nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-020-1215-5

Check for updates

Determinants of genetic variation across eco-evolutionary scales in pinnipeds

Claire R. Peart[1,2,19], Sergio Tusso[1,2,19], Saurabh D. Pophaly[2,3,19], Fidel Botero-Castro[2], Chi-Chih Wu[1], David Aurioles-Gamboa[4], Amy B. Baird[5], John W. Bickham[6], Jaume Forcada[7], Filippo Galimberti[8], Neil J. Gemmell[9], Joseph I. Hoffman[7,10], Kit M. Kovacs[11], Mervi Kunnasranta[12,13], Christian Lydersen[11], Tommi Nyman[12,14], Larissa Rosa de Oliveira[15], Anthony J. Orr[16], Simona Sanvito[8], Mia Valtonen[17], Aaron B. A. Shafer[1,18] ✉ and Jochen B. W. Wolf[1,2] ✉

The effective size of a population (N_e), which determines its level of neutral variability, is a key evolutionary parameter. N_e can substantially depart from census sizes of present-day breeding populations (N_C) as a result of past demographic changes, variation in life-history traits and selection at linked sites. Using genome-wide data we estimated the long-term coalescent N_e for 17 pinniped species represented by 36 population samples (total $n = 458$ individuals). N_e estimates ranged from 8,936 to 91,178, were highly consistent within (sub)species and showed a strong positive correlation with N_C ($R^2_{adj} = 0.59$; $P = 0.0002$). N_e/N_C ratios were low (mean, 0.31; median, 0.13) and co-varied strongly with demographic history and, to a lesser degree, with species' ecological and life-history variables such as breeding habitat. Residual variation in N_e/N_C, after controlling for past demographic fluctuations, contained information about recent population size changes during the Anthropocene. Specifically, species of conservation concern typically had positive residuals indicative of a smaller contemporary N_C than would be expected from their long-term N_e. This study highlights the value of comparative population genomic analyses for gauging the evolutionary processes governing genetic variation in natural populations, and provides a framework for identifying populations deserving closer conservation attention.

Biodiversity is declining at an alarming rate[1]. Intraspecific genetic variation provides the raw material for adaptive evolution and is an important component of biodiversity, as reflected by targets in the international treaty for the conservation of biodiversity[2]. Genetic variation is continuously generated by novel mutations entering a population, but is eroded by purifying selection and random genetic drift. The expected rate of loss of neutral or nearly neutral genetic variability is inversely related to the classic population genetic metric known as the effective population size, or N_e (refs. [3,4]).

In long-term stable populations, N_e is expected to scale proportionally with the observed number of potentially breeding individuals (N_C) in the population[5]. Nevertheless, across taxonomic scales the empirical relationship between the level of genetic variability and census population size is often weak, and N_e tends to be orders of magnitude lower than N_C (ref. [6]). Several explanations for this observation, often referred to as Lewontin's paradox[7], have been proposed[8]. One explanation involves the apparent inverse relationship between mutation rate and population size[9], resulting in a diminishing returns relation between genetic diversity and population size. Another explanation invokes selection at linked sites whereby both the spread of beneficial mutations and the removal of deleterious alleles reduce genetic variation of closely linked neutral polymorphisms[10]. Both forms of selection at linked sites, genetic hitch-hiking and background selection, have been shown to constrain neutral diversity, with the diversity-reducing effect being more pronounced the larger a population[11,12]. Another explanation for the common empirical disconnect between N_e and N_C can be sought in factors influencing the amount of genetic drift. In the absence of selection at linked sites, these include constitutive factors such as the species' mating system or other life-history traits introducing reproductive skew[13], as well as demographic perturbations reflecting historical deviations from constant population sizes.

Estimates of N_e can be contrasted to N_C, with the ultimate goal of predicting one from the other[14] or identifying populations with a low N_e/N_C ratio, which, although a contentious issue[15], has been

[1]Science of Life Laboratories and Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden. [2]Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Planegg-Martinsried, Germany. [3]Max Planck Institute for Plant Breeding Research, Cologne, Germany. [4]Laboratorio de Ecología de Pinnípedos 'Burney J. Le Boeuf', Centro Interdisciplinario de Ciencias Marinas, Instituto Politécnico Nacional, Baja California Sur, México. [5]Department of Natural Sciences, University of Houston-Downtown, Houston, TX, USA. [6]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA. [7]British Antarctic Survey, Natural Environment Research Council, Cambridge, UK. [8]Elephant Seal Research Group, Sea Lion Island, Falkland Islands. [9]Department of Anatomy, University of Otago, Dunedin, New Zealand. [10]Department of Animal Behaviour, Bielefeld University, Bielefeld, Germany. [11]Norwegian Polar Institute, Fram Centre, Tromsø, Norway. [12]Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland. [13]Natural Resources Institute Finland (Luke), Joensuu, Finland. [14]Department of Ecosystems in the Barents Region, Norwegian Institute of Bioeconomy Research, Svanhovd Research Station, Svanvik, Norway. [15]Laboratory of Mammal Ecology, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brazil. [16]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Marine Mammal Laboratory, Seattle, WA, USA. [17]The Saimaa Ringed Seal Genome Project, Institute of Biotechnology, University of Helsinki, Helsinki, Finland. [18]Forensic Science & Environmental Life Sciences, Trent University, Peterborough, Ontario, Canada. [19]These authors contributed equally: Claire R. Peart, Sergio Tusso, Saurabh D. Pophaly. ✉e-mail: aaronshafer@trentu.ca; j.wolf@biologie.uni-muenchen.de

related to a population's adaptive potential[16,17]. N_e/N_C can vary substantially among species with differences in life-history traits, such as the sex ratio. Factors increasing reproductive skew tend to result in lower N_e relative to current census sizes[5,18,19]. There are multiple approaches to estimating N_e, which differ in important aspects. Non-genetic methods based on demographic data[5] and pedigrees[20] generally provide a time-sensitive estimate over the course of a single or few generations. Methods using molecular data exploiting excess in heterozygosity over Hardy–Weinberg expectation, temporal sampling or linkage disequilibrium[21] integrate over slightly longer timescales, ranging from several dozen generations[22] to several thousand (although «N_e) generations[23].

In contrast, estimates of the coalescent effective population size capture the interaction of evolutionary forces shaping genetic variation over extended periods in the order of $4N_e$ generations, which corresponds to the expected time to the most recent common ancestor of the alleles sampled from (diploid) current-day individuals. In vertebrates, this time frame spans tens of thousands to millions of years[23–25]. The coalescent N_e scales with the harmonic mean of a (rapidly fluctuating) sequence of values of effective sizes of each generation back to the most recent common ancestor. It thus encapsulates factors that have shaped the ancestral genetic variation that is still observed in a contemporary population sample[26]. Consequently, genomic data sampled from putatively neutral sites of remnant individuals provide a simple means for obtaining diversity (θ_π)-based estimates of the coalescent $N_e = \theta_\pi/2c\mu$ if the mutation rate, μ, and ploidy, c, are known. This relationship assumes that genetic diversity of the species' populations approximates an equilibrium between mutation and genetic drift, and holds within the limits of the infinite sites mutation model that is applicable for most species, except for 'hyper-diverse' organisms such as viruses[27]. It allows the investigation of processes affecting genetic diversity in the past and explores their relationship to current population sizes.

Investigation of the relationship between N_e and N_C has mostly been limited to single-species estimates or meta-analyses across large taxonomic scales[6,11]. Contrasting diversity patterns among closely related species, however, allows the inference of relevant factors shaping genetic diversity while circumventing phylogenetic inertia that may hamper the interpretation of the relevant life-history differences. Pinnipeds, comprising true seals, sea lions, fur seals and walruses, are a group of marine mammals that display considerable variation in life-history traits and breeding habitats[28]. Pinniped population abundances have fluctuated in response to changes in sea surface temperatures, as these are thought to impact the availability of prey and the location and extent of suitable breeding areas (for example, refs. [29,30]). Populations of many pinniped species were substantially reduced by commercial sealers from the late eighteenth to early twentieth century and which, coupled with climate change, has led to almost 50% of recognized taxa being considered threatened or endangered[31].

Here we estimate coalescent N_e across pinniped species, assess the impact of ecological factors and population-specific demographic fluctuations and discuss the roles of drift and selection at linked sites in shaping diversity estimates. We compared genome-wide estimates of the coalescent N_e in 17 pinniped species to current N_C estimates, and inferred how this ratio is influenced by life-history traits and demographic perturbations preceding anthropogenic impact. The study has fundamental implications for understanding the influence of a species' life history and historical processes on contemporary genetic diversity and population size. It provides a basis to inform present-day conservation initiatives aimed at maintaining or restoring N_C and N_e.

Results and discussion

Mutation rate μ, genetic variation θ_π and N_e/N_C ratio. We generated double-digest restriction site-associated DNA (ddRAD) data for a total of 458 individuals sourced from 36 populations of 17 species (20 taxa, including subspecies) distributed across the pinniped phylogeny (Fig. 1a,b). From an average of 31,956 (range 11,826–45,438) mapped loci per population remaining after rigorous filtering, we estimated the basic population genetic summary statistics Watterson's theta (θ_w), nucleotide diversity (θ_π) and Tajima's *D*. In addition, we inferred genome-wide mutation rates using 8,864 orthologous genes extracted from draft genome sequences of representatives for the three pinniped families: Otariidae (Antarctic fur seal, Californian sea lion), Odobenidae (walrus) and Phocidae (Weddell seal, Hawaiian monk seal) (Fig. 1b and Supplementary Table 1; see Methods for accession numbers). Mutation rate estimates of μ per base pair (bp) and generation ranged from 5.00×10^{-9} to 1.32×10^{-8} ($0.4–1.21\times10^{-9}$ bp^{-1} year^{-1}), which is comparable to other mammalian lineages[9] including humans[32]. The similarity of these estimates, with no apparent effect of population size, precludes a strong influence of mutation rate on the variation in N_e across species.

Using the relationship $N_e = \theta_\pi/4\mu$, we inferred N_e for each population separately (Supplementary Table 2). These resulting estimates ranged from 8,936 to 91,178 (median, 18,993) and were thus comparable to values for other large mammalian species, including humans[33,34]. We were interested in obtaining a global estimate of N_e for the entire (sub)species while facing the empirical limitation of access to local population sample(s). Under a broad set of migration models, theory suggests that the mean within-population diversity is most appropriate for comparison of properties across species[4]. In species with strong population structure, however, the mean within-population diversity may deviate markedly from the pooled population estimate, as quantified by *F*-statistics. A literature review suggests that the fixation index, F_{ST} is generally low for most of these species, findings supported in this study for species where multiple populations were available (Supplementary Table 3). Therefore, single-population samples are expected to contain robust information on genetic variation relevant at the level of the (sub)species. In line with this prediction, N_e estimates were consistent among population samples with little or no genetic differentiation (coefficient of variation (CV): New Zealand fur seal, *Arctocephalus forsteri* (*ArcFor*), 0.002; Galápagos sea lion, *Zalophus wollebaeki* (*ZalWol*), 0.002; Supplementary Table 3 and Extended Data Figs. 1–3). These estimates were also similar among moderately stratified populations (South American fur seal, *Arctocephalus australis* (*ArcAus*), 0.005; California sea lion, *Zalophus californianus* (*ZalCal*), 0.012), but could differ substantially among formally recognized subspecies with a high degree of evolutionary independence (ringed seal, *Pusa hispida* (*PusHis*), 0.029). We thus concluded that θ_π-based N_e estimates approximated the amount of long-term genetic drift experienced by the (sub)species as a whole. Consequently, we extrapolated the 'species' N_e from the population best representing the core distribution of the nominate subspecies. We note that this estimate will differ from the mean value over all interconnected subpopulations[35] and will accordingly introduce additional variance in N_e/N_C estimates.

