Supplementary information

mtDNA variation of humpback whales in their wintering grounds of Guerrero, southern Mexican Pacific

Luis Medrano-González, Katherina Audley, Charles Scott Baker, Debbie Steel

1. QUICK BRIEF ON THE PAPER'S POPULATION GENETICS

1.1. Mitochondrial DNA population structure and neutrality

For haploid and diploid genomes, haplotypes are closely linked genetic markers on a chromosome that tend to be inherited together. Mitochondrial (mt) DNA is a circular chromosome inherited through the ovum cells of animals (ca. 16 kilobases in cetaceans) whose variations in nucleotide sequence across geography define maternal lineages or haplotypes that provide information about the ancestry and historical flow of females and are thus relevant for the conservation of animal populations. Haplotypes are equivalent to alleles in diploid populations (e.g. Avise 1994). Nucleotide sequences of haplotypes allow to determine their phylogeny, i.e., their ancestry-descendance relationships. Clades, this is, sets of haplotypes in one phylogenetic branch (lineages with shared ancestry) are often referred to as haplogroups, such as the four human-Amerindian mtDNA groups A, B, C, and D (Torroni et al. 1993). Humpback whale mtDNA exhibits three major clades worldwide named AE, the more basal and apparently originated in the North Pacific Ocean, CD, and IJ. In the North Pacific, clade A is subdivided into haplogroups A and E, whilst clade CD is represented only by haplogroup F (Baker et al. 1993). Which clades or haplogroups are chosen for comparison, depends on the analytical goals.

Two main aspects of genetic variation in populations define their structure: diversity and differentiation among units that are first approached as spatial or temporal sets of samples. For population genetics, it is also relevant to characterize neutrality, this is, the comparison of observed and expected measures of polymorphism under models of neutral evolution and population equilibrium. From the number and frequency of alleles, the Ewens (1972)-Watterson (1978) method estimates the probability that the fixation (endogamy) index (F) expected under neutrality (F_{rand}), is larger than its observed value (F_{obs}). Similarly, at the nucleotide level, Tajima's D statistic (1989) compares the population size estimated by nucleotide diversity as $n\pi$, with the effective population size estimated by the number of variable sites as S/a_1, being $D = (n\pi - S/a_1) / (V(n\pi - S/a_1))^{1/2}$), where n is sample size, π is the mean heterozygosity per nucleotide, S is the number of variable sites, $a_1 = \sum_{i=1}^{n-1}(1/i)$, and V stands for the variance of $n\pi - S/a_1$.

Under neutrality, as is for animal mtDNA, positive *D* values indicate large nucleotide differences among alleles, suggesting an ancient or extended period of population growth, whilst negative *D* values stand for alleles exhibiting many variable sites but low nucleotide differences, a feature typical of population growth after a recent bottleneck. Statistical significance is usually set for $F_{rand} < F_{obs}$ and negative *D*-values with probabilities under 0.05, as well as for $F_{rand} \geq F_{obs}$ and *D*-values with probability over 0.95.

1.2. Estimation of allele richness

(translated and adapted excerpt from Medrano González 2023)

The apparently most simple way to measure diversity is the number of different elements, i.e., richness, such as alleles or haplotypes in a population, species in an ecosystem, or any set of discrete elements in a system. However, comparing richness becomes complicated when treating sets with different sample sizes. In most cases, samplings are not sufficient to register all richness and therefore, comparisons cannot be made directly from the richness observed. We present three approaches to compare richness in samples with different sizes.

1) Estimate how many alleles are expected in different samples adjusting their size to the smaller sample. This is the rarefaction method developed by Hurlbert (1971) to compare species richness in ecosystems that was adopted by El Mousadik & Petit (1996) to compare allele or haplotype richness. The sample size to compare different samples is a subsample of size *n* within a sample of size *N*, with $n \leq N$. For every recorded allele, *i*, the probability that such allele is found in all possible samples of size *n* is calculated first from how many combinations of *n* alleles subsampled from *N* do not contain *i* (q_i), this is $N - N_i$, to later estimate the complementary probability ($1 - q_i$) that *i* is contained in set *n*. The richness estimated for the subsample *n* (R_n) is therefore:

$$R_n = \sum_{i=1}^{k} \left(1 - \frac{\binom{N-Ni}{n}}{\binom{N}{n}} \right)$$

where N_i is the number of alleles *i* in the sample of size *N* ($\Sigma N_i = N$) and *k* is the number of different alleles in the sample of size *N* (observed richness). The number of combinations of *n* in *N* and of *n* in $N - N_i$ are respectively:

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

$$\binom{N - Ni}{n} = \frac{(N - Ni)!}{n!\,(N - Ni - n)!}$$

Every term in the summation of the richness formula (what is inside the big parenthesis) is the probability that a subsample of size n contains each of the alleles indicated in the sum. When $n > N$ or $n > N - N_i$, their respective combinations are zero, i.e., there is no combination with more than N elements contained in N or larger than $N - N_i$ elements contained in $N - N_i$. There is no solution to the richness formula when $n > N$, meaning that the rarefaction method cannot estimate richness for a sample size larger than the analyzed one. When $n > N - N_i$ and $n \leq N$, the numerator in the richness formula becomes zero making the probability of i being contained in the subsample n equal 1.

