

OPEN

Long-term isolation at a low effective population size greatly reduced genetic diversity in Gulf of California fin whales

Vania E. Rivera-León¹, Jorge Urbán², Sally Mizroch³, Robert L. Brownell Jr.⁴, Tom Oosting¹, Wensi Hao¹, Per J. Palsbøll^{1,5} & Martine Bérubé^{1,5}

The Gulf of California, Mexico is home to many cetacean species, including a presumed resident population of fin whales, *Balaenoptera physalus*. Past studies reported very low levels of genetic diversity among Gulf of California fin whales and a significant level of genetic differentiation from con-specifics in the eastern North Pacific. The aim of the present study was to assess the degree and timing of the isolation of Gulf of California fin whales in a population genetic analysis of 18 nuclear microsatellite genotypes from 402 samples and 565 mitochondrial control region DNA sequences (including mitochondrial sequences retrieved from NCBI). The analyses revealed that the Gulf of California fin whale population was founded ~2.3 thousand years ago and has since remained at a low effective population size (~360) and isolated from the eastern North Pacific ($N_e m$ between 0.89–1.4). The low effective population size and high degree of isolation implied that Gulf of California fin whales are vulnerable to the negative effects of genetic drift, human-caused mortality and habitat change.

Population genetic data hold the potential to provide insights into the basis of genetic variation in natural populations¹, which in turn is a function of past and present effective population sizes, migration rates, selection, and mutation rates. A population acquires genetic variation by mutation and immigration. Genetic variation is lost by random genetic drift resulting in higher rates of loss in small and isolated populations compared to larger, connected populations². Possible consequences of low genetic variation include increased genetic load and loss of adaptive traits, which in turn can elevate short- and long-term extinction rates^{3–6}. Consequently, understanding the processes that shaped contemporary levels of genetic variation may add vital insights to inform the conservation and management of local populations.

Population genetic assessments have proven particularly useful to study elusive species, such as baleen whales^{7–10}. Baleen whales typically undertake long-range seasonal migrations to forage in cold nutrient-rich waters at high latitudes during the summer and winter in warm tropical waters during the breeding season¹¹. Exceptions to this pattern are Omura's (*Balaenoptera omurai*)¹² and Bryde's whales (*Balaenoptera brydei*)¹³, which remain at low latitudes all-year as well as bowhead whales (*Balaena mysticetus*) that are confined to polar waters¹⁴. A few populations among those baleen whale species that undertake seasonally migrations, do not migrate but remain in the same area year-round, such as blue whales (*Balaenoptera musculus*)^{15,16} and humpback whales (*Megaptera novaeangliae*) in the Arabian Sea^{17–20}.

Baleen whales are also observed year-round in the Gulf of California, Mexico. Baleen whales are commonly observed around the Midriff Islands and in the Canal de Ballenas between the east coast of the Baja California Peninsula and Isla Angel de la Guarda^{21–23}. These whales were probably first noted by, and the canal named by,

¹Marine Evolution and Conservation, Groningen Institute of Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands. ²Departamento de Ciencias Marinas y Costeras, Universidad Autónoma de Baja California Sur, Km 5.5 Carretera al Sur, 23081, La Paz, Baja California Sur, Mexico. ³Blue Sea Research PO Box 15805, Seattle, WA, 98115, United States of America. ⁴Southwest Fisheries Science Center, NOAA Fisheries, 34500 Highway 1, Monterey, CA, 93940, United States of America. ⁵Centre for Coastal Studies, 5 Holway Avenue, Provincetown, Massachusetts, 02657, United States of America. Per J. Palsbøll and Martine Bérubé jointly supervised this work. Correspondence and requests for materials should be addressed to V.E.R.-L. (email: v.e.rivera.leon@rug.nl) or P.J.P. (email: p.j.palsboll@rug.nl) or M.B. (email: m.berube@rug.nl)

Received: 13 September 2018

Accepted: 6 August 2019

Published online: 27 August 2019

the Croatian Jesuit cartographer Fernando Consag during his 1746 expedition to the Upper Gulf of California undertaken to prove that Baja California was not an island. The whales in the Canal de Ballenas were identified as fin whales as early as the late 18th Century²⁴ and were later “rediscovered” during the 1940 expedition on the Western Flyer in the Sea of Cortez^{23,25}. Gilmore^{23,26} was the first to suggest that the fin whales in Gulf of California were all-year residents, a notion later repeated by Norris²⁷. Since the 1980s, regular surveys aimed at baleen whales in the Gulf of California have been conducted in the waters off the Midriff Islands, in particular Canal de Ballenas^{21,23,28–30}. Fin whales, and their vocalizations have been recorded throughout the year in the Gulf of California and appear most abundant in the Upper Gulf^{21,23,31,32}.

The Gulf of California fin whales appear isolated from con-specifics in the eastern North Pacific. Fin whales have not been observed in the entrance of the Gulf of California and sightings off the southern tip (south of 24°N) of the Baja California Peninsula have been scarce^{31,33–35}, although some fin whales were sighted south of 24°N in February during the annual surveys between 1965 and 1967³⁶. The presumed residency of this fin whale population was also supported by movement data from 11 satellite-radio tagged fin whales, which remained in the Gulf of California while transmitting location data^{34,37}. Further supporting the notion of isolation were the 20 Hz fin whale calls recorded in the Gulf of California that were unique and distinct from fin whale calls recorded in the North Pacific and North Atlantic³⁸. However, some occasional tempo-spatial overlap was noted in fin whale calls recorded in the Gulf of California and off Southern California, indicating the possibility of some connectivity between these two areas³⁹.

Genetic analyses conducted by Bérubé *et al.*^{40,41} confirmed the presumed reproductive isolation of the Gulf of California fin whales. Bérubé *et al.*⁴⁰ compared mitochondrial control region (mtCR) DNA sequences and genotypes at 16 nuclear microsatellite loci in 12 samples from the eastern North Pacific with 77 samples from the Gulf of California. Bérubé *et al.*⁴⁰ detected a high degree of genetic differentiation between the two areas; Weir's θ^{42} was estimated at 0.24 and 0.22 for mtCR DNA sequences and microsatellite genotypes, respectively. The high degree of genetic differentiation estimated by Bérubé and co-workers^{40,41}, in combination with the very low levels of genetic diversity observed among the Gulf of California fin whales, strongly suggested that this population was reproductively isolated from con-specifics in the eastern North Pacific. However, to date no study has attempted to estimate migration rates between the Gulf of California and the eastern North Pacific or when the Gulf of California population was founded and presumably became isolated. Nigenda-Morales *et al.*⁴³ also detected low levels of genetic diversity in the Gulf of California fin whales at the usually highly diverse major histocompatibility complex (MHC) DQB-1 locus. Nigenda-Morales and colleagues⁴³ detected only three alleles in a sample of 36 fin whales, a level of MHC diversity comparable to that observed in the severely bottlenecked northern elephant seal, *Mirounga angustirostris*⁴⁴. The low genetic diversity at the mtCR DNA as well as at nuclear microsatellite and MHC loci suggested that effective population size (N_e) of the Gulf of California was small, but so far, no actual estimates of N_e have been reported.

The low level of genetic diversity among the Gulf of California fin whales is a function of N_e as well as the duration and degree of isolation. Extended periods at a low N_e increase random genetic drift possibly leading to inbreeding depression⁴⁵ as well as loss of adaptive traits and potentially increasing the overall risk of extinction⁴⁶. The aims of this study were to estimate when the Gulf of California fin whale population was founded, the census (N_c) and N_e , as well as the degree of connectivity with con-specific populations in the eastern North Pacific. Towards this end, a population genetic assessment of the Gulf of California and eastern North Pacific fin whale specimens was undertaken based upon genotypes at 18 nuclear microsatellite loci in 402 samples and 565 mtCR DNA sequences (including previously published mtCR DNA sequences retrieved from the National Center for Biotechnology Information, NCBI). The estimates suggested that the Gulf of California fin whale population was founded approximately 2.3 thousand years ago (kya), and has since remained at a low N_e . The estimates of historic and contemporary connectivity with the eastern North Pacific fin whales were low as well.

Results

A total of 402 tissue samples were collected from individuals off Kodiak Island, off the Californian coast and in the Gulf of California. The sample sizes per year were: Kodiak Island, 14 samples in 2002; coastal California, 16 samples in 1996 and Gulf of California, 372 samples collected in 1993–1995 and 1997–2004 (Table 1 and Fig. 1). The sex, the genotype at 18 nuclear microsatellite loci and the DNA sequence of the 5'-end of the mtCR was determined in all the tissue samples. Two samples from Coastal California and seven from the Gulf of California samples were removed from subsequent data analyses because data were missing at five or more microsatellite genotypes. The assessment of the consistency rate per multi-locus genotype identified a single inconsistent genotype among 470 microsatellite genotypes, yielding an estimate at ~0.2% per genotype. The actual consistency rate in the final dataset was lower since all samples mismatching at four or fewer loci were checked, possibly re-amplified, and corrected if necessary.