As a complement to the above approach, we calculated the harmonic mean of N_e estimates obtained as the rate of coalescence using multiple sequentially Markovian coalescent analyses for a single genome (PSMC')[36] where genome assemblies of sufficient quality were available. PSMC'-based values were similar to the ddRAD-derived estimates for the California sea lion, *ZalCal* (18,057 versus 16,398–22,766) and higher for the Antarctic fur seal, *Arctocephalus gazella*, *ArcGaz* (29,133 versus 22,550). Higher values for the PSMC'-based estimates are consistent with population structure decreasing the rate of coalescence, although this will depend on the timing and level of gene flow and population size changes[37]. For walrus, *Odobenus rosmarus* (*OdoRos*), PSMC' was conducted on an individual from the Pacific subspecies (*O. r. divergens*; $N_C = 100{,}000$), whereas ddRAD estimates were based on a population sample from

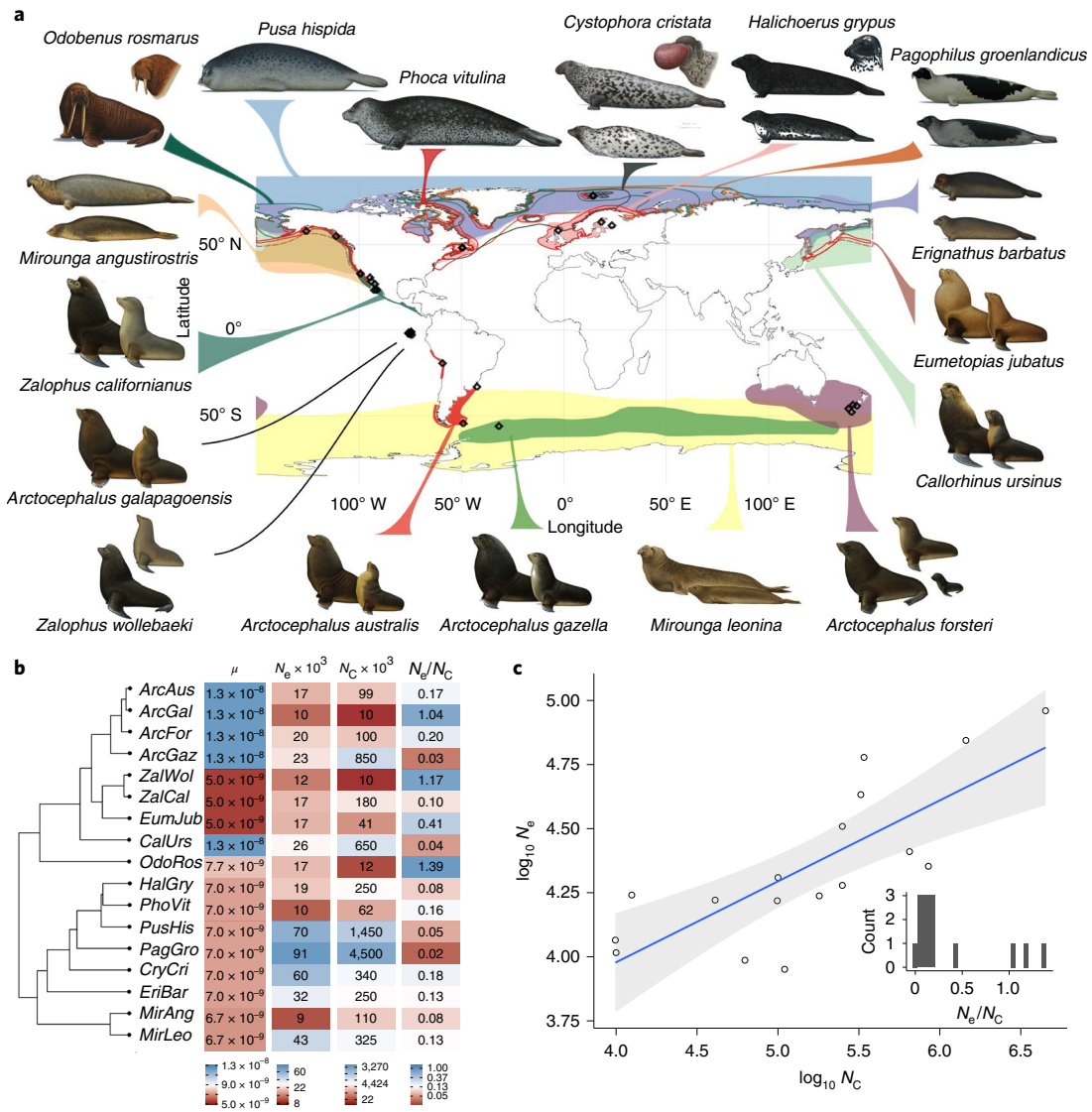


Fig. 1 | Sampling set-up and core population parameters of the 17 pinniped species investigated in this study. a, Sampling locations and geographic ranges of focal seal species and populations. **b**, Phylogenetic relationships among focal species following Higdon et al.[48]. Species abbreviations are the first three letters of both genus and species names. Mutation rate estimates, μ (bp^{-1} per generation) derived from five genomes were used for all species within a clade. Population genetic summary statistics and estimates of census population sizes and N_e/N_C shown in the heatplot are based on the nominal subspecies and populations representing the core species range. **c**, Relationship between N_e and N_C shown on a logarithmic scale (base 10). The blue line depicts the regression line, the shaded area its 95% confidence intervals. The inset shows the frequency distribution of binned N_e/N_C ratios. Illustrations in **a** courtesy of Brett Jarrett; map retrieved from IUCN[74].

the Atlantic subspecies (*O. r. rosmarus*; $N_C = 12{,}500$). Despite the substantial difference in current-day N_C, estimates of N_e were similar across the two subspecies and methods (19,808 versus 15,491), which suggests that walruses in the Atlantic might still share substantial genetic variation with those in the Pacific, a finding consistent with mitochondrial results[38]. Overall, these results support the previous notion that the coalescent N_e contains information on the effects of long-term genetic drift, or selection at linked sites, shared among populations within species[25].

Selection at linked sites probably affects N_e (see Main); however, in-depth investigation of its effects would require detailed knowledge of local mutation and recombination rates along the genomes of all species considered here. To assess the magnitude of the possible effect, we quantified the potential target size for selection (approximated by protein-coding genes) for walrus where both an annotated reference genome and ddRAD data were available. In total, 43.5% of ddRAD loci in this species overlapped at least partially with annotated genes. Consistent with the diversity-reducing effects of selection both against amino acid substitutions and at linked sites, estimates of nucleotide diversity, θ_π, within genic regions were lower than in intergenic regions (0.0005197 and 0.0005504, respectively; genome-wide mean, 0.0005371). This translates to a change in N_e from 16,984 in genic regions to 17,987 in intergenic regions (genome-wide, 17,551). The small magnitude of this difference is consistent with comparative analyses suggesting moderate effects of selection at linked sites for species with a relatively low N_e, such as pinnipeds[11]. It is also compatible with the proposition that, in the relatively large mammalian genomes, much intergenic sequence is far from selection targets. Within the range of N_e across species, changes that may be attributable to selection at linked sites are thus expected to be minor and are unlikely to affect the relationships with N_C, demographic change and the life-history variables investigated below.

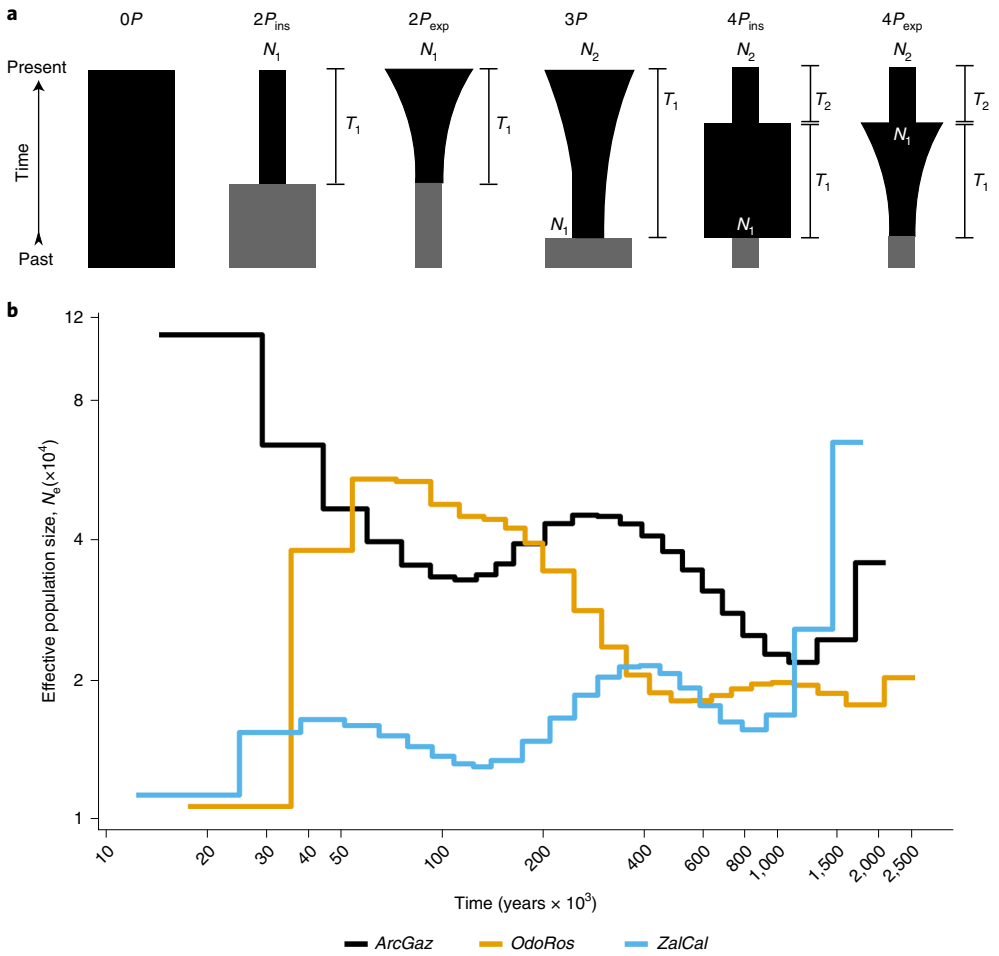


Fig. 2 | Historical changes in population size. a, Demographic scenarios including (1) a null model of constant population size (0*P*), (2) a single, instantaneous ($2P_{ins}$) or gradual ($2P_{exp}$) population size change, (3) an instantaneous population size change with subsequent gradual increase or reduction of population size (3*P*) and (4) models incorporating two independent instantaneous ($4P_{ins}$) or gradual ($4P_{exp}$) population size changes. The magnitude of change is parameterized as N_1 and N_2 (size relative to ancestral population size N_{anc}, depicted in grey). The timepoints of change are accordingly parameterized as T_1 and T_2. **b**, Changes in effective population size (N_e) over time as inferred by PSMC' shown on a logarithmic scale.

Next, we compiled estimates for the number of breeding individuals, N_C, for each subspecies and explored their relationship to N_e (Fig. 1c, Extended Data Fig. 4 and Supplementary Table 2). Despite uncertainty in both estimates, the correlation was remarkably strong, explaining about 60% of the overall variance ($P=0.0002$, $R^2_{adj}=0.59$). This indicates that both independently derived variables share a substantial amount of information on species abundance. The N_e/N_C ratio was consistently below 1.0, with a median of 0.13 (mean, 0.31; exceptions: Galápagos sea lion (*ZalWol*), 1.17; Atlantic walrus (*OdoRos*), 1.39). The inferred N_e/N_C ratio was strikingly similar to published median estimates of 0.11 (ref. [5]) and 0.14 (ref. [18]) derived across a much larger taxonomic range and based on the temporal method, which capitalizes on the idea that variance in neutral allelic frequencies across generations is inversely proportional to N_e. Frankham[5] identified life-history parameters increasing the variance in reproductive success and population oscillation as the main factors responsible for low N_e/N_C ratios. Non-genetic approaches estimating N_e/N_C ratios on the basis of age-specific survival and fecundity rates have likewise identified life-history traits as the main contributor to reduced N_e estimates[19]. Despite the apparent similarities between N_e/N_C ratios obtained using these approaches and this study, they differ in terms of timescale and the information they carry. N_e estimates based on vital rates or allelic composition shifts integrate at most a handful of generations, and thereby are insensitive to population oscillations across larger timescales. The estimate used in this study scales in coalescent units of $4N_e$ generations which, in our dataset, translates to a median of approximately 1,584,549 years (range, 310,995–5,726,017 years). Assuming detailed information of dynamic changes in N_C across all generations in the past, the harmonic mean of N_C would similarly accommodate demographic change across a species' history[39]. Nevertheless, present-day estimates of N_C reflect only a snapshot of population history: we therefore expect a disproportionate role of past population oscillations on N_e/N_C through a reduction in N_e.