2) An alternative method is the extrapolation developed by Foulley & Ollivier (2006), which consists of adding to all alleles observed in a population, the expected probability of missing alleles that are observed in all the populations sampled. This method assumes that the alleles in all populations belong to an ancestral population or to a metapopulation and that every population contains all alleles with a specific probability depending on sample size. The probability of an allele i of being omitted in a population j with sample size n_j is $(1 - f_i)^{nj}$ where f_i is the allele frequency among all populations sampled. Extrapolated allele richness can then be calculated in a population j (R_j) by adding to the alleles registered (k_j), the probability of omission of the non-recorded alleles (O_j). R_j is then estimated as follows:

$$R_j = k_j + \sum_{i \in Oj} (1 - f_i)^{n_j}$$

Richness estimated by this method in a population thus varies between its observed richness and the richness observed in all populations. Given the sharing of most mtDNA haplotypes among humpback whales in their Northeast Pacific wintering grounds and the total sharing of the most abundant haplotypes, E1 and F2, we used this method to estimate and compare haplotype richness.

3) Another method to estimate allele richness is the calculation of the asymptotic value in a curve of richness accumulation along sampling. There are basically three accumulation models whose application depends on how much richness has been accumulated with reference to its asymptotic value. The smaller the sample, the lesser fraction of richness registered, and the lesser accuracy in estimating the asymptotic value (Soberón & Llorente 1993).

3.1) For a sample with richness close to its asymptote, the probability of finding a new allele decreases proportionally to the alleles list in the sample. Exponential saturation is the most adequate model to fit in this situation. This is:

$$R_n = \frac{a}{b}(1 - e^{bn})$$

where *n* is sample size, *a* is the start discovery rate for new alleles, and *b* is the rate at which the discovery rate decreases as the alleles list grows. Notice that for *n* tending to infinity, the asymptotic value is $R_\infty = a/b$.

3.2) When the accumulation curve is not close to its asymptote but points to it fairly, the hyperbolic Michaelis-Menten model (introduced to ecology by Clench in 1979; Soberón & Llorente 1993) becomes the best model to fit. This is:

$$R_n = R_\infty \frac{n}{b + n}$$

where R_∞ is the asymptotic richness to estimate and *b* is the sample size at which half R_∞ is reached. Notice that when *n* tends to infinity, R_n also tends to R_∞ (V_{max} in the model's original phenomenology of enzyme kinetics).

3.3) An accumulation curve is far from its asymptote when the sample size is too low or when, for a phenomenological circumstance, increasing the sample size decreases the chances of finding new alleles. This may happen in large and poorly known populations or in populations that change whilst they are sampled. In cases like this, non-asymptotic accumulation curves may have information on the phenomena that are changing the population. Therefore, the analysis becomes to estimate the sample size needed to register a given richness. Soberón & Llorente (1993) have provided the following model for non-asymptotic accumulation curves.

$$R_n = \frac{1}{z} ln(1 + zan)$$

where *a* and $z = 1 - e^{-b}$ are the parameters described for the exponential saturation.

1.3. Four differentiation indices

Genetic differentiation can be conceived as the fraction of total diversity in a group of populations owed to the differences between them. Differentiation is thus a component of endogamy (*F*) between the populations under comparison (*s*) within the whole set of populations (*t*). This factor of endogamy or difference between populations is thus named *Fst* and is conceptually defined as follows:

$$Fst = \frac{Ht - Hs}{Ht} = 1 - \frac{Hs}{Ht}$$

where *Ht* and *Hs* are the total diversity (expected heterozygosity) for all populations and their average diversity, respectively (e.g. Nei 1987). There are different approaches to estimating genetic differentiation depending on data sorts. Cockerham (1969) devised a calculation for *Fst* from a variance analysis (ANOVA) that Excoffier et al. (1992) extended to consider nucleotide differences among alleles for an index named *Φst*. These authors also extended the variance analysis to a two-level hierarchical design (AMOVA for Analysis of Molecular Variance) to estimate differentiation between groups of populations and differentiation between populations within groups.

As part of their estimation of allelic richness, Foulley & Ollivier (2006) defined a differentiation index based on the allelic richness estimated from their extrapolation method or from the rarefaction method. Allele richness differentiation (*ρst*) is calculated as follows:

$$\rho_{ST} = 1 - \frac{R-1}{K-1}$$

where *R* is the average allelic richness within populations and *K* is the total allelic richness among populations.

