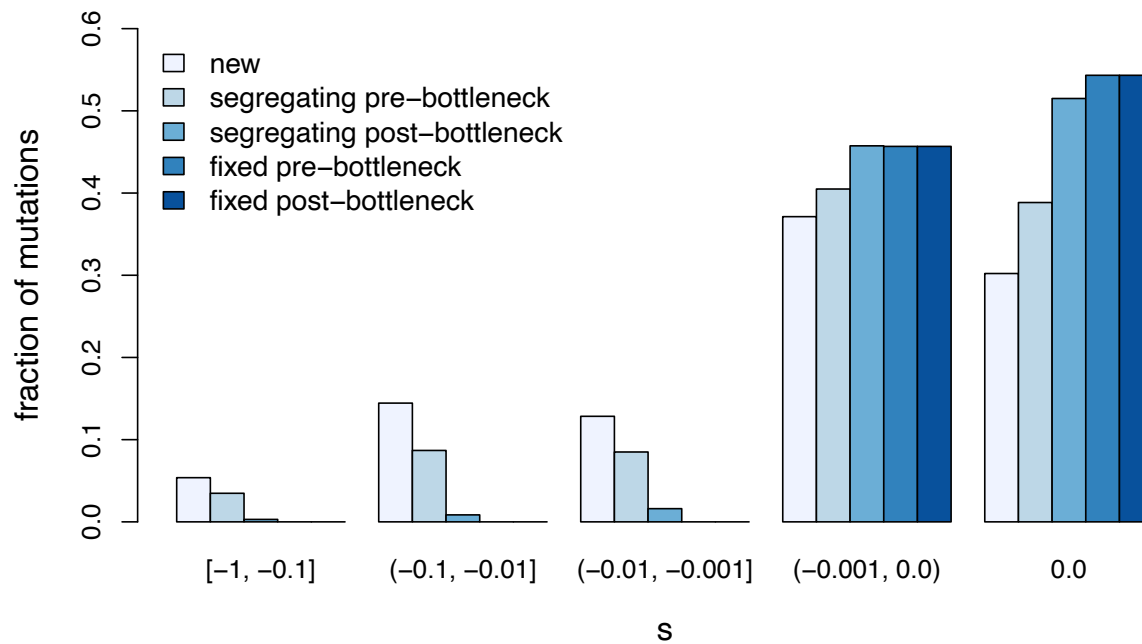


Fig. S14.

Comparison of the distribution of selection coefficients for new, segregating, and fixed nonsynonymous mutations in simulations. Note that a lower fraction of strongly and moderately deleterious mutations segregate compared to the fraction that enter the population, due to the effects of purifying selection. This effect increases during the contemporary bottleneck, as increased genetic drift and purifying selection against these mutations leads to their relative loss compared to weakly deleterious mutations. In addition, only weakly deleterious and neutral mutations become fixed, both before and after the bottleneck. Finally, note that the distribution of s for new mutations here differs from that presented in fig. S13, due to differences in parameterization between $\text{fit}\partial\text{a}\partial\text{i}$ and SLiM (see Materials and Methods).

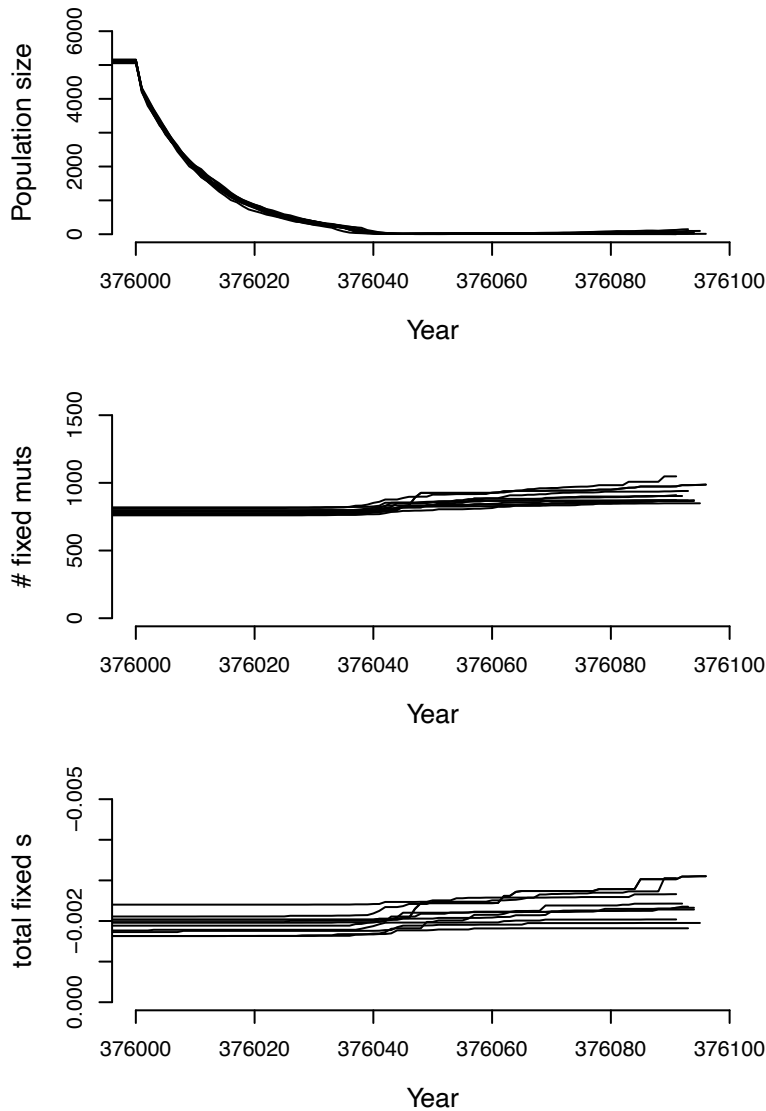


Fig. S15.

