

# Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure

Kristen Ruegg · Howard C. Rosenbaum ·  
Eric C. Anderson · Marcia Engel · Anna Rothschild ·  
C. Scott Baker · Stephen R. Palumbi

Received: 11 May 2012 / Accepted: 23 November 2012 / Published online: 13 December 2012  
© Springer Science+Business Media Dordrecht 2012

**Abstract** Once hunted to the brink of extinction, humpback whales (*Megaptera novaeangliae*) in the North Atlantic have recently been increasing in numbers. However, uncertain information on past abundance makes it difficult to assess the extent of the recovery in this species. While estimates of pre-exploitation abundance based upon catch data suggest the population might be approaching pre-whaling numbers, estimates based on mtDNA genetic diversity suggest they are still only a fraction of their past abundance levels. The difference between the two estimates could be accounted for by inaccuracies in the catch record, by uncertainties surrounding the genetic estimate, or by differences in the timescale to which the two estimates apply. Here we report an estimate of long-term

population size based on nuclear gene diversity. We increase the reliability of our genetic estimate by increasing the number of loci, incorporating uncertainty in each parameter and increasing sampling across the geographic range. We report an estimate of long-term population size in the North Atlantic humpback of ~112,000 individuals (95 % CI 45,000–235,000). This value is 2–3 fold higher than estimates based upon catch data. This persistent difference between estimates parallels difficulties encountered by population models in explaining the historical crash of North Atlantic humpback whales. The remaining discrepancy between genetic and catch-record values, and the failure of population models, highlights a need for continued evaluation of whale population growth and shifts over time, and continued caution about changing the conservation status of this population.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10592-012-0432-0) contains supplementary material, which is available to authorized users.

K. Ruegg (✉) · S. R. Palumbi  
Department of Biology, Hopkins Marine Station,  
Stanford University, 120 Oceanview Boulevard,  
Pacific Grove, CA 93950, USA  
e-mail: kruegg@ucsc.edu

H. C. Rosenbaum  
Ocean Giants Program, Global Conservation,  
Wildlife Conservation Society, Bronx, NY, USA

H. C. Rosenbaum · A. Rothschild  
Sackler Institute for Comparative Genomics,  
American Museum of Natural History,  
79th Street and Central Park West,  
New York, NY, USA

E. C. Anderson  
Fisheries Ecology Division, Southwest Fisheries Science Center,  
National Marine Fisheries Service, NOAA, 110 Shaffer Road,  
Santa Cruz, CA 95060, USA

E. C. Anderson  
Department of Applied Math and Statistics,  
University of California, Santa Cruz,  
CA 95060, USA

M. Engel  
Instituto Baleia Jubarte/Humpback Whale Institute,  
Rua Barao do Rio Branco 26, Caravelas,  
BA 45900-000, Brazil

C. S. Baker  
Marine Mammal Institute, Hatfield Marine Science Center,  
Oregon State University, 2030 SE Marine Science Drive,  
Newport, OR 97365, USA

**Keywords** Effective population size · Humpback whale · Census population size · Population structure

## Introduction

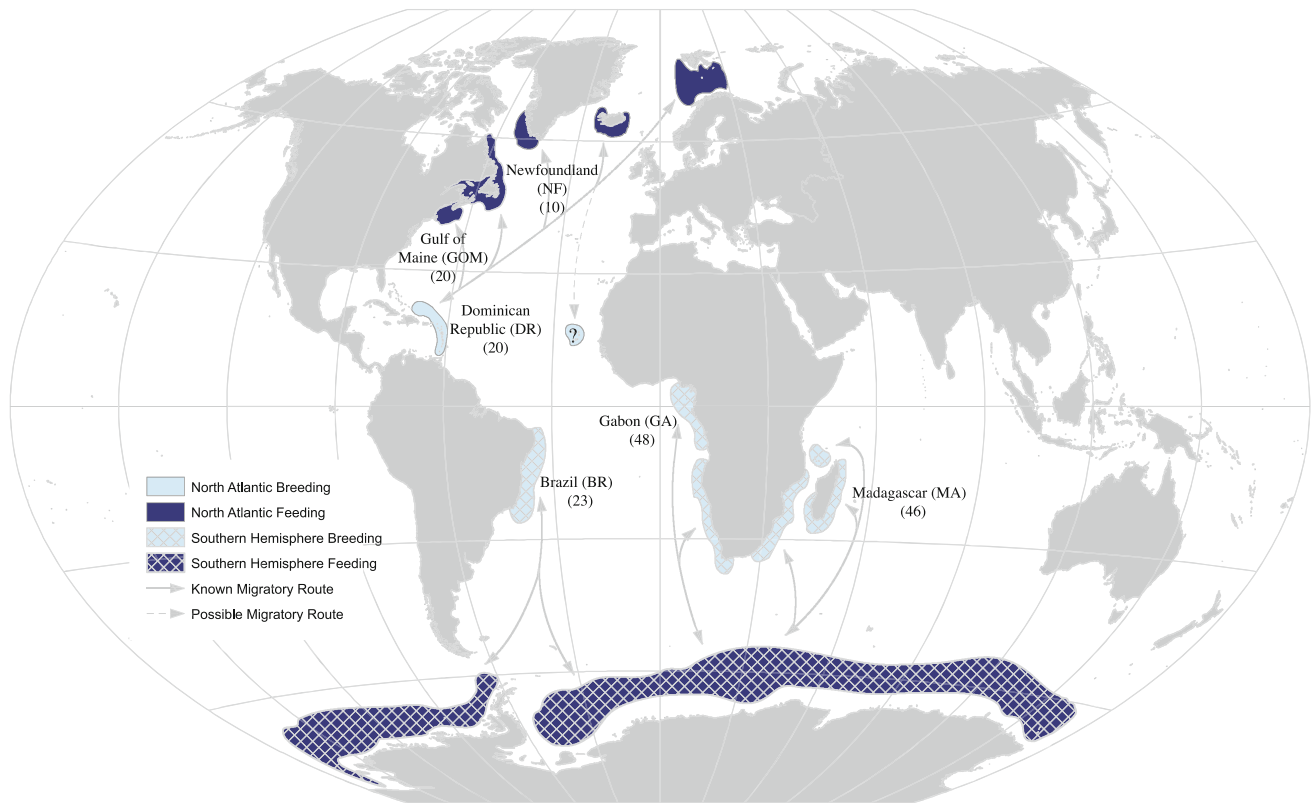
Over-exploitation has resulted in the collapse of many marine populations (Pauly et al. 1998; Myers and Worm 2003; Estes et al. 2006). In some cases, however, national or international protection has led to the recovery of previously threatened or endangered species (reviewed within Scott et al. 2005). Humpback whales (*Megaptera novaengliae*) in the North Atlantic were severely depleted as a result of intense hunting during the 19th and 20th centuries (Mitchell and Reeves 1983; Braham 1984; Winn and Reichley 1985) and are currently listed as ‘endangered’ or ‘vulnerable’ by various governments and international conservation organizations (Klinowska 1991). Before the International Whaling Commission (IWC) banned commercial whaling in the North Atlantic in 1955, it was estimated that this population was reduced to <1,000 individuals (Mitchell and Reeves 1983; Katona and Beard 1990). After many decades of legal protection, humpback whales have increased in numbers (Stevick et al. 2003) and recent survey estimates suggest that they may be approaching 20,000 animals (Smith and Pike 2009). Such increases in population size within the North Atlantic have led the IUCN and the US to re-evaluate their conservation status.