Probability of identity and sex ratios. The estimates of the probability of identity (PI)⁴⁷ in the three samples were sufficiently low to discern among unrelated individuals. PI was estimated at 2.15×10^{-22} in the Kodiak Island sample, 1.75×10^{-22} in the coastal California sample and 9.08×10^{-12} in the Gulf of California sample. Accordingly, the expected numbers of pairs of unrelated individuals with identical genotypes by chance in the sample were estimated at 9.24×10^{-21} , 8.57×10^{-21} and 3.04×10^{-7} , respectively in the above mentioned samples. Accordingly, all specimens with identical multi-locus microsatellite genotypes, sex and mtCR DNA sequences were inferred as duplicate specimens collected from the same individual, and only data from a single tissue sample were retained in each sample partition during data analysis. No statistically significant deviation from parity was observed for the sex ratio among samples from Kodiak Island, coastal California and the Gulf of California.

The final data set comprised 284 unique individuals totalling 5,092 single-locus genotypes. Six and 14 single-locus microsatellite genotypes were missing among the 25 and 259 unique individuals detected in the

	Kodiak Island Alaska	Coastal California	Eastern North Pacific	Gulf of California
Tissue samples	14	16	30	372
Nuclear microsatellite genotypes				
<i>n</i>	14	11	25	259
<i>PI</i>	2.1×10^{-22}	1.4×10^{-22}	3.1×10^{-24}	4.8×10^{-11}
N_a	8.8 ± 3.2	8 ± 2.6	10.8 ± 3.7	7.6 ± 2.4
H_o	0.76 ± 0.2	0.82 ± 0.14	0.79 ± 0.15	0.48 ± 0.21
H_e	0.81 ± 0.14	0.83 ± 0.08	0.82 ± 0.1	0.49 ± 0.21
A_R	7.3 ± 2.36	7.5 ± 2.27	7.5 ± 2.09	3.8 ± 1.32
F_{IS}	0.057	0.015	0.047	0.024
Mitochondrial control region DNA sequences				
<i>n</i>	152 ^{&}	160 ^{&}	312 ^{&}	253
<i>S</i>	22	19	24	5
h_n	25	21	31	4
θ	0.014 ± 0.004	0.012 ± 0.004	0.014 ± 0.003	0.003 ± 0.001
π	0.008 ± 0.0005	0.006 ± 0.0005	0.007 ± 0.0003	0.0006 ± 0.0001
<i>h</i>	0.86 ± 0.02	0.77 ± 0.02	0.82 ± 0.015	0.15 ± 0.03

Table 1. Samples, individuals and genetic diversity of fin whales in three sampling locations. Notes: *n*, sample size after removing poor quality samples and duplicates; *PI*, probability of identity; N_a , the number of alleles; H_o , the observed heterozygosity; H_e , the expected heterozygosity and A_R , allelic richness (based upon 18 individuals); F_{IS} , the inbreeding coefficient; *S*, the number of segregating sites; h_n , the number of haplotypes; θ per site from *S*; π , nucleotide diversity and *h*, haplotype diversity. [±]Denotes the standard deviation. The Eastern North Pacific is comprised by the Kodiak Island, Alaska and the Coastal California. [&]Additional mtCR DNA sequences obtained from NCBI.

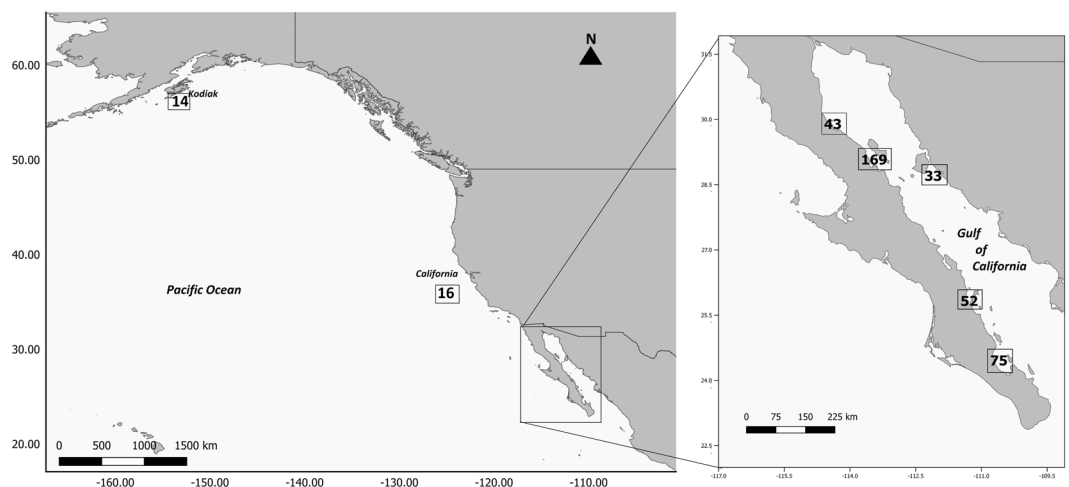


Figure 1. Map of the sampling areas and sample sizes.

eastern North Pacific and Gulf of California sample, respectively. The final mtCR DNA sequence dataset comprised 565 sequences (incl. 284 retrieved from NCBI, see Table S1), each of a length of 280 base pairs. The amounts of data per sampling area, after removal of specimens with insufficient data and duplicate data, are summarised in Table 1.

Deviations from Hardy-Weinberg genotype proportions and linkage equilibrium. Among the Kodiak Island samples, a single statistically significant deviation (locus EV001) from the expected Hardy-Weinberg genotype proportions was detected, whereas significant deviations were detected at eight loci in the Gulf of California sample after applying a false discovery rate (FDR)⁴⁸ at 0.05 (Table S2). After removing two individuals that were inferred as recent immigrants (see below), the deviations in the Gulf of California sample persisted at five loci (loci EV037, GATA053, GATA098, GGAA520 and GT211, Table S2). Unsurprisingly, the assessment of null alleles suggested the presence of null alleles mostly at the loci that deviated from the expected Hardy-Weinberg genotype frequencies, i.e., Kodiak Island (locus EV001) and the Gulf of California (loci EV037, GATA053, GATA098 and GT575). No significant deviations from linkage disequilibrium were detected after applying FDR.

	Kodiak Island, Alaska	Coastal California	Gulf of California
Kodiak Island, Alaska		0.027**	0.25***
Coastal California	0.016* [0.004–0.03]		0.19***
Gulf of California	0.25*** [0.17–0.34]	0.21*** [0.13–0.29]	

Table 2. Population differentiation estimates. Notes: Above the diagonal; ϕ_{ST} estimated from the mtCR DNA sequences. Below the diagonal; F_{ST} estimated from the microsatellite genotypes along with the 95% confidence interval in brackets. * $P \leq 0.05$; ** $P \leq 0.01$ and *** $P \leq 0.001$.

	$N_{e(ENP)}$ *	$N_{e(GC)}$ *	$N_{e(A)}$ *	$T^{\&}$	$m_{(GC \rightarrow ENP)}$	$m_{(ENP \rightarrow GC)}$	$N_{e(GC)}m_{(ENP \rightarrow GC)}$
	MSVAR				BAYESASS		
msat	NA	253 (126–516)	12,500 (5,900–26,800)	19,000 (9,200–38,700)	0.020 (0–0.06)	0.004 (0.0005–0.009)	1.4 (0.12–5.39)
	IMA2P						
mtCR	28,300 (16,400–47,900)	330 (0–2,300)	21,700 (0–125,300)	1,370 (0–2,700,000)	0.0008 (0.0002–0.002)	0.0002 (0–0.0047)	0.89 (0.04–3.13)
mtCR & msat	6,500 (2,300–60,800)	40 (0–700)	7,400 (4,900–12,700)	2,300 (0–62,500)	0.0003 (0–0.002)	0.002 (0–0.015)	—

Table 3. Estimates of effective population size, divergence times and immigration rates. Notes: *ENP, GC and A, denotes the eastern North Pacific, Gulf of California and ancestral population, respectively. $\&$ Estimated timing of the event in years. N_e denotes the effective population size and m the immigration rate. mtCR, denotes the mitochondrial control region DNA sequences and msat, denotes microsatellite genotypes. The contemporary immigration rates from BAYESASS represent the fraction of individuals in the eastern North Pacific that are immigrants from the Gulf of California per generation ($m_{(GC \rightarrow NP)}$) and the fraction of individuals in the Gulf of California that are immigrants from the eastern North Pacific ($m_{(NP \rightarrow GC)}$). In order to facilitate a comparison with the estimates of the long-term immigration rates per generation obtained with IMA2P (Tables S6 and S7) which are scaled to the mutation rate, IMA2P estimates were transformed to m . The mutation rate for the transformation of the mtCR data was 5.2×10^{-8} per site per year, considering 280 base pairs and a generation time of 25.9 years. The mutation rate for the transformation of the combined dataset was 5×10^{-4} per generation, which is the geometric mean of the mtCR and msat mutation rate. The contemporary number of immigrants per generation from the North Pacific to the Gulf of California ($N_{e(GC)}m_{(NP \rightarrow GC)}$) was estimated using the estimates obtained with BAYESASS and the estimate of N_e obtained with NEESTIMATOR. The long-term $N_e m$ was estimated using IMA2P. $N_{e(GC)}m_{(NP \rightarrow GC)}$ for the combined data (mtCR and msat) was not estimated (see Fig. S8) but the mutation rate scaled by the mutation rate was (m , Fig. S9).