Demographic changes. To quantify past population size changes, we formulated a set of possible demographic models for each population, ranging from constant population size to the incorporation of two independent size changes (Fig. 2a). We then calculated their likelihood based on the diffusion approximation to the allele frequency spectrum[40]. Parameter estimates for the model with the best fit are reported in Fig 3. Similar to estimates of N_e, model inference was highly consistent among interconnected populations with a shared evolutionary history. For example, for all eight populations of the Galápagos sea lion *ZalWol*, we inferred an exponential reduction to an average of 0.26 of the ancestral population size (range, 0.09–0.42; CV = 0.39). For all four New Zealand fur seal *ArcFor* populations, the three-parameter model was preferred,

Species	Population	Ind.	Model	N_1	N_2	T_1	T_2	Taj's D
ArcAus	Peru	8	2Pins	0.002		0.000		0.334
	Brazil	11	3P	14.046	0.405	0.142		0.007
ArcGal	Isabela	12	3P	1.185	0.496	0.438		0.197
ArcFor	CFoulwind	8	3P	73.800	0.545	0.723		-0.026
	OBay	7	3P	54.878	0.609	0.568		-0.139
	OhauP	6	3P	42.943	0.470	0.456		-0.033
	VictoryB	6	3P	53.466	0.682	0.554		-0.178
ArcGaz	BirdIs	84	4Pins	22.785	0.305	6.154	0.017	-0.062
ZalWol	Baltra	10	2Pexp	0.330		0.516		0.311
	Espanola	10	2Pexp	0.107		0.039		0.312
	Fernandina	20	2Pins	0.088		0.007		0.413
	Genovesa	9	2Pexp	0.292		0.249		0.225
	IsabelaB	15	2Pexp	0.319		0.127		0.353
	IsabelaV	10	2Pexp	0.418		0.481		0.202
	Pinta	10	2Pexp	0.223		0.090		0.367
	SantaFe	9	2Pins	0.310		0.077		0.293
ZalCal	GCal1	12	3P	10.804	0.971	1.176		-0.010
	GCal2	13	2Pexp	2.238		0.171		-0.266
	GCal3	23	2Pins	12.613		0.073		-0.119
	GCal4	14	2Pexp	56.806		0.984		-0.110
	SMargarita	14	2Pins	0.012		0.000		0.203
	SMiguel	9	2Pins	4.023		0.234		-0.555
EumJub	GrassyIs	8	2Pins	0.308		0.352		0.018
CalUrs	SMiguel	8	2Pins	2.803		0.554		-0.576
OdoRos	Svalbard	12	3P	51.251	0.672	0.487		-0.208
HalGry	Scotland	3	2Pexp	0.008		0.138		-0.096
	StLawrence	6	2Pexp	0.067		0.346		0.158
PhoVit	Svalbard	13	4Pexp	1.237	1.292	0.180	0.035	-0.237
PusHis	Baltic	9	3P	41.192	0.808	0.412		-0.394
	Saimaa	10	2Pins	0.131		0.265		0.412
	Svalbard	14	3P	0.921	4.437	0.546		-0.664
PagGro	StLawrence	9	3P	32.197	1.653	0.419		-0.652
CysCri	Svalbard	8	3P	59.262	1.187	0.588		-0.511
EriBar	Svalbard	12	2Pins	0.858		0.399		-0.029
MirAng	NAtlantic	15	3P	0.039	0.534	0.155		0.066
MirLeo	SAtlantic	11	2Pexp	5.083		0.266		-0.299

Fig. 3 | Demographic parameter estimates. Summary of parameter estimates for demographic population histories according to the model best fitting the data for each pinniped species and population. For parameter nomenclature, see Fig. 2a. N_1, N_2, Tajima's *D*: blue background colour indicates population growth, red population decline. Ind represents the number of samples used for the analysis, with populations used for subsequent model analysis (see Fig. 2a) shown in grey.

portraying an initial 52-fold increase in population size (N_1 range, 42–74; CV = 0.22) followed by a bottleneck (N_2 range, 0.47–0.68; CV = 0.15). Demographic reconstructions were also similar in more stratified populations: subspecies of the ringed seal *PusHis* showed signatures of a large population reduction in both the Saimaa ringed seal (*P. h. saimensis*, $N_1 = 0.13$) that colonized Lake Saimaa at least 9,500 years ago[41] and the Baltic sea population (*P. h. botnica*, $N_1 = 41.19$, $N_2 = 0.81$), although in the Baltic this was preceded by a large population increase.

Single-population estimates of the demographic history are probably an oversimplification because they ignore the confounding effect of population structure and gene flow[42] (Extended Data Figs. 1–3). Nevertheless, in the absence of comprehensive sampling of all possible populations for each species, this is the only operational way forward. Moreover, PSMC' analyses, based on individual genome assemblies that assume panmixia, were broadly consistent with the demographic scenarios inferred from ddRAD population data (Fig. 2b). In walrus *OdoRos*, both analyses showed a population increase followed by a large decline to below the ancestral population size (PSMC' time estimate ≈ 500,000–40,000 years ago, estimate from δaδi software $T_1 = 0.49$), despite using data from distinct subspecies (Fig. 2b and Fig. 1). In the Antarctic fur seal *ArcGaz*, both methods recovered an increase in population size followed by a more recent decrease (PSMC' time estimate ≈ 1,000,000–100,000 years ago, δaδi $T_1 = 6.1$). PSMC' then recovered a successive population increase. For the California sea lion *ZalCal*, the demographic histories inferred from the ddRAD data covered different timescales (T_1, Fig 1), with some populations showing greater population growth than the PSMC' (Fig. 2b). A multi-population model incorporating migration might ultimately improve demographic estimates for the species and provide a better comparison to the PSMC'.

Tajima's *D* is a summary statistic that contains information on the demographic history of a population by comparing two measures of genetic diversity, θ_π and θ_w. The latter is more sensitive to the presence of rare alleles, which accumulate following population expansion. Assuming neutrality, negative values of Tajima's *D* are consistent with a scenario of recent population expansion whereas positive values indicate a recent period of demographic contraction. Tajima's *D* co-varied significantly with our estimates concerning recent population size changes N_1 and N_2 ($P = 0.0002$, $R^2_{adj} = 0.59$)

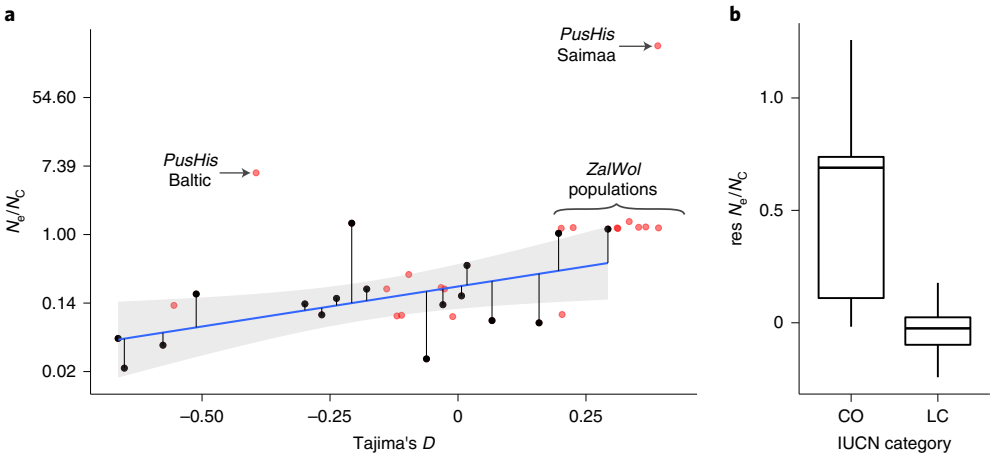


Fig. 4 | The relationship between N_e/N_C ratio and Tajima's *D*. a, Relationship between N_e/N_C and Tajima's *D* reflecting the demographic history of species. Note that the regression line (blue) refers only to the depicted relationship and does not consider the influence of other predictor variables. The grey shaded area depicts 95% confidence intervals of the regression line. Black points, populations used for model inference; red points, additional subspecies and populations. The *y* axis follows a natural logarithmic scale. **b**, Relationship between the residual variance from **a** (resN_e/N_C on *y* axis) and IUCN Red List category. Residual values per species (black points in **a**) were grouped into categories of concern (CO, $n=5$) or least concern (LC, $n=12$).

and, hence, captured basic information of the explicit demographic models. Using Tajima's *D* as a proxy for demographic history, we investigated whether demographic processes were shared among species with similar ecological niches and life histories. Variables analysed included breeding habitat (ice or land), species range, length of breeding season and reproductive skew based on harem size (Supplementary Table 2). Stoffel et al.[22] suggested that recent human-induced bottlenecks are related to the mating system and breeding habitat in pinnipeds, with bottlenecks linked to dense agglomerations in polygynous species breeding on land. In our models, breeding habitat likewise received statistical support (cumulative weight of evidence ($wAIC_c$) = 0.59; Supplementary Table 4). This similarity with Stoffel et al.[22] is surprising considering the different timescales investigated. Stoffel et al.[22] based their inferences on summary statistics from microsatellite data exploring recent demographic changes in the range of hundreds of years, as opposed to the scale of several hundred thousand years considered in this study. In contrast to Stoffel et al.[22], mating system had a relatively small impact as expressed in Akaike's weights (mating system $wAIC_c = 0.14$). Models including a species' geographic range size were more strongly supported on average ($wAIC_c = 0.85$; Supplementary Table 4); Tajima's *D* was negatively correlated with species range ($P = 0.008$, $R^2_{adj} = 0.34$), implying that species with large current distributions were more likely to have experienced a population expansion in the past. Overall, these results motivate two key interpretations. First, a species' ecological condition and life history affects its demographic stability through evolutionary time. Second, present distribution ranges contain information on past demographic processes—just as current species abundance contains information on long-term population sizes in the order of $4N_e$ generations (see Fig. 1c). On a cautionary note, our estimate of long-term effective population size, N_e, using the relationship $N_e = \theta_\pi/4\mu$, is based on the expectation of stationary allele frequency distributions; hence, it assumes that populations have been approximately stable or had sufficient time for mutations to accumulate and reach equilibrium. In pinnipeds, recent human-induced bottlenecks[22] and demographic changes (this study) challenge this assumption. Despite the potential deviation from the equilibrium assumption, N_e estimates were correlated with estimates of population size (N_1 or N_2) from the demographic models ($P < 0.05$, with phylogenetic correction $P < 0.001$), supporting the notion that N_e estimates based on θ_π constitute a simple, but meaningful, measure of long-term population size.

The effect of demographic change and life-history variables on N_e/N_C. We next explored the relationship between N_e/N_C, demographic change as inferred from Tajima's *D* and life-history trait variables, including breeding habitat (ice or land), breeding latitude, length of breeding season and reproductive skew as inferred by harem size. The best-supported model, explaining 30% of the variance in N_e/N_C, included only Tajima's *D* ($P = 0.02$), which was also the best-supported predictor variable overall ($wAIC_c = 0.60$; Supplementary Table 5). While models including life-history parameters such as the length of the breeding season ($wAIC_c = 0.42$) and breeding habitat ($wAIC_c = 0.19$) also received some support, demographic change was the major variable explaining variation in N_e/N_C (Fig. 4a and Extended Data Fig. 5). Species inferred to have experienced recent expansions, such as the harp seal, *Pagophilus groenlandicus* (*PagGro*), had the lowest N_e/N_C ratio of 0.02, whereas both Galápagos species (*A. galapagoensis and Z. wollebaeki* (*ArcGal* and *ZalWol*), with genetic signatures indicative of a bottleneck, had values around 1.0. This relationship appears to run counter to the common interpretation of the N_e/N_C ratio based on short-term measures of N_e: a population with small N_e/N_C will lose genetic diversity and fix deleterious mutations faster than an equally sized population with a higher ratio. Accordingly, a low N_e/N_C value is usually interpreted as a signal for increased relative genetic risk of that population. The interpretation based on the coalescent N_e estimate used in this study is necessarily different, as N_e and N_C react on different timescales. The diversity-based measure of $N_e = \theta_\pi/4\mu$ reflects the mean coalescent times for a sample of two (that is, θ_π). It is therefore less sensitive to very recent and rapid changes in population size than the actual size of a breeding population, N_C, which can change substantially in the course of a few generations. Hence, a N_e/N_C ratio close to 1.0, as seen for the Galápagos pinnipeds, is entirely consistent with a recent bottleneck scenario strongly reducing current-day N_C while not affecting long-term N_e (Extended Data Fig. 6).

Following this logic, we can expect that recent anthropogenic activity, such as active harvest or habitat degradation[22,31], will have a strong impact on N_C but less so on the coalescent N_e. Substantial human-associated population size changes are expected to be in the range of hundreds of years, which corresponds to $<10^{-3} N_e$

generations. Even heavy exploitation during such a time frame is not expected to leave visible traces in estimates of the coalescent N_e. Very recent changes in N_C should introduce residual variation in N_e/N_C not accounted for by the long-term demographic history of the population and by life-history characteristics of the species. To quantify this potential anthropogenic component of changes in N_C, we subtracted N_e/N_C estimates from the values predicted by the model that controlled for demographic change (the residuals from Fig. 4a, resN_e/N_C). We reason that species with positive residuals (res$N_e/N_C > 0$) have reduced N_C relative to the expectation of the model and may therefore deserve specific conservation attention. For example, resN_e/N_C of the isolated ringed seal *PusHis* subspecies reached 6.0 for the Baltic Sea and >220 for Lake Saimaa. These populations have been separated for only a relatively short period from the core species range and thus harbour more (ancestral) genetic variation than would be expected by their current-day N_C. A similar pattern is seen in the Galápagos sea lion *ZalWol*, which has a disproportionately high amount of genetic variation given its N_C. The large resN_e/N_C for the small population of Atlantic walrus, *OdoRos*, similarly indicates that its relatively high genetic diversity may be reminiscent of a large ancestral population or continued cohesion with the large Pacific stock. In line with this assumption, species classified by the International Union for Conservation of Nature (IUCN) as of 'concern' had significantly higher resN_e/N_C values than species classified as 'least concern' ($P = 0.001$, $R^2_{adj} = 0.47$; Fig. 4b).

Conclusions

Genetic diversity is the raw material of evolution. The coalescent N_e reflects the rate at which neutral or nearly neutral genetic variation is lost, and thus constitutes a useful measure for intraspecific biodiversity, which may contribute to adaptive potential in changing environments. Here, we demonstrated for a mammalian clade that the long-term dynamics of N_e co-vary strongly with the actual number of extant breeding individuals, N_C (Fig. 1). The strong correlation between N_e and N_C contributes to the debate on the relative roles of selection at linked sites and population size in shaping diversity (for example, refs. [11,12,43]), and implies a strong role for classical genetic drift in species with small to medium effective population sizes.