Assessing the recovery of previously depleted populations requires knowledge of past population sizes, but robust estimates of past abundance can be difficult to attain. Different approaches to estimating pre-whaling population sizes can lead to starkly different conclusions about the extent of recovery in the North Atlantic humpback whale (Roman and Palumbi 2003; Holt and Mitchell 2004; Punt et al. 2006). Traditionally, the IWC has relied upon population dynamic models that use a combination of information on current abundance, catch records, rates of increase, and population structure to estimate changes in population size through time (Punt et al. 2006). Recent model estimates for the North Atlantic humpback whale suggest a pre-whaling population size of between 20,000 and 46,000 individuals, depending upon the catch data used (Punt et al. 2006). Given current abundance estimates of ~17,700 individuals (Smith and Pike 2009), population model-based estimates suggest that humpbacks in the North Atlantic are approaching the lower boundary of their pre-whaling numbers. Alternatively, genetic-based estimates of pre-whaling abundance use the relationship between genetic diversity ( $\theta$ ) and effective population size ( $N_e$ ) ( $\theta = 4N_e\mu$ , where  $\mu$  is the average mutation rate) to estimate the long-term population size of North Atlantic Humpback whales (Roman and Palumbi 2003). Genetic estimates calculated using mitochondrial

DNA (mtDNA) control region sequence, suggest a pre-whaling abundance of 150,000–240,000 depending upon the mutation rate employed (Roman and Palumbi 2003; Alter and Palumbi 2009). These genetic estimates suggest that there were substantially more whales prior to whaling than previously believed.

The discrepancy between estimates of pre-whaling abundance based upon catch records and estimates based upon genetic variability has been the subject of vigorous debate (Lubick 2003; Holt and Mitchell 2004; Clapham et al. 2005). Some argue that unavoidable uncertainties in the catch record may have led to underestimates in the number of whales removed from the North Atlantic due to whaling (Palumbi and Roman 2007). However, a recent review and re-reading of whaling records revealed only slight increases in the numbers of North Atlantic humpback whales estimated to be killed as result of whaling (from 29,000 to 30,852 total catches) (IWC 2002, 2003; Smith and Reeves 2010). Others argue that genetic estimates of long-term abundance may be inaccurate as a result of reliance on a single locus, uncertainty surrounding mutation rates and generation times, the potential influence of incomplete sampling, and the evolutionary time-scale to which a genetic estimate applies (Lubick 2003; Holt and Mitchell 2004; Clapham et al. 2005). While recent estimates of long-term population size in gray and minke whales have reduced some of these uncertainties through a variety of methodological improvements (Alter et al. 2007; Alter and Palumbi 2009; Ruegg et al. 2010), humpback whales are particularly challenging because of their complex oceanic and worldwide population structure (Baker et al. 1993; Palsboll et al. 1995; Rosenbaum et al. 2009).

Population structure may affect estimates of long-term effective population size ( $N_e$ ) in a variety of ways depending upon the extent of isolation between populations. Theoretical models suggest meta-population  $N_e$  and sub-population  $N_e$  converge as migration between sub-populations increases (Hudson 1991; Waples 2010). It has also been shown that even low migration rates between sub-populations can cause an estimate of long-term  $N_e$  that is based upon samples from one sub-population to approximate the long-term  $N_e$  of the whole meta-population (Hudson 1991). Thus, if there is migration between sub-populations, then the absence of samples from one sub-population should have very little effect on an estimate of long-term  $N_e$  for the whole meta-population because the two values will be equivalent. Alternatively, if there is no migration between populations and one uses samples from an isolated sub-population, then an estimate of long-term  $N_e$  for the meta-population will be downwardly biased. Thus, in order to identify the impact of population structure on  $N_e$  it is important to also measure levels of population structure.



