Rapid increase in southern elephant seal genetic diversity after a founder event

Mark de Bruyn[1], Malin Pinsky[2], Brenda Hall[3], Paul Koch[4], Carlo Baroni[5], A. Rus Hoelzel[6]

1. Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Deiniol Road, Bangor, UK
2. Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, New Jersey, USA
3. Climate Change Institute and School of Earth and Climate Sciences, University of Maine, Orono, Maine, USA
4. Department of Earth and Planetary Sciences, University of California, Santa Cruz, California, USA
5. Dipartimento di Scienze della Terra, Università di Pisa, Pisa, Italy
6. School of Biological and Biomedical Sciences, University of Durham, Durham, UK

Corresponding authors: Mark de Bruyn, Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Deiniol Road, Bangor LL57 2UW, UK Fax: +44 1248 370731, email: markus.debruyn@gmail.com

A. Rus Hoelzel, School of Biological and Biomedical Sciences, University of Durham, Durham DH1 3LE, UK. Fax: +44 191 3341201, email: a.r.hoelzel@durham.ac.uk

Summary

Genetic diversity provides the raw material for populations to respond to changing environmental conditions. The evolution of diversity within populations is based on the accumulation of mutations and their retention or loss through selection and genetic drift, while migration can also introduce new variation. However, the extent to which population growth and sustained large population size can lead to rapid and significant increases in diversity has not been widely investigated. Here we assess this empirically by applying Approximate Bayesian Computation to a novel ancient DNA dataset that spans the life of a southern elephant seal (*Mirounga leonina*) population, from initial founding ~7000 years ago, to eventual extinction within the past millenium. We find that rapid population growth and sustained large population size can explain substantial increases in population genetic diversity over a period of several hundred generations, subsequently lost when the population went to extinction. Results suggest that the impact of diversity introduced through migration was relatively minor. We thus demonstrate, by examining genetic diversity across the life of a population, that environmental change could generate the raw material for adaptive evolution over a very short evolutionary time scale through rapid establishment of a large, stable population.

Key words (3-6): ancient DNA, approximate Bayesian computation, extinction, founder event, genetic diversity, population growth.

Introduction

Understanding how micro-evolutionary processes lead to changes in evolutionary potential in the form of genetic diversity is a fundamental goal of population genetics. Population genetic diversity can be gained either *in situ* through mutation, through immigration, or some combination of the two (Maruyama 1970, Nei & Feldman 1972). Variation in the form of new mutations is expected to enter a population at a rate of $2\ N_e\ u$, where N_e is the effective population size and u the neutral mutation rate per generation, while genetic drift decays variation at the rate of $1/2\ N_e$ (Kimura 1983). New mutations within a population can either be fixed, or more commonly lost (Kimura and Ohta 1969, Crow and Kimura 1970). Fixation time of new unlinked neutral mutations within a randomly mating population is strongly reliant on N_e, taking some $4N_e$ generations when N_e is constant (Kimura and Ohta 1969, 1973). However, change in population size, a fundamental characteristic of many wild populations, can significantly impact on this timeframe (Kimura and Ohta 1969, 1973). Changes in genetic diversity can also result from the movement of individuals (migration) and their genes (gene flow) into or out of a population. Thus, rapid changes in population genetic diversity may result from demographic processes such as immigration, or *in situ* processes such as mutation and genetic drift. These processes will also be influenced by changes in N_e, which may result from varying physical conditions, resource availability, habitat availability, predator density, human impacts (Otto & Whitlock 1997), or skewed effective breeding ratios.

While theoretical models predict changes in levels of genetic diversity within a population over time, it is uncommon to be able to track such changes in vertebrate populations empirically, particularly over timeframes longer than that of pedigree studies. However, molecular analyses of ancient DNA allow these processes to be assessed more fully, especially

63 when ancient DNA samples span the life of the population. Here we investigate such a dataset,
64 provided by the founding and subsequent extinction of a colony of southern elephant seals. This
65 population existed along the Victoria Land Coast (VLC), Antarctica for much of the Holocene,
66 resulting from the expansion of suitable habitat due to altered environmental conditions (Hall et
67 al. 2006, de Bruyn et al. 2009). Our study is inspired by the observation that this short-lived
68 population had apparently accumulated extensive novel variation, most of which was lost when
69 the population went to extinction (de Bruyn et al. 2009), but the relative contribution of
70 migration and *in situ* mutation to this increase in genetic diversity remains unknown.
71 Radiocarbon dates from 223 individual seal remains provided upper- and lower-bound
72 age estimates on the population of 7,200 – 270 years before present (YBP; Hall et al. 2006, de
73 Bruyn et al. 2009). Subsequent analyses utilizing a southern elephant seal-specific substitution
74 rate estimated from these dated samples supported the notion of an accelerated short-term time-
75 dependent rate (Ho et al. 2005), and agreed with estimates obtained from various other aDNA
76 datasets (e.g. Ho et al. 2007, Burridge et al. 2008), and the mean human mitochondrial control
77 region pedigree rate estimate (Howell et al. 2003). Importantly, population genetic analyses of
78 VLC founder dynamics conducted using this substitution rate fitted well the timeframe recovered
79 from geological dating of Holocene raised beaches along the VLC, and the fate of the Ross Sea
80 Ice Sheet, which indicated that this coastline would have been inhospitable to southern elephant
81 seals prior to around 8,000 YBP (Conway et al. 1999, Hall et al. 2006). Elephant seals require
82 breeding habitat with access to open water, and the VLC coastline was beneath the Ross Sea Ice
83 Sheet prior to this time (Hall et al. 2006). The retreat of the ice shelf was followed by a period of
84 relative warmth when the seal colony apparently thrived and expanded, before more recent

cooling around 1,000 YBP lead to an increase in land-fast sea-ice - a likely cause of the demise and eventual extinction of this population (Hall et al. 2006, de Bruyn et al. 2009).

