Minimum bottleneck abundance of Antarctic blue whales based on current mtDNA diversity

T.A. BRANCH[#], J.A. JACKSON[*]

ABSTRACT

A lower bound for the bottleneck population size of Antarctic blue whales is estimated using the current mitochondrial DNA haplotype diversity in the extant population. The method relies on the observation that each of the 26 haplotypes in the current population represents at least one surviving female at the time of the bottleneck. Conservative correction factors are used to extrapolate from these 26 haplotypes to a lower bound on the bottleneck population. Correcting for low sample effort increases the number of haplotypes to 51, to which are applied multiplicative factors of 1.29 for haplotypes missed because of short mtDNA sequences, 2.11 for the sex ratio in the bottleneck population, 1.50 for overlapping generations in the bottleneck population, and 1.03 for haplotypes that went extinct after the bottleneck. The resulting lower bound of 214, if included in a revised assessment of Antarctic blue whales, would probably increase the estimated bottleneck population size, reduce the estimated population rate of increase and increase the estimated pre-exploitation abundance, resulting in a lower ratio of current abundance to pre-exploitation abundance for Antarctic blue whales.

INTRODUCTION

Antarctic blue whales (*Balaenoptera musculus intermedia*) currently number just 2280 (95% CI 1160–4500) (Branch, accepted) and are just a small fraction of the 202,000–311,000 that originally existed (Branch *et al.*, 2004). This assessment of their current status was based entirely on recent abundance estimates and historical catches, from which it was inferred that they are increasing at 7.3% per annum (95% credibility interval 1.4–11.6%) from minimum abundance levels in 1973 of 360 (95% credibility intervals of 150–840) (Branch *et al.*, 2004). An independent estimate of minimum abundance would reduce the uncertainty in the historical trajectory and relative status of Antarctic blue whales.

In a recent review Baker and Clapham (2004) suggested that this minimum abundance could be estimated using contemporary diversity in mitochondrial DNA (mtDNA). Males do not contribute any mtDNA to the next generation, therefore the number of mtDNA haplotypes in a population reflects the number of maternal lineages that must have survived the population bottleneck (Patenaude, 2002), and if appropriately scaled up, can be used to estimate not just maternal bottleneck size but also the minimum population size of all males and females. These methods have been developed and applied to southern right whales (*Eubalaena australis*) to show that the genetic minimum abundance is substantially larger than that predicted by assessment models based on catches and recent abundance estimates (Jackson *et al.*, 2008). Here, we apply the same methods to obtain a minimum abundance estimate for Antarctic blue whales.

Jackson *et al*. (2008) previously calculated an estimate of the lower bound on minimum abundance which also included a correction for the sample size required to capture all haplotypes at the bottleneck, given a neutral pre-bottleneck distribution. No information on pre-exploitation genetic diversity is presently available for blue whales, and so we employ the conservative assumption that all haplotype lineages have been captured at a frequency of one at the bottleneck.

In brief, the method employed in this paper involves scaling up the number of mtDNA haplotypes in samples from the current population to account for the following factors: (1) haplotypes in the current population that were not sampled, (2) unique haplotypes in the samples that were not detected because only a section of the mtDNA was sequenced, (3) haplotypes that survived the bottleneck but were subsequently lost from the population before any sampling occurred, (4) overlapping maternal generations at the bottleneck, (5) males at the bottleneck period. We use conservative assumptions at each step to obtain a lower bound on the minimum abundance, which can be used as a constraint for assessments of Antarctic blue whales.

METHODS

Data

LeDuc *et al*. (2007) sequenced 420 base pairs of the mitochondrial control region of 47 Antarctic samples, finding 26 unique haplotypes, listed in Table 1. The Antarctic samples were collected during 1993–2002.

[#] *20504 86th Pl, Edmonds WA 98026, USA, e-mail: tbranch@gmail.com*

[*] *Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport OR 97365, USA.*

Extrapolating from sample to population: discovery curve

The total number of haplotypes in the population is greater than in the available sample of 47. To extrapolate from the sample haplotype richness to the population haplotype richness we used to discovery-curve approach of Jackson *et al*. (2008). This involved taking subsamples of between 5 and 42 without replacement from the pool of 47 samples. A plot of the subsample size against haplotype richness is then fitted by a Clench curve, and this curve is extrapolated to the estimated current abundance. The Clench curve is given by:

$$\hat{h} = \frac{a \cdot n}{1 + b \cdot n},$$

where $\hat{h}$ is the predicted haplotype richness for a sample of size *n*, and *a* and *b* are parameters to be estimated.