**Fig. 1** Approximate breeding and feeding distributions of the North Atlantic humpback whale and 3 stocks of the Southern Hemisphere humpback whale (as described in Rosenbaum et al. 2009; Johnson

and Wolman 1984). Arrows represent hypothesized migratory pathways. Sampling location names are followed by the number of samples in parentheses

Presently, humpback whales are divided into three oceanic populations, the North Atlantic, the North Pacific and the Southern Hemisphere, based on genetic and tagging data suggesting limited migration between ocean basins (Mackintosh 1965; Baker et al. 1993). Previous analysis of worldwide population structure based upon mtDNA suggests that humpback whales from the North Atlantic are most strongly differentiated from those in the North Pacific and less strongly differentiated from those in the Southern Hemisphere (Table 2, Baker et al. 1993). Strong divergence between North Atlantic and North Pacific humpback whales is thought to result from the fact that sea ice has likely blocked the main northern migratory corridor between the two groups since the Sangamonian Interglacial period (~140,000 years ago). As a result, genetic diversity within the North Atlantic is unlikely to be strongly influenced by past migration from the North Pacific. Thus, while we will test the assumption that gene flow with the North Pacific does not influence  $N_e$  in the North Atlantic, the main focus of our analysis will be on populations from the North Atlantic and the Southern Hemisphere.

Humpback whales within the North Atlantic and Southern Hemisphere exhibit varying degrees of within-

ocean sub-population structure resulting from complex patterns of breeding, feeding, and migration specific to each ocean region (Fig. 1). North Atlantic humpback whales show site fidelity to several discrete feeding areas extending from the Gulf of Maine to the Barents Sea off the northern coast of Norway, but individuals from all known feeding areas congregate on a common breeding area in the West Indies (Katona and Beard 1990; Smith et al. 1999; Stevick et al. 1999). Despite overlap on the West Indies breeding grounds, significant population structure between eastern and western North Atlantic feeding aggregations have been identified using mtDNA ( $K_{st} \sim 0.04$ ) (Palsboll et al. 1995) and nuclear loci ( $F_{st} \sim 0.036$ ) (Valsecchi et al. 1997).

Patterns of migratory connectivity in the Southern Hemisphere are less well understood, but recent evidence based upon mtDNA suggests low, but significant sub-population structure between Southwestern Atlantic, Southeastern Atlantic, and Southwestern Indian Ocean groups (Breeding Stocks A, B, and C respectively;  $F_{st}$  range 0.0029–0.0166) with the Northern Indian Ocean (stock X) falling out as strongly differentiated from all other groups ( $F_{st}$  range from 0.0797 to 0.1473) (Rosenbaum et al. 2009). In this study, we will use multiple

nuclear loci and increased sampling from within each ocean basin to gain a better perspective on the impact of population structure on long-term population size in the North Atlantic.

We calculate the long-term population size of the North Atlantic humpback whale within the context of the worldwide population structure. We focus on an in-depth analysis of the North Atlantic and the Southern Hemisphere, with particular focus on the South Atlantic, because previous data indicate that these two populations are the most likely to have exchanged migrants during a time period that may impact an estimate of effective population size. We identify strongly differentiated populations within and between the North Atlantic and Southern Hemisphere using nine nuclear loci and a multi-locus genetic clustering method. We then estimate long-term population size of the North Atlantic humpback whale, while accounting for the possibility of migration with other strongly differentiated groups. Our new estimate of long-term population size in the North Atlantic is compared with previous genetic (mtDNA) and catch-based estimates in order to highlight remaining uncertainties in estimates of pre-whaling abundance and discuss important areas for future research.

## Methods

### Sample collection and sequencing

Genetic samples representing 173 individuals were collected from humpback whales across the Southern Hemisphere (South Atlantic and Indian Oceans) and the North Atlantic Ocean (for regional sample sizes see Fig. 1). Biopsy samples from living whales were collected with appropriate national permits using protocols approved by the American Museum of Natural History and the Oregon State University's Animal Care and Use Committees. Samples were preserved in 70 % ethanol or salt saturated 20 % dimethyl sulfoxide solution (DMSO) and later stored at  $-20^{\circ}\text{C}$  until processed. Total genomic DNA was extracted using a standard phenol/chloroform extraction method or using a DNAeasy tissue kit (Qiagen).

Nine nuclear loci were amplified and sequenced using standard PCR and sequencing protocols (Saiki et al. 1988; Palumbi 1995) and published primers (Lyons et al. 1997) (Table 1; doi:10.5061/dryad.bj506). Individuals were sequenced in both directions for 8 of 9 loci and sequences were trimmed so that only the highest quality sequences were included in the consensus. We found that the inclusion of the reverse direction for RHO lowered the overall sequence quality. Thus, in order to avoid the possibility of artificially inflating our estimate of genetic diversity by including low quality sequence in our analysis, we restricted our analysis of

RHO to the 186 bp forward direction sequence. All variable sites for the 9 loci were checked by eye using Sequencher ver. 4.8 (Gene Codes Corporation). SNPs were verified through visual confirmation in forward and reverse sequences and/or in multiple individuals. SNPs that only occurred in one individual, could not be verified with reverse complement sequences, or could not be called with confidence were removed from the analysis. In order to ensure that our dataset did not contain replicate samples, we confirmed that no individual had the same sequence across all loci. Despite multiple attempts, not all individuals sequenced successfully for every locus, resulting in variation in the final sample sizes for each locus (NA mean 42, range 27–56; SH mean 101, range 80–117).

PHASE 2.1 (Stephens et al. 2001) was used to reconstruct gametic phase, defined as the original allele combination that an individual received from each of its parents, using a burn in of 10,000 iterations and a run length of 10,000 iterations. Using Arlequin ver. 3.0 (Excoffier et al. 2005) we found no significant linkage disequilibrium among loci after correcting for multiple comparisons. To determine if our sequences were evolving in a manner consistent with equilibrium and neutrality, Tajima (1989) and Fu (1997) tests were performed using DnaSP (Rozas et al. 2003). In neutrally evolving sequences, both values will be approximately equal to zero, while balancing selection or population expansion will result in values that are significantly greater or less than zero, respectively. We also used DnaSP to calculate the minimum number of recombination events in the sample (Hudson and Kaplan 1985) and found that 3 of 9 loci showed evidence of recombination. As a result, in loci with evidence for recombination, coalescent simulations ( $n = 1,000$ ) incorporating the per gene recombination parameter ( $R$ ) were used to generate 95 % confidence intervals (CI) for both Tajima's  $D$  and Fu's  $F_s$  statistics.