A consequent index of differentiation between populations is based on the variable sites in allele sequences (*vst*) and is simply calculated as follows:

$$vst = 1 - \frac{vs}{vt}$$

where *vs* is the average number of variable sites in the sequences within populations and *vt* is the number of variable sites among the sequences of all populations (Arbanasić et al. 2024).

1.4. Isolation by distance

When populations are distributed continuously over a large region such as the wintering grounds of humpback whales from Baja California to Central America (ca. 3000 km), including the offshore Revillagigedo Islands, the limited mobility of individuals at least for reproduction, raises genetic differentiation between the organisms of distant areas for differentiation being proportional to geographic distance. This is the so-called isolation-by-distance model of Wright (1943). There is isolation by distance when genetic differentiation is proportional to geographic distance. The number of population comparisons of geographic distances and genetic differentiation is for both $n(n-1)/2$ which is half under or over the diagonal of the n x n populations matrix. However, care must be taken to consider in the analysis proper statistically independent comparisons

(e.g. Harvey & Pagel 1991). Usually, the Pearson's correlation is employed in a permutational test of significance developed by Mantel (1967) but other approaches, such as the general linear regression, may also be adequate when comparisons do not include complete pairwise comparisons.

We examined isolation by distance with the four differentiation indices described above among the wintering grounds of humpback whales MxBC, MxAR, MxML, MxCP, MxGue, MxOax, and CAm. We made the linear regression of differentiation with respect to geographic distance (Tables S1, S2) and tested their correlation with the Mantel test available in Arlequin 3.5.2.2 (Excoffier et al. 2005; Table S3; Figure S1).

Indices *Fst* and ρst, which are dependent upon the number and frequencies of different haplotypes, were bounded under 0.16 whilst Φst and *vs*t, which depend also on the number of variable sites, were limited under 0.40, meaning that nucleotide differences enlarge the differences in haplotypes frequency. The four indices showed significant isolation by distance according to the Mantel test, though exhibiting low regression indices due to the dispersion of comparisons that indicate either differential gene flow across geography or population subdivision (Table S3; Figure S1).

Table S1. Straight geographic distances among seven wintering areas of humpback whales in the Northeast Pacific Ocean.

	MxBC	MxAR	MxML	MxCP	MxGue	MxOax	CAm
Geographic distance (km)							
MxBC	-						
MxAR	490	-					
MxML	475	612	-				
MxCP	678	815	203	-			
MxGue	1079	1216	604	401	-		
MxOax	1653	1790	1178	975	574	-	
CAm	2803	2940	2328	2125	1724	1150	-

Table S2. MtDNA differentiation of humpback whales in their wintering grounds of the Northeast Pacific Ocean. Bold numbers indicate statistical significance at $p < 0.05$.

	MxBC	MxAR	MxML	MxCP	MxGue	MxOax	CAm
Fst \ *Φst*							
MxBC	-	0.003	0.000	**0.084**	**0.137**	**0.268**	**0.156**
MxAR	0.002	-	**0.028**	**0.130**	**0.222**	**0.352**	**0.240**
MxML	0.008	0.004	-	**0.039**	0.070	**0.188**	**0.087**
MxCP	**0.028**	**0.040**	**0.018**	-	0.000	**0.049**	0.000
MxGue	**0.023**	**0.043**	**0.034**	0.000	-	0.000	0.000
MxOax	**0.109**	**0.134**	**0.119**	**0.038**	0.000	-	0.003
CAm	**0.071**	**0.096**	**0.081**	**0.022**	0.000	0.000	-
ρst[a] \ *vst*							
MxBC	-	**0.278**	0.222	**0.130**	0.259	**0.296**	**0.296**
MxAR	0.041	-	0.241	**0.148**	0.278	**0.315**	**0.315**
MxML	0.053	0.062	-	0.093	0.222	**0.259**	**0.259**
MxCP	**0.106**	**0.132**	0.085	-	0.130	**0.167**	0.167
MxGue	0.049	0.055	0.076	0.065	-	0.296	0.296
MxOax	**0.152**	**0.143**	**0.132**	**0.138**	0.072	-	0.333
CAm	**0.119**	**0.142**	**0.133**	**0.118**	0.066	0.043	-

[a] *ρst* determined from the estimated number of haplotypes.

The four indices exhibited relatively high and significant isolation between MxAR and MxCP despite being in the same latitude. Instead, MxBC, MxAR, and MxML showed low and significant differentiation per distance among them for *Fst*, *ρst*, and *Φst*. However, the *vst* values among these regions were high respective to their distances, being significant for the comparison MxBC-MxAR (Tables S2, S3; Figure S1). This indicates similar mtDNA variation in the three regions with important differences in variable sites, meaning that the three regions are differentiated by similar, but not the same haplotypes, besides the frequency differences among shared haplotypes.