Extrapolating from sample to population: statistical models

An alternative approach to extrapolate from the sample haplotype richness to population haplotype richness, is to assume that the frequency of haplotypes in the current population follows a particular statistical distribution. A variety of statistical distributions were examined: uniform, Poisson, normal, binomial, lognormal, gamma, and negative binomial. The process followed several steps. First, the observed frequency of haplotypes in the sample was ordered from most common to least common. Second, 10,000 sets of 47 numbers were drawn from the statistical distribution, and rounded or truncated so that each number represented a haplotype. These simulated haplotypes were then ordered by frequency, and the predicted probability was obtained by averaging the ordered frequencies over the 10,000 sets. Third, a search was conducted for the parameter values for each statistical distribution that would best match the observed frequency in the data, by minimizing the multinomial negative log likelihood, ignoring constants this is:

$$-\ln L = -\sum_{i=1}^{46} n_i \ln p_i$$

Where n_i is the observed frequency of haplotypes at ordered position *i* and p_i is the simulated probability of a particular frequency at ordered position *i*.

The statistical distribution with the lowest negative log likelihood was then chosen for the best prediction of the number of haplotypes in the population. The predicted number of haplotypes was then obtained by drawing 1160, 2280 and 4500 haplotypes from this distribution to match the current population size.

Accounting for sequence length

About 420 base pairs (bp) in the mitochondrial control region were sequenced in LeDuc *et al*. (2007). Sequencing a greater length of the control region may have yielded more haplotypes among the 47 samples. Longer sequences were not read for Antarctic blue whales, but some longer sequences were read for New Zealand southern right whales.

Accounting for males in the bottleneck population

The number of mtDNA haplotypes reflects only the number of females in the population, and needs to be adjusted by the sex ratio in the population to account for the number of males alive at the bottleneck point.

Accounting for overlapping female generations in the bottleneck population

At the time of the bottleneck, each haplotype may have been present in more than one female. Assuming that mutations are rare, all surviving females sharing a common female ancestor would have the same haplotype. This factor is difficult to account for with any degree of accuracy. Jackson *et al*. (2008) scaled estimates up by a factor of two by assuming that breeding females would be balanced by at least an equal proportion of immature and non-breeding individuals at the bottleneck. However, in a heavily exploited population such as Antarctic blue whales, very few females may survive exploitation, which would greatly reduce the number of females in each lineage at the time of the bottleneck.

To estimate the correction for this factor we ran an individual-based simulation of female lineages. Each lineage was started 100 years before the bottleneck, with a single female born in 1873 (sensitivities included starting in 1853 and 1893). The following assumptions were made: survival in the first year of life was 0.9, survival of all other ages was 0.975 (Ramp *et al.*, 2006), age at first parturition was 11 yr (Branch, 2008), and the fraction of births that are female was 0.473 (Branch *et al.*, 2004). The inter-calf interval is assumed to be either 2 yr or 3 yr with probability 0.50 assigned to each. In each year, a random number was drawn to see if the female survived, then another random number determined if it gave birth (if two years had elapsed since the previous birth), and a third random number determined the sex of the calf, and if the calf was female, it was added to the lineage. During the years of whaling mortality, adult survival was further decreased by the estimated whaling mortality in that year, according to Branch *et al.* (2004). A large number of lineages had to be simulated (1,000,000) because the high whaling mortality resulted in the extinction of more than 99% of all lineages before the

bottleneck. The average number of females in the surviving lineages is an estimate of the multiplicative factor to account for overlapping female generations in the bottleneck population.