Molecular analyses of these ancient DNA samples from the VLC, and from contemporary samples from seven putative populations, representing all four major extant southern elephant seal breeding stocks (Leboeuf & Laws 1994), have provided considerable insight into the origin and demographic history of the VLC seals (de Bruyn et al. 2009). Our earlier research demonstrated that the VLC seals were an independent breeding colony - most likely founded by seals from Macquarie Island (MQ) - which grew in size through the mid-Holocene, before declining drastically around 1,000 YBP, resulting in eventual extinction and the possible return of some VLC animals to the source population at MQ (de Bruyn et al. 2009). Inferred timings of these demographic events derived from molecular data are consistent with inferences based on fossil radiocarbon dates and geological data. However, it was not clear if this scenario was sufficient to explain the very high levels of diversity of the VLC population, or the relative importance of migration compared to *in situ* mutation.

We therefore applied Approximate Bayesian Computation (ABC) methods to the VLC and MQ dataset to test hypotheses about the timing and relative importance of different processes, thereby testing hypotheses about life history and behavior (such as the likelihood of the VLC being a breeding, as opposed to a molting population, and the degree of insularity of the newly founded population). ABC enables the joint assessment of a large number of increasingly complex demographic models, and allowed us to examine in detail processes leading to significant changes in genetic diversity, from initial founding to the eventual extinction of a population.

Materials and Methods

Data

The VLC and MQ dataset analyzed here comprised 223 ancient VLC samples, ranging in age from ±7,200 – 270 YBP, and 49 contemporary samples from MQ (Table S1; Genbank accession numbers: FJ168073–FJ168343). Details of laboratory DNA methodology (including extractions, PCR, sequencing of a 325 bp fragment of the HVR1 section of the mitochondrial control region, cloning, 'independent replication' and other strict aDNA protocols), as well as calculation of aDNA calendar dates from radiocarbon dating that incorporated a time-dependent Southern Ocean marine reservoir effect, are presented elsewhere (de Bruyn et al. 2009).

Coalescent models and summary statistics

Based on earlier population genetic results, including Isolation-with-Migration analyses (de Bruyn et al. 2009), a series of increasingly complex models were examined within an ABC framework. These ranged from a simplified one population (Model 1) or model with two populations that diverged 12,000 ybp (Model 2), to those incorporating exponential population growth and bi-directional migration (e.g. Model 7) (Fig. 1). Models of intermediate complexity examined colonization from VLC to MQ (Model 3, the opposite of what we believe occurred), colonization from MQ to VLC (Model 4), ongoing migration from MQ to VLC after colonization (Model 5), and a single migration event from VLC to MQ when the VLC population was extirpated (Model 6). We also added alternative versions of Models 4-7 that included exponential growth (consistent with known demographic potential) in VLC for the first 500 years after colonization, followed by a stable population (Models 4b-7b). For models with colonization, we specified that the timing of colonization was 8000 ybp, consistent with the

retreat of ice from VLC and the first available breeding habitat. This was a conservative choice given that a more recent colonization would require more rapid population growth.

We used serial coalescent theory, as implemented in Bayesian Serial Simcoal (Anderson et al. 2005), to simulate 1 million datasets from each of our eleven models. We set the mutation rate to 9.8×10^{-7} mutations site^{-1} yr^{-1}, as estimated from dated aDNA samples, with a transition/transversion ratio of 0.8564 and a mutation rate gamma distribution shape parameter of 0.2003 (de Bruyn et al. 2009). The sample sizes, sample dates, and length of the locus (325bp) matched our collected data (Table S1). We used uniform priors on population size and migration rates (Tables 1 & S2) so as to specify little prior knowledge about these quantities. Uniform priors for population size and migration rates have been commonly recommended or used in ABC (Bertorelle et al. 2010, Li & Jakobsson 2012, Robinson et al. 2013). The prior on back-migration in Models 6 and 7 was also uniform 0-100% (Table S2). For both on-going migration from MQ to VLC and the back-migration event from VLC to MQ, 100% migration corresponded to full replacement of the sink population with genes from the source population, while 50% migration corresponded to replacement of only half the sink population. In addition, we evaluated the sensitivity of our conclusions to alternative priors by generating 333,333 simulations from each of the 11 models with 1) log uniform migration rates from e^{-9} to e^{0}; or with 2) wider priors on the size of the colonizing population (uniform 0 to 10,000). Colonizing populations up to 10,000 in a single generation are highly unlikely given that MQ only has 10,000 females.

From the real data and from the simulations, we then calculated a series of summary statistics on each of three statistical groups: MQ (n=49), VLC 3000-7100 ybp (n=64), and VLC 0-3000 ybp (n=159). Statistical groups were chosen to examine temporal dynamics in VLC while

maintaining a sufficiently large sample size for the ancient group. However, we also conducted a sensitivity analysis after adding an additional statistical group for the earliest VLC years (6000-7100 ybp) and trimming the other ancient VLC statistical group to 3000-6000 ybp. The sensitivity analysis was conducted with 333,333 simulations from each of our 11 models.