Total quantity and impact on fitness of fixed mutations during the simulated bottleneck. Top row depicts population sizes; middle row depicts total count of fixed mutations in the simulated population; bottom row depicts total sum of s mutations that became fixed until that timepoint. Note that although nearly 1,000 mutations fix by the end of the simulation, their cumulative impact on fitness remains relatively small (on the order of -0.2%, with a mean s for fixed mutations of -2.4×10^{-5}). Each line represents 1 of 10 simulation replicates, each with a burn-in duration of 350,000 years.

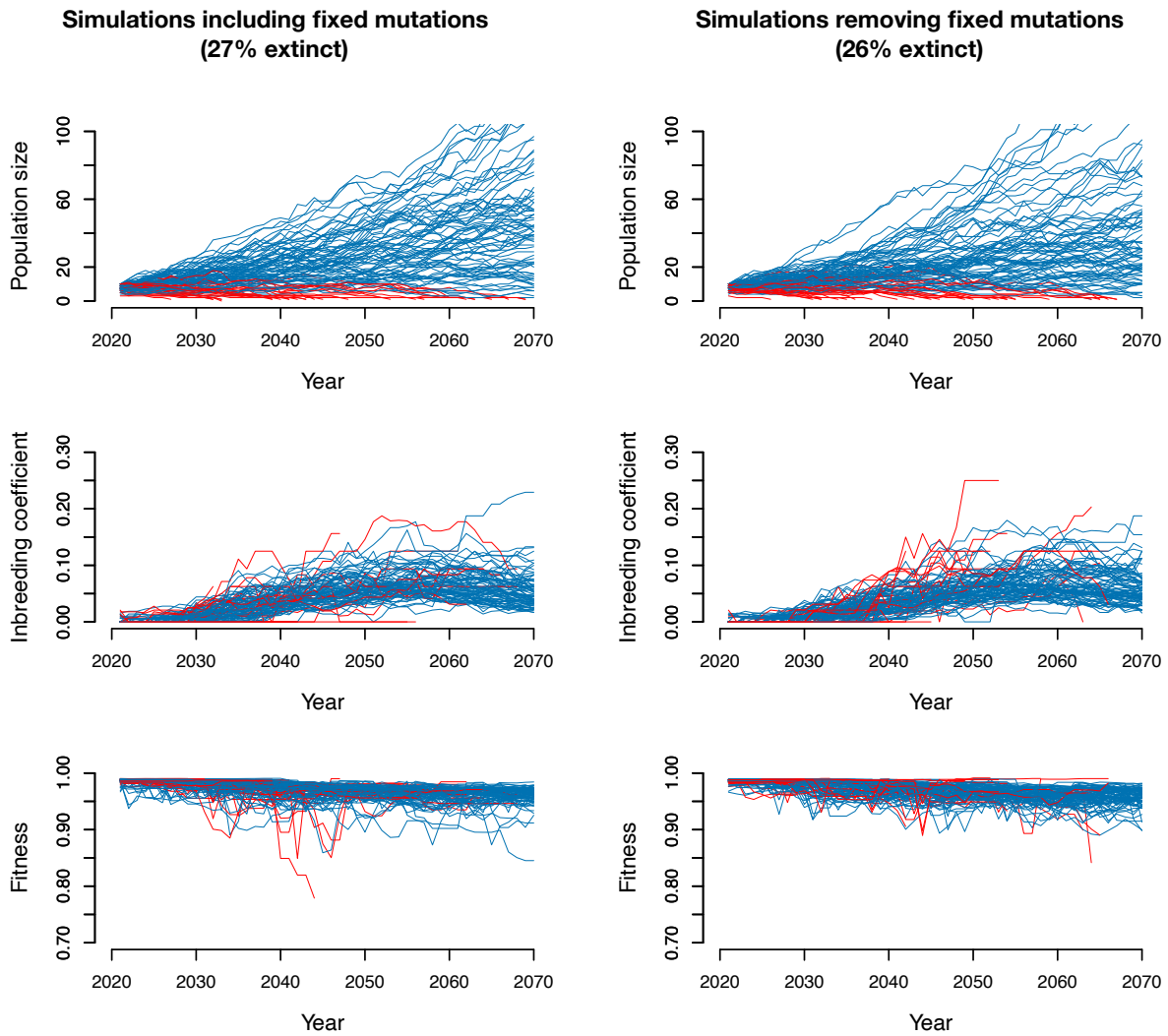


Fig. S16. Comparison of simulation results when retaining vs. removing fixed mutations. Both sets of simulations assume a 90% reduction in bycatch mortality and threshold population size of 10. Note the highly similar extinction rates, suggesting that fixed weakly deleterious mutations do not greatly impact model behavior.