### Testing for population structure

An analysis of population structure was performed to investigate whether or not the major divisions within and between the North Atlantic and the Southern Hemisphere humpback whales remained with increased sampling in both regions. In order to properly account for migration that may impact our estimate of genetic variation ( $\theta$ ,  $\theta$ ) for the North Atlantic humpback whale, we test 3 population structure scenarios: (1) populations with no migration over recent evolutionary history (i.e.  $4 N_e$  generations), (2) genetically distinct populations connected by very limited migration, and (3) sub-populations that may be biologically meaningful, but are exchanging migrants at a high enough rate that they cannot be distinguished using multi-locus clustering methods. To estimate long-term population size, sub-populations (scenario 3) were lumped

**Table 1** Summary statistics for 9 introns sequenced in North Atlantic and Southern Hemisphere humpback whales

Intron	Seq. length	Ocean basin	N	N <sub>S</sub>	N <sub>H</sub>	R <sub>m</sub>	π	Tajima's D	Fu's Fs
ACT	886	NA	27	3	7	2	0.0016	<b>2.224</b>	−0.801
		SH	90	5	10	2	0.0015	1.089	−1.706
CAT	500	NA	40	1	2	0	0.0010	1.691	2.138
		SH	96	4	6	1	0.0012	−0.190	−1.390
FGG	941	NA	51	1	2	0	0.0004	0.070	1.300
		SH	116	5	6	0	0.0006	−0.589	−1.490
ESD	598	NA	38	5	6	1	0.0025	1.041	<b>0.656</b>
		SH	107	6	11	2	0.0017	−0.037	−3.873
GBA	298	NA	56	1	2	0	0.0002	−0.809	−1.146
		SH	117	2	3	0	0.0004	−0.841	−1.662
LAC	560	NA	31	1	2	0	0.0009	1.563	1.943
		SH	81	2	3	0	0.0008	0.231	0.555
PLP	810	NA	29	1	2	0	0.0004	0.579	1.088
		SH	107	3	4	0	0.0005	−0.296	−0.468
PTH	267	NA	55	2	3	0	0.0014	−0.006	0.128
		SH	112	2	3	0	0.0013	0.067	0.348
RHO	186	NA	56	3	5	1	0.0037	0.381	−0.495
		SH	80	3	6	2	0.0054	1.440	−0.008

*N* number of individuals, *N<sub>S</sub>* number of polymorphic sites, *N<sub>H</sub>* number of distinct haplotypes as determined by PHASE, ver. 2.1 (Stephens et al. 2001), *R<sub>m</sub>* minimum number of recombination events, *π* nucleotide diversity (Nei 1987)

\* Numbers in bold refers to a significant deviation from neutral expectation before a bonferroni correction for multiple comparison (*p* < 0.05), as determined by coalescent simulations of the null distribution using DNAsp (Rozas et al. 2003)

into respective population categories (scenario 2) and populations with no possibility of migration with the North Atlantic over recent evolutionary history were considered separately (scenario 1).

Pairwise *F<sub>st</sub>* within and between ocean basins at each locus as well as across all loci were calculated using the program Arlequin ver. 3.0 (Excoffier et al. 2005). A null distribution of *F<sub>st</sub>* was generated through 1,000 permutations of the haplotypes between populations and the *p* value represents the proportion of permutations leading to an *F<sub>st</sub>* larger than or equal to the observed value. To assess the potential for within and between ocean basin population structure within a multi-locus framework, we used the program Structure ver. 2.2 (Pritchard et al. 2000). Preliminary runs indicated that the power for assigning individuals to clusters dropped off significantly when individuals had missing data for more than 2 of the 9 loci. Therefore, individuals with missing data for more than 2 loci were removed from the multi-locus analysis in order to ensure that there was sufficient statistical power for assignment of all individuals to clusters. Structure requires unlinked markers, so the maximum a posteriori haplotypes from PHASE at each locus were recoded as alleles. We performed 3 independent runs at each *K* value (*K* = 1–5) using a burn-in period of 100,000 iterations and a run length of 500,000. The structure analysis

was run using the admixture model with correlated allele frequencies with and without the location prior.

The location prior is intended to use location information to help identify more subtle population structure, without detecting structure that is not present (Hubisz et al. 2009) and we implement it here in an attempt to identify a signature of population structure in our data that may influence our subsequent estimate of *q*. Locations included the Gulf of Maine (GOM), Dominican Republic (DR), Gabon (GA), Brazil (BR) and Madagascar (BA) (Fig. 1, SI Table 1). Individuals from NF were grouped within the GOM location due to low sample size from NF, geographic proximity between NF and GOM and the lack of significant *F<sub>st</sub>* values between NF and GOM (see “Results” section). We determined support for the number of clusters (*K*) by plotting the average *ln* [P(X|K)] of each model as a function of *K* and using the ad hoc  $\Delta K$  statistic proposed by Evanno et al. (2005).

#### Estimating $\theta$

Using our knowledge of population structure, we employed genetic models (Kuhner 2006) that estimate long-term *N<sub>e</sub>* while explicitly accounting for the possibility of migration between populations deemed distinct according to multi-

locus clustering methods. We use LAMARC ver. 2.1.3 to simultaneously estimate  $\theta$  while incorporating recombination and migration between ocean regions into the model. In contrast to summary statistic estimates of  $\theta$  ( $\theta_s$ ,  $\theta_\pi$ , etc.), LAMARC accounts for uncertainty in the data by integrating over the space of possible genealogies using a Markov chain Monte Carlo (MCMC) procedure. In order to account for uncertainty in the data resulting from unknown gametic phase and to accommodate inter-locus variation in mutation rate, we followed the methods described in Ruegg et al. (2010). In short, to account for unknown gametic phase, LAMARC was run on 15 realizations from PHASE's posterior distribution for each of 9 introns. In addition, as recommended by the LAMARC manual, we subsampled our data to restrict the input size for each LAMARC run to 20 sequences from each major population. Thus, for each of the 15 realizations from PHASE's posterior, LAMARC was run on a different subsample of 10 randomly chosen individuals from each population. The final result from these 15 LAMARC runs was obtained by catenating the summaries from all the runs following the recommendations in the LAMARC manual for "poor man's parallelization." (Initially we attempted 3 random subsamples from each phasing—45 total LAMARC runs—but this exceeded the memory available to LAMARC). To accommodate interlocus variation in mutation rate, we implemented the gamma model for mutation rate variation within a Bayesian framework using an extension of the LAMARC package known of as GUF-BUL (Gamma Updating for Bayesians Using LAMARC; Ruegg et al. 2010).