Both N_e and N_C were affected by historical processes, though at largely different timescales (Fig. 2 and Fig. 1). Considering absolute values, N_C exceeded N_e by one order of magnitude and showed considerably higher variation across taxa: estimates of N_C varied by up to 450-fold, whereas those of N_e were remarkably consistent, differing at most tenfold. This result fits with predictions of population genetic theory: N_e is expected to scale approximately with the harmonic mean of single-generation population size 'snapshots'. Mean genetic diversity is thus most strongly influenced by historical periods of small population size with below-average numbers of breeding individuals. As a corollary, mean genetic diversity has reduced variance relative to N_C and is expected to be less sensitive to recent, short-term population changes during the Anthropocene[44,45]. Strong, recent contractions of N_C will, however, cause losses or rare variants[22] which, in species with moderate population sizes, may constitute an important component of adaptive potential.

When controlling for the influence of demographic oscillation, which interacted with a species' life history, residual variation of the close relationship between N_e and N_C contained information on population vulnerability (Fig. 4). This result may serve as proof of concept in helping to identify species at risk. A low N_e/N_C ratio considered per se without invoking information on long-term population dynamics and estimates for related species is not an immediate cause for concern. On the contrary, it may point towards current-day population levels exceeding numbers of effectively breeding individuals as inscribed in the long-term genetic record of the species. However, residuals in N_e/N_C relative to other species might flag up unaccounted population stratification or admixture[46] and human-mediated population crashes.

The approach used herein is readily applicable to other clades, and requires only three sets of data: estimates of mutation rate, genome-wide genetic diversity and current-day population size estimates. Mutation rate estimates, if not inferred from the target group, can generally be extrapolated from closely related species. Neither should estimation of the genetic parameters be a limitation, as it requires estimation of only two central population genetic parameters, θ_π and Tajima's D, the latter a composite of θ_w and θ_π. Accurate estimates of N_C are arguably the most difficult to obtain. Nevertheless, extracting data from the IUCN database, which is noisy and with high levels of uncertainty[47], still yielded a remarkably high correlation with N_e.

Methods

Species and population sampling. A total of 17 globally distributed pinniped species were collected from across the phylogenetic range[48] (Fig. 1). Samples were chosen to cover the two major pinniped families, Otariidae ($n = 8$) and Phocidae ($n = 8$), in even proportion, and to include the walrus (*O. rosmarus*), which is the only extant member of the family Odobenidae ($n = 1$). Several subspecies or populations were sampled for six of the species. To attribute taxonomic and conservation status, we followed the current classification from IUCN v.2018-1. For the South American fur seal, a population sample from the subspecies *A. australis australis* from Brazil was included, and a population from Peru with contentious subspecific status[49]. The Steller sea lion was represented by the subspecies Loughlin's Steller Sea Lion (*Eumetopias jubatus monteriensis*), and samples from the walrus come from the Atlantic stock *O. rosmarus rosmarus*. Samples from the grey seal were obtained from the subspecies *Halichoerus grypus grypus* in the Western Atlantic, and from *H. g. macrorhynchos* from the United Kingdom. For the ringed seal, three different subspecies from Svalbard (*P. hispida hispida*), the Baltic Sea (*P. h. botnica*) and Lake Saimaa (*P. h. saimensis*) were included. Multiple populations were sampled for *A. forsteri* (four populations) and *Z. californianus* (six populations). In addition, the *Z. wollebaeki* dataset from Shafer et al.[50], consisting of 94 individuals from eight rookeries (93 after filtering in this study), was used, with each rookery analysed separately. See Supplementary Table 6 for details.

Reduced-representation library preparation and data processing. Samples consisted of frozen blood, flipper or muscle tissue extracted using either an UltraClean Blood DNA Isolation Kit (MoBio) or a DNeasy Blood and Tissue Kit (QIAGEN). We generated ddRAD libraries using a modified protocol from Brelsford et al.[51] and described in Shafer et al.[50] using *Mse*I and *Sbf*I restriction enzymes followed by size selection targeting 200–430-bp insert size. Samples were distributed across a total of ten lanes and were 125-bp paired-end sequenced on an Illumina HiSeq2500 machine. After trimming adaptors using cutadapt v.1.8 (ref. [52]), reads were de-multiplexed using the process_radtags module of STACKS[53]. Overlapping reads were merged using PEAR 0.9.6 (ref. [54]) with default parameters, and all merged read-pairs of length >50 bp were retained. Previous work has shown that mapping of restriction site-associated read data to genomes of closely related species yields more reliable genotypes than de novo-based approaches[55], particularly with reference to demographic inference[50]. At the beginning of our study, four pinniped reference genomes were available to us: walrus (*O. rosmarus*, Odobenidae, genome v.1.0, GCF_000321225.1)[56], Antarctic fur seal (*A. gazella*, Otariidae, genome v.1.4. GCA_900500725)[57], Weddell seal (*Leptonychotes weddellii*, Phocidae, genome v.1.0 GCF_000349705.1) and a draft assembly of the California sea lion (*Z. californianus*, genome v.1.2). Reads were mapped to the closest available reference genome using STAMPY 1.0.23 (ref. [58]), with the substitution-rate parameter set based on the distance between each species and its closest relative with a reference genome in the phylogeny of Higdon et al.[48]. Because the northern fur seal, *Callorhinus ursinus*, is equally distant in this phylogeny to both the Antarctic fur seal and the California sea lion, for this species the analysis was repeated using both reference genomes. Results were highly congruent, so only the statistics based on the Antarctic fur seal genome are reported.

Sequence alignment files (.bam) resulting from the mapping procedure were filtered with a wrapper developed for this study, ClustRAD, described in Supplementary Information (see also Extended Data Fig. 7). The resulting sequence clusters were distributed across the entire genome assembly, and thus represent genome-wide genetic variation (Extended Data Fig. 8). To reduce genotype errors mimicking rare alleles with a strong impact on demographic inference, we generated folded-site frequency spectra (SFS) incorporating genotype uncertainty and up to 20% missing data using ANGSD 0.921 (ref. [59]). Due to the difficulty of inferring genotypes in hemizygous sex chromosomes, partially missing sex information and expected discordance in demographic histories between the X chromosome and autosomes, loci mapping to X-chromosomal scaffolds were excluded. X-linked scaffolds were inferred by

synteny to the dog (*Canis familiaris*) genome. Lastz v.1.04 (ref. [60]) was used to align each pinniped reference genome to the dog genome (CanFam3.1) using matches of at least 10 kb (parameters $M=254$, $K=4{,}500$, $L=3{,}000$, $Y=15{,}000$, $C=2$, $T=2$, ambiguous = iupac, matchcount = 10,000). Any scaffolds with matches to the dog X chromosome were classified as putatively X linked and excluded from further analysis. In addition, scaffolds <400 kb, for which power to establish synteny was low, were likewise excluded. We investigated population structure by performing a principal component analysis for each species with NGScovar in NGStools[61,62] using genotype probabilities estimated with ANGSD 0.921 (ref. [59]). Only putative polymorphic ($P<0.001$), bi-allelic sites with a maximum of 20% missing data, a minimum site quality and mapping quality of 20 were included in the analysis.

Estimation of summary statistics. For each population, we derived estimates of nucleotide diversity θ_π (ref. [63]), Watterson's theta θ_w (ref. [64]) and Tajima's D (ref. [65]) from the folded-site frequency spectra as inferred by ANGSD. Because the ddRAD clusters in this dataset are small, often containing no segregating sites, we estimated parameters in two ways. First, we calculated the mean for each scaffold excluding scaffolds with <3,000 cumulative sites located in ddRAD clusters. This approach incorporates variation across the genome. Second, we obtained a single, genome-wide estimate concatenating all ddRAD clusters and using the sum of the per-site estimates of θ_π and θ_w to calculate Tajima's D. The results of both approaches were highly correlated (θ_π, $R^2_{adj}=0.997$; θ_w, $R^2_{adj}=0.998$; Tajima's D, $R^2_{adj}=0.939$), and we proceeded with the first approach. F_{ST} estimates were calculated using the approach used in ref. [66] for species with multiple populations sampled that were sequenced in the same ddRAD libraries.

Under the simplifying assumption of mutation-drift equilibrium, N_e can be inferred from the relationship $N_e=\theta_\pi/4\mu$. Because mutation rates of pinnipeds are unknown, they were inferred from branch-specific substitution rates of putatively neutrally evolving, synonymous sites in coding sequences. We first compiled a set of 11,256 orthologues for the two most closely related outgroup species, dog and panda *(Ailuropoda melanoleuca)*, available from the OrthoMam v.9 database[67]. We then screened GenBank for corresponding orthologues in the three species of pinnipeds for which annotations were available: Weddell seal, Hawaiian monk seal (*Neomonachus schauinslandi*) and walrus. In addition, we extracted 1/1 orthologues for the Antarctic fur seal genome v.1.4 (ref. [57]) and the California sea lion genome v.1.2. To minimize the risk of including paralogues, for each species of pinniped we next performed a reciprocal TBLASTN to the dog reference and retained only sequences with unique hits. We then aligned the obtained sequences using MACSE[68] (with default parameters except for –gap op = 3 and –ext gap ratio = 0.75) and filtered the amino acid alignments using HMMcleaner[69] to remove misaligned regions in sequences, which can heavily impact estimation of substitution rates. We applied a very conservative threshold of a maximum of five consecutive positions differing from the correspondent profiles. Sites filtered in this step were then masked in the nucleotide sequences and, within each alignment, only sequences where >50% of the original sites remained unmasked were retained. We then kept only those alignments in which at most one taxon was missing, which left us with a total of 8,864 alignments.

Separate branch-specific substitution rates were calculated for the phocid clade, represented by the Hawaiian monk seal and Weddell seal, estimated to have diverged 18.2 million years ago (Ma); the otariid clade, represented by the California sea lion and Antarctic fur seal, which diverged from each other around 9.2 Ma; and the odobenid clade, represented by walrus, which diverged from otariids 22 Ma. The divergence between phocids and the clade including the other three pinnipeds has been estimated to have occurred around 27 Ma (Extended Data Fig. 9, modified from ref. [48]). To obtain estimates of neutral substitution rates of synonymous sites, we used a substitution mapping approach[70]. This approach first performed a maximum likelihood optimization of branch lengths and then mapped substitutions with the inferred model parameters using the *bppml* and *mapnh* programmes available in the Program Suite (BppSuite) and testnh libraries of the Bio++ libraries[71]. We additionally performed a correction to take into account the opportunity for mutation using the *kappa* parameter estimated in the previous step[72,73]. Results for all genes were averaged for each species and then reported as substitution rate per site and per year following divergence dates in ref. [48]. For the estimation of N_e, the substitution rate of the branch leading to the closest relative of the target species was chosen, and mutation rate was approximated as the product of the estimated substitution rate per site and per year and the harmonic mean of generation times of all species sampled in the clade (see Supplementary Table 2 and ref. [74]).

Both mutation rate estimates, μ, and estimates of θ_π are subject to sampling variance and systematic variation across the genome. In the absence of positional information of the loci and recombination rate governing the correlation between parameters and neighbouring loci, sensible confidence intervals around N_e cannot be provided. We therefore report only genome-wide averages of N_e based on the mean of θ_π estimates and the point estimate of μ, as described above.

Demographic inference. Inference of demographic history was based on the folded SFS for each species derived as described above, and modelled using a composite likelihood approach implemented in δaδi[40]. Previous works simulating RAD-like genotyping-by-sequencing data demonstrated that demographic parameters are well estimated for simple models using this approach[75] and δaδi can use the SFS directly from ANGSD, avoiding the need for genotyping and reducing bias[76]. Several demographic models were considered (see Fig. 2a):

(1) Constant population size (null model (model 0 P)).
(2) Two-parameter models (2 P) characterizing a single-population size change at time T_1 (in units of $2N_{anc}$ generations). Population size change was modelled either as (a) an instantaneous change ($2P_{ins}$) resulting in rapid population expansion or bottleneck) or (b) gradual exponential change in population size ($2P_{exp}$), gradual growth or contraction). The parameter characterizing population size change was expressed as relative population size after the change N_1 (relative to the ancestral population size N_{anc}). A reduction in population size was indicated if $N_1<1$, and expansion if $N_1>1$. In the case of gradual population size change, change was characterized by an exponential growth parameter.
(3) A three-parameter model including one rapid population change of size N_1 at time T_1, followed by gradual change to the present population size N_2 (3P).
(4) Four-parameter models incorporating two independent population size changes, again considering (a) instantaneous change ($4P_{ins}$) or (b) gradual change following an exponential distribution ($4P_{exp}$). Accordingly, a total of four parameters were required to describe the timing (T_1 and T_2) of population size changes N_1 and N_2 (both relative to the ancestral population size, N_{anc}). Each model was run independently four times, with confidence intervals based on 100 bootstrap replicates. Only the replicate with the highest mean log-likelihood for each model was considered for comparison between models. A statistical test of the goodness of fit between models was performed using likelihood-ratio tests with the R package extRemes v.2.0.8, using a significance level of $P=0.05$ and degrees of freedom equal to the difference in the number of model parameters. Nested models were tested in pairs, starting with the null model (0P) followed by pairs of models with increasing numbers of parameters. If models had the same number of parameters, the model with the highest mean log-likelihood was selected. Parameter estimates are reported for the model with the best fit.