Patterns of population genetic diversity and differentiation. Estimates of genetic diversity at nuclear loci and mtCR DNA sequences in sampling areas outside the Gulf of California yielded values approximately twice the levels observed in the Gulf of California (Table 1). The level of genetic differentiation was high for the mtCR DNA sequences and nuclear loci in comparisons involving the Gulf of California (Table 2). The lowest level of population genetic differentiation was detected between the Kodiak Island and the coastal California samples ($\phi_{ST} = 0.027$ for the mtCR DNA sequences and $F_{ST} = 0.016$, 95% confidence interval (CI): 0.004–0.03 for the microsatellite genotypes; Table 2). Since the degree of population genetic differentiation between the two eastern North Pacific samples was one order of magnitude lower than comparisons including the Gulf of California, the two eastern North Pacific samples were combined into a single “eastern North Pacific” sample during subsequent data analyses. The eastern North Pacific samples were included in the analysis specifically as a means to gain insights into the history and isolation of the Gulf of California population. The aim of this study was not to make inferences concerning potential connectivity and divergence times among eastern North Pacific sample areas *per se*, given the modest sample sizes available from the eastern North Pacific region.

Contemporary estimates of $N_e m$, N_e and N_c . Contemporary immigration rates were estimated using the approach implemented in BAYESASS. The Markov chain Monte Carlo (MCMC) mixing parameters for the migration rates was at 0.07, the allele frequencies at 0.2 and inbreeding coefficients at 0.15 (for convergence, Fig. S1). The number of immigrants per generation ($N_e m$) from the eastern North Pacific into the Gulf of California (Table 3 and Fig. S2) was estimated at ~ 1.4 (95% High Posterior Density (HPD): ~ 0.12 –5.39, effective sample size (ESS): 2,444) employing an estimate of N_e at 360 (see below). Excluding loci with possible null alleles from the assessment did not alter the estimates (Table S3). A single first (migration ancestry = 1) and a single second-generation immigrant (migration ancestry = 0.99) were identified among the Gulf of California samples.

The linkage disequilibrium-based estimate of N_e was 360 [95% CI: 230–665] in the Gulf of California sample. N_c in the Gulf of California was estimated from 259 unique individuals, among which one, five and 40 individuals were sampled in four, three, or two different years, respectively (Table S4). The remaining individuals were

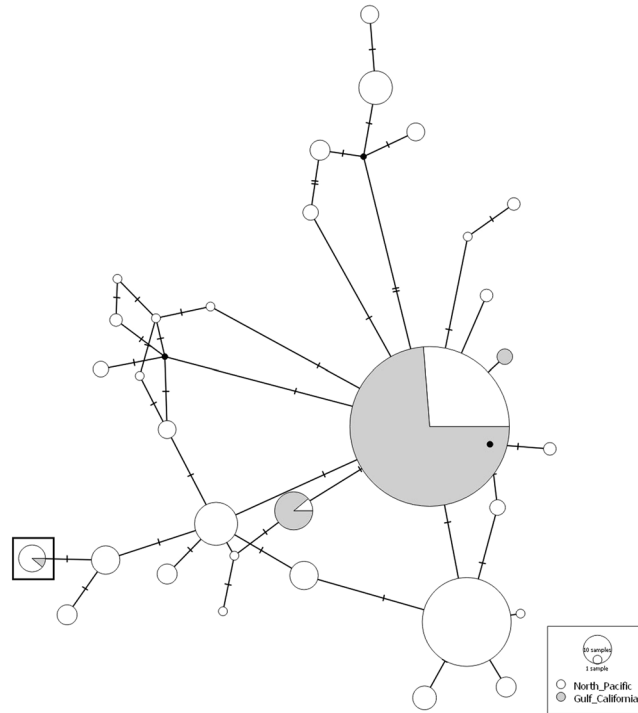


Figure 2. Media Joining network inferred from fin whale mtCR DNA sequence haplotypes. The haplotype of the individual sampled in the Gulf of California inferred to be a first-generation immigrant is marked with a square.

sampled only once. The above re-capture histories yielded an estimate of N_e at 730 [95% CI: 530–930] individuals after model averaging (Table S5). The N_e estimate refers to the total number of individuals that existed during the study period.

Effective population sizes and divergence time. A possible change in the effective population size in the Gulf of California fin whales was estimated using MSVAR⁴⁹. The MSVAR estimate (microsatellite genotypes only) obtained from the Gulf of California samples yielded an estimate of an ancestral N_e at 12,000 (ESS: 5,220), which was reduced to an N_e at ~250 (ESS: 3,007) approximately 19 kya [95% HPD: 9.22–38.7 kya, ESS: 3,075] (Table 3; Figs S3 and S4). The population divergence time, long-term N_e , and immigration rates were estimated using the parallelised version of the isolation-with-migration model⁵⁰ implemented in IMA2P⁵¹. The IMA2P parameter estimates differed among the data employed in the analysis but were consistent with the MSVAR results in terms of a low N_e for the Gulf of California. The IMA2P assessment based solely on mtCR DNA sequences yielded a N_e for the Gulf of California at ~330 [95% HPD: 0–2,300; ESS: 75,588]. The IMA2P assessment based upon both mtCR DNA sequences and microsatellite genotypes resulted in an estimate of N_e at ~40 [95% HPD: 0–700; ESS: 1,202]. The estimate of the divergence time between the Gulf of California and the eastern North Pacific was estimated at ~1.3 kya [95% HPD: 0–2,700 kya; ESS: 559,148] when employing only mtCR DNA sequences and ~2.3 kya [95% HPD: 0–62.5 kya; ESS: 4,615] for the combined data (Tables 3, S6 and S7; Figs S5 and S6). $N_e m$ following population divergence was estimated at 0.89 [95% HPD: 0.03–3.13; ESS: 49,663] for the mtCR DNA sequences (Tables 3 and S6; Fig. S7). However, obtaining an estimate of $N_e m$ from the combined data was not feasible (Tables 3 and S7; Figs S8 and S9).

Because the 95% HPD interval of the IMA2P estimate of the population divergence time included zero, the relative probabilities of a recent and an older population divergence time were estimated from the mtCR DNA sequences and microsatellite genotypes using the approximate Bayesian computation (ABC)⁵² approach implemented in the software DIYABC⁵³. The scenario with an older population divergence time (>1 kya) was preferred with a posterior probability at 0.89 [95% CI: 0.88–0.9] using the logistic approach (Table S8, Figs S10 and S11). The average type I and II error rates were estimated at 6.6% and 2.8%, respectively.

MtCR DNA sequence haplotype network. Only four mtCR DNA sequence haplotypes were detected in the Gulf of California sample, of which three were also detected in the eastern North Pacific sample (Fig. 2). One of the mtCR DNA sequence haplotypes that was detected both in the eastern North Pacific and the Gulf of California samples was identified in a single Gulf of California individual, an individual inferred to be a first-generation immigrant (haplotype marked with a square in Fig. 2). A single mtCR DNA sequence haplotype accounted for ~92% (238 of 259) of the Gulf of California individuals.

Discussion

The analyses in this study revealed that the Gulf of California fin whale population was founded approximately 2.3 kya and subject to very limited gene flow from the neighbouring populations in the eastern North Pacific. The contemporary N_e and N_e estimates for the Gulf of California population (~360 and 730, respectively) were low as was the IMA2P estimate of the “long-term” N_e (~40) suggesting that this fin whale population has remained small since it was founded, which in turn explains the very low levels of contemporary genetic diversity in this population. The contemporary, highly productive, oceanographic conditions in the Gulf of California were established ~2.8 kya^{54,55}, which correspond remarkably well with the IMA2P estimate of the divergence time between the Gulf of California and eastern North Pacific fin whales. The primary productivity in the Upper Gulf of California can reach very high levels (364 mgC m⁻³ D⁻¹)⁵⁶ seemingly resulting in conditions that can sustain a resident population.

Our results supported the notion that the establishment of the Gulf of California fin whale population was not a recent phenomenon. The DIYABC-based assessment yielded strong support for a divergence time between the Gulf of California population and the eastern North Pacific more than 40 generations ago (>1 kya). The IMA2P divergence time estimate agreed with the DIYABC-based assessment, suggesting a reduction from an ancestral N_e at ~7,000 to ~40 approximately 2.3 kya. The MSVAR assessment also suggested a similar reduction, although the timing of the change in N_e inferred from the MSVAR assessment was much older than the IMA2P estimate, but included in the 95% HPD of the IMA2P estimate. The IMA2P estimates accounted for migration, which was not the case for the estimates obtained using DIYABC or MSVAR. The MSVAR estimate could also reflect another, older change in N_e . Despite the quantitative differences, the qualitative conclusion among these slightly different analytical approaches was similar; i.e., the Gulf of California fin whale population was founded more than one kya, and most likely around 2.3 kya.