We used the number of haplotypes (*H*), number of segregating sites (*S*), and average number of pairwise differences within each of the statistical groups (π) as summary statistics (Tables 2 and S3). We also used F_{ST} values calculated between MQ and each of the two VLC statistical groups. We chose these summary statistics because they reflect the demographic changes in which we were interested (Bertorelle et al. 2010), namely changes in population size (*H*, *S* and π at two time points) (Nei & Takahta 1993, Tajima 1989) and the exchange of migrants through time between MQ and VLC (F_{ST}s) (Slatkin 1985) (Table S3). Preliminary simulations also suggested that they were sufficient to distinguish among our competing hypotheses.

Approximate Bayesian Computation (ABC)

After simulating the models and calculating the summary statistics, we conducted our ABC analysis in three steps: model selection, parameter estimation, and quality control. For model selection, we used a weighted multinomial logistic regression to calculate the relative support for each model (Beaumont 2008) from the 5% of all 11 million simulations for which δ (the Euclidean distance between the observed and simulated summary statistics) was smallest. Parameter estimation followed a similar process. From the 0.5% of simulations from each model for which δ was smallest, we used weighted local linear regression to estimate the posterior distributions of each parameter (Beaumont et al. 2002). Before the parameter calculations, we

log-transformed MQ and final VLC population sizes, and logit-transformed migration rates and the size of the VLC colonizing population. Both transformations were applied to facilitate parameter estimation and projection back onto untransformed axes. We report the median, mode, and 95% highest probability density (HPD) for each parameter. Calculations were done in R 2.15.3 with the abc package (Csillery et al. 2012).

For quality control, we followed the recommendations of Bertorelle et al. (2010) and generated pseudo-observed datasets (PODs). We sampled 1000 parameter sets from the posterior distributions of each model and used these to simulate PODs. By sampling from the posteriors, we account for our full knowledge and uncertainty about the true model parameters. Previous authors have used only the best parameter estimates (Bertorelle et al. 2010) or the full prior distributions (Fagundes et al. 2007). We re-ran our model selection calculation on each of 100 PODs from each model (1100 total PODs) to evaluate our model selection procedure when the true model was known. This process allowed us to calculate model choice error rates. In addition, these results also allowed us to calculate a *corrected* relative support for each model following the procedure of Fagundes et al. (2007)

$$\Pr(Model = X \mid P_{Model\ A} = a) = \frac{\Pr(P_{Model\ A}=a \mid Model=X)}{\sum_{i \in Models} \Pr(P_{Model\ A}=a \mid Model=i)}$$

for each potential model X, the originally identified best model A, and the originally observed relative support *a* for Model A. The procedure calculates the probability that X is the correct model given our initial (uncorrected) observation that Model A was chosen as the best model with relative support *a*. The procedure accounts for difficulties differentiating between the models while also accounting for the fact that Model A might have been chosen with strong

support (a close to 1). We implemented the procedure by calculating the relative support for Model A from each of the 100 PODs produced by each of our eleven models. For each model, we then calculated a probability density across the range of relative supports (0 to 1). We expected that models that were rarely confused with Model A would have densities centered near zero relative support, while models similar to Model A (including Model A itself) would have densities centered closer to 1. We then evaluated these densities at the observed value (a) of relative support for Model A (e.g., $\Pr(P_{Model\ A} = a \mid Model = X)$ to evaluate the above equation.

We also assessed model fit against our observed data using tail-area probability, or *p*-value, tests for each summary statistic (Cornuet et al. 2010). We used false discovery rate (FDR) corrections on the statistics within each model (Benjamini & Hochberg 1995), as well as the method of Ghirotto et al. (2010) to combine *p*-values across statistics. When combining *p*-values, we assigned 0.001 to those *p*-values at the edge of the empirical distribution in order to avoid infinite values in the calculation. Low *p*-values indicated that the observed data were unlikely given a particular model. Finally, we performed a principal components analysis (PCA) on the summary statistics and plotted the observed data along with simulated datasets from the prior and from the posterior (Cornuet et al. 2010). Ideally, the observed data would fall within the posterior for the selected model.

To test the accuracy of our parameter estimation procedure, we estimated model parameter ($\hat{\theta}_i$) for each POD *i* and compared the estimates against the known, true parameters (θ_i) used to simulate each POD. We then calculated the bias and root mean square error (RMSE) for each parameter across all *n* PODs:

$$bias = \frac{1}{n}\sum_{i-1}^{n}\frac{\hat{\theta}_i - \theta_i}{\theta_i}$$

$$RMSE = \frac{1}{\theta}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2}$$

We also calculated the Factor 2 statistic, which is the proportion of estimated values that fall between 50% and 200% of the true value (Neuenschwander et al. 2008). We used coverage to assess the accuracy of our confidence intervals, defined as the proportion of simulations that fall within the 50% or 90% HPD around the estimated parameter. Finally, we calculated the coefficient of determination (r^2) for each parameter, which is the proportion of parameter variance that can be explained by the summary statistics (Bertorelle et al. 2010).

Results

Model Selection

The raw posterior probabilities from our ABC analysis suggested the highest support (0.70) for Model 4b (Table 3), the scenario that involved rapid growth of the VLC population after colonization from MQ, but that did not include ongoing migration from MQ or back-migration to MQ. We found lower support for Model 6b (0.24), which did include migration back to MQ near the collapse of VLC. There was very little support for scenarios in which VLC and MQ were panmictic, where they were completely independent, where MQ was colonized by VLC, or where VLC grew slowly.