As a complement to the above-mentioned approach, we used PSMC' (ref. [36]) to estimate changes in effective population size over time for all species where both ddRAD and whole-genome sequencing data were available. PSMC' was run on the whole-genome shotgun sequencing data generated for each of the pinniped reference genome assemblies, including the Antarctic fur seal (SRR2658532–SRR2658561), California sea lion (10.5281/zenodo.3741488) and walrus (SRR575502–SRR575518). We mapped the short-read data to their respective reference genome (versions as above for ddRAD analyses) using bwa[77] and called single-nucleotide polymorphisms with Samtools mpileup/bcftools v.1.3 (ref. [78]), with a minimum quality score of 20. Scaffolds inferred to be X-linked were excluded. Regions with low mapping success were masked using a mask generated with SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml), as were areas with excessively low or high coverage when the input files were formatted using scripts from the MSMCtools repository (https://github.com/stschiff/msmc-tools). Each analysis was performed for a single individual using unphased data (PSMC' approach). The substitution rates calculated above and generation times were used to rescale both time and effective population sizes. Long-term effective population size was calculated as the harmonic mean of population sizes inferred at each coalescent event.

Statistical inference. We were interested in assessing the relationship between the genetic variation and other biological traits of the species. We limited our considerations to explanatory variables for which estimates were available for all species, and compiled the following information.

(1) Estimates of the current census population size (N_C) of species and subspecies were extracted from the IUCN website (https://www.iucn.org/ accessed 15 January 2019)[74]. These numbers mostly represent the number of mature individuals. Where ranges were presented, we chose the midpoint. IUCN population estimates were not available for *Erignathus barbatus*, but past estimates put the total species at 700,000, and *E. b. barbatus*, the subspecies that we have sampled, at 250,000 individuals[79].
(2) Similarly, for the assessment of conservation status we followed the classification of the IUCN[74]. Following Stoffel et al.[22], we grouped species listed as 'near threatened', 'vulnerable' or 'endangered' into a 'concern' category and contrasted them to species of 'least concern'.
(3) Data on species-specific life-history traits were extracted from ref. [28]. Variables with right-skewed distributions were transformed logarithmically. Several of the predictor variables were strongly intercorrelated (Extended Data Fig. 10), which is expected to lead to suppression effects and make problematic the interpretation of regression coefficients of linear models. Moreover, the number of predictor variables was high relative to sample size, given by the number of populations or species. We therefore eliminated intercorrelated predictor variables with high Pearson regression coefficients

a priori to reduce the effect of collinearity and to avoid overfitting models. We kept the variables with presumed lower measurement error and more immediate biological relevance. This led to the following four predictor variables:

(a) Harem size rather than mean sexual size dimorphism ($r=0.91$) as a more direct measure of the mating system
(b) Length of reproductive season rather than breeding latitude ($r=-0.61$) and length of lactation period ($r=0.47$)
(c) Species distribution range rather than breeding latitude ($r=0.77$); because the former was significantly associated with Tajima's *D*, it was replaced by the latter in models in which Tajima's *D* was used as a predictor variable
(d) Breeding habitat (ice or land), which has been shown to be an important predictor of recent human-induced population size changes[22].

Linear models exploring the relationship between genetic and predictor variables were fitted using R[80]. We fitted all possible models with a maximum of two predictor variables. Model selection was based on Akaike's information criterion for small sample sizes (AIC_c), which provides information about the goodness of fit of the respective model to the observed data. Following common convention, models with $\Delta AIC_c < 2$ were regarded as having statistical support. We further provide for each model information-theory based model selection statistics (AIC_c, ΔAIC_c and $wAIC_c$) and R^2 adjusted for the number of parameters. The sum of Akaike weights ($\Sigma wAIC_c$, ranging from 0 to 1) was used to estimate and compare the relative contribution of explanatory parameters. Different numbers of populations were sampled for each species and, in some cases, populations or subspecies from the same species were sequenced in different ddRAD libraries. To mitigate the effects of this, only the population that best represents the core species range was used for model fitting (Extended Data Fig. 5).

The statistical model employed assumes independence of data points. To test for possible effects of genealogical non-independence, we also conducted all correlations in a phylogenetically explicit framework using the package nlme v.3.1–131.1. (refs. [80,81]). This did not qualitatively change the results.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated for this study are archived in the sequence read archive under bioproject no. PRJEB37019 at the National Centre of Biotechnology Information (www.ncbi.nlm.nih.gov/sra). For individual accession numbers see also Supplementary Table 6. All code used for the analyses and the alignments used to infer substitution rates are available at 10.5281/zenodo.3741488.

Received: 17 May 2019; Accepted: 28 April 2020;
Published online: 08 June 2020

References

1. Ceballos, G. et al. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015).
2. Secretariat of the Convention on Biological Diversity *Global Biodiversity Outlook 4* (World Trade Centre, 2014).
3. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
4. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
5. Frankham, R. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* **66**, 95–107 (1995).
6. Leffler, E. M. et al. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
7. Lewontin, R. C. *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
8. Ellegren, H. & Galtier, N. Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433 (2016).
9. Lynch, M. et al. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016).
10. Cutter, A. D. & Payseur, B. A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* **14**, 262–274 (2013).
11. Corbett-Detig, R. B., Hartl, D. L. & Sackton, T. B. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* **13**, e1002112 (2015).
12. Coop, G. Does linked selection explain the narrow range of genetic diversity across species? Preprint at *bioRxiv* https://doi.org/10.1101/042598 (2016).
13. Romiguier, J. et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014).
14. Ferchaud, A.-L. et al. Making sense of the relationships between *Ne*, *Nb* and *Nc* towards defining conservation thresholds in atlantic salmon (*Salmo salar*). *Heredity* **117**, 268–278 (2016).
15. Lande, R. Genetics and demography in biological conservation. *Science* **241**, 1455–1460 (1988).
16. Palstra, F. P. & Fraser, D. J. Effective/census population size ratio estimation: a compendium and appraisal. *Ecol. Evol.* **2**, 2357–2365 (2012).
17. Waples, R. S. Making sense of genetic estimates of effective population size. *Mol. Ecol.* **25**, 4689–4691 (2016).
18. Palstra, F. P. & Ruzzante, D. E. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol. Ecol.* **17**, 3428–3447 (2008).
19. Waples, R. S., Luikart, G., Faulkner, J. R. & Tallmon, D. A. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proc. R. Soc. B* **280**, 20131339 (2013).
20. Leroy, G. et al. Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. *Genet. Sel. Evol.* **45**, 1 (2013).
21. Wang, J. Estimation of effective population sizes from data on genetic markers. *Phil. Trans. R. Soc. B* **360**, 1395–1409 (2005).
22. Stoffel, M. A. et al. Demographic histories and genetic diversity across pinnipeds are shaped by human exploitation, ecology and life-history. *Nat. Commun.* **9**, 4836 (2018).
23. Tenesa, A. et al. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* **17**, 520–526 (2007).
24. Nordborg, M. & Krone, S. M. in *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot* (eds Slatkin, M. & Veuille, M.) 194–232 (Oxford Univ. Press, 2002).
25. Vijay, N. et al. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Mol. Ecol.* **26**, 4284–4295 (2017).
26. Wakeley, J. *Coalescent Theory: An Introduction* (W. H. Freeman, 2008).
27. Charlesworth, B. & Jain, K. Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. *Genetics* **198**, 1587–1602 (2014).
28. Krüger, O., Wolf, J. B. W., Jonker, R. M., Hoffman, J. I. & Trillmich, F. Disentangling the contribution of sexual selection and ecology to the evolution of size dimorphism in pinnipeds. *Evolution* **68**, 1485–1496 (2014).
29. de Oliveira, L. R., Meyer, D., Hoffman, J., Majluf, P. & Morgante, J. S. Evidence of a genetic bottleneck in an El Niño affected population of South American fur seals, *Arctocephalus australis*. *J. Mar. Biol. Assoc. U.K.* **89**, 1717–1725 (2009).
30. Soto, K. H., Trites, A. W. & Arias-Schreiber, M. The effects of prey availability on pup mortality and the timing of birth of South American sea lions (*Otaria flavescens*) in Peru. *J. Zool.* **264**, 419–428 (2004).
31. Kovacs, K. M. et al. Global threats to pinnipeds. *Mar. Mammal Sci.* **28**, 414–436 (2012).
32. Scally, A. The mutation rate in human evolution and demographic inference. *Curr. Opin. Genet. Dev.* **41**, 36–43 (2016).
33. Takahata, N. Allelic genealogy and human evolution. *Mol. Biol. Evol.* **10**, 2–22 (1993).
34. Brüniche-Olsen, A. et al. The inference of gray whale (*Eschrichtius robustus*) historical population attributes from whole-genome sequences. *BMC Evol. Biol.* **18**, 87 (2018).
35. Nei, M. & Takahata, N. Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* **37**, 240–244 (1993).
36. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919 (2014).
37. Mazet, O., Rodríguez, W., Grusea, S., Boitard, S. & Chikhi, L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity* **116**, 362–371 (2016).
38. Andersen, L. W. et al. Walruses *(Odobenus rosmarus rosmarus*) in the Pechora Sea in the context of contemporary population structure of Northeast Atlantic walruses. *Biol. J. Linn. Soc.* **122**, 897–915 (2017).
39. Kalinowski, S. T. & Waples, R. S. Relationship of effective to census size in fluctuating populations. *Conserv. Biol.* **16**, 129–136 (2002).
40. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
41. Nyman, T. et al. Demographic histories and genetic diversities of Fennoscandian marine and landlocked ringed seal subspecies. *Ecol. Evol.* **4**, 3420–3434 (2014).
42. Chikhi, L., Sousa, V. C., Luisi, P., Goossens, B. & Beaumont, M. A. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186**, 983–995 (2010).
43. Mackintosh, A. et al. The determinants of genetic diversity in butterflies. *Nat. Commun.* **10**, 3466 (2019).
44. Wright, S. Size of population and breeding structure in relation to evolution. *Science* **87**, 430–431 (1938).
45. Slatkin, M. Gene genealogies within mutant allelic classes. *Genetics* **143**, 579–587 (1996).

46. Lancaster, M. L., Gemmell, N. J., Negro, S., Goldsworthy, S. & Sunnucks, P. Ménage à trois on Macquarie Island: hybridization among three species of fur seal (*Arctocephalus* spp.) following historical population extinction. *Mol. Ecol.* **15**, 3681–3692 (2006).
47. Akcakaya, H. R. et al. Making consistent IUCN classifications under uncertainty. *Conserv. Biol.* **14**, 1001–1013 (2000).
48. Higdon, J. W., Bininda-Emonds, O. R. P., Beck, R. M. D. & Ferguson, S. H. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol. Biol.* **7**, 216 (2007).
49. de Oliveira, L. R. & Brownell, R. L. Taxonomic status of two subspecies of South American fur seals: *Arctocephalus australis australis* vs. *A. a. gracilis*. *Mar. Mammal Sci.* **30**, 1258–1263 (2014).
50. Shafer, A. B. A. et al. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* **8**, 907–917 (2017).
51. Brelsford, A., Dufresnes, C. & Perrin, N. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity* **116**, 177–181 (2016).
52. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
53. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
54. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
55. Nadeau, N. J. et al. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res.* **24**, 1316–1333 (2014).
56. Foote, A. D. et al. Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).
57. Humble, E. et al. RAD sequencing and a hybrid Antarctic fur seal genome assembly reveal rapidly decaying linkage disequilibrium, global population structure and evidence for inbreeding. *G3 (Bethesda)* **8**, 2709–2722 (2018).
58. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
59. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
60. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA* (The Pennsylvania State Univ., 2007).
61. Fumagalli, M., Vieira, F. G., Linderoth, T. & Nielsen, R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–1487 (2014).
62. Fumagalli, M. et al. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* **195**, 979–992 (2013).
63. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273 (1979).
64. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
65. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
66. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583 (1992).
67. Ranwez, V. et al. OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol. Biol.* **7**, 241 (2007).
68. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS ONE* **6**, e22594 (2011).
69. Di Franco, A., Poujol, R., Baurain, D. & Philippe, H. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol. Biol.* **19**, 21 (2019).
70. Romiguier, J. et al. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PLoS ONE* **7**, e33852 (2012).
71. Guéguen, L. et al. Bio++: efficient extensible libraries and tools for computational molecular evolution. *Mol. Biol. Evol.* **30**, 1745–1750 (2013).
72. Figuet, E. et al. Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Mol. Biol. Evol.* **33**, 1517–1527 (2016).
73. Botero-Castro, F., Figuet, E., Tilak, M.-K., Nabholz, B. & Galtier, N. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. *Mol. Biol. Evol.* **34**, 3123–3131 (2017).
74. *The IUCN Red List of Threatened Species. Version 2017-3* (IUCN, 2017).
75. Shafer, A. B. A., Gattepaille, L. M., Stewart, R. E. A. & Wolf, J. B. W. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. *Mol. Ecol.* **24**, 328–345 (2015).
76. Warmuth, V. M. & Ellegren, H. Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. *Mol. Ecol. Resour.* **19**, 586–596 (2019).
77. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
78. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
79. *Bearded Seal* (Greenland Institute of Natural Resources, 2018); http://www.natur.gl/en/birds-and-mammals/marine-mammals/bearded-seal/
80. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).
81. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. nlme: Linear and nonlinear mixed effects models. R package version 3.1 (2017).