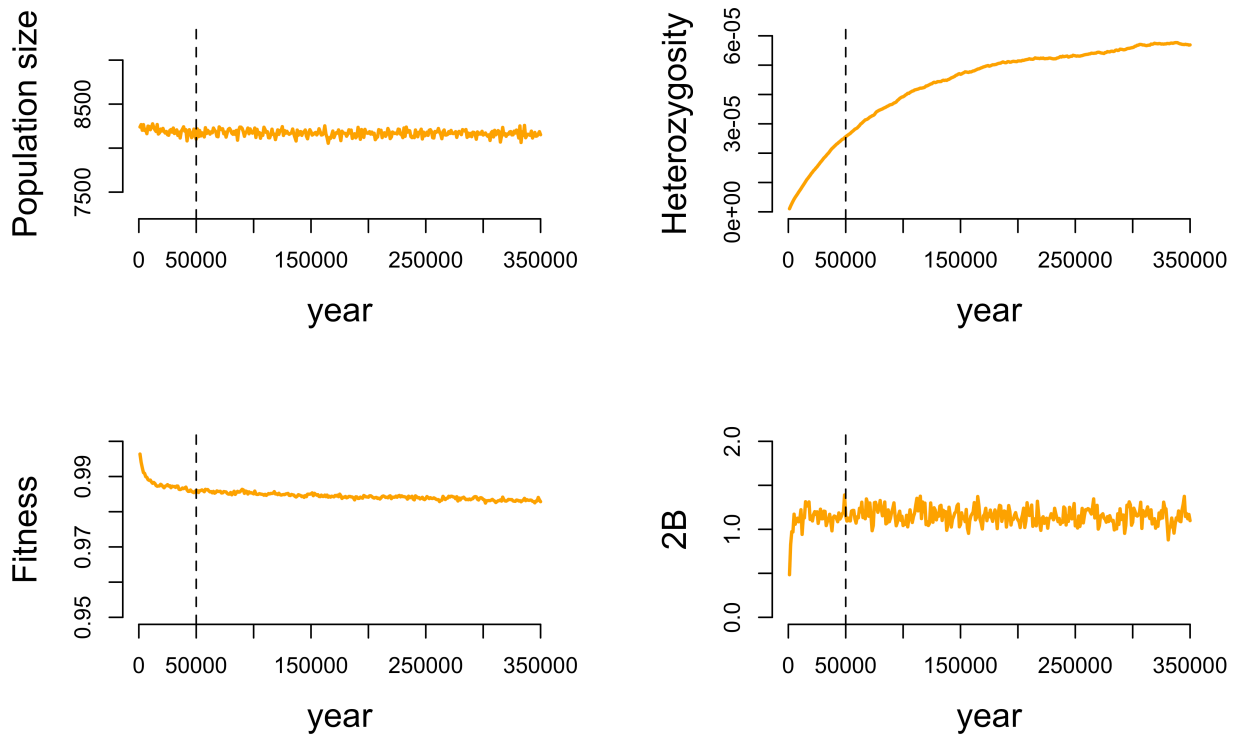


Fig. S17.

Simulation dynamics during a burn-in of 350,000 years. The population size, mean heterozygosity, mean fitness, and mean inbreeding load ($2B$) are shown as a function of time in the simulations. Statistics were recorded every 1,000 years from a sample of 60 individuals. The dotted line denotes 50,000 years, the duration of burn-in that was used for all simulation replicates. Note that, although heterozygosity does not reach equilibrium by 50,000 years, fitness and inbreeding load have stabilized by that time. See fig. S18 for a comparison of extinction dynamics with burn-in durations of 50,000 vs 350,000 years.