However, when we simulated datasets to validate our model choice results, we found that distinguishing between certain models could be difficult, at least after such models had been fit to the observed data (a more stringent test than used in many previous studies) (Table S3). In particular, Models 1, 3, 5, 6, 7, 5b, and 7b were misidentified as a different model more than

50% of the time (Table S3). Model 4b was the most accurately identified scenario, with 90% of simulations correctly chosen as Model 4b. Most of the remaining Model 4b simulations were mistaken for Model 6b, a model that included 4b as a limiting case when back-migration was low. Encouragingly, however, and of particular relevance to our study, few scenarios were misidentified as Model 4b with high support (e.g., ≥0.70) (Fig. S1). The exception to this general rule was Model 5b, which, like 6b, included 4b as a limiting case when the migration rate was low. We corrected for the possibility of misidentification by calculating Pr(Model X | $P_{\text{Model 4a}}$ = 0.70) for each model (Table 3, Fig. S1). We found that most of the posterior support, once corrected for errors in model choice, was split between Models 4b and 5b, with lower support for Models 6b and 7b. This result suggested that rapid growth soon after VLC colonization was highly likely, but determining whether or not migration continued after colonization could be more difficult. Back-migration to MQ near the collapse of VLC was less well supported (probability of 0.27 that the true scenario was Models 6b or 7b).

Further model checking with tail probabilities indicated that at least one summary statistic failed the test ($p < 0.05$) for each of Models 1-7, but none of the statistics failed for Models 4b-7b (those with rapid growth) (Table S4). The combined *p*-value statistic indicated particularly poor fits for Models 1 and 2 ($p < 0.001$), but did not reject any of the other models. The PCA analysis on the summary statistics indicated poor fits for Models 1-4, with the observed data clearly falling outside the simulations from the posterior (Fig. S3). The other models provided better fits to the data, according to this test.

Under alternative model priors or alternative statistical group definitions, Model 4b continued to be selected as the best model (Table 3). However, under log normal priors for migration rate, the model choice procedure indicated lower support for Model 4b than before,

with the remaining support distributed primarily among Models 5b, 6b, and 7b. This result is to be expected because very low values of migration rates (close to 0) were commonly selected under log normal priors, and so the functional differences among Models 4b, 5b, 6b, and 7b were smaller. Under wider priors for the size of the colonizing population, support for Model 4b remained high (68%). Support also remained substantially higher for Model 4b (51%) than for alternative models (≤18%) with four statistical groups.

Parameter estimation

We then estimated the posterior distributions for both of the well-supported models (4b and 5b) and found similar parameter estimates from both (Table 1, Fig. 2). In particular, we found that MQ was about two orders of magnitude smaller than VLC, consistent with the lower genetic diversity in this population (Table 1, Fig. 2). Posterior parameter estimates for the migration rate between MQ and VLC also clarified why Model 5b received similar support to Model 4b: the migration rate posterior for Model 5b was nearly 0 (95% HPD of 0 to 0.003; Table 1, Fig. 2). With a very low to zero migration rate, Model 5b was functionally similar to Model 4b, and the two were therefore difficult to distinguish. From the width of the posterior, it was clear that we did not have much power to detect the effective number of females that colonized VLC, though a founding size <200 seals was unlikely (Fig. 2).

In testing the accuracy of our parameter estimates, we found that the relative bias and RMSE error were quite low for the MQ and ancient VLC effective size (<50%), but substantially higher for the final VLC effective size with a positive bias of up to 100% (Table 4). Our estimates of final VLC size, however, were in the right order of magnitude. This accuracy was also reflected in the factor 2 calculations, which were highest for MQ size and lowest for final

VLC size. The coverage of our 50% and 90% HPDs were quite accurate, with mean values across parameters of 44% and 88%. The high values for our coefficients of determination suggested that our choices of summary statistics were informative. We concluded that our estimates of MQ and ancient VLC population sizes were most accurate, but we could only approximate the final VLC size to within an order of magnitude.

When we compared the fit of simulated datasets based on our posterior distributions, we found that they were a substantially better fit to the observed data than were the prior distributions (Fig. 3). However, the best model continued to predict haplotype diversity in the ancient VLC population (H_{VLC1}) that was somewhat too low, while the predicted number of segregating sites in the recent VLC population remained somewhat too high (S_{VLC2}). On the other hand, the key prediction of final haplotype number in the VLC was a strong match. The fit for Model 4b was also substantially better than for a similar model that did not include rapid initial population growth in VLC (Model 4). The latter model substantially under-predicted haplotype diversity in VLC (H_{VLC1} and H_{VLC2}), suggesting that rapid growth was important for the generation and maintenance of genetic diversity (Fig. S2).

Discussion

Our analysis indicated that the high levels of genetic diversity attained within the VLC southern elephant seal colony (Table 2) most likely arose largely by rapid mitochondrial control region *in situ* mutation as a function of early rapid population growth to a large effective population size, which effectively reduces the probability of extinction for new (single-copy) alleles (Crow and Kimura 1970). This followed the release of a large expanse of suitable breeding habitat along the Victoria Land Coast, after the retreat of the Ross Sea Ice Sheet in the

early- to mid-Holocene (Hall et al. 2006). Approximate Bayesian Computation provides strong support for the idea that Macquarie Island was the source of founders of the VLC colony, consistent with previous findings (de Bruyn et al. 2009; though the mtDNA marker provides long-term inference only about the movement of females). Further, our analysis did not support the idea that seals from VLC were purely a non-breeding moulting colony, that VLC was the source of the MQ population, or that seals from the two locations were panmictic. We could also rule out any of the other major extant southern elephant seal breeding colonies as sources of either founders, or observable levels of gene flow (de Bruyn et al. 2009). Thus, the significant levels of genetic diversity observed in the VLC population likely resulted from a combination of rapid substitution rate at the mitochondrial control region, and rapid population growth to a large effective population size, soon after colonization of newly-available habitat resulting from Holocene climate change.