Acknowledgements

Funding and research support were provided by The Royal Physiographic Society in Lund (to A.B.A.S. and J.B.W.W.), Swedish Research Council Formas (no. 231-2012-450 to J.B.W.W.), the Natural Science and Engineering Research Council of Canada (to A.B.A.S.), the Wenner-Gren Foundation (to A.B.A.S.), the Royal Swedish Academy of Sciences (to A.B.A.S.), the German Research Foundation (no. HO 5122/3-1 to J.I.H. and J.B.W.W.) and LMU Munich (to J.B.W.W.). SciLifeLab Uppsala provided sequencing support. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project no. SNIC 2018-3-658. This work contributes to the Ecosystems project of the British Antarctic Survey, Natural Environmental Research Council, and is part of the Polar Science for Planet Earth Programme. We thank F. Gulland for sample collection.

Author contributions

Study design was performed by J.B.W.W., A.B.A.S. and C.R.P. Laboratory work was done by A.B.A.S. and C.-C.W. Data analysis was carried out by C.R.P., S.T., S.D.P., F.B.-C., J.B.W.W. and A.B.A.S. Samples were provided by A.B.B., A.J.O., C.L., D.A.-G., F.G., J.W.B., J.F., J.I.H., J.B.W.W., K.M.K., L.R.O., M.K., M.V., N.J.G., S.S. and T.N. The initial manuscript draft was written by C.R.P., A.B.A.S. and J.B.W.W. All authors edited the manuscript, contributed to interpretation of results and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the US Government, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. Additionally, the findings and conclusions in the paper are those of the author(s) and do not necessarily represent the views of the National Marine Fisheries Service, NOAA.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-020-1215-5.

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-1215-5.

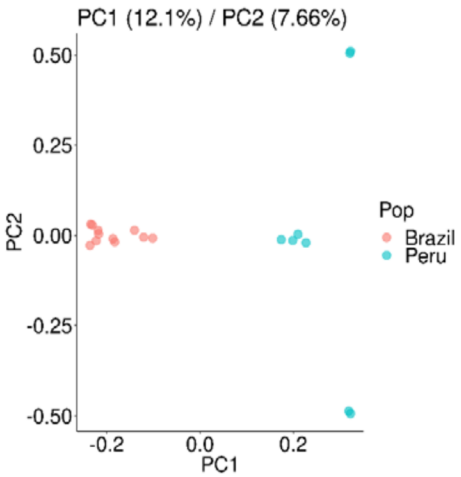
Correspondence and requests for materials should be addressed to A.B.A.S. or J.B.W.W.

Reprints and permissions information is available at www.nature.com/reprints.

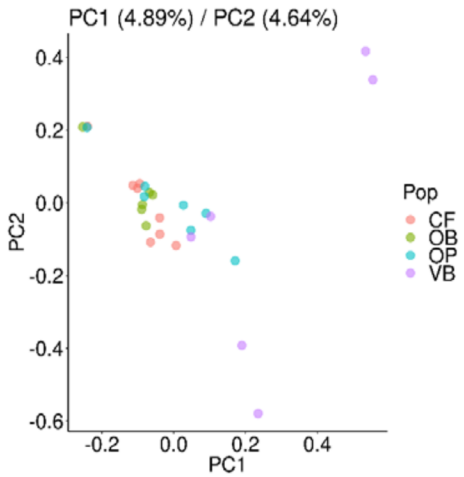
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

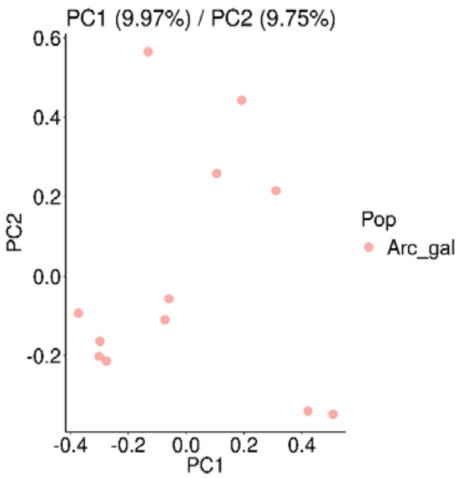
Arctocephalus australis



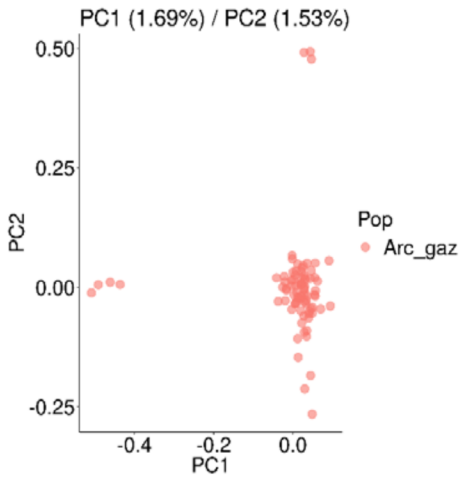
Arctocephalus forsteri



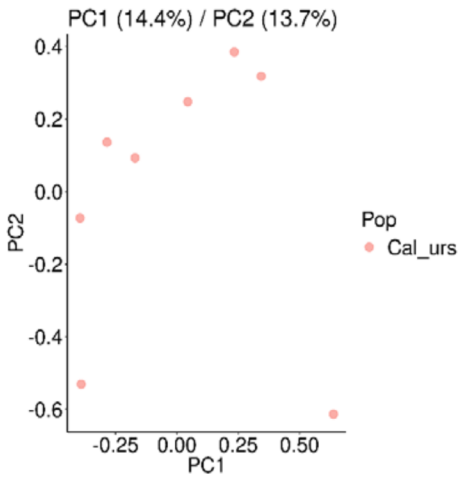
Arctocephalus galapagoensis



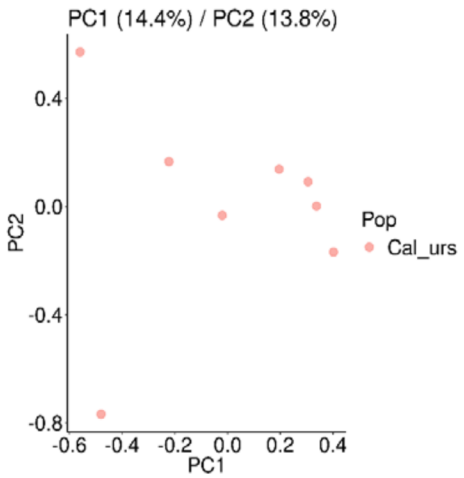
Arctocephalus gazella



Callorhinus ursinus (mapped to ArcGaz)

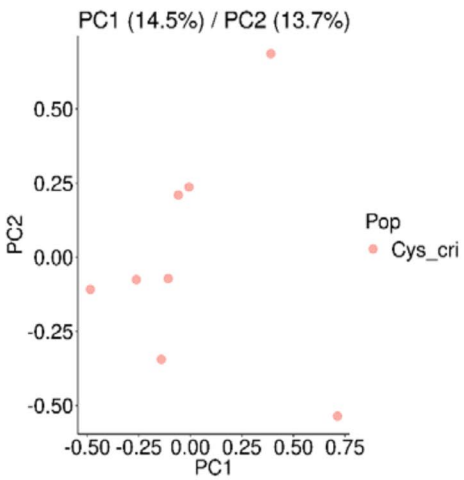


Callorhinus ursinus (mapped to ZalCal)

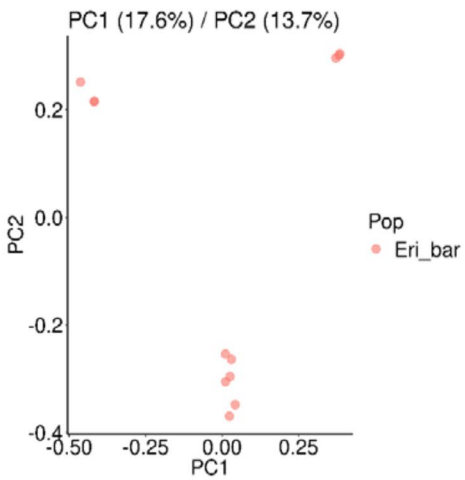


Extended Data Fig. 1 | Principal component analysis (part1). Principal component analysis (first two axes) for Arctocephalus australis, Arctocephalus forsteri, Arctocephalus galapagoensis, Arctocephalus gazella and Callorhinus ursinus.

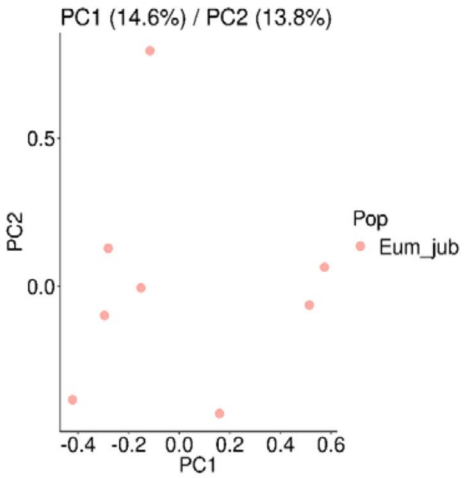
Cystophora cristata



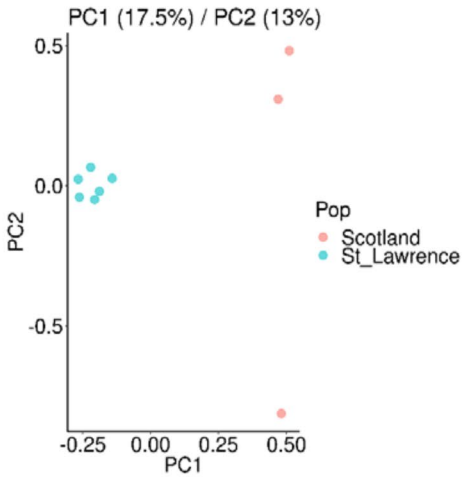
Erignathus barbatus



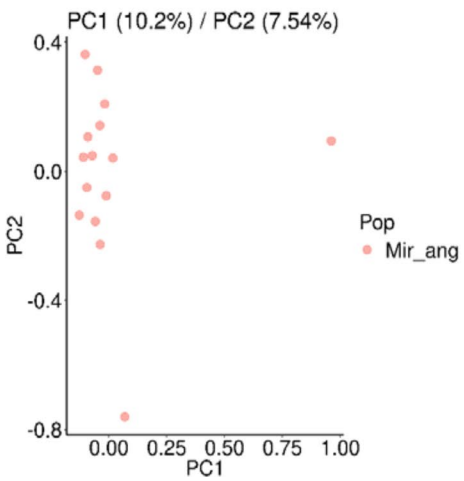
Eumetopias jubatus



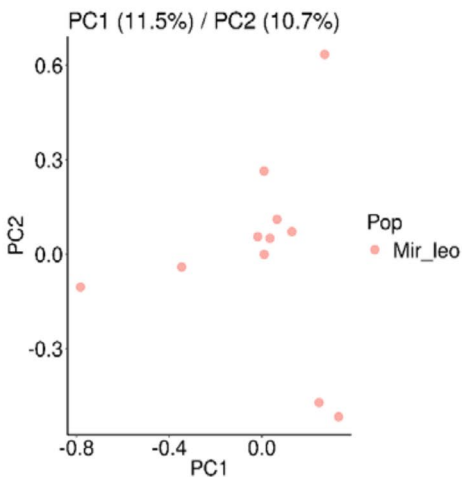
Halichoerus grypus



Mirounga angustirostris

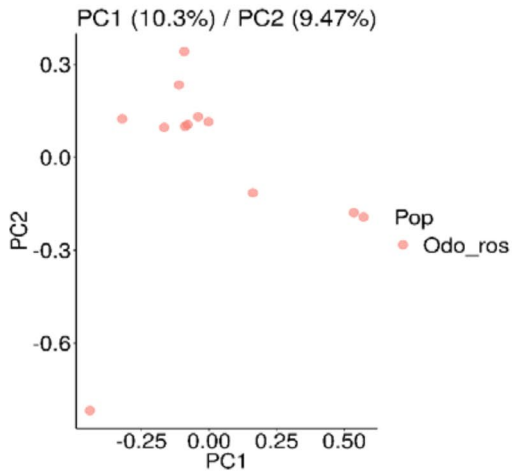


Mirounga leonina

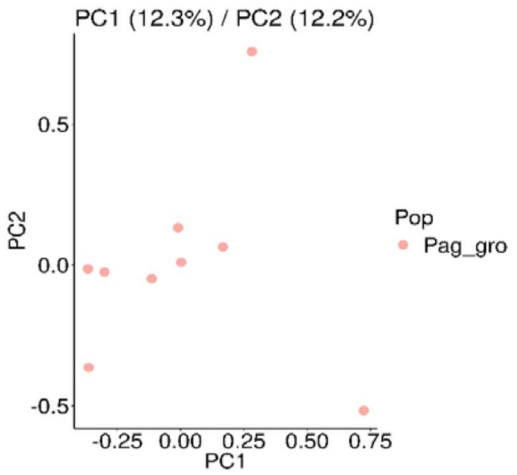


Extended Data Fig. 2 | Principal component analysis (part2). Principal component analysis (first two axes) for Cystophora cristata, Erignathus barbatus, Eumetopias jubatus, Halichoerus grypus, Mirounga angustirostris and Mirounga leonina.