The estimates of migration rates between the Gulf of California and the eastern North Pacific were well below the values viewed as evidence for evolutionary independence^{57–59}. The contemporary and “long-term” $N_e m$ were estimated at ~1.4 and 0.9 respectively. All estimates of migration rates (and effective population sizes) are sensitive to ghost populations (current unsampled, and even extinct populations)^{59,60}. However, the distinct and very low levels of genetic diversity in the Gulf of California fin whales compared to the adjacent eastern North Pacific fin whale populations (see Archer *et al.*⁶¹) suggest that immigration from these adjacent populations has been minimal, and hence likely not of major concern in this case. Moreover, the Gulf of California is a semi-enclosed sea, hence immigration from other populations besides the eastern North Pacific was unlikely.

The estimated contemporary immigration rates were roughly consistent with the identification of only two individuals with an immigrant ancestry among the 259 individuals sampled in the Gulf of California during the 12-year sampling period. One of the two individuals with an immigrant ancestry, was a male inferred as a first-generation immigrant, whereas the second “immigrant” individual was inferred as a F1, i.e., the offspring of an immigrant individual. The latter individual was previously reported by Bérubé and colleagues⁴⁰. Overall the results obtained during this study were in agreement with the inferences drawn in previous genetic^{40,41} and acoustic assessments^{38,39,60}, as well as the restricted range observed in satellite-tagged fin whales in the Gulf of California^{34,37} and the scarcity of fin whales sightings at the southern tip (south of 24°N) of the Baja California Peninsula^{31,33–35}.

The degree of population genetic differentiation between the Gulf of California and the eastern North Pacific was relatively high ($F_{ST} \sim 0.2$ for microsatellite genotypes and $\Phi_{ST} \sim 0.2$ for mtCR DNA sequence data), compared to levels observed among other con-specific populations in other baleen whale species in the same oceanic basin⁴⁰. Fin whales in the Mediterranean Sea also appeared to be (at least in part) a resident population⁴¹. The degree of differentiation using mtCR DNA sequences between the Mediterranean Sea fin whales and their eastern North Atlantic con-specifics was slightly lower (ranging from 0.09 to 0.15)⁴¹ compared to the Gulf of California and eastern North Pacific in this study. $N_e m$ between the North Atlantic and the Mediterranean Sea fin whales was estimated at two (female) migrants per generation⁶¹. However, the degree of genetic differentiation between the North Atlantic and the Mediterranean fin whale populations at nuclear microsatellite genotypes was much lower ($F_{ST} \sim 0.007$), perhaps due to occasional male-mediated gene flow^{61,62}.

The results presented here suggested that the low genetic diversity in the Gulf of California fin whales was a product of a low, long-term N_e and a low $N_e m$. The low genetic diversity among the Gulf of California fin whales has been noted earlier^{40,41,43}. Other studies have reported low estimates of N_e in specific baleen whale populations, such as eastern Canada and Sea of Okhotsk bowheads⁶³, as well as among right whales in the South Atlantic and Indo-Pacific ocean basins⁶⁴. However, the levels of genetic diversity reported for these populations were much higher ($H_o = 0.71–0.83$) than the genetic diversity observed among Gulf of California fin whales. An illustration of the very low genetic diversity among the Gulf of California fin whales was the observation that a single mtCR DNA sequence haplotype accounted for ~92% of all the sampled individuals. The mtCR DNA sequence haplotype (h) and nucleotide diversity (π) was estimated at 0.14 and 0.0006, respectively, which were among the lowest values reported for cetaceans^{40,41,65–69}. For instance, reported estimates of mtCR DNA sequence diversity in the highly endangered western North Atlantic right whale (*Eubalaena glacialis*) were substantially higher ($h = 0.69$ and $\pi = 0.006$)⁶⁶.

The very low level of genetic diversity among the Gulf of California fin whales is of potential concern in terms of loss of adaptive traits and a possible increase in genetic load^{5,46,70–72}. A few recent genomic-based studies have detected an increase in genetic load at a low N_e . Pedersen *et al.*⁷³ identified elevated levels of detrimental alleles among Greenlandic Inuit (*Homo sapiens*), which underwent a bottleneck for 15–20 kya at a N_e at ~1,000. Rogers and Slatkin⁷⁴ identified an apparent increase in the number of detrimental mutations in a small isolated population of woolly mammoth (*Mammuthus primigenius*) on Wrangel Island, which underwent a bottleneck of a similar extent and duration to that inferred for the Greenlandic Inuit. Both examples confirmed the expected increase in detrimental traits following a protracted population bottleneck and hence a reduction in population

fitness. In both examples, the negative effects were observed in populations with a N_e larger than the estimated in the Gulf of California fin whale population. However, the duration of the bottleneck in the two mentioned examples was much longer.

On the other hand, there have also been examples of mammal populations that persisted for thousands of years despite low levels of genetic diversity, such as the endangered Iberian lynx (*Lynx pardinus*), which appeared to have persisted at a N_e at 200–300 for 8.1–50 kya⁷⁵. Among marine mammals very low levels of mtCR DNA sequence diversity has been reported in the narwhal (*Monodon monoceros*)⁶⁹ with no apparent adverse effects and a large contemporary population size (~24,000 individuals⁷⁶). In contrast to the narwhal, N_e of the Gulf of California fin whale population was estimated at only 730 individuals. The genetic tagging⁸ estimate of N_e for the Gulf of California fin whale population was concordant with previous N_e estimates based upon line transect surveys (ranging from 297 to 820)³⁴ and a single mark-recapture estimate (600) based upon individual identification from natural markings⁷⁷.

The results presented here implies that Gulf of California fin whales could be subject to elevated rates of genetic load and loss of adaptive variation, although the population appeared to have persisted at a low N_e for ~2.3 kya. Assessing the contribution of low genetic variation upon the population extinction rate in a species, such as fin whales is non-trivial. Unlike most other baleen whale populations, the Gulf of California fin whales were not subjected to commercial exploitation. However, this small population is currently exposed to a multitude of recent anthropogenic pressures, including bycatch (Urban and Brownell, unpublished data), entanglement in fishing gear, ship strikes, possible disturbance by whale-watching boats, anthropogenic sound^{78,79} and habitat degradation due to chemical pollution, microplastics⁸⁰ and urban development³⁵.

Such anthropogenic effects on mortality and reproductive rates may add to the genetic effects stemming from a low N_e through several thousand years. The low N_e and N_e of the Gulf of California fin whale population, coupled with the very low connectivity with the eastern North Pacific fin whales documented here, indicates that a cautionary management scheme should be considered in order to avoid this population joining the other small baleen whale populations currently declining from anthropogenic threats^{81–83}, especially bycatch⁸⁴, entanglements and ship strikes.

Methods

Sample collection. Tissue samples were collected as skin biopsies⁸⁵ and sloughed skin⁸⁶ from free-ranging fin whales, or during necropsies of beached fin whales. Tissue samples were collected by Universidad Autónoma de Baja California Sur with the approval and permits issued by the Mexican Wildlife Agency (Dirección General de Vida Silvestre, Subsecretaría de Gestión para la Protección Ambiental, Secretaría del Medio Ambiente y Recursos Naturales). Samples from US waters were collected by the Southwest Fisheries Science Center and by the National Marine Mammal Lab in accordance with national guidelines and regulations. The research permits also included the necessary ethical approval in terms of sample collection. All sample collection was undertaken in accordance with relevant national guidelines and regulations. Tissue samples were preserved in a saturated sodium chloride solution with 20% dimethyl sulfoxide⁸⁷ and stored at –20 degrees Celsius (°C).

Laboratory analyses. Genomic DNA was extracted by either standard phenol-chloroform extractions⁸⁸ or using QIAGEN DNEasy™ extraction columns for animal tissue, following the manufacturer's instruction (QIAGEN Inc.). The quality of the extracted DNA was visually checked by electrophoresis through a 0.7% agarose gel, and quantified with a Qubit™ following the manufacturer's instructions (Thermo Fisher Scientific Inc.). DNA extractions were normalized to a concentration of 20 ng DNA/μL. Sex was determined by the differential amplification of the zinc finger coding regions, ZFX and ZFY⁸⁹ as well as by co-amplification of a SRY-specific region with one or more autosomal microsatellite loci (Bérubé, *in prep*).

The genotypes were determined at 18 microsatellite loci: GATA028⁹⁰, GATA053⁹⁰, GATA098⁹⁰, GATA417⁹⁰, GGAA520⁹⁰, TAA023⁹⁰, GT011⁴¹, EV001⁹¹, EV037⁹¹, EV094⁹¹, GT023⁹², GT575⁹², GT195⁹², GT211⁹², GT271⁹², GATA25072, GATA43950 and GATA6063862 (last three loci, Bérubé, *in prep*). Negative and positive controls were included with each set of polymerase chain reaction⁹³ (PCR) amplifications. Individual PCR amplifications were carried out as described by Palsbøll, *et al.*⁹⁰ and Bérubé, *et al.*^{41,92}. PCR fragments were separated by electrophoresis on an ABI 3730 DNA Analyzer™ (Applied Biosystems Inc.) with GeneScan™-500 ROX (Applied Biosystems Inc.) as size standard. The length of each PCR product was determined with GENEMAPPER™ (ver. 4.1, Applied Biosystems Inc.). The consistency rate per multi-locus genotype was estimated by re-amplifying ten loci (GATA028, GATA053, GATA098, GATA417, GGAA520, TAA023, GT011, GT575, GT195, and GT211) in 47 randomly selected DNA extractions.