Indices *Fst*, *ρst*, and *Φst* also presented several high and significant differentiation values that indicated a major division of haplogroups, haplotypes, and their frequencies between MxCP and MxML. Instead, index *vst* showed its largest values in the ranges of CAm, MxOax, and MxGue, though being non-significant. With reference to geographic distance, the ratio *vst*/km exhibited its larger values in the comparisons MxOax-MxGue and MxGue-MxCP. None of the four differentiation indices between MxGue and MxCP was high or significant (Table S2; Figures 1, S1). Similarly to MxBC, MxML, and MxAR, differentiation south of MxGue given by *vst*, appears owed to different but closely related haplotypes, making thus MxCP and MxGue seem a transitional region. This yields a structure more complicated than a sole division of the Mexico and Central America distinct population segments in southern Mexico (Martien et al. 2021; Taylor et al. (2021).

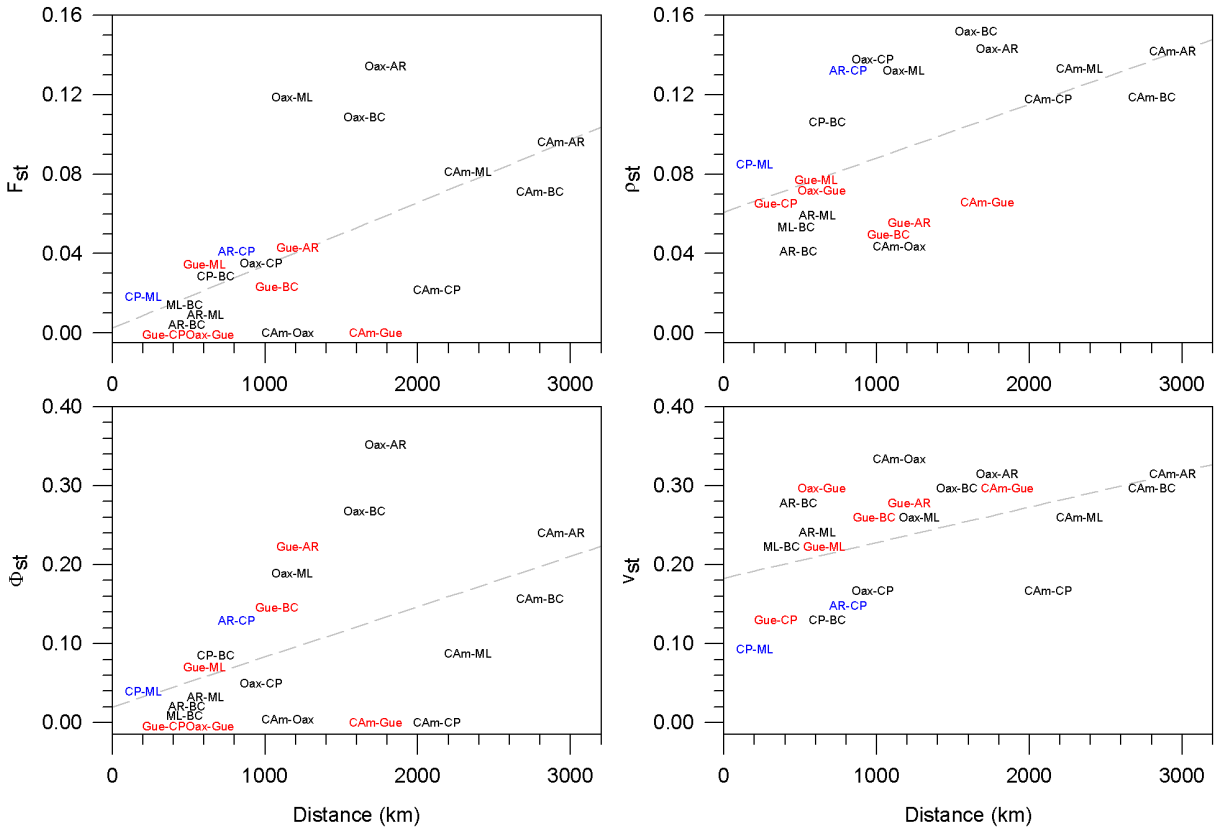

Figure S1. MtDNA isolation by distance for four differentiation indices in the wintering grounds of humpback whales in the Northeast Pacific Ocean. Comparisons involving MxGue are indicated in red. Comparisons involving the neighbouring regions of MxCP are indicated in blue. The abbreviation Mx of regions has been retired to ease plot examination. A few pairwise comparisons were slightly moved to separate them according to their relative positions. Dashed lines indicate anyway the exact linear regressions with the parameters shown in Table S3.

Table S3. Regression parameters for the mtDNA-isolation-by-distance plots of humpback whales in their wintering grounds between Baja California and Central America (Figure S1).