Loss of haplotypes after the bottleneck

Some of the haplotypes in the population may have been lost after the bottleneck. Estimates of haplotype loss via drift subsequent to the 1972 bottleneck were explored using the Bottlesim model (Kuo and Janzen, 2003). Female abundance trajectories starting in 1972 (the 'bottleneck' year) were simulated under a simple logistic model with initial population sizes of n = 150, 250 and 400. Each abundance trajectory was extended forward with two population growth rates, corresponding to the expected rate of increase based on biological parameters (4.3%) and the estimated rate of increase in the Antarctic blue whale population (7.3%) obtained by Branch *et al*. (2004). Rates of haplotype loss along these trajectories were explored for initial haplotype richness of either h = 65 and 70, under two scenarios representing the extremes of haplotype distributions. For the first scenario ("uniform"), h was uniformly distributed among all individuals in the 1972 population (frequency n/h); for the second scenario ("skewed"), the haplotype distribution was highly skewed, with $h - 1$ haplotypes occurring once only, and one haplotype occurring $n-(h-1)$ times. Bottlesim ran these scenarios over 1000 replicates, under the assumptions of fully overlapping generations, a generation time of 30 yr and an age at first parturition of 11 yr (Branch, 2008).

RESULTS

Extrapolating from sample to population

The Clench curve fitted nearly perfectly to the mean haplotype richness from the subsampled data (Figure 1), with estimates of a = 1.033 and b = 0.0186, which asymptotes as $n \to \infty$ at $a/b = 55$. The predicted haplotype richness was 53–55 for a population of 1160–4500 (Table 2).

The uniform distribution provided a slightly worse fit to the observed sample ($-\ln L$ = 147.3) and is not considered further. All of the other statistical distributions provided very similar fits ($-\ln L$ = 147.1 for all). For a population of 1160–4500 the mean predicted number of haplotypes in the population was similar for the normal, Poisson, binomial and negative binomial distributions (51–60), but substantially higher for the gamma (68–87) and lognormal (65–84) distributions (Table 2).

The predictions of the number of haplotypes in the current population using the statistical distributions were either similar or higher than the predictions from the discovery curve method. A conservative prediction of the number of haplotypes in the current population is taken to be 51, which is the lowest mean prediction for a population of 1160, and coincidentally also the minimum lower 95% percentile for a population of 2280.

Accounting for sequence length

Jackson *et al*. (2008) assumed that increasing the sequence length for southern right whales from 275 to 500 bp would increase the number of haplotypes detected by 25%, based on results in Carroll (2006). Carroll (2006) also showed an increase from 7 to 9 haplotypes when extending the sequence from 450 to 900 base pairs, an increase of 29%.

Accounting for males in the bottleneck population

Branch *et al*. (2004) examined the sex ratio in catches of Antarctic blue whales, finding that 5637 out of 11,942 fetuses (47.2%) and 87,098 out of 184,280 adults (47.3%) were female (Mackintosh, 1942; Nishiwaki and Oye, 1951; Tomilin, 1967). To account for males would therefore involve multiplying the number of females by a factor of 2.11.

Accounting for overlapping female generations in the bottleneck population

Out of 1,000,000 simulated lineages starting in 1873, only 1,243 had any surviving (female) members. The lineages contained between 1 and 4 females each, with a mean of 1.50. When the simulations were started in 1853, there were nearly twice as many surviving lineages at the bottleneck but the mean lineage size was similar (1.52); while when the simulations were started in 1893, only 593 lineages survived to the bottleneck but again the mean lineage size was similar (1.53). The correction for this factor was therefore taken to be 1.50.

Loss of haplotypes after the bottleneck

Bottlesim simulations suggest that loss of diversity since the bottleneck is 0–12% for a population growth rate of 7.3% (Table 3). With haplotypes uniformly distributed at the bottleneck, loss of diversity was minimal (0–6%) while 12% of haplotypes were lost under the skewed haplotype distribution scenario. For the more conservative population growth rate of 4.3%, haplotype loss was greater, with 0–11% of haplotypes lost under the uniform scenario and 21% of haplotypes lost under the skewed scenario. In the absence of data regarding the true haplotype distribution at the bottleneck, we take a conservative approach and assume that haplotypes lost since the bottleneck period fall in the realm of 0–6%. Given the biological parameters of Antarctic blue whales and the relatively recent time period of the bottleneck, it is not surprising that only ~3% of the haplotypes were lost under conservative assumptions. To account for this factor, the estimate should be multiplied by 1.03.

Minimum population size
Given these factors, an approximate lower bound for the minimum historical population size of Antarctic blue whales is 51×1.29×2.11×1.50×1.03 = 214.