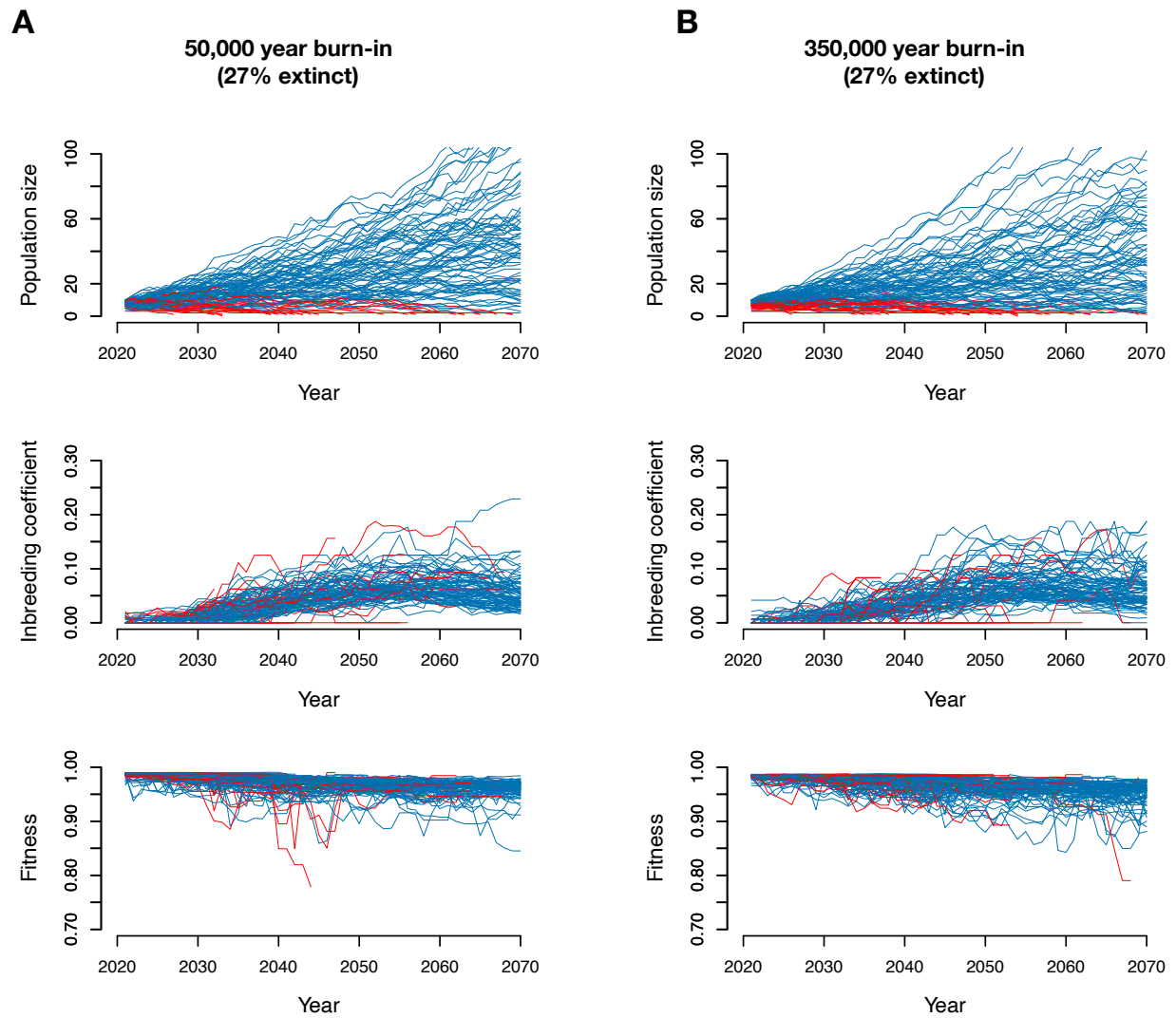


Fig. S18. Comparison of extinction rates and recovery trajectories using 50,000 year burn-in (A) vs a 350,000 year burn-in (B). Plotted is the size, average inbreeding coefficient, and average fitness of the simulated population across 100 simulation replicates. Replicates that persisted after 50 years are colored blue, whereas those that went extinct are colored red. Note the similar trends for (A) and (B), as well as the identical extinction rates.

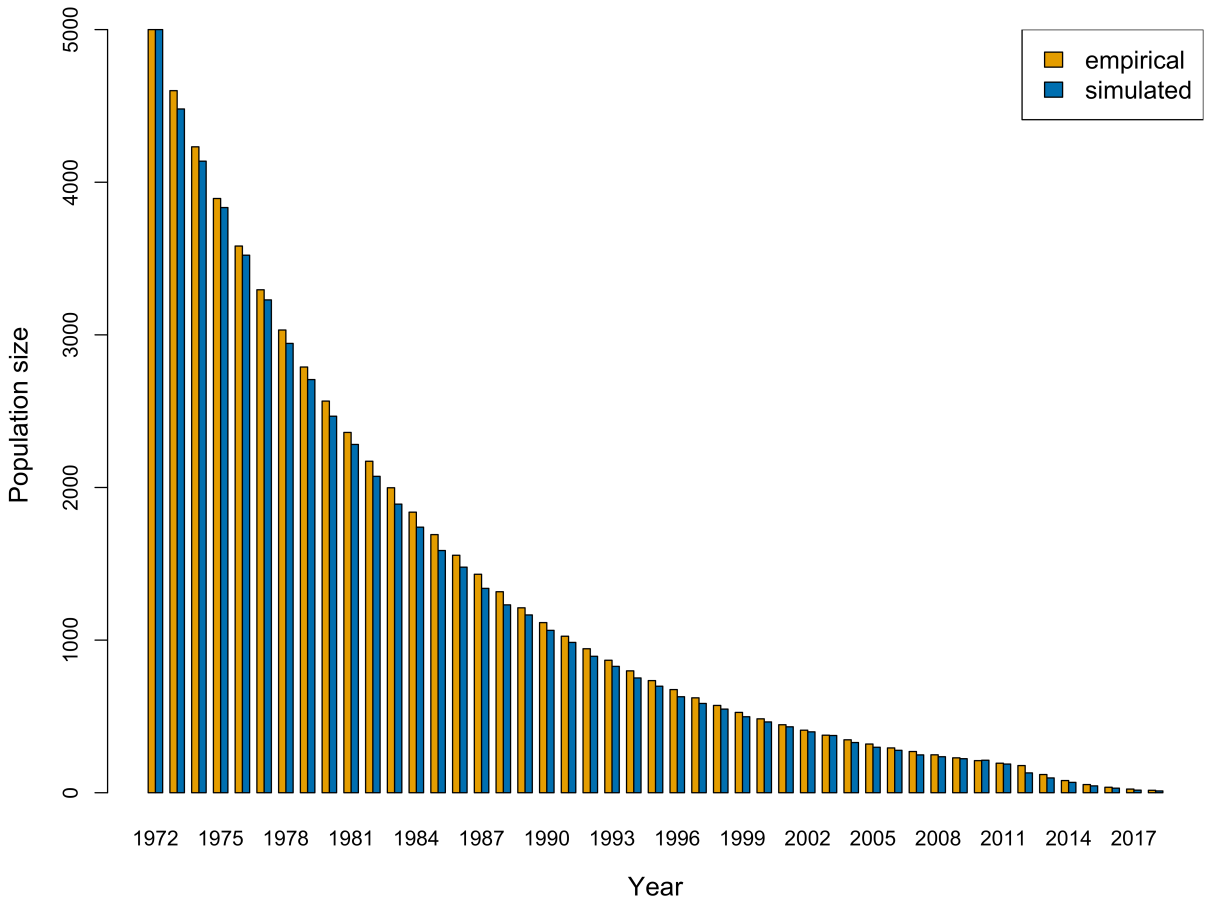


Fig. S19. Comparison of simulated and empirical bottleneck dynamics. Empirical population size trajectory is based on a model informed by census records (see Materials and Methods). Note the increase in the rate of decline starting in 2012.