The first 285 base pairs of the 5'-end of the mtCR were amplified using the oligo-nucleotides MT4F⁹⁴ and BP16071R⁹⁵. The PCR amplifications were performed in 20 μL reaction volumes with 0.2 μM of each dNTP, 67 mM Tris-HCl (pH 8.8), 2 mM MgCl₂, 17 mM NH₃SO₄, 10 mM β-mercaptoethanol, 0.1 μM of each oligo-nucleotide, 0.4 units of *Taq* DNA polymerase (New England BioLabs® Inc) and 10–20 ng of DNA extraction. The thermo-cycling profile was 2 min at 94 °C, followed by 25 cycles each with 15 seconds at 94 °C, 30 seconds at 54 °C and 120 seconds at 72 °C. Unincorporated nucleotides and excess oligo-nucleotides were enzymatically removed using *shrimp alkaline phosphatase* and *exonuclease I* as described by Werle, *et al.*⁹⁶. Cycle-sequencing was conducted according to the manufacturer's instructions using Big Dye™ (ver. 3.1, Life Technologies Inc.) with the oligo-nucleotides employed in the initial PCR amplification. The cycle-sequencing products were precipitated with 96% ethanol⁸⁸ and re-suspended in 10 μL deionized formamide (Calbiochem® Inc.). The order of cycle-sequencing products was resolved by electrophoresis on an ABI 3730 DNA Analyzer™ (Applied Biosystems Inc.). The reverse strand was sequenced in DNA extractions with unique mtCR DNA sequence haplotypes. Additional mtCR DNA sequences from Bérubé, *et al.*⁴¹ and Archer, *et al.*⁹⁷ were retrieved from NCBI (Table S1).

Data analysis. *Probability of identity and sex ratios.* PI^{47} was estimated using CERVUS (ver. 3.0.7)⁹⁸ assuming that all individuals were unrelated, in order to determine the probability that two unrelated individuals had identical multi-locus genotypes.

The expected number of pairs of unrelated individuals with identical genotypes by chance in the sample was estimated by multiplying the number of pairwise comparisons by PI . The PI was also used to estimate the number of loci needed to identify an individual⁹⁹, and this threshold was used to check the samples with few mismatches. The probability of an observed deviation from parity of the sex ratio was estimated as the fraction of 2,000 random permutations of the data in which χ^2 was equal to or larger than the observed value of χ^2 by means of a χ^2 test in R (ver. 3.0.1)¹⁰⁰.

Deviations from Hardy-Weinberg genotype proportions and linkage equilibrium. Departures from Hardy-Weinberg equilibrium and linkage disequilibrium were estimated using ARLEQUIN (ver. 3.5.2.2)¹⁰¹. Departures from Hardy-Weinberg equilibrium were assessed for each microsatellite locus in each sample from 10,000 dememorization MCMC steps followed by 100,000 steps. The probability of linkage disequilibrium between all pairs of loci per population were estimated from 10,000 permutations. The significance of the departures from Hardy-Weinberg equilibrium and linkage equilibrium was corrected using FDR⁴⁸ at 0.05. The possible presence of null alleles, large allele dropout and scoring errors due to stuttering was assessed using MICRO-CHECKER (ver. 2.2.3)¹⁰².

Patterns of population genetic diversity and differentiation. The number of alleles (N_a) as well as the observed (H_o) and expected heterozygosity (H_e) were estimated using ARLEQUIN. Allelic richness (A_R) was estimated with the package ADZE (ver.1.0)¹⁰³ for all 18 microsatellite loci in sample sizes of 18 (the smallest sample size). The inbreeding coefficient (F_{IS})¹⁰⁴ was estimated as implemented in FSTAT (ver. 2.9.3.2)¹⁰⁵. The mtCR DNA sequences were aligned using the Clustal W algorithm¹⁰⁶ implemented in MEGA (ver. 6.06)¹⁰⁷ with the default settings. DNASP (ver. 5.10.1)¹⁰⁸ was employed to estimate the number of segregating sites (S) and haplotypes (h_n) as well as π , h and θ (per site) from S ¹⁰⁹.

The degree of population genetic differentiation was estimated as ϕ_{ST} from the mtCR DNA sequences using ARLEQUIN¹⁰¹ from the number of pairwise differences. In case of the microsatellite genotypes, the population genetic differentiation was estimated as F_{ST} using FSTAT (ver.2.9.3.2)¹⁰⁵. The 95% CI was estimated from 15,000 permuted datasets by bootstrapping loci.

Contemporary estimates of $N_e m$, N_e and N_c . “Contemporary” immigration rates were estimated from the microsatellite genotypes employing the Bayesian approach implemented in the software BAYESASS (ver. 3.0.4)¹¹⁰. Preliminary estimations were performed to adjust the MCMC mixing parameters of migration rates, allele frequencies, and inbreeding coefficients to acceptance rates at 20–60%. A total of 11 independent MCMCs were performed, each with different random starting seeds and a burn-in of 1,000,000 steps followed by 10,000,000 steps sampling every 1,000th step. The final parameter estimates were inferred from the estimation with the lowest Bayesian deviance using the R script authored by Meirmans¹¹¹.

Contemporary N_e was estimated from the degree of linkage disequilibrium among the microsatellite loci using the bias-corrected approach by Waples and Do¹¹² as implemented in NEESTIMATOR (ver. 2.1)¹¹³. The estimation was conducted excluding the first-generation immigrant detected during the BAYESASS assessment because the linkage disequilibrium approach in NEESTIMATOR assumes that all genotypes originate from a single panmictic population. The second-generation migrant was included in the dataset because the individual was part of the Gulf of California population. The critical minimum allele frequency was set to 0.02. The 95% CI of each point estimate was estimated by jack-knifing among samples¹¹⁴.

The census population size (N_c) was estimated using the POPAN derivation of the Jolly-Seber model as implemented in MARK (ver. 6.1)¹¹⁵ via the *RMark* package¹¹⁶. Each year was employed as a single capture event (Table S4). The Akaike’s information criterion with a correction for small sample size was employed to assess the model support. Model averaging was performed in MARK¹¹⁵.

Population sizes and divergence time. A possible change in effective population size in the Gulf of California sample was estimated using MSVAR (ver. 1.3)⁴⁹. Loci with “micro-variant” alleles¹¹⁷ causing imperfect repeats were excluded: GATA053, GATA417, GGAA520 and GATA25072). Two duplicate runs were conducted, each with different random seeds using the exponential population size change model. Starting values of each duplicate run were initiated from the upper and lower range of the prior distribution of each parameter (current population size, $N_0 = 300$, ancestral population size, $N_1 = 10,000$, mutation rate, $\mu = 0.0005$ and time of population size change, $T_\alpha = 16,000$). Each MCMC comprised of 10,000,000 steps with sampling at each 1,000th step. The first 5,000 of the 10,000 recorded steps were discarded before assessing convergence and parameter estimation.

The population divergence time, long-term N_e , and immigration rates were estimated using the parallelised version of the isolation-with-migration model⁵⁰ implemented in IMA2P⁵¹. Two different datasets were analysed; a dataset comprised solely of mtCR DNA sequences and another dataset including both mtCR DNA sequences and microsatellite genotypes. Microsatellite loci with “micro-variant” alleles were excluded from this estimation as well. The Hasegawa-Kishino-Yano, HKY, substitution model¹¹⁸ was employed with a mutation rate at 5.2×10^{-8} per site per year¹¹⁹ for the mtCR DNA sequences. The mutation rate per locus was based upon 280 base pairs. A stepwise mutation model was employed for the microsatellite genotypes using a mutation rate at 7×10^{-4} per generation (estimated in the MSVAR analysis). A generation time at 25.9 years¹²⁰ was used for converting the generational mutation rate into an annual mutation rate. Preliminary estimations were conducted to identify the optimal parameter prior ranges. The final estimates for the mtCR DNA sequences were based upon three replicate estimations:

each with ten MCMCs consisting of 400,000 steps, of which the first 200,000 were discarded, and sampling each 100th step. In addition, one run was started using the Markov chain state generated in a previous run.

For the dataset including both mtCR DNA sequences and microsatellite genotypes a random sample of 30 individuals from the Gulf of California was employed in the analysis, in order to avoid effects stemming from unequal sample sizes between the eastern North Pacific and the Gulf of California. The estimations for the combined data were based upon four replicates: each with ten MCMCs consisting of 250,000 steps, of which the first 150,000 were discarded, and a sampling interval at every 100th step. In addition, two runs were started using the Markov chain state generated in previous runs. Reported estimates were those obtained from the run with the highest ESS.