DISCUSSION

The number of mtDNA haplotypes in the extant population provides useful information about minimum population abundance size, that could be incorporated in assessment models of large whales. For Antarctic blue whales, only about 2% of the estimated current population has been genetically sampled, but this has yielded a relatively high number of unique haplotypes (26 haplotypes from 47 samples). Each haplotype would need to have been represented in the bottleneck population by at least one female since mtDNA is matrilineal. The following factors are taken into account to convert the 26 haplotypes into a lower bound for the minimum abundance at the point of the bottleneck: additional unsampled haplotypes in the current population; undiscovered haplotypes in portions of the mtDNA that were not sequenced; accounting for males in the bottleneck population; accounting for overlapping generations of females in the bottleneck population; and the potential loss of haplotypes from the population after the bottleneck. Each of these factors is discussed in turn.

Extrapolating from sample to population
Two types of extrapolation were conducted, resampling from the observed data, as done by Jackson *et al*. (2008), and comparing predicted haplotype frequency distributions with the observed data, assuming that the haplotypes in the current population can be characterized by a parametric distribution. A wide range of possible types of haplotype distributions were equally able to explain the observed haplotype distribution, providing predictions that ranged from 51 to 84 haplotypes in the current population of 2280 whales. The lowest value of 51 was assumed for this factor.

Accounting for sequence length
Little is known about how many more haplotypes would be discovered if a greater length of mtDNA than 450 base pairs were to be sequenced. We assumed that 29% more haplotypes would be discovered, based on data where the sequence length was doubled for southern right whales, which is conservative in one sense because sequencing an even greater length of mtDNA may further increase the haplotype richness. On the other hand, basing the estimate on data from southern right whales adds some uncertainty, as there is doubtless some species-to-species differences in variability at different points on the mtDNA control region. Although the section of control region that was sequenced is the more variable end, some species have a second region of variability on the other end (R. LeDuc, pers. comm., 4 February 2008). Additional sequencing of the blue whale mtDNA control region would be needed to obtain more reliable estimates of this factor.

Accounting for males in the bottleneck population
This factor is well defined by both fetal and adult sex ratios in the catch, which are nearly identical, and show strong evidence that the population is slightly male-dominated. There is little uncertainty in this correction factor of 2.11.

Accounting for overlapping female generations in the bottleneck population
The individual-based simulation model returns an estimated mean lineage size of 1.50 at the time of the bottleneck. This is substantially lower than the factor of 2 assumed by Jackson *et al*. (2008), who assumed that the ratio of effective population size to actual population size should be 0.50 or lower, based on previous papers. The main reason for the lower correction factor obtained from the simulations was the high whaling mortality experienced by Antarctic blue whales: 20–40% in most years over a period spanning four decades (Branch *et al.*, 2004). As a result, very few whales survived and the average lineage size was greatly reduced. The estimate of 1.50 is a conservative estimate of this correction factor because the simulations started only 100 years before the bottleneck with a single female. Given a population size of 239,000 before whaling, <100 haplotypes, and females comprising 0.473 of the population, the average lineage must have contained thousands of females prior to whaling, not a single female.

Loss of haplotypes after the bottleneck
This factor is estimated to be negligible, adding only 0–6% to the final answer under conservative assumptions, primarily because of the current high rate of increase and relatively long average generation time of 30 yr.

Minimum population size
The lower bound estimate of population size at the bottleneck was 214. In the most recent assessment, the bottleneck population was estimated to be 360 with 95% credibility intervals of 150–840 (Branch *et al.*, 2004). This genetic estimate could be included in a revised Bayesian assessment as a lower bound on population size, which would increase the model-based population estimate at the bottleneck. Including this lower bound in future assessments would also eliminate some of the posterior probability associated with high rates of increase and low pre-exploitation abundance. As a result, the current population would be estimated to be increasing less rapidly, and to be more depleted, than the estimates in Branch *et al.* (Branch *et al.*, 2004).

ACKNOWLEDGEMENTS

T.A.B. is grateful for funding for this project from the International Whaling Commission. R. LeDuc assisted in interpreting mtDNA data from previous blue whale studies.