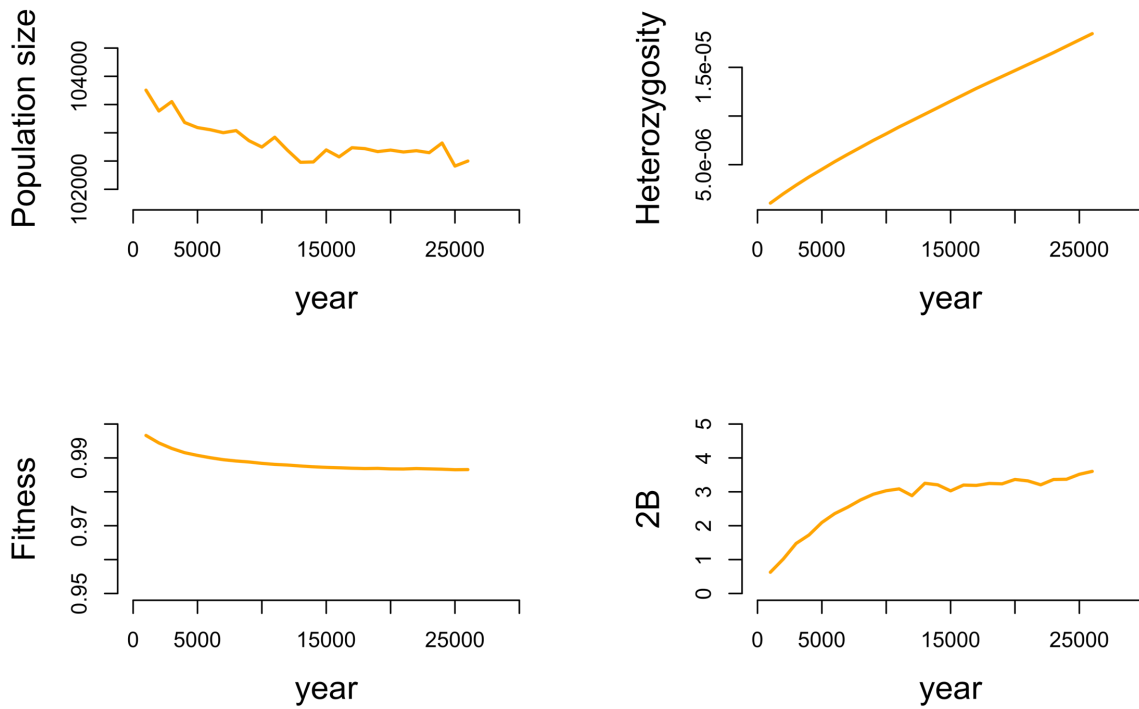


Fig. S20.

Burn-in dynamics for simulations with the historical population size increased by a factor of 20. The population size, mean heterozygosity, mean fitness, and mean inbreeding load ($2B$) during a 26,000 year burn-in are shown as a function of time in the simulations. Statistics were recorded every 1,000 years from a sample of 60 individuals. Note that, although heterozygosity does not reach equilibrium by 26,000 years, fitness and inbreeding load have stabilized by that time.