Because the 95% HPD interval of the IMA2P estimate of the population divergence time included zero, DIYABC (ver. 2.1)⁵³ was used to test between a recent (scenario 1: 0–20 generations ago) versus an older (scenario 2: 40–900 generations ago) population divergence time. The point estimate of the population divergence time given by IMA2P analysis using only mtCR DNA sequences was used as the starting point for the divergence time distribution of scenario 2 (40 generations ago, which is ~1,000 years ago). The divergence time distribution of scenario 1 starts at zero and ends at 20 generations in order to not overlap with scenario 2. DIYABC assessments were based upon two million simulated datasets for each scenario using mtCR DNA sequences and microsatellite genotypes. Uniform prior distributions were used for the demographic parameters: North Pacific effective population size, N_1 [10,000–80,000]; Gulf of California effective population size, N_2 [1–1,000], scenario 1 divergence time, T_1 [0–20] and scenario 2 divergence time, T_2 [40–900]. A HKY mutation model was used for the mtCR DNA sequences with a uniform distribution from 1×10^{-7} to 1×10^{-5} and 96% of invariable sites. Default values were used for the other mutation model parameters. The posterior probability of each scenario was estimated from a logistic regression performed on one percent of the simulated datasets closest to the observed values after a linear discriminant analysis¹²¹. The confidence in the scenario choice was evaluated by estimating the type-I and type-II error rates for 1,000 simulated datasets in a linear discriminant analysis. The model checking (Table S8) was performed as implemented in DIYABC.

MCMC-based estimations were deemed to have converged when estimates were consistent among independent replicate MCMC runs (minimum three, except MSVAR) and when effective sample sizes (ESS) were above 200. The ESS, trace plots, and Gelman-Rubin diagnostic were estimated for MSVAR and BAYESASS using the R package *coda*¹²².

MtCR DNA sequence haplotype network. A median-joining mtCR DNA sequence haplotype network¹²³ was inferred using POPART (ver. 1.7)¹²⁴ with ϵ set at zero.

Data Availability

The datasets generated during the current study are available at Datadryad.org under accession: <https://doi.org/10.5061/dryad.s07g211>.

References

- Allendorf, F. W. Genetics and the conservation of natural populations: allozymes to genomes. *Mol. Ecol.* **26**, 420–430, <https://doi.org/10.1111/mec.13948> (2016).
- Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
- Franklin, I. R. & Frankham, R. How large must populations be to retain evolutionary potential? *Animal Conservation* **1**, 69–70, <https://doi.org/10.1111/j.1469-1795.1998.tb00228.x> (1998).
- Kimura, M., Maruyama, T. & Crow, J. F. The Mutation Load in Small Populations. *Genetics* **48**, 1303–1312 (1963).
- Lacy, R. C. Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy* **78**, 320–335, <https://doi.org/10.2307/1382885> (1997).
- O'Brien, S. J. *et al.* Genetic Basis for Species Vulnerability in the Cheetah. *Science* **227**, 1428–1434, <https://doi.org/10.1126/science.2983425> (1985).
- Huijser, L. A. E. *et al.* Population structure of North Atlantic and North Pacific sei whales (*Balaenoptera borealis*) inferred from mitochondrial control region DNA sequences and microsatellite genotypes. *Conservation Genetics* **19**, 1007–1024 (2018).
- Palsbøll, P. J. *et al.* Genetic tagging of humpback whales. *Nature* **388**, 767–769, <https://doi.org/10.1038/42005> (1997).
- Rooney, A. P., Honeycutt, R. L. & Derr, J. N. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. *Evolution* **55**, 1678–1685 (2001).
- Alter, S. E., Rynes, E. & Palumbi, S. R. DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proceedings of the National Academy of Sciences* **104**, 15162–15167, <https://doi.org/10.1073/pnas.0706056104> (2007).
- Mackintosh, N. A. The Natural History of Whalebone Whales. *Biological Reviews of the Cambridge Philosophical Society* **21**, 60–74, <https://doi.org/10.1111/j.1469-185X.1946.tb00453.x> (1946).
- Cerchio, S. & Yamada, T. K. In *Encyclopedia of Marine Mammals (Third Edition)* (eds Bernd Würsig, J. G. M. Thewissen & Kit M. Kovacs) 656–659 (Academic Press, 2018).
- Kato, H. & Perrin, W. F. In *Encyclopedia of Marine Mammals (Third Edition)* (eds Bernd Würsig, J. G. M. Thewissen, & Kit M. Kovacs) 143–145 (Academic Press, 2018).
- Rugh, D. *et al.* A review of bowhead whale (*Balaena mysticetus*) stock identity. *Journal of Cetacean Research and Management* **5**, 267–280 (2003).
- Branch, T. A. *et al.* Past and present distribution, densities and movements of blue whales *Balaenoptera musculus* in the Southern Hemisphere and northern Indian Ocean. *Mammal Review* **37**, 116–175, <https://doi.org/10.1111/j.1365-2907.2007.00106.x> (2007).
- Yochem, P. K. & Leatherwood, S. In *Handbook of Marine Mammals* Vol. 3 (eds Sam. H. Ridgway & Richard J. Harrison) 362 (Academic Press Limited, 1985).
- Pomilla, C. *et al.* The World's Most Isolated and Distinct Whale Population? Humpback Whales of the Arabian Sea. *Plos One* **9**, e114162, <https://doi.org/10.1371/journal.pone.0114162> (2014).
- Whitehead, H. Humpback whale songs from the North Indian Ocean. *Invest. Cetacea*, 157–162 (1985).
- Reeves, R. R., Leatherwood, S. & Papastavrou, V. Possible stock affinities of humpback whales in the northern Indian Ocean. *UNEP Marine Mammal Technical Reports* **3**, 259–269 (1991).
- Mikhalev, Y. A. Humpback whales *Megaptera novaeangliae* in the Arabian Sea. *Marine Ecology Progress Series* **149**, 13–21 (1997).