REFERENCES

Baker, C.S. and Clapham, P.J. 2004. Modelling the past and future of whales and whaling. *Trends. Ecol. Evol.* 19:365-371.

Branch, T.A. 2008. Biological parameters for pygmy blue whales. *IWC Paper* SC/60/SHxx:15pp?

Branch, T.A. accepted. Abundance of Antarctic blue whales south of 60°S from three complete circumpolar sets of surveys. *J. Cetacean Res. Manage.*

Branch, T.A., Matsuoka, K. and Miyashita, T. 2004. Evidence for increases in Antarctic blue whales based on Bayesian modelling. *Mar. Mamm. Sci.* 20:726-754.

Carroll, E.L. 2006. The demographic and genetic bottleneck of the New Zealand southern right whale. MSc thesis. dissertation.

Jackson, J.A., Patenaude, N.J., Carroll, E.L. and Baker, C.S. 2008. How few whales were there after whaling? Inference from contemporary mtDNA diversity. *Molecular Ecology* 17:236-251.

Kuo, C.-H. and Janzen, F.J. 2003. *BOTTLESIM*: a bottleneck simulation program for long-lived species with overlapping generations. *Molecular Ecology Notes* 3:669-673.

LeDuc, R.G., Dizon, A.E., Goto, M., Pastene, L.A., Kato, H., Nishiwaki, S., LeDuc, C.A. and Brownell, R.L. 2007. Patterns of genetic variation in Southern Hemisphere blue whales and the use of assignment test to detect mixing on the feeding grounds. *J. Cetacean Res. Manage.* 9:73-80.

Mackintosh, N.A. 1942. The southern stocks of whalebone whales. *Discovery Reports* 22:197-300.

Nishiwaki, M. and Oye, T. 1951. Biological investigation on blue whales (*Balaenoptera musculus*) and fin whales (*Balaenoptera physalus*) caught by the Japanese Antarctic whaling fleets. *Sci. Rep. Whales Res. Inst., Tokyo* 5:91-167.

Patenaude, N.J. 2002. Demographic and genetic status of southern right whales at the Auckland Islands, New Zealand. PhD dissertation.

Ramp, C., Bérubé, M., Hagren, W. and Sears, R. 2006. Survival in blue whales *Balaenoptera musculus* in the Gulf of St. Lawrence, Canada. *Mar. Ecol. Prog. Ser.* 319:287-295.

Tomilin, A.G. 1967. *Balaenoptera musculus* L. Blue whales. Pages 76-112 Cetacea. Israel Program for Scientific Translations, Jerusalem.

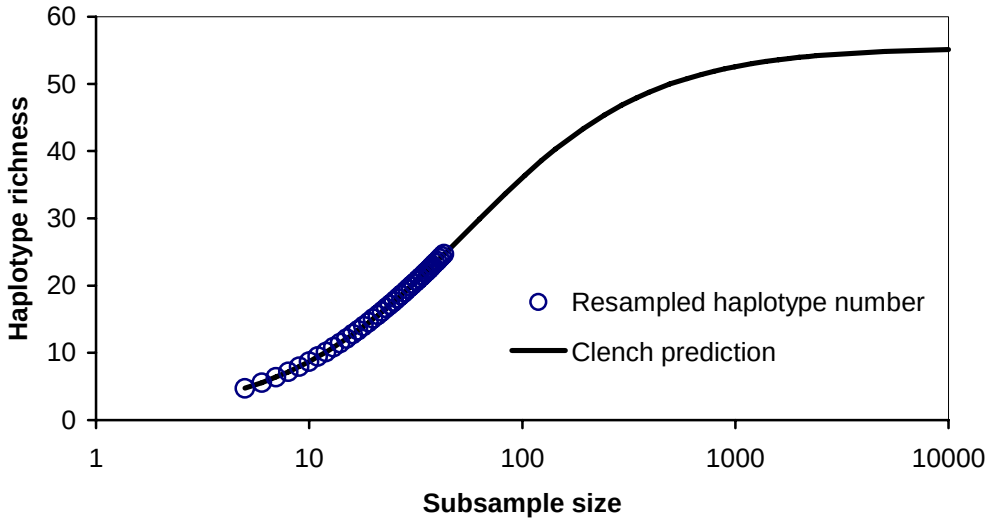

Figure 1. Discovery curve method of estimating the number of haplotypes in the population. The circles are mean number of haplotypes in subsamples of different size of the data, sampled without replacement; the line is the Clench curve fitted to the circles and extended to a large sample size.