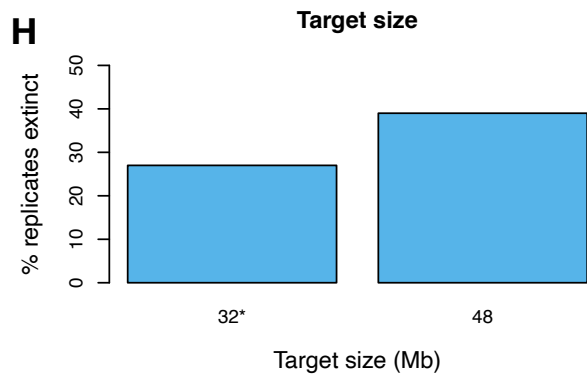
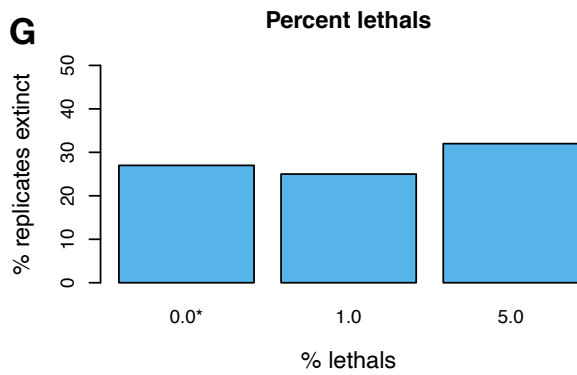
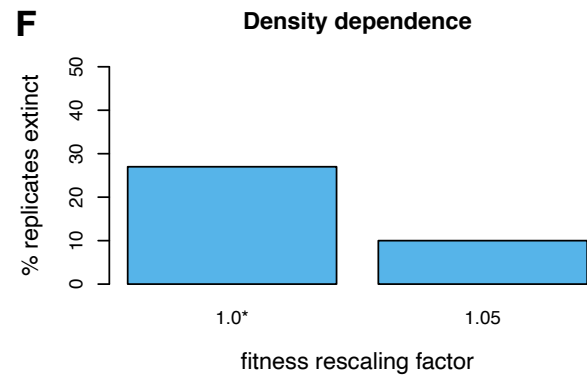
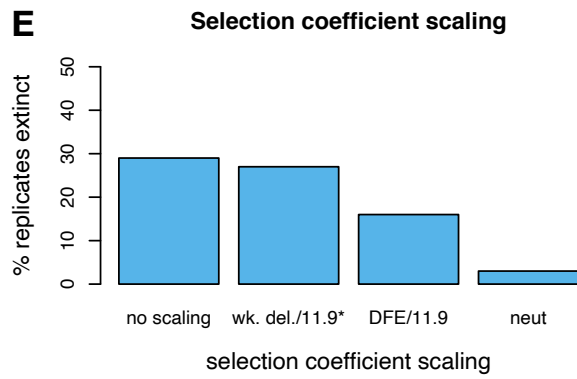
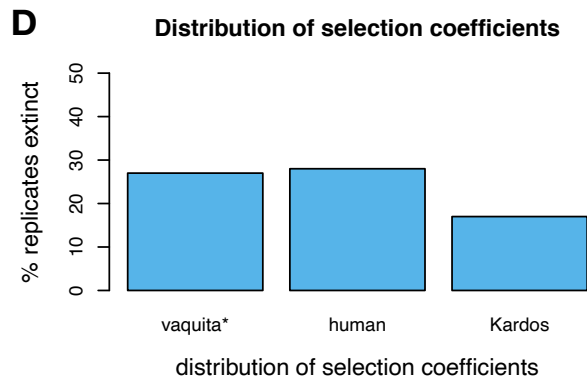
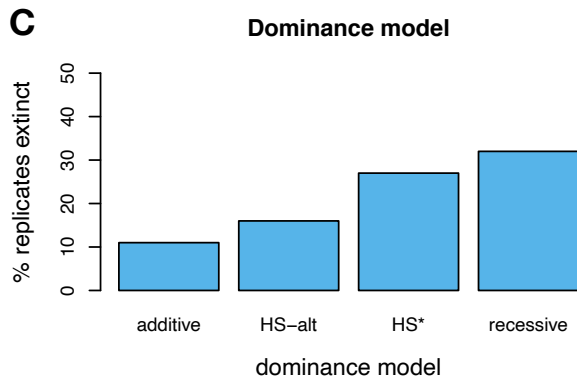
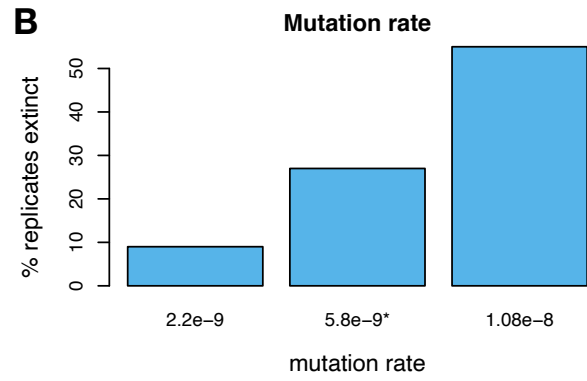
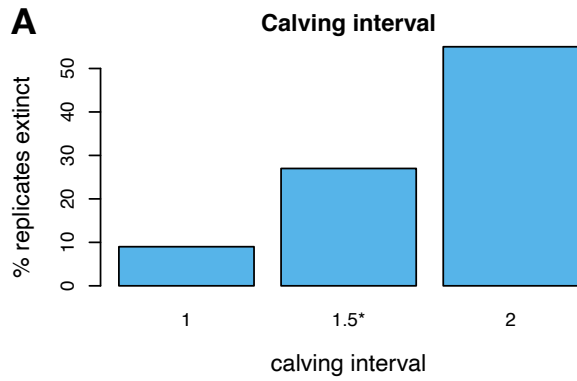


Fig. S21.

Sensitivity analysis of model parameters. (A) Extinction rates with calving intervals of 1, 1.5, or 2 years. (B) Extinction rates with varying mutation rates of 2.2×10^{-9} , 5.8×10^{-9} , or 1.08×10^{-8} mutations/site/generation. (C) Extinction rates under varying models of dominance. Under the recessive model, all mutations are assumed to be fully recessive, and under the additive model, all mutations are assumed to be fully additive. See Materials and Methods for details on the *hs* and *hs-alt* models. (D) Extinction rates under varying distributions of selection coefficients. The “human” distribution is based on results from (80) and the “Kardos” distribution is based on (85). Note that the “Kardos” distribution also includes a different distribution of dominance coefficients (see Materials and Methods). (E) Extinction rates under different scaling of selection coefficients to convert from per-generation to per-year selection coefficients. For “wk. del./11.9” only weakly deleterious mutations ($s > -0.01$) were scaled by the generation time (11.9 years), whereas for “DFE/11.9” the mean of the entire DFE was divided by the generation time. (F) Extinction rates under varying impacts of density-dependent fitness rescaling. For “1.0”, no upward rescaling of fitness was allowed when $N < K$, and for “1.05”, up to a 5% upward rescaling of fitness was allowed when $N < K$. (G) Extinction rates when adding a small fraction of recessive lethal mutations (1.0% or 5.0%) to the nonsynonymous distribution of selection coefficients. (H) Extinction rates when increasing the deleterious mutation target size from ~ 32 Mb to ~ 48 Mb. For all results, the empirically-inferred demographic parameters were used and we assumed a reduction in bycatch mortality rates of 90% and a threshold population size of 10. In each panel, the default model presented in the main text is denoted with an asterisk.

Table S1.

Vaquita sample information. We generated whole genome sequence data from 19 vaquitas and incorporated previously generated sequence data from a single individual (z0186935) (12). Genomic DNA libraries were prepared using one of two approaches: the Swift Biosciences Inc. Accel-NGS single-strand 1S DNA Library Kit or a blunt-end ligation method with double indexing (M&K) (30, 31). The library for the previously sequenced individual was generated with a 10X Genomics Chromium linked-reads library kit (12).