Our main objective was to estimate  $\theta$  in the North Atlantic while accounting for the possibility of migration with the Southern Hemisphere. To this end we used a 2-population migration model in LAMARC on the full dataset that included nine loci (Table 1). The fact that ice has blocked the main northern migratory corridor between the North Pacific and the North Atlantic since the Sangamonian Interglacial period ( $\sim 140,000$  years ago) makes gene flow between the two populations unlikely. However, to further investigate the possibility that genetic diversity in the North Atlantic is influenced by migration with the North Pacific, we ran a LAMARC analysis using a 3-population migration matrix on 6 of the nine loci for which we had sequence data.  $\theta$  values generated using the 3-population model were compared to values calculated using the 2-population model for each of the 6 loci.

#### Calculating census population size from $\theta$

The conversion of  $\theta$  into effective population size ( $N_e$ ) is based upon the relationship  $\theta = 4N_e\mu$  where  $\mu$  is the average mutation rate. To calculate an average  $\mu$  for North Atlantic Humpback whales, and to estimate uncertainty surrounding

our estimate, we followed the methods described in Ruegg et al. (2010). In short, we sampled with replacement from among 9 previously published individual locus mutation rates for humpback whales; 1 of the individual locus mutation rates (PLP) was from Alter et al. (2007), while the remaining 8 were taken from a Bayesian analysis of baleen whale phylogeny and fossil history (Jackson et al. 2009). For each re-sampled locus, a sample mutation rate was drawn from the posterior distribution of the estimated mutation rate or, for PLP, uniformly from the 95 % confidence intervals on the mutation rate. This was repeated 9 times for each bootstrap replicate, and we performed 100,000 bootstrap replicates. The mean  $\mu$  and the variability around that mean was obtained from these bootstrap replicates. To convert  $\mu$  from units of mutations per base pair per year into mutations per base pair per generation requires an estimate of the generation length. To approximate generation length we sampled uniformly from within a range of possible values for North Atlantic humpback whales of between 12 and 24 years (Chittleborough 1965; Roman and Palumbi 2003; Taylor et al. 2007). While this lower bound on generation time, taken from Chittleborough's (1965) estimate, may be low because of age-estimate inaccuracies, it is similar to the 14.5 year estimate for modern humpback whales from Taylor et al. (2007). Here we maintain the 12–24 year range in order to stay consistent with previous estimates of long-term population size in the North Atlantic humpback whale (Roman and Palumbi 2003), and discuss the implications of different generation times on estimates of  $N_e$ .

To convert  $N_e$  to census population size ( $N_c$ ) requires an estimate of the ratio of mature adults to the effective number of adults ( $N_{\text{mature}}/N_e$ ) and the proportion of juveniles in the population. Although  $N_{\text{mature}}/N_e$  is difficult to calculate in most natural populations, theory suggests this ratio approaches 2 in most populations with constant size (Nunney and Elam 1994). We based our estimate of  $N_{\text{mature}}/N_e$  on equation (1) in Nunney and Elam (1994):  $N_e = N/(2-T^{-1})$ , where  $T$  = generation length. To approximate juvenile abundance we used catch and survey data to calculate (no. of adults + juveniles)/(no. adults) (Chittleborough 1965; Roman and Palumbi 2003). To incorporate uncertainty in juvenile abundance we sampled uniformly from within a range of likely values for North Atlantic humpback whales.

## Results

### Tests for neutrality and equilibrium

Among the 9 nuclear introns, nucleotide diversity averaged 0.0014 (range 0.0002–0.0054), with an average of 5 haplotypes per locus (range 2–11) and an average of 43 samples from the North Atlantic and 101 samples from the Southern Hemisphere (Table 1). These values were similar

to other baleen whale species for which data are available (gray whales: range 0.0031–0.00016; Alter et al. 2007). While Tajima's *D* for ACT and Fu's *FS* for ESD were significantly different from the simulated null distribution given  $p < 0.05$ , neither remained significant after Bonferroni correction for multiple comparison (corrected  $p = 0.05/18$  tests = 0.003). The results of the Tajima's *D* and Fu's *FS* tests suggest the loci are evolving in a manner consistent with neutrality and equilibrium (Table 1).