21. Tershy, B. R., Breese, D. & Strong, C. S. Abundance, seasonal distribution and population composition of balaenopterid whales in the Canal de Ballenas, Gulf of California, Mexico. *Reports of the International Whaling Commission (Special Issue)* 369–375 (1990).
22. Tershy, B. R. Body size, diet, habitat use, and social behavior of *Balaenoptera* whales in the Gulf of California. *Journal of Mammalogy*, 477–486 (1992).
23. Rojas-Bracho, L. *Presencia y distribución del rorcual común, Balaenoptera physalus (Linnaeus, 1758; Cetacea: Balaenopteridae) en el Golfo de California, México* Licenciatura thesis, Universidad Nacional Autónoma de México, (1984).
24. Clavijero, F. X. *Historia de la antigua o Baja California*. First edition edn, (Editorial Porrúa, 1789).
25. Steinbeck, J. & Ricketts, E. F. *Sea of Cortez: A leisurely journal of travel and research*. (Viking Press 1941).
26. Gilmore, R. M. Whales Aground in Cortés' Sea: Tragic Strandings in the Gulf of California. *Pacific Discovery* **10**, 22–27 (1957).
27. Norris, K. *Some observations on the migration and orientation of marine mammals*. (Oregon State University Press, 1966).
28. Tershy, B. R., Breese, D. & Alvarez-Borrogo, S. Increase in cetacean and seabird numbers in the Canal de Ballenas during an El Niño-Southern Oscillation event. *Marine Ecology Progress Series* **69**, 299–302 (1991).
29. Breese, D. & Tershy, B. R. Relative abundance of cetacea in the Canal de Ballenas, Gulf of California. *Marine Mammal Science* **9**, 319–324 (1993).
30. Vidal, O., Findley, L. T. & Leatherwood, S. Annotated checklist of the marine mammals of the Gulf of California. **28**, 1–16 (1993).
31. Tershy, B. R., Urbán, J., Bréese, D., Rojas-B. L. & Findley, L. T. Are fin whales resident to the Gulf of California. *Rev. Invest. Cient., Univ. Auton. de Baja California Sur* **1**, 69–71 (1993).
32. Širović, A., Williams, L. N., Kerosky, S. M., Wiggins, S. M. & Hildebrand, J. A. Temporal separation of two fin whale call types across the eastern North Pacific. *Marine Biology* **160**, 47–57, <https://doi.org/10.1007/s00227-012-2061-z> (2013).
33. Hamilton, T. A. *et al.* Atlas of Cetacean Sightings for Southwest Fisheries Science Center Cetacean and Ecosystem Surveys, 1986–2005 (2009).
34. Urbán, J., Rojas-Bracho, L., Guerrero-Ruiz, M., Jaramillo-Legorreta, A. & Findley, L. Cetacean diversity and conservation in the Gulf of California. In *Biodiversity, ecosystems, and conservation in Northern Mexico* (eds Jean-Luc E. Cartron, Gerardo Ceballos & Richard Stephen Felger), 276–297 (Oxford University Press, 2005).
35. Urbán, J., Cárdenas Hinojosa, G. & Gómez-Gallardo Unzueta, A. Cetaceans of the Gulf of California's Southwest Coast. In *Los Cabos: Prospective for a Natural and Tourism Paradise* (eds Ganster, P., Arizpe, C. & Ivanova, A.) Ch. 6, 101 (San Diego State University Press, 2012).
36. Mizroch, S. A., Balcomb, K. C. & Rice, D. W. Historical winter and summer distribution of large whales along the eastern North Pacific coast based on data from the US whale marking program, 1962–1969. *Report to the Scientific Committee of the International Whaling Commission*, 1–28 (2018).
37. Jiménez López, M. E., Palacios, D. M., Jaramillo Legorreta, A., Urbán, R. J. & Mate, B. R. Fin whale movements in the Gulf of California, Mexico, from satellite telemetry. *Plos One* **14**, e0209324, <https://doi.org/10.1371/journal.pone.0209324> (2019).
38. Thompson, P. O., Findley, L. T. & Vidal, O. 20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico. *The Journal of the Acoustical Society of America* **92**, 3051–3057 (1992).
39. Širović, A., Oleson, E. M., Buccowich, J., Rice, A. & Bayless, A. R. Fin whale song variability in southern California and the Gulf of California. *Scientific reports* **7**, 10126, <https://doi.org/10.1038/s41598-017-09979-4> (2017).
40. Bérubé, M., Urban, J., Dizon, A. E., Brownell, R. L. Jr. & Palsbøll, P. J. Genetic identification of a small and highly isolated population of fin whales (*Balaenoptera physalus*) in the Sea of Cortez, Mexico. *Conservation Genetics* **3**, 183–190, <https://doi.org/10.1023/a:1015224730394> (2002).
41. Bérubé, M. *et al.* Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus 1758): analysis of mitochondrial and nuclear loci. *Mol. Ecol.* **7**, 585–599 (1998).
42. Weir, B. S. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. (Sinauer Associates, Inc., 1996).
43. Nigenda-Morales, S., Flores-Ramirez, S., Urban-R, J. & Vazquez-Juarez, R. MHC DQB-1 polymorphism in the Gulf of California fin whale (*Balaenoptera physalus*) population. *Journal of Heredity* **99**, 14–21, <https://doi.org/10.1093/hered/esm087> (2008).
44. Weber, D. S., Stewart, B. S., Schienman, J. & Lehman, N. Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Mol. Ecol.* **13**, 711–718, <https://doi.org/10.1046/j.1365-294X.2004.02095.x> (2004).
45. Crnokrak, P. & Roff, D. A. Inbreeding depression in the wild. *Heredity* **83**, 260–270, <https://doi.org/10.1038/sj.hdy.6885530> (1999).
46. Frankham, R. *et al.* *Genetic Management of Fragmented Animal and Plant Populations*. (Oxford University Press, 2017).
47. Paetkau, D. & Strobeck, C. Microsatellite analysis of genetic variation in black bear populations. *Mol. Ecol.* **3**, 489–495, <https://doi.org/10.1111/j.1365-294X.1994.tb00127.x> (1994).
48. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
49. Beaumont, M. A. Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013–2029 (1999).
50. Hey, J. & Nielsen, R. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747–760, <https://doi.org/10.1534/genetics.103.024182> (2004).
51. Sethuraman, A. & Hey, J. IMA2p-parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. *Molecular ecology resources* **16**, 206–215, <https://doi.org/10.1111/1755-0998.12437> (2016).
52. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410–418, <https://doi.org/10.1016/j.tree.2010.04.001> (2010).
53. Cornuet, J.-M. *et al.* DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187–1189, <https://doi.org/10.1093/bioinformatics/btt763> (2014).
54. Barron, J. A., Bukry, D. & Dean, W. E. Paleooceanographic history of the Guaymas Basin, Gulf of California, during the past 15,000 years based on diatoms, silicoflagellates, and biogenic sediments. *Marine Micropaleontology* **56**, 81–102, <https://doi.org/10.1016/j.marmicro.2005.04.001> (2005).
55. Barron, J. A., Bukry, D. & Bischoff, J. L. High resolution paleoceanography of the Guaymas Basin, Gulf of California, during the past 15 000 years. *Marine Micropaleontology* **50**, 185–207, [https://doi.org/10.1016/S0377-8398\(03\)00071-9](https://doi.org/10.1016/S0377-8398(03)00071-9) (2004).
56. Mercado-Santana, J. A. *et al.* Productivity in the Gulf of California large marine ecosystem. *Environmental Development*, <https://doi.org/10.1016/j.envdev.2017.01.003> (2017).
57. Mills, L. S. & Allendorf, F. W. The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology* **10**, 1509–1518, <https://doi.org/10.1046/j.1523-1739.1996.10061509.x> (1996).
58. Wang, J. Application of the One-Migrant-per-Generation Rule to Conservation and Management. *Conservation Biology* **18**, 332–343, <https://doi.org/10.1111/j.1523-1739.2004.00440.x> (2004).
59. Waples, R. S. & Gaggiotti, O. E. INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* **15**, 1419–1439, <https://doi.org/10.1111/j.1365-294X.2006.02890.x> (2006).
60. Bonilla-Garzón, A., Romero Vivas, E. & Urbán, J. In *The Journal of the Acoustical Society of America* Vol. 138 1790–1790 (2015).
61. Palsbøll, P. J., Bérubé, M., Aguilar, A., Notarbartolo-Di-Sciara, G. & Nielsen, R. Discerning between Recurrent Gene Flow and Recent Divergence under a Finite-Site Mutation Model Applied to North Atlantic and Mediterranean Sea Fin Whale (*Balaenoptera physalus*) Populations. *Evolution* **58**, 670–675, <https://doi.org/10.1111/j.0014-3820.2004.tb01691.x> (2004).