Sample	Year	Sex	Depth of coverage (X)	Library preparation method	SRA accession	Notes
z0000703	1985	F	53.3	Swift Biosciences	SRR15435925	
z0001649	1993	M	96.0	M&K	SRR15435924, SRR15435913	
z0001654	1992	F	95.8	M&K	SRR15435900, SRR15435899	
z0001660	1993	F	48.5	Swift Biosciences	SRR15435902	Fetus of z0001663
z0001663	1993	F	64.2	M&K	SRR15435897, SRR15435896	Mother of z0001660
z0004379	1990	F	41.8	Swift Biosciences	SRR15435898	
z0004380	1990	F	45.3	M&K	SRR15435895	Mother of z0004393
z0004381	1990	M	45.3	M&K	SRR15435921	
z0004382	1990	M	96.7	M&K	SRR15435923, SRR15435922	
z0004390	1991	F	103.0	M&K	SRR15435919, SRR15435918	
z0004393	1990	M	48.8	M&K	SRR15435920	Fetus of z0004380
z0004394	1991	M	47.0	Swift Biosciences	SRR15435917	
z0183496	2016	M	45.3	M&K	SRR15435914	
z0184983	2004	F	47.4	M&K	SRR15435910	
z0184984	2004	F	46.9	M&K	SRR15435907	
z0185383	2017	M	61.6	M&K	SRR15435916, SRR15435915	
z0185384	2017	F	24.0	M&K	SRR15435912, SRR15435911	
z0185385	2017	F	84.8	M&K	SRR15435909, SRR15435908	
z0186934	2017	F	44.5	M&K	SRR15435906	Live-caught individual; calf
z0186935	2017	F	61.3	10X Genomics Chromium	SRR15435905, SRR15435904, SRR15435903, SRR15435901	Individual used to generate reference genome (12)

Table S2.

Summary of parameter estimates and confidence intervals for each demographic model inferred with $\partial a \partial i$ and fastsimcoal2. The population size parameter values represent numbers of diploids, while the time parameters represent numbers of generations. For time parameters estimated with $\partial a \partial i$ for the 3- and 4-epoch models, the raw values were added to convert them to the total time elapsed from present to the time of the demographic event, for better comparison with fastsimcoal2 results. All times are presented forward in time. See fig. S4 for model illustrations. The 1- and 2-epoch models show high consistency in the parameter values estimated with both methods, while the 3- and 4-epoch models show greater discordance.

Model	Parameter	$\partial a \partial i$		fastsimcoal2	
		Estimate	95% CI	Estimate	95% CI
1-epoch	<i>Nanc</i>	3675	NA	3665	3589 – 3692
2-epoch	<i>Nanc</i>	4485	3468 – 5503	4439	4342 – 4536
	<i>Ncur</i>	2807	1832 – 3982	2784	2727 – 2841
	<i>T</i>	2162	0 – 5965	2045	1842 – 2248
3-epoch	<i>Nanc</i>	4336	3815 – 4857	4344	4279 – 4409
	<i>Nbot</i>	294	181 – 429	2253	1921 – 2584
	<i>Ncur</i>	3034	2131 – 4085	3131	2826 – 3435
	<i>Tbot</i>	710	67 – 1504	1124	964 – 1283
	<i>Tcur</i>	668	46 – 1437	264	176 – 351
4-epoch	<i>Nanc</i>	3798	2934 – 4662	5360	4958 – 5762
	<i>Nbot</i>	10747	0 – 29081	1674	1074 – 2273
	<i>Nrec</i>	95	49 – 156	2982	2895 – 3068
	<i>Ncur</i>	3044	1949 – 4376	398	39 – 756
	<i>Tbot</i>	3541	0 – 10562	4170	3968 – 4371
	<i>Trec</i>	863	12 – 2654	3520	3254 – 3785
	<i>Tcur</i>	830	0 – 2591	2	0.6 – 3

Table S3.

Performance of the best run for each demographic model under each inference method. K : the number of estimated parameters; Data Log-likelihood: the best possible log-likelihood derived from the data site frequency spectrum; Log-likelihood: the best log-likelihood estimated for the model; AIC: The Akaike information criterion, calculated using equation $AIC = 2 \cdot K - 2 \cdot \text{Log-likelihood}$. The convergence of parameters and log-likelihood is reported.

Model	<i>∂a∂i</i>					fastsimcoal2				
	K	Data Log-likelihood	Log-likelihood	AIC	Converged	K	Data Log-likelihood	Log-likelihood	AIC	Converged
1-epoch	0	-71.984	-1297.820	2595.641	Yes	1	-774592.691	-775127.249	1550256.498	Yes
2-epoch	2	-71.984	-94.761	193.521	Yes	3	-774592.691	-774600.596	1549207.192	Yes
3-epoch	4	-71.984	-89.958	187.915	Partial	5	-774592.691	-774599.411	1549208.822	Partial
4-epoch	6	-71.984	-88.047	188.094	Partial	7	-774592.691	-774602.510	1549219.02	Partial

Table S4.

Likelihood-ratio test (LRT) statistics between nested demographic models. The LRT was computed as: $-2 \times [\log\text{-likelihood}(\text{simple}) - \log\text{-likelihood}(\text{complex})]$. We only compared nested models with the closest complex model, that is, with the one that differed by only two more parameters. The LRT values and their significance are shown above the diagonal for *ada*i models and below the diagonal for fastsimcoal2 models. For both methods, the 2-epoch model has significantly better log-likelihood than the 1-epoch model, while the 3-epoch model log-likelihood is marginally significant only for *ada*i. Significance level: ** < 0.001, * < 0.01, the absence of asterisk represents non-significant values. The negative LRT statistic for the 3- and 4 epoch fastsimcoal2 comparison is likely due to variance in the simulations used to calculate the likelihood combined with the partial convergence of the optimization. In any case, this does not suggest an improvement in fit of the 4-epoch model over the 3-epoch model.

Model	1-epoch	2-epoch	3-epoch	4-epoch
1-epoch	–	2406.119**		
2-epoch	1053.306**	–	9.606383*	
3-epoch		2.37	–	3.8209
4-epoch			-7.698	–

Table S5.