	Slope ($x10^{-5}$/km)	Ordinate	r^2	*p* (Mantel)
Fst	3.164	0.002	0.346	0.009
Φst	6.363	0.020	0.234	0.028
ρst	2.718	0.061	0.317	0.014
vst	4.488	0.183	0.251	0.005

2. MATERIALS AND METHODS

2.1. Sampling of tissues and mtDNA sequences

We conducted our surveys along the northern coast of the State of Guerrero (17.30°N – 17.67°N) in partnership with local people looking to cultivate an ethos of stewardship toward whales in a nascent whale-watching community. Data and samples were collected under scientific research permits, giving priority to responsible whale-watch practices over data collection and, thus, abandoning whales when they exhibited repeated evasive behaviour. In alignment with this principle, we chose to collect sloughed skin from photo-identified animals rather than biopsies. We used aquarium nets attached to a pole to collect sloughed skin in the water from photo-identified whales after vigorous behaviours such as breaches and lobtailing. Skin pieces were taken from the net using flamed dissection forceps to put the tissue in cryotubes containing 95% molecular-grade ethanol. Two samples were obtained from dead individuals. Samples were kept in a refrigerator until they arrived at a laboratory where they were stored at -86°C. We collected 40 samples from which we extracted genomic DNA and amplified a 419 bp segment of the mitochondrial control region following the laboratory procedures described by Baker et al. (2013). Only 22 amplicons yielded sequences with adequate quality control scores (Table S4).

Table S4. Identity data of the 22 identified individuals for which we obtained adequate-quality sequences of the mtDNA control region (419 bp). Sample dates are indicated in the sample ID as year-day-month.

Sample ID	Haplotype	Longitude	Latitude	WGRP ID	CRC ID
Mnov-2015-27-01-Gue-01	E1	-101.4714	17.5062	WGRP_HB_065	10704
Mnov-2015-29-01-Gue-02	A3	-101.6504	17.6515	WGRP_HB_070	16141
Mnov-2015-09-02-Gue-04	E1	-101.2590	17.2965	WGRP_HB_089	-
Mnov-2015-24-02-Gue-05	F2	-101.6411	17.6443	WGRP_HB_107	-
Mnov-2015-26-02-Gue-06	A+	-101.2766	17.3408	WGRP_HB_113	-
Mnov-2015-02-03-Gue-07	E1	-101.4824	17.5368	WGRP_HB_127	-
Mnov-2015-03-03-Gue-08	F4	-	-	-	-
Mnov-2015-05-03-Gue-10	F2	-101.5014	17.5813	WGRP_HB_131	16198
Mnov-2017-13-01-Gue-11	F2	-101.4768	17.5715	WGRP_HB_200	17201
Mnov-2017-29-01-Gue-12	F3	-101.3941	17.4900	WGRP_HB_250	15879
Mnov-2017-04-02-Gue-13A	A-	-101.3613	17.4936	Dead calf	-
Mnov-2017-10-03-Gue-14A	E1	-101.4344	17.5121	WGRP_HB_287	10878
DW(Mnov)-2018-21-01-Gue-32	F2	-	-	Dead whale	-
Mnov-2018-22-01-Gue-16	F3	-101.4142	17.5010	WGRP_HB_213	16164
Mnov-2018-02-02-Gue-18	F2	-101.4711	17.5265	WGRP_HB_339	15204
Mnov-2018-07-02-Gue-19	E4	-101.5206	17.5702	WGRP_HB_345	12492
Mnov-2018-23-02-Gue-22	E1	-101.4540	17.5194	WGRP_HB_370	18399
Mnov-2018-28-02-Gue-23	E3	-101.4799	17.5968	WGRP_HB_372	10800
Mnov-2018-06-03-Gue-25	F2	-101.3900	17.4948	WGRP_HB_380	10763
Mnov-2018-15-03-Gue-28	E13	-101.454	17.5238	WGRP_HB_386	10729
Mnov-2018-15-03-Gue-30	F2	-101.551	17.6105	WGRP_HB_390	18405
Mnov-2021-01-02-Gue-32	E1	-101.5018	17.6026	WGRP_HB_554	10641

2.2. Sufficiency analysis of the MxGue sample

Even having a sample representative of the population it is drawn off, the sample still can be insufficient, i.e., the statistics that describe the sample do not attain stable values as the data accumulate. Sampling sufficiency depends on the statistic measured, its variation, the question it is targeted to answer, and the confidence needed for it (DePatta Pillar 1998; Ramsey & Hewitt 2005; Schefler 1980).

To better assess the statistical power of our comparisons, we examined the sufficiency of the small sample of mtDNA haplotypes in the Guerrero humpback whales (n = 22) by determining the stability of the observed haplotype richness (*Ro*), gene diversity (H), and nucleotide diversity (π) as these statistics are involved in the four measures of differentiation examined. We estimated the values of *Ro*, H, and π along the sampling series, calculating also a measure of their variability along sampling (Aguirre-Samudio et al. 2014). For two consecutive data in the sampling series, x and x-1, variability for the xth datum (V_x) is, in the case of H:

$$V_x = 2abs\left(\frac{H_x - H_{x-1}}{H_x + H_{x-1}}\right)$$

meaning the absolute difference between consecutive data in the series divided by their average. As the data accumulate, diversity measures attain a stable value and their consecutive variability tends to zero (Figure S2).