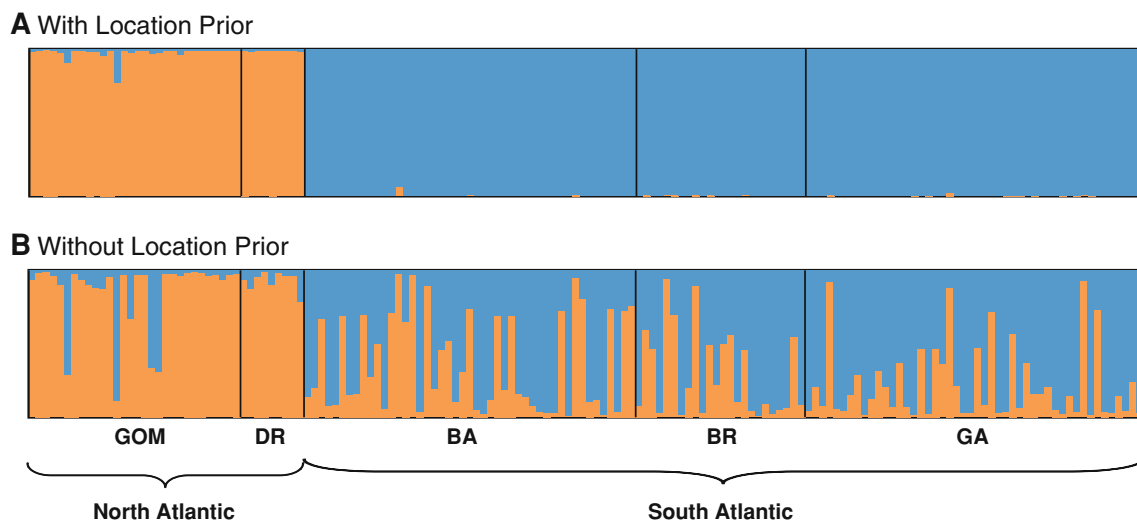
### Population structure

Across the 9 loci, *F*<sub>st</sub> ranged from 0 to 0.36 (SI Table 1), with 69 % (18 of 26) of the significant pairwise *F*<sub>st</sub> values being between North Atlantic and Southern Hemisphere populations, 31 % of the significant comparisons being between populations in the Southern Hemisphere, and 0 % coming from comparisons between populations in the North Atlantic. When the North Atlantic and Southern Hemisphere populations were grouped into two groups, the overall *F*<sub>st</sub> across all loci was 0.14. For the multi-locus analysis of population structure, inspection of the average log probability of the data ( $\ln [P(X|K)]$ ) and the *ad hoc*  $\Delta K$  statistic of Evanno et al. (2005) indicated  $K = 2$  was the most likely number of clusters in the data (SI Fig. 1a and b). A plot of the average  $\ln [P(X|K)]$  of each model as a function of  $K$  showed the likelihood increased substantially with an increase in  $K$  from 1 to 2, but increased to a lesser extent or decreased thereafter (SI Fig. 1a). Similarly,  $\Delta K$  was substantially greater for a  $K$  of 2 than for any other value of  $K$  (SI Fig. 1b). The results were the same without using the location prior (results not shown).

Summary plots of *Q*, the estimated membership fraction for each individual, for  $K = 2$  indicated that most individuals from the North Atlantic were assigned to cluster 1, while most individuals from the Southern Hemisphere were assigned to cluster 2 (Fig. 2). When the data were run with the location prior for all five populations, the only emergent multi-locus signal of population structure was between the North Atlantic and Southern Hemisphere. Without the location prior, the main signal was also between the North Atlantic and Southern Hemisphere, but it is clear that the two groups are connected by some migration.

### Estimating genetic diversity ( $\theta$ )

Estimating  $\theta$  in LAMARC using all 9 loci and allowing for migration between the North Atlantic and Southern Hemisphere resulted in a posterior mean  $\theta$  for the North Atlantic of 0.00096 (95 % CI 0.00048–0.0017; Table 2). From locus to locus,  $\theta$  ranged from 0.0003 to 0.0026 for the North Atlantic (Table 2) and from 0.0004 to 0.0031 for the Southern Hemisphere (SI Table 2), presumably reflecting variation among loci in mutation rate or coalescent history. A comparison between the two-population migration model (North Atlantic and Southern Hemisphere) and the three-population migration model (North Atlantic, Southern Hemisphere, and North Pacific) at the 6 loci for which we had sequence data confirmed that  $\theta$  in the North Atlantic was not significantly influenced by ancient migration with the North Pacific (SI Fig. 2). While the estimates of  $\theta$  from the two-population, 6 locus model (MPE 0.000843, 95 % CI 0.000519–0.003239) were slightly higher than the estimates from the three-population, 6 locus model (MPE 0.000747, 95 % CI 0.000495–0.004318) (SI Fig. 2), it



**Fig. 2** Results of the multi-locus population structure analysis conducted using STRUCTURE. **a** Despite using location information for all five populations, the only emergent multi-locus signal of population structure is between the North Atlantic and the Southern

Hemisphere populations of humpback whales. **b** Without the use of a location prior there is a weak, but consistent multi-locus signal of population structure between the North Atlantic and the Southern Hemisphere humpback whales

**Table 2** Theta values for the North Atlantic estimated using a two population (NA and SH) migration matrix

Marker	$\theta$	Min	Max
ACT	0.0012	0.00024	0.00353
CAT	0.0004	0.00002	0.00223
ESD	0.0011	0.00022	0.00433
FGG	0.0003	0.00001	0.00131
GBA	0.0005	0.00000	0.00323
LAC	0.0005	0.00003	0.00235
PLP*	0.0003	0.00002	0.00150
PTH	0.0018	0.00009	0.00745
RHO	0.0026	0.00019	0.01296
Posterior mean	0.0007	0.0005	0.0043

\* Located on the X chromosome

is clear that inter-locus variation is much greater than variation between the two models. Thus, we conclude that the 2-population, 9 locus model adequately captured variation in  $\theta$  within the NA.

#### Estimate of census population size from $\theta$

Using a mutation rate of  $4.40 \times 10^{-10}$  (95 % CI  $3.66 \times 10^{-10}$ – $5.29 \times 10^{-10}$ ) and a range of generation lengths from 12 to 24 years we calculated  $N_e$  for the North Atlantic humpback whale to be 31,900 (95 % CI 13,200–66,100). To convert  $N_e$  to  $N_c$  we estimated juvenile abundance and variation in reproductive success. We estimated juvenile abundance or the ratio of total population size to total adults to be between 1.6 and 2.0 based upon survey and catch data for humpbacks (Chittleborough 1965; Roman and Palumbi 2003). Using the ratio of  $N_{\text{ma-ture}}/N_e$  of 2 (Nunney and Elam 1994), we multiplied the

product of the two ratios by our estimate of effective population size for an estimate of census population size of 112,000 individuals (Fig. 3). Bootstrap re-sampling across the variation in mutation rate, generation lengths, the ratio of total population size to total adults and from the posterior distribution of effective size yields a 95 % CI for census size from 45,000 to 235,000.