The model with the highest posterior probability is one of initial founding of VLC by a cohort of MQ seals (95% HPDs of female effective size: 157 – 953), followed by rapid population growth and a lack of gene flow between the two locations until final extinction of the VLC colony, some 7000 years later (de Bruyn et al. 2009). Interestingly, inclusion of an initial rapid growth phase would appear critical to fitting the models to our data (Table 3, Fig. 3, Fig. S2). The next best-fit model is otherwise the same, also including a rapid growth phase, but distinguished by very low levels of ongoing migration from MQ to VLC (95% HPDs of migration rate 0 – 0.003). Although immigration from an unsampled 'ghost' population cannot be fully excluded, there are no apparent candidate populations not already directly or indirectly sampled, and data from de Bruyn et al. (2009), which sampled all major extant populations, illustrate a clear connection between VLC and MQ in the mtDNA networks. One other possible

interpretation is that migrants to VLC were from other now-extinct colonies, not from Macquarie. The most likely candidate, based on geographic proximity, is the population from the north-western coast of Tasmania, extinct since prehistoric times (Bryden et al. 1999), though 1500 km further from VLC than MQ. In this context, it seems unlikely that the VLC population would be founded by the Tasmanian (or the even more distant historical colony in New Zealand) rather than by the MQ population. The colonies on Juan Fernandez and St. Helena were likely small and exterminated by hunting in the 19th and 20th centuries. Both were very distant from VLC, over 4,000km to Juan Fernandez and over 6,000 km to St. Helena (the northern most record for this species; Fraser 1935). The chance that either of these would form a useful/relevant reference sample for this study seems very remote.

Simulated datasets based on posterior distributions fitted the observed data far better than the prior distributions (Fig. S2), including a strong fit between observed recent VLC haplotypic diversity and that predicted (Fig. 3). There were two exceptions, with the model prediction for haplotype diversity being low for the older VLC sample, and high for the number of segregating sites in the more recent VLC sample. The lower than expected observed number of segregating sites may simply reflect higher than expected variance of substitution rate among sites. While the best-supported model was one of no post-founder gene flow, one possible explanation for higher than expected haplotypic diversity early on is a post-founder 'connection event(s)' (see Alcala et al. 2013). Although this may be consistent with our second (Model 5b) and third (Model 6b) best-fitting models (Table 3), the level of migration indicated in those models is very small. Other possible explanations include stochastic sampling effects, and sequence error in the older samples due to post-mortem damage. The latter is unlikely given the controls and replication undertaken to ensure accurate genotyping, including multiple replicate extractions-through-to-

sequencing, and cloning (see de Bruyn et al. 2009 for further details).

The key result remains the match between the very high haplotype diversity at VLC, including extensive novel diversity compared to the source population MQ (de Bruyn et al. 2009), and model expectations derived using ABC and ancient DNA. While the generation of genetic diversity in this case is presumed to be neutral and the mutation rate relatively high (mitochondrial control region), our results suggest that given the right conditions, environmental change could generate the raw material for adaptive evolution (affecting multi-genic phenotypic traits) through rapid population growth over a very short evolutionary time scale.

Acknowledgements

We thank M. Blum for assistance with calculating the coefficient of determination, and Oscar Gaggiotti for helpful comments that improved the manuscript. MLP was supported by National Science Foundation Graduate Research and David H. Smith Conservation Research fellowships. Research funding was provided by National Science Foundation grant PLR-0807292, and support from the Italian Antarctic Programme (PNRA).

References

Alcala N, Streit D, Goudet J, Vuilleumier S. 2013 Peak and persistent excess of genetic diversity following an abrupt migration increase. *Genetics* **193,** 953–971.

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. 2005 Serial SimCoal: A population genetic model for data from multiple populations and points in time. *Bioinformatics*, **21**, 1733–1734.

Beaumont MA. 2008 Joint determination of topology, divergence time and immigration in population trees. In *Simulation, Genetics and Human Prehistory* (eds. S. Matsamura, P. Forster, C. Renfrew), pp. 135–154. Cambridge, UK: McDonald Institute for Archaeological Research.

Beaumont MA, Zhang W, Balding DJ. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162,** 2025–2035.

Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* **57,** 289-300.

Bertorelle G, Benazzo A, Mona S. 2010 ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* **18,** 2609–2625.

Bryden MM, O'Connor S, Jones R. 1999 Archaeological evidence for the extinction of a breeding population of elephant seals in Tasmania in prehistoric times. *Int. J. Osteoarchaeol.* **9,** 430–437.

Burridge CP, Craw D, Fletcher D, Waters JM. 2008 Geological dates and molecular rates: fish DNA sheds light on time dependency. *Mol. Biol. Evol.* **25,** 624–633.

Conway H, Hall BL, Denton GH, Gades AM, Waddington ED. 1999 Past and future grounding-line retreat of the West Antarctic ice sheet. *Science* **286,** 280–283.