Since the actual sampling series is contingent respective to the final diversity value, we built 100 random series of the data to determine a confidence interval for the diversity stabilization and the variability decay with attention to the 95% confidence interval of the last V_x value which was up to 0.105 for *Ro*, 0.022 for H, and 0.026 for π. This means that the Guerrero sample is sufficient for H, π, and thus, for *Fst*, Φst, and *vst*, under the criterion of $V_x < 0.05$, whilst *Ro*, and therefore ρst, are not. These results thus made it necessary to estimate haplotype richness and ρst, correcting for the different sample sizes (Figure S2).

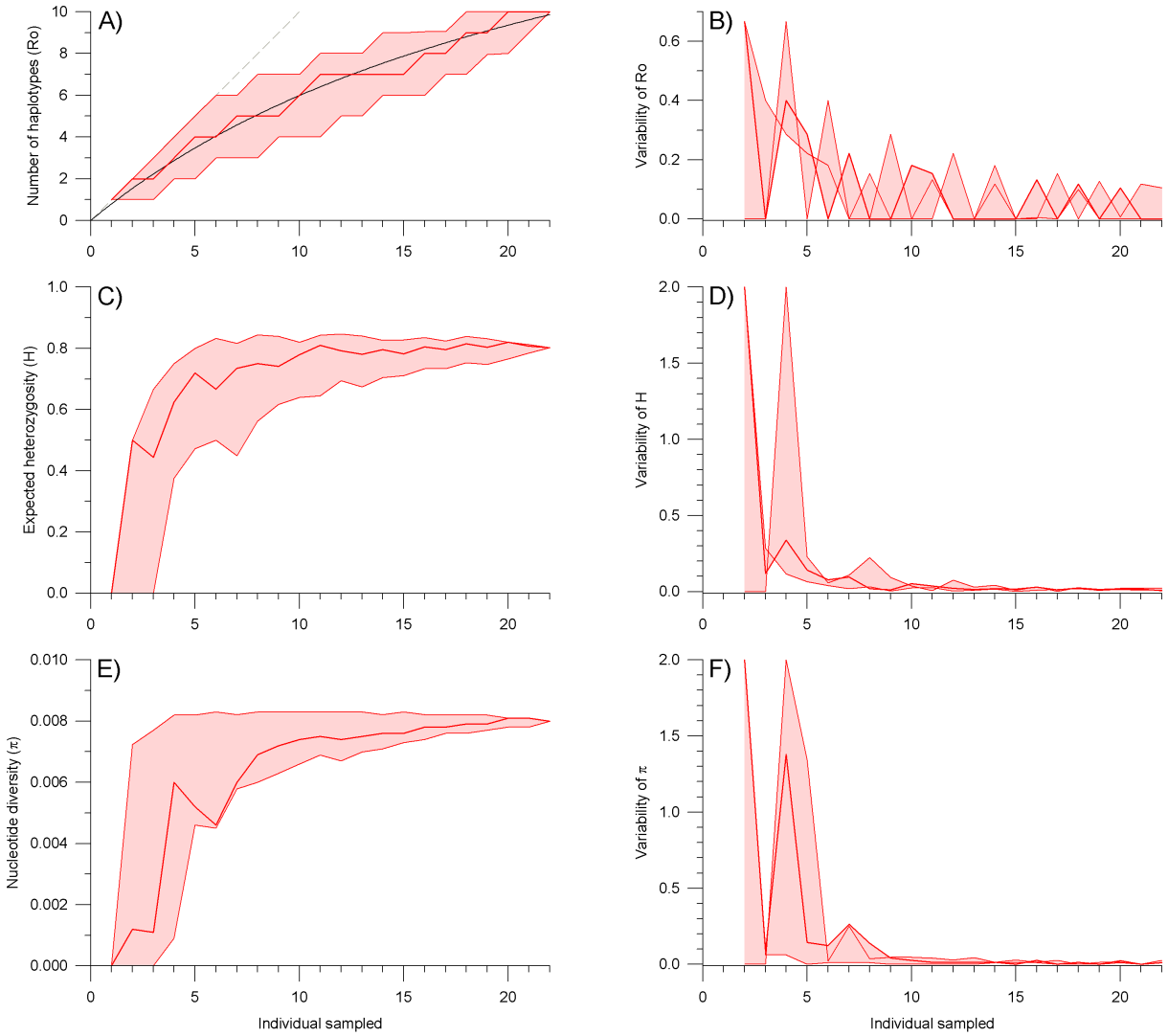

Figure S2. Stabilization of the diversity indices *Ro* (panel A), *H* (panel C), and π (panel E) of mtDNA-haplotypes along the sampling of 22 humpback whales from Guerrero. The dashed grey line in the accumulation of *Ro* indicates the diagonal, whilst the black line is the fit to the hyperbolic model of saturation (Soberón & Llorente 1993) with an asymptote parameter of 21.5, Michaelis-Menten constant of 25.8, and r^2 = 0.978. Profiles of diversity variability along the sampling are panel B for *Ro*, panel D for *H*, and panel F for π. Bold lines indicate the profiles of the actual sampling, whilst the surrounding pink areas stand for the 95% confidence of 100 shuffled sampling series.