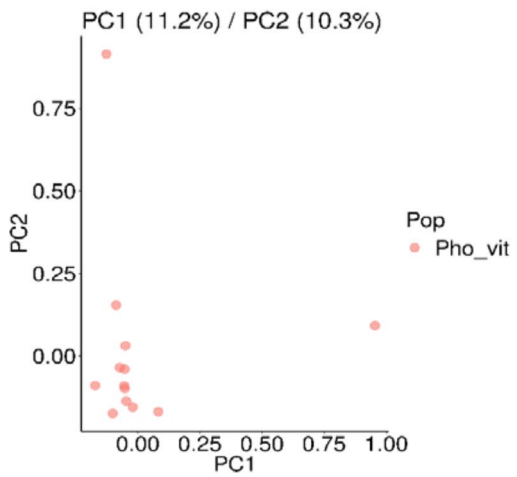
Odobenus rosmarus



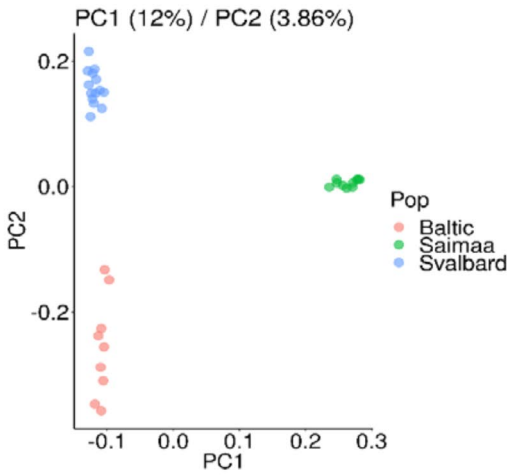
Pagophilus groenlandicus



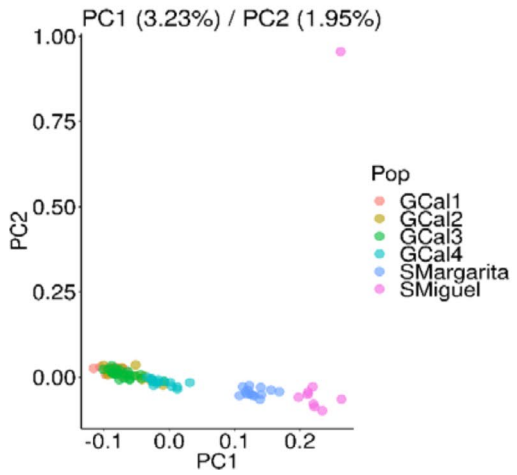
Phoca vitulina



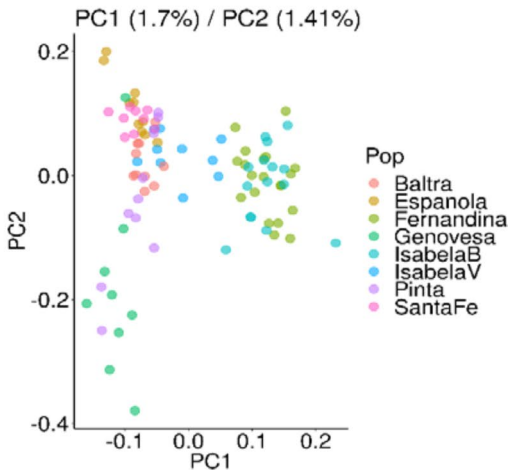
Pusa hispida



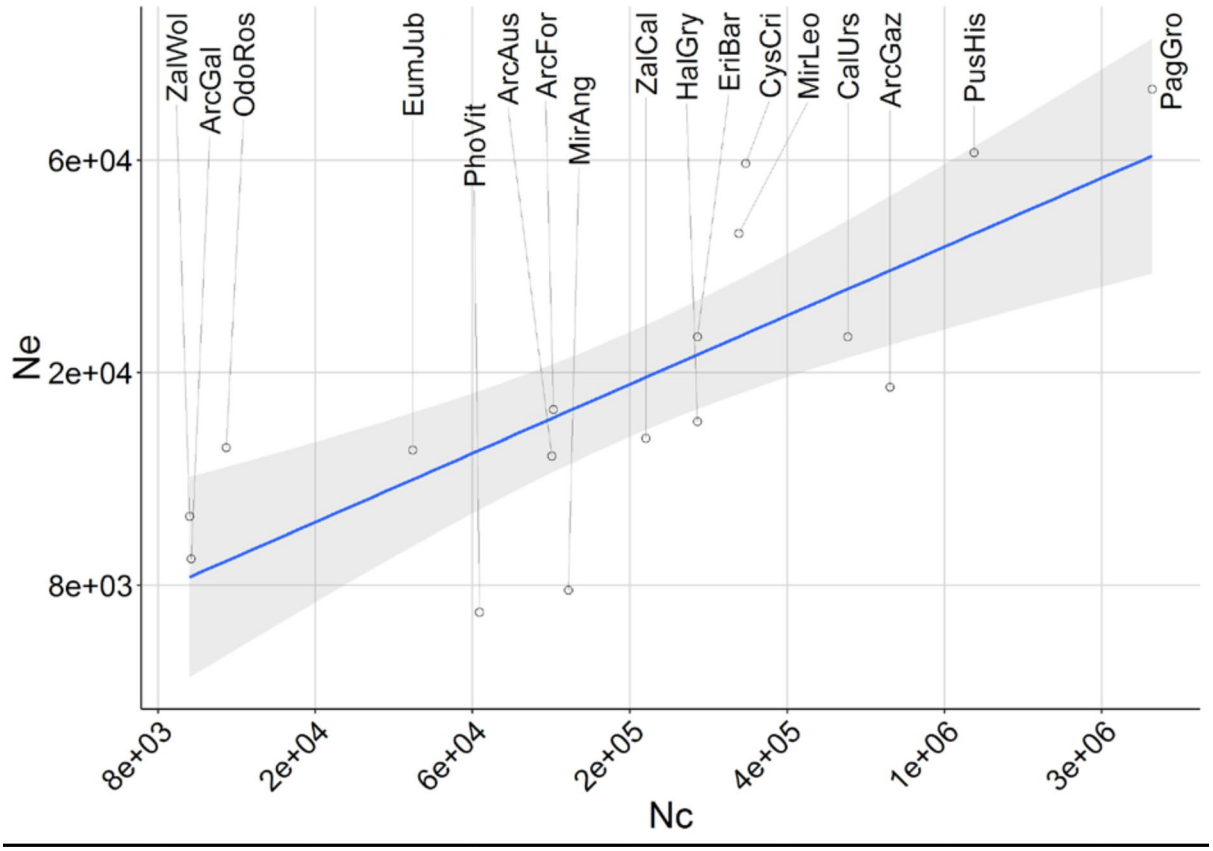
Zalophus californianus



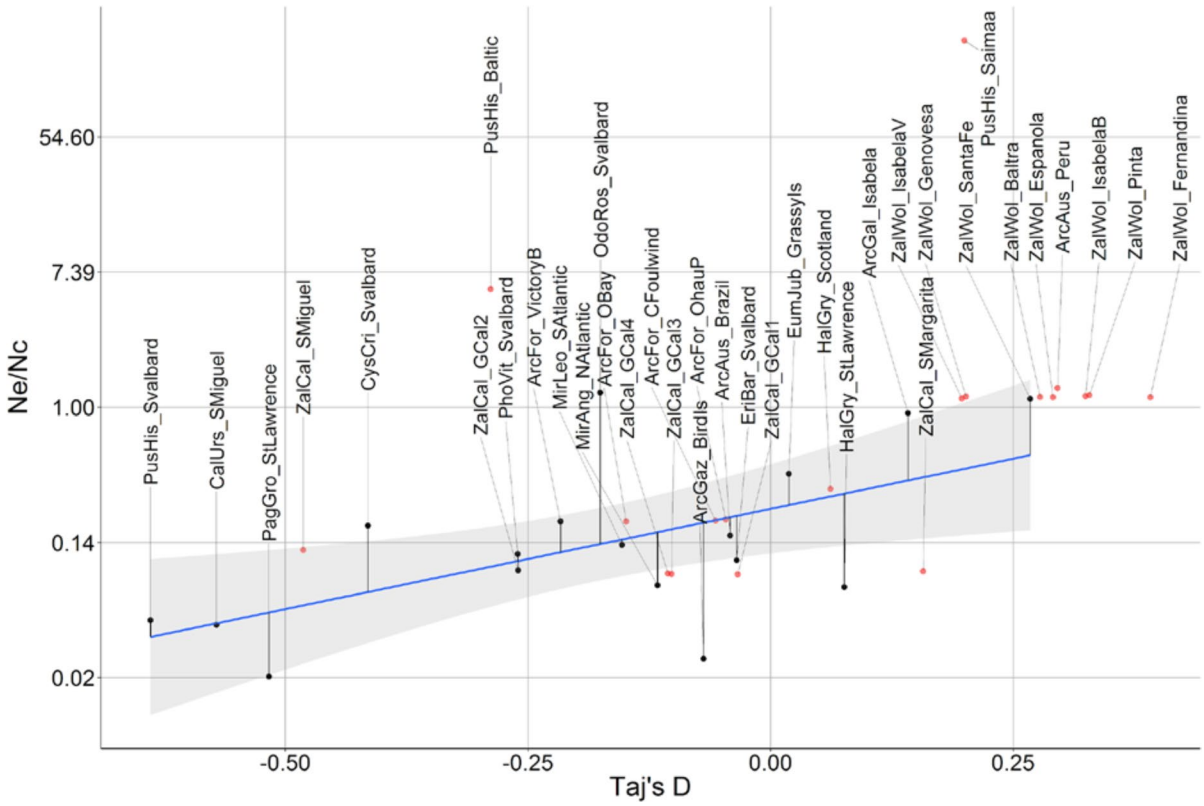
Zalophus wollebaeki



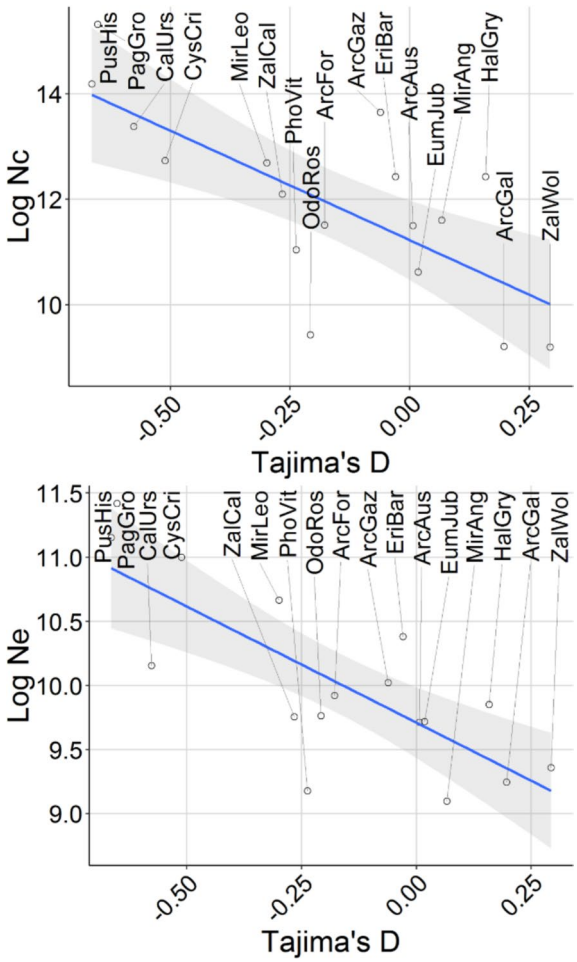
Extended Data Fig. 3 | Principal component analysis (part3). Principal component analysis (first two axes) for Odobenus rosmarus, Pagophilus groenlandicus, Phoca vitulina, Pusa hispida, Zalophus californianus and Zalophus wollebaeki.



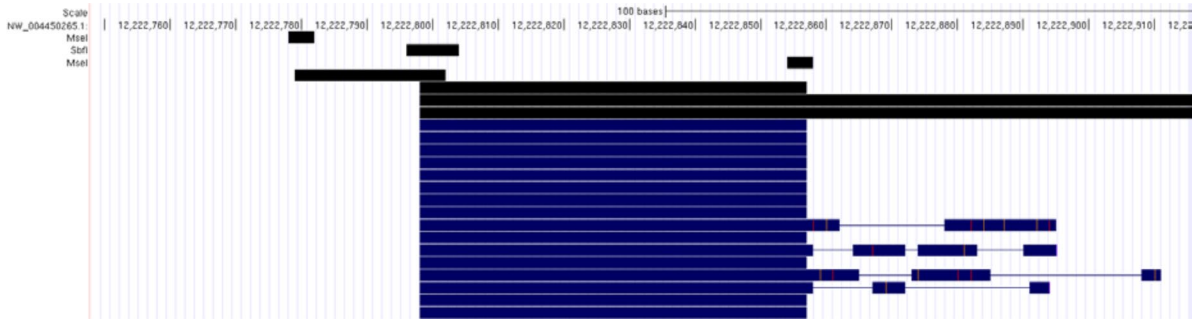
Extended Data Fig. 4 | Relationship between effective population size and census population size. Relationship between effective population size (N_e) and census population size (N_C) per species as in Fig. 1c (main text) including species labels. Blue line: lineal regression line; shade: 95% confidence interval of regression. The y-axis follows a natural logarithmic scale.