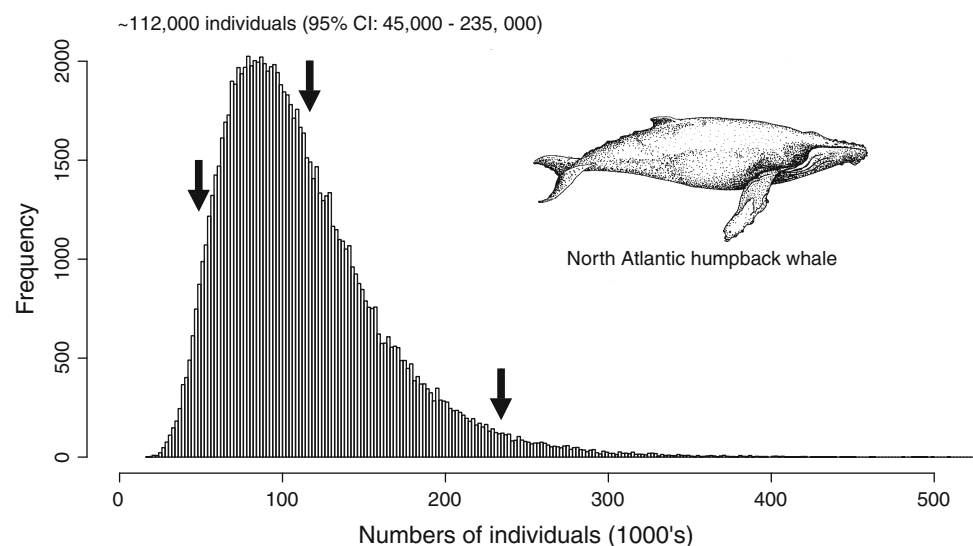
## Discussion

To improve estimates of long-term population size in the North Atlantic humpback whale, we have addressed recommendations for larger numbers of genetic loci, a better perspective on the impact of population structure, greater confidence in the mutation rate, and a greater focus on the historical timeframe of genetic population estimates (Clapham et al. 2005). Our new estimate of long-term population size of  $\sim 112,000$  individuals (95 % CI 45,000–235,000) is less than half of the previous mtDNA-based estimate of  $\sim 240,000$  (95 % CI 156,000–401,000) (Roman and Palumbi 2003), but is very similar to a revised population number of 150,000 (95 % CI 45,000–180,000) based on a more accurate estimate of the mutation rate (Alter and Palumbi 2009). However, the median of our most recent estimates remains far higher than the highest pre-whaling abundance estimate based upon catch data (notional upper limit: 40,000–47,000) (Smith and Pike 2009) and the discrepancy between the estimates warrants further discussion.

#### Population structure

Because genetic diversity within populations is strongly influenced by migration between populations, estimates of

**Fig. 3** Distribution of long-term census population size estimates, taking account of uncertainty in  $\theta$ , mutation rate, generation time, and the ratio of total population size to total adults. The arrows represent the mean value and the upper and lower 95 % confidence intervals



long-term population size must account for population structure.

We re-evaluated population structure within the humpback whale based upon previous work (Baker et al. 1993; Valsecchi et al. 1997; Olavarria et al. 2007; Rosenbaum et al. 2009) and our own multi-locus analysis. Our results confirm that  $\theta$  in the North Atlantic has not been significantly influenced by migration with the North Pacific (SI Fig. 2), unlike recent reports for Bowhead whales (Alter et al. 2012). Thus, our main analysis focused upon the North Atlantic and the Southern Hemisphere. While the locus-by-locus analysis revealed some signal of sub-population structure within the Southern Hemisphere (SI Table 1), the results of our within ocean basin multi-locus analysis indicate a lack of significant population structure overall, even when a strong prior for the presence of multiple sub-populations was included (Hubisz et al. 2009) (Fig. 2a). Overall, both the analysis of the average log probability of the data and the *ad hoc*  $\Delta K$  statistic indicate that  $K = 2$  is the most likely number of clusters in the data (SI Fig. 1). Consistent with previous research (Valsecchi et al. 1997; Olavarria et al. 2007; Rosenbaum et al. 2009), our results suggest that humpback whales within each ocean basin consist of two distinct populations connected by some migration.

One limitation with our study was the lack of samples from the eastern North Atlantic where previous research suggests the existence of a genetically distinct sub-population (Valsecchi et al. 1997). It is possible that additional samples from this region may have increased the number of distinct clusters found within the North Atlantic. However, because even small amounts of migration will cause sub-population  $N_e$  and whole population  $N_e$  to converge (Waples 2010; Hudson 1991), the absence of samples from the eastern North Atlantic sub-population is not likely to have influenced our estimate of long-term  $N_e$ . If, contrary to previous research, there is no migration between eastern and western North Atlantic feeding groups, then including samples from the eastern North Atlantic would increase our estimate of long-term population size.

#### Mutation rates

Attaining accurate estimates of mutation rates is a challenge common to all studies that use genetics to infer past population process (Ho et al. 2005; Emerson 2007). The difference between the original mtDNA-based estimate of  $\sim 240,000$  (Roman and Palumbi 2003), the updated mtDNA-based estimate of  $\sim 150,000$  (Alter and Palumbi 2009) and our multi-locus estimate of  $\sim 112,000$  individuals highlights the importance of mutation rates to estimates of long-term population size. In their revised estimate, Alter and Palumbi (2009) recalibrated the control

region mutation rate used in Roman and Palumbi (2003) by implementing a cytochrome b clock. Their analysis suggested that the previous mutation rate estimate was low by about two-fold because of multiple substitutions in the quickly evolving mtDNA control region. When the recalibrated control region mutation rate is employed, the mtDNA-based estimate becomes statistically indistinguishable from the multi-locus estimate of long-term population size.