3. HUMPBACK WHALE DISTRIBUTION IN THE MEXICAN PACIFIC OCEAN

The Marine Mammalogy Group at Facultad de Ciencias UNAM surveyed 61 067 km of the Mexican Pacific Ocean during the years 1981 – 2019, registering 166 sightings of humpback whales (Figure S3; Medrano González et al. 2022).

Between November and April, humpback whales and their newborns congregated around the Revillagigedo Islands (18°N latitude, ca. 600 km off the mainland the closest), Southern Baja California (22°N – 25°N latitude), Mainland Mexico coasts between Colima and Sinaloa (19°N – 24°N latitude), and the coasts between Oaxaca and Guerrero (16°N – 18°N latitude). We found few whales between these regions. From May to September, humpback whales were sighted along the mid-Pacific coast of Baja California (28°N latitude) and around the Midriff Islands in the Gulf of California (28°N – 30°N latitude), where a few calves have also been recorded (Figure S3). Notice that the genetic boundaries that we described along the coasts of Southern Jalisco and Southern Guerrero, correspond to apparent congregation regions and not to the species' distribution hiatus along the coast of Michoacan.

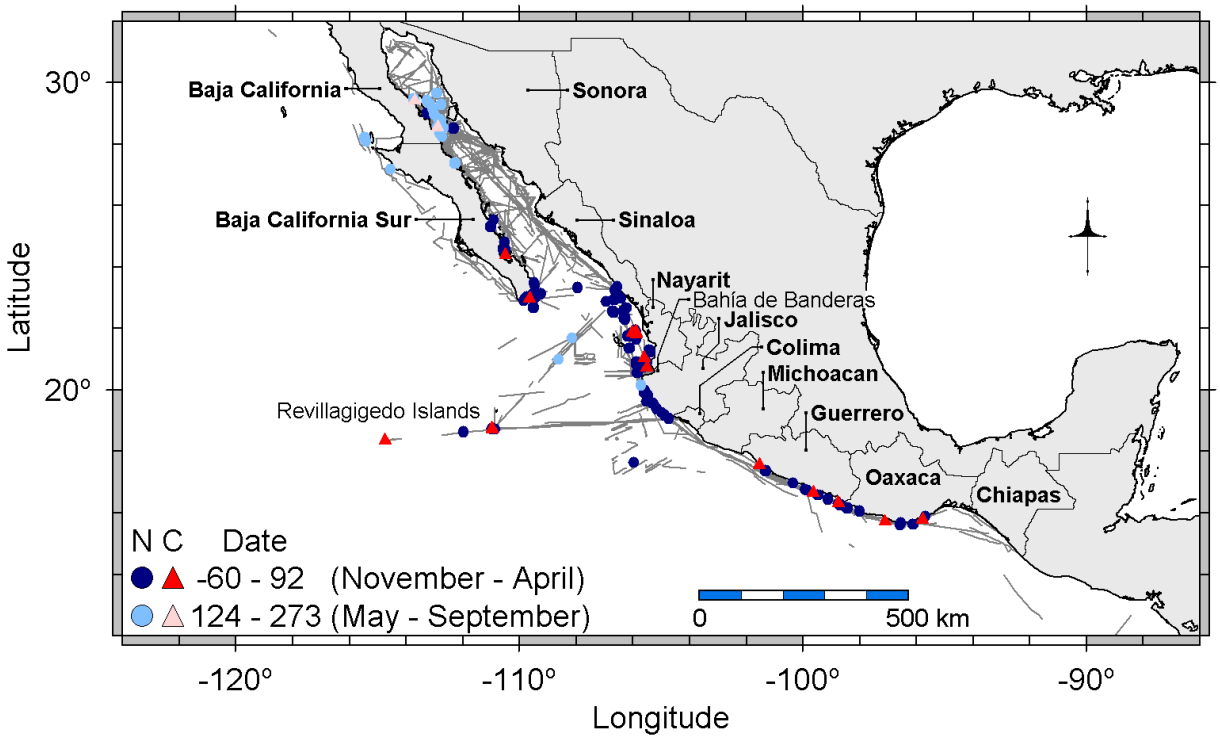

Figure S3. Seasonal records of humpback whales in the Mexican Pacific Ocean. Grey lines indicate 5926 survey transects. We show the 145 non-calf pods records (N) in blue circles and the 21 records of humpback whale pods with calves (C) in red triangles. Hue for circles and triangles indicates seasonality. States on the Mexican Pacific coast are shown in bold type.