62. Gauffier, P. *et al.* Contemporary migration of fin whales through the Strait of Gibraltar. *Marine Ecology Progress Series* **588**, 215–228 (2018).
63. Morin, P. A. *et al.* Empirical comparison of single nucleotide polymorphisms and microsatellites for population and demographic analyses of bowhead whales. *Endangered Species Research* **19**, 129–147 (2012).
64. Carroll, E. L. *et al.* Incorporating non-equilibrium dynamics into demographic history inferences of a migratory marine species. *Heredity* **122**, 53–68, <https://doi.org/10.1038/s41437-018-0077-y> (2019).
65. Lyrholm, T., Leimar, O. & Gyllenstein, U. Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: Implications for estimates of time since common ancestry. *Molecular Biology and Evolution* **13**, 1318–1326 (1996).
66. Rosenbaum, H. C. *et al.* World-wide genetic differentiation of *Eubalaena*: questioning the number of right whale species. *Mol. Ecol.* **9**, 1793–1802, <https://doi.org/10.1046/j.1365-294x.2000.01066.x> (2000).
67. Sremba, A. L., Hancock-Hanser, B., Branch, T. A., LeDuc, R. L. & Baker, C. S. Circumpolar Diversity and Geographic Differentiation of mtDNA in the Critically Endangered Antarctic Blue Whale (*Balaenoptera musculus intermedia*). *Plos One* **7**, <https://doi.org/10.1371/journal.pone.0032579> (2012).
68. Alter, S. E., Newsome, S. D. & Palumbi, S. R. Pre-Whaling Genetic Diversity and Population Ecology in Eastern Pacific Gray Whales: Insights from Ancient DNA and Stable Isotopes. *Plos One* **7**, <https://doi.org/10.1371/journal.pone.0035039> (2012).
69. Palsbøll, P. J., Heide-Jørgensen, M. P. & Dietz, R. Population structure and seasonal movements of narwhals, *Monodon monoceros*, determined from mtDNA analysis. *Heredity* **78**, 284–292 (1997).
70. Reed, D. H. & Bryant, E. H. Experimental tests of minimum viable population size. *Animal Conservation* **3**, 7–14, <https://doi.org/10.1017/s13679430000069x> (2000).
71. Frankham, R. Genetics and extinction. *Biological Conservation* **126**, 131–140, <https://doi.org/10.1016/j.biocon.2005.05.002> (2005).
72. Soulé, M. E. & Wilcox, B. A. *Conservation biology: an evolutionary-ecological perspective*. (Sinauer, 1980).
73. Pedersen, C.-E. T. *et al.* The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. *Genetics* **205**, 787–801, <https://doi.org/10.1534/genetics.116.193821> (2017).
74. Rogers, R. L. & Slatkin, M. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genetics* **13**, e1006601, <https://doi.org/10.1371/journal.pgen.1006601> (2017).
75. Rodríguez, R. *et al.* 50,000 years of genetic uniformity in the critically endangered Iberian lynx. *Mol. Ecol.* **20**, 3785–3795, <https://doi.org/10.1111/j.1365-294X.2011.05231.x> (2011).
76. Heide-Jørgensen, M. P. *et al.* Abundance of narwhals (*Monodon monoceros*) on the hunting grounds in Greenland. *Journal of Mammalogy* **91**, 1135–1151, <https://doi.org/10.1644/09-mamm-a-198.1> (2010).
77. Díaz-Guzmán, C. F. *Abundancia y movimientos del rorcual común, Balaenoptera physalus, en el Golfo de California*. Master Thesis, Universidad Nacional Autónoma de México (2006).
78. Caruso, F. *et al.* Detection and tracking of fin whales during seismic exploration in the Gulf of California. *Proceedings of Meetings on Acoustics* **27**, 070021, <https://doi.org/10.1121/2.0000424> (2016).
79. Croll, D. A. *et al.* Only male fin whales sing loud songs. *Nature* **417**, 809, <https://doi.org/10.1038/417809a> (2002).
80. Fossi, M. C. *et al.* Fin whales and microplastics: The Mediterranean Sea and the Sea of Cortez scenarios. *Environmental Pollution* **209**, 68–78, <https://doi.org/10.1016/j.envpol.2015.11.022> (2016).
81. Kraus, S. D. *et al.* North Atlantic Right Whales in Crisis. *Science* **309**, 561–562, <https://doi.org/10.1126/science.1111200> (2005).
82. LeDuc, R. G. *et al.* Genetic analysis of right whales in the eastern North Pacific confirms severe extirpation risk. *Endangered Species Research* **18**, 163–167 (2012).
83. Wade, P. R. *et al.* The world's smallest whale population? *Biology Letters* **7**, 83–85, <https://doi.org/10.1098/rsbl.2010.0477> (2011).
84. Brownell, R. L. Jr. & Mallette, S. D. Global baleen whale bycatch: the most threatened populations. 1–11 (2018).
85. Palsbøll, P. J., Larsen, F. & Sigurd Hansen, E. Sampling of Skin Biopsies from Free-Raging Large Cetaceans in West Greenland: Development of new Biopsy Tips and Bolt Designs. *International Whaling Commission Special Issue Series* (1991).
86. Clapham, P. J., Palsbøll, P. J. & Mattila, D. K. High-energy behaviors in humpback whales as a source of sloughed skin for molecular analysis. *Marine Mammal Science* **9**, 213–220, <https://doi.org/10.1111/j.1748-7692.1993.tb00448.x> (1993).
87. Amos, B. & Hoelzel, A. R. *Long-term preservation of whale skin for DNA analysis*. Vol. 13 99–104 (1991).
88. Sambrook, J. & Russell, D. Molecular cloning: a laboratory manual. *Quarterly Review of Biology* **76**, 348–349 (2001).
89. Bérubé, M. & Palsbøll, P. J. Identification of sex in Cetaceans by multiplexing with three ZFX and ZFY specific primers. *Mol. Ecol.* **5**, 283–287 (1996).
90. Palsbøll, P. J., Bérubé, M., Larsen, A. H. & Jørgensen, H. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Mol. Ecol.* **6**, 893–895 (1997).
91. Valsecchi, E. & Amos, W. Microsatellite markers for the study of cetacean populations. *Mol. Ecol.* **5**, 151–156, <https://doi.org/10.1111/j.1365-294X.1996.tb00301.x> (1996).
92. Bérubé, M., Jørgensen, H., McEwing, R. & Palsbøll, P. J. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Mol. Ecol.* **9**, 2181–2183 (2000).
93. Mullis, K. B. & Faloona, F. A. Specific synthesis of DNA invitro via a polymerase-catalyzed chain-reaction. *Methods in Enzymology* **155**, 335–350 (1987).
94. Arnason, U., Gullberg, A. & Widgren, B. Cetacean Mitochondrial DNA Control Region: Sequences of All Extant Baleen Whales and Two Sperm Whale Species. *Molecular Biology and Evolution* **10**, 960–970 (1993).
95. Drouot, V. *et al.* A note on genetic isolation of Mediterranean sperm whales (*Physeter macrocephalus*) suggested by mitochondrial DNA. *Journal of Cetacean Research and Management* **6**, 29–32 (2004).
96. Werle, E., Schneider, C., Renner, M., Volker, M. & Fiehn, W. Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research* **22**, 4354–4355, <https://doi.org/10.1093/nar/22.20.4354> (1994).
97. Archer, F. I. *et al.* Mitogenomic Phylogenetics of Fin Whales (*Balaenoptera physalus* spp.): Genetic Evidence for Revision of Subspecies. *Plos One* **8**, <https://doi.org/10.1371/journal.pone.0063396> (2013).
98. Wagner, A. P., Creel, S., Frank, L. G. & Kalinowski, S. T. Patterns of relatedness and parentage in an asocial, polyandrous striped hyena population. *Mol. Ecol.* **16**, 4356–4369, <https://doi.org/10.1111/j.1365-294X.2007.03470.x> (2007).
99. Rew, M. B., Robbins, J., Mattila, D., Palsbøll, P. J. & Bérubé, M. How many genetic markers to tag an individual? An empirical assessment of false matching rates among close relatives. *Ecological Applications* **21**, 877–887 (2011).
100. R Development Core Team. (R Foundation for Statistical Computing, 2013).
101. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**, 564–567, <https://doi.org/10.1111/j.1755-0998.2010.02847.x> (2010).
102. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–538, <https://doi.org/10.1111/j.1471-8286.2004.00684.x> (2004).
103. Szpiech, Z. A., Jakobsson, M. & Rosenberg, N. A. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–2504, <https://doi.org/10.1093/bioinformatics/btn478> (2008).
104. Weir, B. S., Cockerham, C. C. & Estimating *F*-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370, <https://doi.org/10.2307/2408641> (1984).
105. FSTAT, a program to estimate and test gene diversities and fixation indices v. 2.9.3.2 (2002).

106. Thompson, J. D., Higgins, D. G., Gibson, T. J. & CLUSTAL, W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680, <https://doi.org/10.1093/nar/22.22.4673> (1994).
107. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729, <https://doi.org/10.1093/molbev/mst197> (2013).
108. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452, <https://doi.org/10.1093/bioinformatics/btp187> (2009).
109. Tajima, F. The Amount of DNA Polymorphism Maintained in a Finite Population When the Neutral Mutation Rate Varies Among Sites. *Genetics* **143**, 1457–1465 (1996).
110. Wilson, G. A. & Rannala, B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177–1191 (2003).
111. Meirmans, P. G. Nonconvergence in Bayesian estimation of migration rates. *Molecular ecology resources* **14**, 726–733, <https://doi.org/10.1111/1755-0998.12216> (2014).
112. Waples, R. S. & Do, C. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular ecology resources* **8**, 753–756, <https://doi.org/10.1111/j.1755-0998.2007.02061.x> (2008).
113. Do, C. *et al.* NESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular ecology resources* **14**, 209–214, <https://doi.org/10.1111/1755-0998.12157> (2014).
114. Jones, A. T., Ovenden, J. R. & Wang, Y. G. Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. *Heredity* **117**, 217–223, <https://doi.org/10.1038/hdy.2016.19> (2016).
115. White, G. C. & Burnham, K. P. Program MARK: Survival estimation from populations of marked animals. *Bird Study* **46**, S120–S139 (1999).
116. Laake, J. & Rexstad, E. In *Program MARK: A Gentle Introduction* (eds Cooch, E. & White, G. C.) C1–C115 (2014).
117. Palsbøll, P. J., Bérubé, M. & Jorgensen, H. Multiple levels of single-strand slippage at cetacean tri- and tetranucleotide repeat microsatellite loci. *Genetics* **151**, 285–296 (1999).
118. Hasegawa, M., Kishino, H. & Yano, T. Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA. *Japanese Journal of Genetics* **59**, 616–616 (1985).
119. Alter, S. E. & Palumbi, S. R. Comparing Evolutionary Patterns and Variability in the Mitochondrial Control Region and Cytochrome b in Three Species of Baleen Whales. *Journal of Molecular Evolution* **68**, 97–111, <https://doi.org/10.1007/s00239-008-9193-2> (2009).
120. Taylor, B. L., Chivers, S. J., Larese, J. & Perrin, W. F. Generation length and percent mature estimates for IUCN assessments of cetaceans. *NOAA, NMFS, Southwest Fisheries Science Center Administrative Report LJ-07-01* **21** (2007).
121. Estoup, A. *et al.* Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular ecology resources* **12**, 846–855, <https://doi.org/10.1111/j.1755-0998.2012.03153.x> (2012).
122. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).
123. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48, <https://doi.org/10.1093/oxfordjournals.molbev.a026036> (1999).
124. Leigh, J. W. & Bryant, D. POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110–1116, <https://doi.org/10.1111/2041-210x.12410> (2015).

Acknowledgements

Financial support by the University of Groningen (PJP) and “Consejo Nacional de Ciencia y Tecnología (CONACYT, VRL)”. Tissue samples from the Gulf of California were collected under the permit of “Dirección General de Vida Silvestre, Secretaría del Medio Ambiente y Recursos Naturales”. Andrea Cabrera and Anna Schleimer are thanked for their advice in regards to data analyses. We kindly thank Lorena Viloria for making the sampling location map. The Center for Information Technology of the University of Groningen provided access and support to its Peregrine High-Performance Computing cluster. We would also like to thank Andrew Dizon, Adam Leiterman, Jooke Robbins, Robin Waples, Emma Carroll and two anonymous reviewers for their valuable comments on this manuscript.

Author Contributions

The research project was conceived and designed by M.B., P.J.P. and V.R. V.R. performed the data analyses and wrote the manuscript with M.B. and P.J.P. Tissue samples were provided by J.U., S.M. and R.B. V.R., M.B. and W.H. performed the laboratory work. T.O. conducted the MSVAR estimation. All authors reviewed and partook in manuscript revisions.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-48700-5>.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019