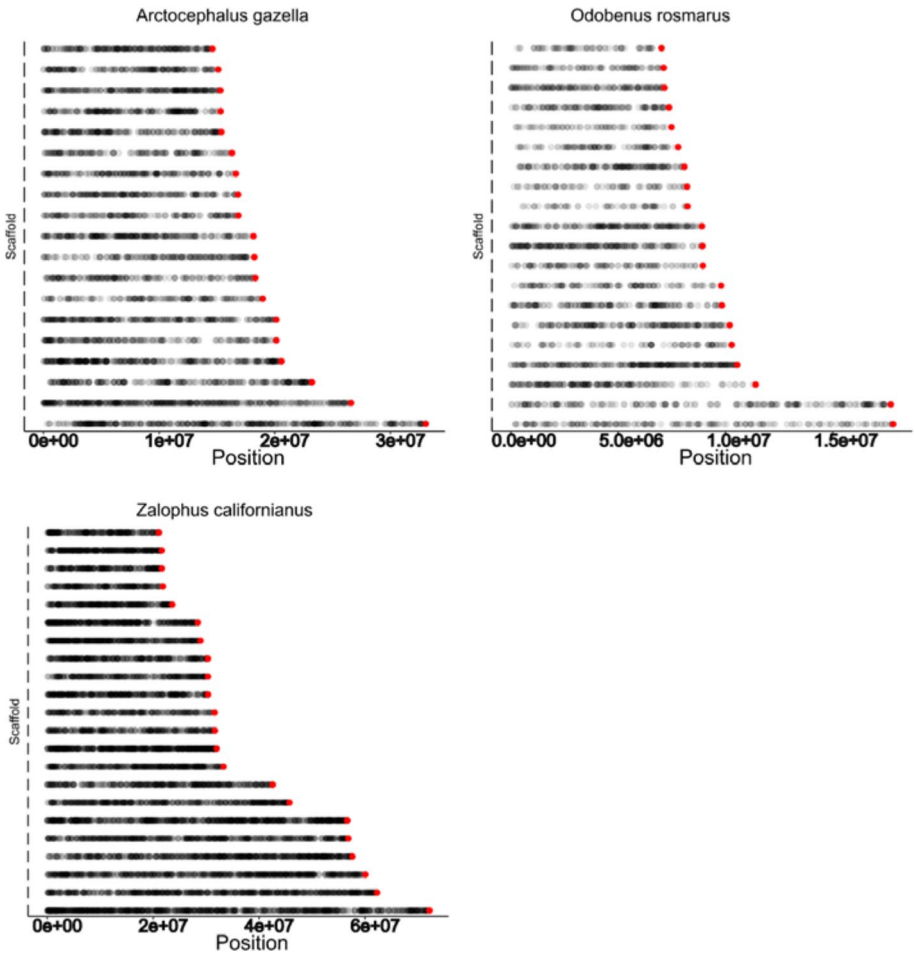
Extended Data Fig. 5 | The ratio of effective population size and census population size in relation to Tajima's *D*. The ratio of effective population size and census population size in relation to Tajima's D representing the species' demographic history; as in Fig. 4a (main text) including species and population labels. Note that the lineal regression (blue line) only refers to the population chosen to represent the species (black points) and the depicted relationship without considering the influence of other predictor variables. 95% confidence interval of the lineal regression is shown in shade. The y-axis follows a natural logarithmic scale.



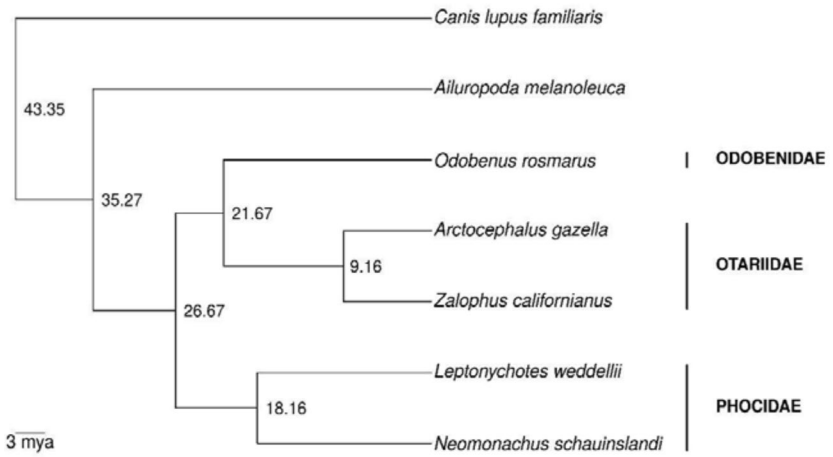
Extended Data Fig. 6 | Census population size (N_c) and population size (N_e) in relation to Tajima's *D* per species. Census population size (N_c) and population size (N_e) in relation to Tajima's D per species (top and bottom panels respectively). The natural logarithm has been used for transformation of N_e and N_c values. Blue line: lineal regression line; shade: 95% confidence interval of regression. Note that the Galápagos seals (ArcGal, ZalWol) lie close to the expected relationship between Tajima's D and N_e, whereas they show strong negative residuals for the relationship with N_c. This is consistent with a very recent reduction in census population size contributing to the high N_e/N_c ratio above one.



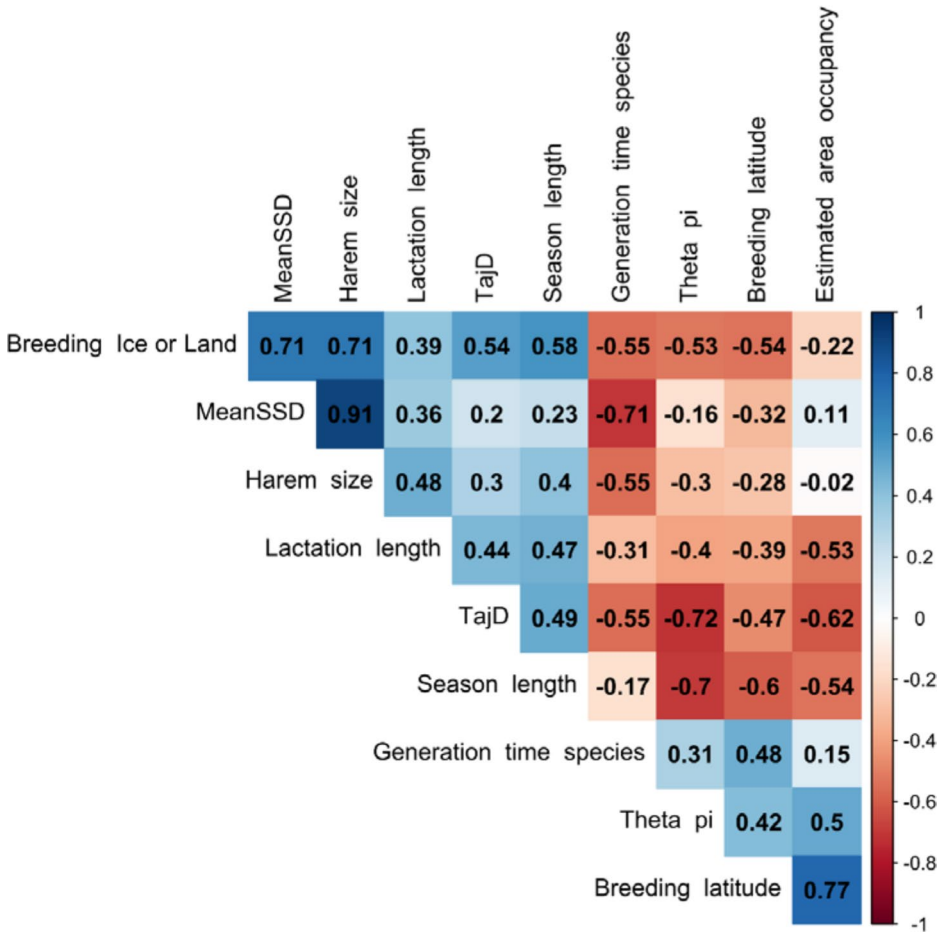
Extended Data Fig. 7 | Examples of problems identified from visual inspection of the mapped ddRAD reads. Screenshot examples of mapped reads displayed in the UCSC browser resulting from a ddRAD library constructed using MseI and SbfI. Black bars joining the SbfI to MSeI cut sites are potential clusters. Blue bar show aligned reads. The extension of reads beyond the MSeI cut site are the 'ghost' PCR extensions with higher visible mismatches and indels.



Extended Data Fig. 8 | Distribution of ddRAD loci across the reference genomes. Distribution of ddRAD loci (black points) across the largest 20 scaffolds in species with both reference genome and ddRAD data. Red points correspond to the end of scaffold. For Zalophus californianus data correspond to population SMargarita.



Extended Data Fig. 9 | Phylogeny of species used to estimate μ. Phylogenetic relationships of five pinniped species (Arctocephalus gazella, Leptonychotes weddellii, Neomonachus schauinslandii, Odobenus rosmarus, Zalophus californianus) and the outgroups dog (Canis lupus familiaris) and panda (Ailuropoda melanoleuca). The phylogeny was obtained by pruning the tree of Higdon et al.[48] to include only the five species above. Branch lengths are proportional to divergence times, numbers next to nodes show estimated node ages.



Extended Data Fig. 10 | The collinearity of life-history traits extracted from ref. [28]. Correlogram of the explanatory variables illustrating the degree of collinearity.

nature research

Corresponding author(s): Aaron Shafer, Jochen Wolf

Last updated by author(s): 03/10/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The sequencing data was generated by us, life history data was extracted from the literature. No software was used to collect the data.
Data analysis	All software and versions used for data pre-processing and downstream analyses are mentioned in the methods section and the supplementary material. Proprietary code is deposited on github and the respective links are provided in the text.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated for this study are archived under bioproject PRJEB37019 in NCBI's sequencing read archive (SRA). See table S6 for detailed information on accession numbers by sample.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Estimation of genetic variation and its relationship to species abundance, life history parameters, and conservation status as described in detail in the methods section.
Research sample	Tissue samples from individuals of pinniped species from across the globe. All samples were available previously and no additional field work was done for the purpose of this study.
Sampling strategy	We aimed for a minimum of 10 chromosomes to obtain a reliable estimate of genetic variation in the population for 17 different species (as e.g. used for simulations in Carling & Brumfield (2007)). Where possible, several populations were included to test for robustness of 'species' genetic variation estimates. 1. M. D. Carling, R. T. Brumfield, Gene Sampling Strategies for Multi-Locus Population Estimates of Genetic Diversity (θ). PLoS ONE. 2, e160 (2007).
Data collection	From each sample, we extracted DNA and generated ddRAD data as made available in the SRA archive of NCBI. Published and unpublished genomes were acquired from identified sources.
Timing and spatial scale	As specified above, the samples have been obtained as part as individual studies from a large group of researchers. They were obtained in different batches and ddRAD data was generated accordingly. Batch effects have been explicitly considered. Once our goal of >15 species was reached we started analyzing the genetic data altogether.
Data exclusions	Data were excluded according to data quality as specified in detail in the methods section.
Reproducibility	We explicitly included several populations for each species to test for reproducibility of estimates among populations with shallow differentiation. As a precursor study for this manuscript, we (i.e. Shafer et al.) ran a series of tests to assess robustness for population genetic estimates from ddRAD data. 1. A. B. A. Shafer et al., Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol. 8, 907–917 (2017).
Randomization	All statistical analyses were also performed controlling for phylogenetic relationships. Demographic analyses were subjected to bootstrapping.
Blinding	Library preparation were done by lab staff who had no insight into the study setup. For the analyses standardized bioinformatic pipelines have been used on all samples excluding human bias.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Samples were obtained as part of independent research programs unrelated to this study. This being a global collection of samples conditions varied across species.
Location	see Table S6.
Access and import/export	Sampling and sample import was conducted in accordance with local legal requirements.
Disturbance	Disturbance includes stress to the animals during capture and biometric measurements, and possible pain during tissue sampling. Restraining was reduced to a time minimum, and tissue samples were reduced in size allowing for sufficient DNA extraction to reduce pain to a minimum.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	n.a.
Wild animals	see Supplementary Table 6.
Field-collected samples	Tissue samples were cryopreserved prior to extraction of DNA.
Ethics oversight	Ethical approval only relates to local sampling which was conducted prior to this study. Sampling permissions are specified in Supplementary Table 2.

Note that full information on the approval of the study protocol must also be provided in the manuscript.