LITERATURE CITED

Aguirre-Samudio AJ, Cruz-Fuentes CS, González-Sobrino BZ, Gutiérrez-Pérez V, Medrano-González L (2014) Haplotype and nucleotide variation in the exon 3-VNTR of the *DRD4* gene from indigenous and urban populations of Mexico. Am J Hum Biol 26: 682–689 (https://doi.org/10.1002/ajhb.22581)

Arbanasić H, Medrano-González L, Hrenar T, Mikelić A, Gomerčić T, Svetličić I, Pavlinec Ž, Đuras M, Galov A (2024) Recent selection created distinctive variability patterns on MHC class II loci in three dolphin species from the Mediterranean Sea. DCI 150: 105079. (https://doi.org/10.1016/j.dci.2023.105079)

Avise JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York NY

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Ramirez JU, Vasquez O, Clapham PJ, Alling A, O'Brien SJ, Palumbi SR (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. PNAS 90: 8239–8243 (https://doi.org/10.1073/pnas.90.17.8239)

Baker CS, Steel D, Calambokidis J, Falcone E, González-Peral U, Barlow J, Burdin AM, Clapham PJ, Ford JKB, Gabriele CM, Mattila D, Rojas-Bracho L, Straley JM, Taylor BL, Urbán J, Wade PR, Weller D, Witteveen DH, Yamaguchi M (2013). Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. Mar Ecol Prog Ser 494: 291–306 (https://doi.org/10.3354/meps10508)

Cockerham CC (1969) Variance of gene frequencies. Evolution 23:72–84 (https://doi.org/10.1111/j.1558-5646.1969.tb03496.x)

DePatta Pillar V (1998) Sampling sufficiency in ecological surveys. Abstr Bot 22: 37–48

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L. Skeels)] endemic to Morocco. Theor Appl Genet 92: 832–836 (https://doi.org/10.1007/BF00221895)

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Pop Biol 3: 87–112 (https://doi.org/10.1016/0040-5809(72)90035-4)

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0. An integrated software package for population genetics data analysis. Evol Bioinf Online 1: 47–50 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658868/)

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479–491 (https://doi.org/10.1093/genetics/131.2.479)

Foulley JL, Ollivier L (2006) Estimating allelic richness and its diversity. Livest Sci 101: 150–158 (10.1016/j.livprodsci.2005.10.02)

Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford UK

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577–586 (https://doi.org/10.2307/1934145)

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27(2): 209–20

Martien KK, Taylor BL, Archer FI, Audley K, Calambokidis J, Cheeseman T, De Weerdt J, Frisch Jordán A, Martínez-Loustalot P, Ortega-Ortiz CD, Patterson EM, Ransome N, Ruvelas P, Urbán Ramírez J, Villegas-Zurita F (2021) Evaluation of Mexico Distinct Population Segment of Humpback Whales as units under the Marine Mammal Protection Act. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-658 (https://doi.org/10.25923/nvw1-mz45)

Medrano González L (2023) El análisis de variación genética y su suficiencia en un caso arqueológico con muestreo reducido (zapotecos) y otro con muestreo amplio (ballenas). In: Aguirre Samudio AJ, Ortíz Díaz E, Higelin Ponce de León R (eds) Población y movilidad entre los zapotecos prehispánicos de Oaxaca. Instituto de Investigaciones Antropológicas UNAM, Cd. México, p 115–152

Medrano González L, Ransome NL, Arroyo Sánchez ER (2022) Evaluación de las interacciones entre el transporte marítimo y grandes cetáceos en el Pacífico Mexicano. Report to World Wildlife Fund, Cd. México

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York NY

Ramsey CA, Hewitt AD (2005) A Methodology for Assessing Sample Representativeness. Environ Forensics 6: 71–75

Schefler WC (1980) Statistics for the Biological Sciences. 2nd ed, Addison-Wesley, Reading MA

Soberón MJ, Llorente BJ (1993) The Use of Species Accumulation Functions for the Prediction of Species Richness. Conserv Biol 7(3): 480–488 (https://doi.org/10.1046/j.1523-1739.1993.07030480.x)

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595 (https://doi.org/10.1093/genetics/123.3.585)

Taylor BL, Martien KK, Archer FI, Audley K, Calambokidis J, Cheeseman T, De Weerdt J, Frisch Jordán A, Martínez-Loustalot P, Ortega-Ortiz CD, Patterson EM, Ransome N, Ruvelas P, Urbán Ramírez J (2021) Evaluation of humpback whales wintering in Central America and southern Mexico as a demographically independent population. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-655 (https://doi.org/10.25923/sgek-1937)

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993) Asian affinities and continental radiation of four Native American mtDNAs. Am J Hum Genet 53(3): 563–590

Watterson GA (1978) The homozygosity test of neutrality. Genetics 88: 405–417 (https://doi.org/10.1093/genetics/88.2.405)

Wright S (1943) Isolation by distance. Genetics 28: 114–138 (https://doi.org/10.1093/genetics/28.2.114)