Cetacean genome information. Species with publicly available annotated reference genomes and short read resequencing data from at least one individual in the SRA were obtained for comparisons with the vaquita. All assemblies and reads are available without restriction from NCBI. Except for the Yangtze finless porpoise, reads from each species were aligned to a reference genome from the same species.

Species common name	Species scientific name	Reference genome		Resequencing						
		Assembly	Source	BioSample	Origin	Same individual used to generate assembly?	Sex	Coverage (X)	Accession	Source
Beluga whale	<i>Delphinapterus leucas</i>	GCF_002288925.2_ ASM228892v3	(98)	SAMN06217832	Churchill, MB, Canada	No (daughter of reference individual)	F	33.9	SRR5197962	(98)
Blue whale	<i>Balaenoptera musculus</i>	GCF_009873245.2_ mBalMus1.pri.v3	*	SAMN07201754	California, USA	No	M	37.0	SRR5665644	(99)
Bottlenose dolphin	<i>Tursiops truncatus</i>	GCF_011762595.1_ mTurTru1.mat.Y	^	SAMN02192671	Captive	No	M	38.2	SRR940825	(52)
Indo-Pacific finless porpoise	<i>Neophocaena phocaenoides</i>	GCF_003031525.1_ Neophocaena_asiae orientalis_V1	(100)	SAMN02192673	South Korea	No	?	27.1	SRR940959	(52)
Orca/killer whale	<i>Orcinus orca</i>	GCF_000331955.2_ Oorc_1.1	(101)	SAMN01180276	Alaska, USA	Yes	F	16.3	SRR1164379	(102)
Long-finned pilot whale	<i>Globicephala melas</i>	GCF_006547405.1_ ASM654740v1	#	SAMN11083132	PEI, Canada	Yes	M	32.5	SRR8867567	#
Minke whale	<i>Balaenoptera acutorostrata scammoni</i>	GCF_000493695.1_ BalAcu1.0	(52)	SAMN02192644	South Korea	No	F	28.7	SRR924087	(52)
Narwhal	<i>Monodon monoceros</i>	GCF_005190385.1_ NGI_Narwhal_1	(103)	SAMN10519625	Greenland	No	F	104.8	SRR8284577, SRR8284578, SRR8284579	(103)
Pacific white-sided dolphin	<i>Lagenorhynchus obliquidens</i>	GCF_003676395.1_ ASM367639v1	#	SAMN09386610	Canada	Yes	F	33.3	SRR7345555	#
Sperm whale	<i>Physeter macrocephalus</i>	GCF_002837175.2_ ASM283717v2	(104)	SAMN06187414	Pacific Ocean	No	?	23.2	SRR5136492, SRR5136494, SRR5136498	(105)
Yangtze finless porpoise	<i>Neophocaena asiaorientalis</i>	GCF_003031525.1_ Neophocaena_asiae orientalis_V1	(100)	SAMN08512128	Hubei, China	No	M	21.3	SRR6923837	(106)

*Provided by Yury Bukhman and the Vertebrate Genomes Project.

^Provided by the Vertebrate Genomes Project.

#Provided by the Canseq150 program supported by CGEN and its partners.

Table S6.

Impact of genetic parameters on inbreeding load and extinction rates. Each row represents a different combination of simulation parameters, including a dominance model, carrying capacity during the second historical epoch (K_2), mutation rate, distribution of selection coefficients, and deleterious mutation target size. For each combination of parameters, we report the resulting levels of strongly deleterious mutations ($s < -0.01$) per individual, very strongly deleterious mutations ($s < -0.1$) per individual, lethals per individual ($s < -0.5$), inbreeding load ($2B$), and percent of replicates going extinct within 50 years of recovery initiation assuming a 90% reduction in bycatch mortality and threshold population size of 10. See Materials and Methods for details on the different simulation parameters. Note that, across different distributions of selection coefficients and dominance effects, the percent of replicates going extinct for the small carrying capacity ($K_2 = 5,200$) is $<50\%$ in all cases, and for many parameter combinations, is $<30\%$. Thus, these results suggest that extinction of the vaquita due to inbreeding depression is not certain.

Dominance model	K_2	Mutation rate	Dist. of sel. coeff.	Target size (Mb)	avg # muts $s < -0.01$	avg # muts $s < -0.1$	avg # muts $s < -0.5$	$2B$	% extinct
hs	5200	5.80E-09	vaquita	31.6	12.3	3	0.14	0.95	27
hs	104000	5.80E-09	vaquita	31.6	30.1	13.1	0.6	3.32	52
recessive	5200	5.80E-09	vaquita	31.6	17.7	3.1	0.14	1.28	32
additive	5200	5.80E-09	vaquita	31.6	0.4	0.02	0	0.016	11
hs-alt	5200	5.80E-09	vaquita	31.6	4.38	0.87	0.014	0.26	16
hs	13709	2.20E-09	vaquita*	31.6	4.63	0.46	0	0.19	9
hs	2793	1.08E-08	vaquita*	31.6	20.86	6.74	0.92	2.4	55
hs	5200	5.80E-09	human	31.6	12.3	1.03	0	0.5	28
hs-kardos	5200	5.80E-09	kardos	31.6	3.5	2.12	1.83	1.9	17
hs	5200	5.80E-09	vaquita + 1.0% lethals	31.6	12.65	3.35	0.51	1.31	25
hs	5200	5.80E-09	vaquita + 5.0% lethals	31.6	13.6	4.64	1.85	2.64	32
hs	5200	5.80E-09	vaquita	48	18.5	4.5	0.21	1.44	39

*Using the vaquita distribution of selection coefficients inferred under the respective assumed mutation rate (fig. S13)