400 Cornuet JM, Ravigné V, Estoup A. 2010 Inference on population history and model checking
401 using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC*
402 *Bioinf.* **11,** 401.
403 Crow JF, Kimura M. 1970 *An Introduction to Population Genetics Theory.* New York: Harper &
404 Row.
405 Csilléry K, Blum M, Francois O. 2012 abc: an R package for Approximate Bayesian
406 Computation (ABC). *Methods Ecol. Evol* **3,** 475–479.
407 de Bruyn M, Hall BL, Chauke LF, Baroni C, Koch PL, Hoelzel AR. 2009 Rapid response of a
408 marine mammal species to Holocene climate and habitat change. *PLoS Genetics* **5(7),**
409 e1000554.
410 Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L.
411 2007 Statistical evaluation of alternative models of human evolution. *Proc. Nat. Acad.*
412 *Sci. USA* **104,** 17614–17619.
413 Fraser FC. 1935 Zoological notes from the voyage of Peter Mundy, 1655-56 ((a) Birds. By Mr.
414 NB. Kinnear) (b) sea elephant on St. Helena; whale at Greenwichs. *Proc. Linn. Soc. Lond.*
415 **1934-1935,** 32-37.
416 Ghirotto S, Mona S, Benazzo A, Paparazzo F, Caramelli D, et al. 2010 Inferring genealogical
417 processes from patterns of Bronze-Age and modern DNA variation in Sardinia. *Mol. Biol.*
418 *Evol.* **27,** 875–886.
419 Hall BL, Hoelzel AR, Baroni C, Denton GH, Le Boeuf BJ, Overturf B, Töpf AL. 2006 Holocene
420 elephant seal distribution implies warmer–than–present climate in the Ross Sea. *Proc.*
421 *Nat. Acad. Sci. USA* **103,** 10213–10217.
422 Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005 Time dependency of molecular rate

estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* **22,** 1561–1568.

Ho SYW, Kolokotronis S-O, Allaby RG. 2007 Elevated substitution rates estimated from ancient DNA sequences. *Biol. Letters* **3,** 702–705.

Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, et al. 2003 The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet.* **72,** 659–670.

Kimura M. 1983 *The Neutral Theory of Molecular Evolution.* Cambridge, UK: Cambridge University Press.

Kimura M, Ohta T. 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61,** 763–771.

Kimura M, Ohta T. 1973 The age of a neutral mutant persisting in a finite population. *Genetics* **75,** 199–212.

LeBoeuf B, Laws R. 1994 *Elephant Seals: Population Ecology, Behaviour, and Physiology.* Berkeley: University of California.

Li S, Jakobsson M. 2012 Estimating demographic parameters from large-scale population genomic data using approximate Bayesian computation. *BMC Genet.* **13,** 22.

Maruyama T. 1970 Effective number of alleles in a subdivided population. *Theor. Popul. Biol.* **1,** 273–306.

Nei M, Feldman MW. 1972 Identity of genes by descent within and between populations under mutation and migration pressures. *Theor. Popul. Biol.* **3,** 460–465.

Nei M, Takahata N. 1993 Effective population size, genetic diversity and coalescence time in subdivided populations. *J. Mol. Evol.* **37,** 240–244.

446 Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlathen P, Excoffier L. 2008
447 Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference
448 under a Bayesian spatially explicit framework. *Mol. Ecol.* **17,** 757–772.
449 Otto SP, Whitlock M. 1997 The probability of fixation in populations of changing size. *Genetics*
450 **146,** 723–733.
451 Robinson JD, Hall DW, Wares JP. 2013 Approximate Bayesian estimation of extinction rate in
452 the Finnish *Daphnia magna* metapopulation. *Mol. Ecol.* **22,** 2627–2639.
453 Slatkin M. 1985 Gene flow in natural populations. *Ann. Rev. Ecol. Syst.* **16,** 393–430.
454 Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA
455 polymorphism. *Genetics* **123,** 585–595.
456

Figure Captions

Figure 1. Diagram of hypotheses about migration and colonization between MQ and VLC over the period in which VLC was extant (8000 to 1000 ybp). The width of the arrows indicates population size through time, while horizontal arrows indicate migration. Alternate versions of Models 4 to 7 (Models 4b to 7b, not shown) specified rapid growth in VLC over the first 500 years (83 generations) to a plateau in population size.

Figure 2. Posterior densities for parameters from Model 4b (top row) and Model 5b (bottom row). Migration rate from MQ to VLC is only present in Model 5b.

Figure 3. Comparison between the posterior distributions (black curve) and the prior distributions (dashed curve) for Model 4b. The summary statistics are from Macquarie (MQ), ancient Victoria Land Coast (VLC1) and more recent Victoria Land Coast (VLC2) and include the number of haplotypes (*H*), the number of segregating sites (*S*), the number of pairwise differences (π), and pairwise F_{ST} values. The observed value of each summary statistic is shown as a vertical line.

Tables and Figures

Table 1: Prior and posterior parameter estimates for the MQ and VLC colonization history.

Parameter	Description	Prior	Median	Mode	95% HPD
Model 4b					
N_{MQ}	MQ female effective size	0 to 10^5	2,050	2,000	1,290 - 3,290
N_{VLC1}	VLC founding female effective size	0 to 10^3	627	719	157 - 953
N_{VLC2}	VLC final female effective size	0 to 10^6	158,000	120,000	65,700 - 395,000
Model 5b					
N_{MQ}	MQ female effective size	0 to 10^5	2610	2410	1220 – 4650
N_{VLC1}	VLC founding female effective size	0 to 10^3	527	664	33 – 998
N_{VLC2}	VLC final female effective size	0 to 10^6	474,000	110,000	21,300 – 986,000
$m_{MQ\text{-}VLC}$	MQ to VLC migration	0 to 100%	0.0001	0.0004	0 – 0.003