Table 1. Frequency distribution of mtDNA haplotypes reported in LeDuc *et al*. (2007). Note that Figure 1 of LeDuc *et al*. (2007) inadvertently listed only one instance of haplotype v instead of two.

Haplotype	Frequency	Haplotype	Frequency
i	4	rr	1
bb	4	ss	1
nn	4	uu	1
u	3	aa	1
rr	3	gg	1
z	3	j	1
l	3	k	1
y	2	ff	1
m	2	hh	1
oo	2	nn	1
v	2	ee	1
h	1	tt	1
c	1	x	1

Table 2. Predicted number of haplotypes for population size (N) numbering 1160–4500 (corresponding to the most recent abundance estimates), under a variety of methods. The "Clench discovery" method is the standard method used in Jackson *et al*. (2008) for extrapolation of a resampled discovery curve by fitting the Clench equation to it. The remaining "statistical" methods involve assuming particular statistical distributions for the frequency of haplotypes in the population, and then finding the parameters that provide the best fit to the sample haplotype frequency. The negative log likelihood ($-\ln L$) of the statistical methods is given.

Method	Parameter estimates	$-\ln L$	N = 1160	N = 2280	N = 4500
Clench discovery	a = 1.033, b = 0.0186	–	53	54	55
Uniform	a = 0, b = 33	147.3	33 (33; 33)	33 (33; 33)	33 (33; 33)
Normal	μ = 0, σ = 8.6	147.1	52 (48; 56)	56 (52; 60)	59 (56; 63)
Poisson	λ = 73.5	147.1	51 (47; 55)	55 (51; 59)	59 (55; 62)
Binomial	n = 305, p = 0.5	147.1	51 (48; 55)	55 (52; 59)	59 (55; 62)
Lognormal	$\ln \mu$ = 2.7, $\ln \sigma$ = 0.6	147.1	65 (59; 71)	74 (68; 80)	84 (78; 90)
Gamma	a = 1.0, b = 13.6	147.1	68 (62; 74)	77 (72; 84)	87 (81; 93)
Negative binomial	μ = 25.7, κ = 12.2	147.1	52 (49; 56)	56 (53; 60)	60 (56; 64)

Table 3. Estimates of haplotype loss (and associated standard errors) via drift subsequent to the bottleneck period, as determined over 1,000 replications in the program Bottlesim. Details of the shape of the uniform and skewed haplotype distributions are provided in the text.

Population rate of increase	Haplotype distribution	Haplotype richness at bottleneck	Bottleneck size (no. females)	Haplotype loss	SE
4.3%	Uniform	65	150	9.05%	0.11%
4.3%	Uniform	65	250	1.02%	0.04%
4.3%	Uniform	65	400	0.09%	0.01%
4.3%	Uniform	70	150	10.69%	0.12%
4.3%	Uniform	70	250	1.41%	0.05%
4.3%	Uniform	70	400	0.18%	0.01%
4.3%	Skewed	65	150	20.93%	0.17%
4.3%	Skewed	65	250	20.71%	0.20%
4.3%	Skewed	65	400	20.65%	0.20%
4.3%	Skewed	70	150	20.71%	0.17%
4.3%	Skewed	70	250	20.55%	0.20%
4.3%	Skewed	70	400	20.62%	0.20%
7.3%	Uniform	65	150	4.49%	0.09%
7.3%	Uniform	65	250	2.46%	0.06%
7.3%	Uniform	65	400	0.35%	0.02%
7.3%	Uniform	70	150	5.69%	0.10%
7.3%	Uniform	70	250	3.34%	0.07%
7.3%	Uniform	70	400	0.51%	0.03%
7.3%	Skewed	65	150	12.06%	0.18%
7.3%	Skewed	65	250	11.86%	0.19%
7.3%	Skewed	65	400	12.10%	0.20%
7.3%	Skewed	70	150	12.33%	0.16%
7.3%	Skewed	70	250	12.03%	0.19%
7.3%	Skewed	70	400	12.11%	0.19%