Table 2: Summary statistics for southern elephant seal populations from Macquarie Island (MQ) and from Victoria Land Coast (VLC) during the 3000-7100 ybp period (VLC1) or the 0-3000 ybp period (VLC2) (from de Bruyn et al. 2009). The statistics include the number of haplotypes (*H*), the number of segregating sites (*S*), the number of pairwise differences (π), and pairwise F_{ST} values.

summary statistic	**value**
MQ **(n=48)**	
H_{MQ}	15
S_{MQ}	23
π_{MQ}	6.58
VLC1 3000-7100 **(n=64)**	
H_{VLC1}	58
S_{VLC1}	49
π_{VLC1}	7.97
VLC2 0-3000 **(n=159)**	
H_{VLC2}	128
S_{VLC2}	79
π_{VLC2}	7.69
$\boldsymbol{F_{ST}}$	
MQ to VLC1	0.20298
MQ to VLC2	0.16809

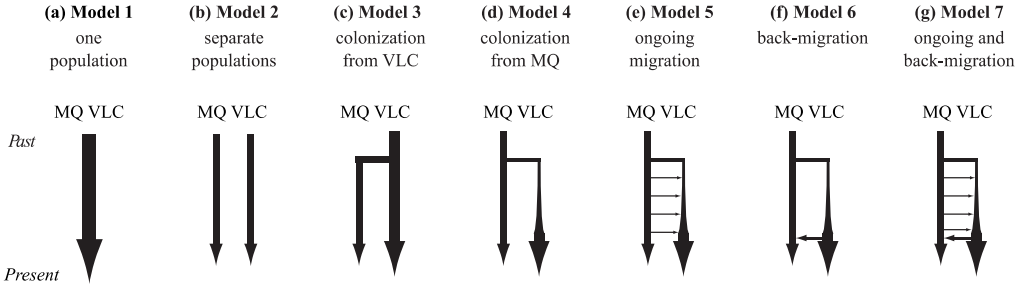
Table 3: Posterior probabilities for each model. Values in bold sum to 90% of posterior model probability within a column. Corrected probabilities account for errors in scenario identification. Posterior probabilities under alternative priors and statistical group definitions are also shown, including log normal migration rate priors (Priors 2), wider priors on the size of the colonizing population (Priors 3), and four statistical groups (4 Groups).

	Standard priors		**Priors 2**	**Priors 3**	**4 Groups**
Model	**posterior probability**	**corrected probability**	**posterior probability**	**posterior probability**	**posterior probability**
1. One population	<0.01	<0.01	<0.01	<0.01	<0.01
2. Separate populations	<0.01	<0.01	<0.01	0.01	**0.04**
3. Colonization from VLC	<0.01	0.06	<0.01	0.01	<0.01
4. Colonization from MQ	<0.01	<0.01	<0.01	**0.07**	**0.09**
4b. with rapid growth	**0.70**	**0.35**	**0.34**	**0.68**	**0.51**
5. Ongoing migration	<0.01	0.01	<0.01	<0.01	0.02
5b. with rapid growth	<0.01	**0.31**	**0.24**	<0.01	0.02
6. Back-migration	<0.01	<0.01	<0.01	0.03	**0.08**
6b. with rapid growth	**0.24**	**0.18**	**0.25**	**0.18**	**0.18**
7. Ongoing and back-migration	0.04	<0.01	<0.01	<0.01	0.02
7b. with rapid growth	<0.01	**0.09**	**0.16**	<0.01	0.03

Table 4: Accuracy of parameter estimation under Model 4b using 1000 simulated datasets.

Parameter	True Median	Estimated Median	Bias	RMSE	Factor 2	50%	90%	R^2
N_{MQ}	2,050	2,180	0.035	0.23	0.983	0.47	0.88	99%
N_{VLC1}	627	454	-0.087	0.44	0.763	0.52	0.92	99%
N_{VLC2}	158,000	330,000	0.95	1.0	0.578	0.33	0.84	99%

497 **Figure 1:**



498

Figure 2:

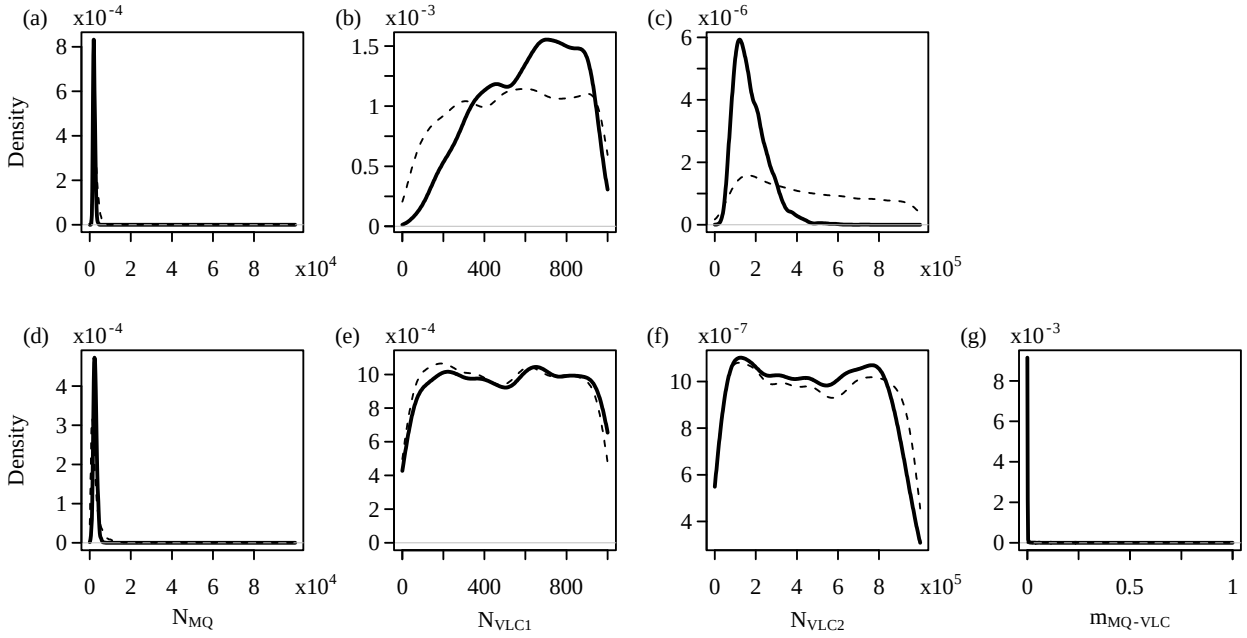
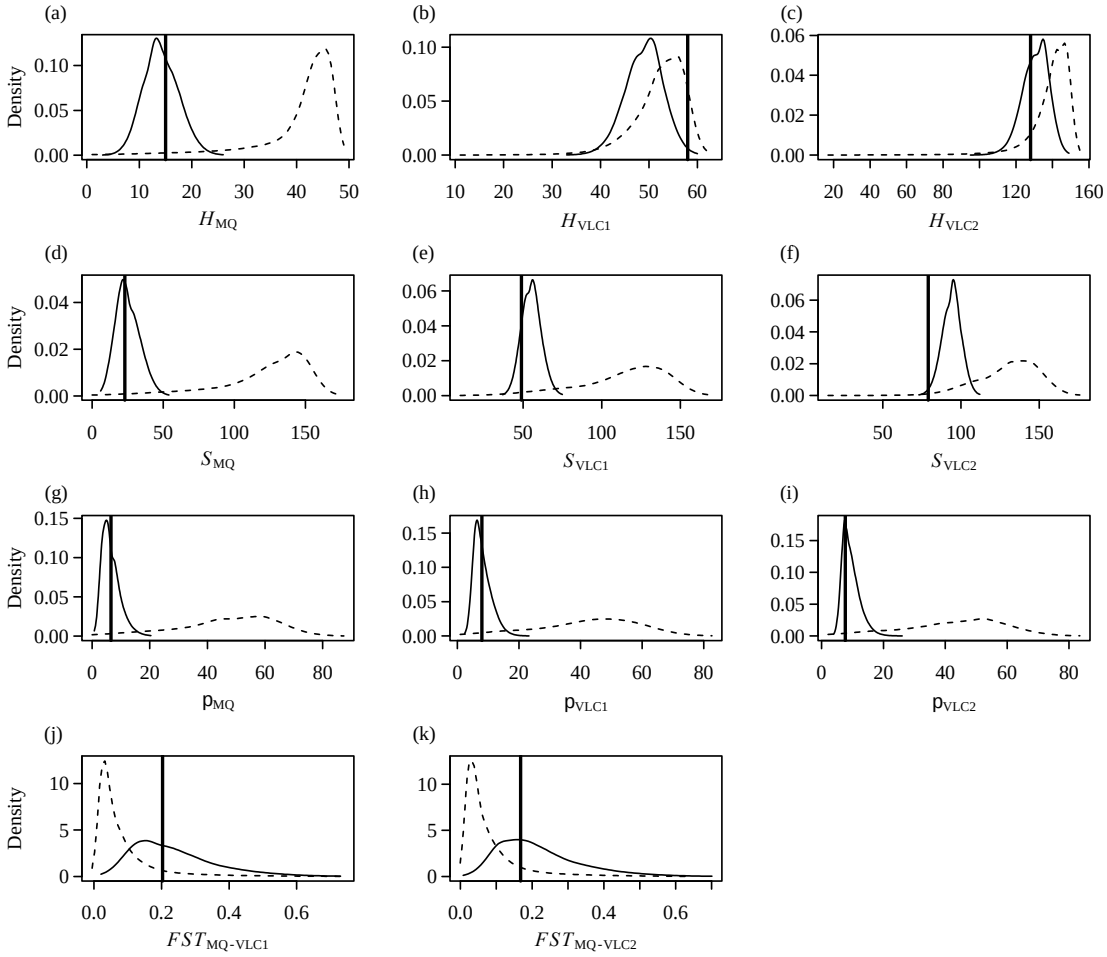

(a)
x10-4
Density
0 2 4 6 8
x10^4
NMQ
(b)
x10-3
0 0.5 1 1.5
0 400 800
NVLC1
(c)
x10-6
0 2 4 6
0 2 4 6 8
x10^5
NVLC2
(d)
x10-4
Density
0 1 2 3 4
0 2 4 6 8
x10^4
NMQ
(e)
x10-4
0 2 4 6 8 10
0 400 800
NVLC1
(f)
x10-7
4 6 8 10
0 2 4 6 8
x10^5
NVLC2
(g)
x10-3
0 2 4 6 8
0 0.5 1
mMQ-VLC

Figure 3:


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
Density
H_{MQ}
H_{VLC1}
H_{VLC2}
S_{MQ}
S_{VLC1}
S_{VLC2}
p_{MQ}
p_{VLC1}
p_{VLC2}
$FST_{MQ\text{-}VLC1}$
$FST_{MQ\text{-}VLC2}$