Here we estimate an average mutation rate across nine nuclear loci using a phylogenetic reconstruction of the baleen phylogeny and fossil history (Jackson et al. 2009). One advantage of our multi-locus nuclear estimate is that whale nuclear DNA has far less saturation of substitutions than the mtDNA control region, and thus is far less likely to be subject to the same rate problems. Furthermore, our multi-locus approach incorporates uncertainty that results from random variation in the coalescent history of each individual locus (Rosenberg and Nordborg 2002). To adequately reflect the uncertainty in mutation rates in our final estimate of long-term population size, we bootstrap resampled across the variation in individual locus mutation rates. Thus, the multi-locus nuclear estimate that we present here should be a more robust approximation of the long-term  $N_e$  than the preceding mtDNA-only estimates.

#### Generation length

Uncertainty surrounding generation lengths interacts with mutation rate to determine estimates of long-term population size. Here we use a wide estimate of generation length for humpback whales ranging from 12 to 24 years (Chittleborough 1965; Roman and Palumbi 2003; Taylor et al. 2007) in order to remain consistent with previous estimates of long-term population size (Roman and Palumbi 2003). However, generation time in whales remains uncertain. Our lower bound of 12 years taken from Chittleborough (1965) is based on female age-size estimates from baleen condition, earplug layers and ovarian cycles. While Chittleborough (1965) provides the most extensive empirical data from which to estimate generation length in humpback whales, his estimates suffer from questions about age estimation (Gabriele et al. 2009; Best 2011) and whether older animals had already been culled (both of which would decrease estimates of generation length).

Taylor et al. (2007) estimated generation length for 58 cetacean species, including humpback whales, using mathematical models based on age at first reproduction and survival. They used an annual adult survival of 96 % and a first breeding age of 6 years to estimate a current generation time of 14.5 years and a stable pre-exploitation generation time of 21.5 years. While both of these estimates fall within our wide range on generation time, the

similarity between the current generation time of 14.5 years and the Chittleborough (1965) estimate of 12 years further highlights how exploitation may skew age patterns towards younger individuals. Furthermore, the absence of empirical data makes model-based estimates such as these especially sensitive to underlying assumptions. Better estimates should come from age distributions of real, unexploited populations, but such data is not readily available.

A longer estimate of humpback whale generation length would decrease our estimate of long-term population size (because the mutation rate per generation would increase). For example, if we use the estimate of 21.5 years taken from Taylor et al. (2007) we would decrease our estimate of population size to  $\sim 90,000$ , bringing it closer to catch based estimates of pre-whaling abundance. If we use the full range of generation times estimated through models by Taylor et al. (2007) for Baleanopterid whales (18–31 years), our mean estimate of long term population size would be  $\sim 81,000$  whales (95 % CI 34,000–163,000). These results highlight the sensitivity of genetic estimates of long-term population size to estimates of generation length. However, generation time would need to be much longer than suggested by previous estimates (in excess of 64 years) in order to bring our genetic estimate of long-term population size down as low as 30,000 whales (see also Roman and Palumbi 2003). If whale generation times were actually this long, it would have far reaching implications beyond the estimation of long-term population size.

#### Differences in time scales

Differences between genetic and catch-based estimates of past population size may arise from the fact that genetic estimates represent an average population size over evolutionary timescales, while catch-based estimates of past population size are calculated over more recent timescales. Long-term estimates of population size based on genetic data represent the weighted harmonic mean of population size over  $4N_e$  generations (e.g., up to 4,000 generations if  $N_e = 1,000$ ), but with greater weight on more recent time scales (Beerli 2009). Therefore it is possible that just prior to whaling, humpback whales were less abundant than their long-term average population size. This explanation would also need to be true of gray whales in the North Pacific (Alter and Palumbi 2009) but not minke whales in the Antarctic (Ruegg et al. 2010).

In the future, it will be important to investigate more fully how past environmental variation may have influenced long-term population size in whales and whether environmental conditions just prior to whaling would have supported a population at, above, or below the long-term average abundance. This information may be especially

helpful in predicting the effect of climate change on whale populations. If, for example, whale populations were generally higher during glacial maxima and lower during glacial minimum, then as the global oceans warm and ice melts, there may be a long-term decline in whale abundance. Such long-term data would be particularly useful in assessments of current and future whale conservation status.

#### Comparison between catch-based and genetic-based estimates of pre-whaling abundance

There has been substantial controversy surrounding the difference between genetic and catch-based estimates of pre-whaling abundance (Lubick 2003; Holt and Mitchell 2004; Clapham et al. 2005). In order to determine whether or not inaccuracies in the catch record lead to an underestimate of the number of whales before whaling, Smith and Reeves (2010) combined previously-used sources of information with additional data from archives to fill some gaps in our understanding of North Atlantic humpback whale removals. The results of their reanalysis indicate a new overall estimate of total removals that is only 6 % higher than that used previously by the IWC Scientific Committee (30,852, SE = 655). Thus, despite a reanalysis on both sides, our multi-locus estimate of average long-term population size remains higher than the pre-whaling estimate of abundance based upon catch records.

One approach to resolving these discrepancies has been population modeling (Baker and Clapham 2004). Here, historical catches, current information about reproductive rates and modern population estimates are joined together in an analytical framework that might be able to reconcile divergent views about past populations. However, population modeling performed by Punt et al. (2006) for the North Atlantic humpback shows a poor ability to explain past population crashes and current population growth. The problem stems from the fact that North Atlantic humpback whale populations can grow so quickly [6–7 % per year (Zerbini et al. 2010)], that their past populations should not have collapsed at the estimated hunting rates. For example, if North Atlantic humpback whales had an original population size of 30,000 animals, and a 6–7 % annual reproductive rate at maximum sustainable yield of 64 % of the original population size (19,200 animals), then the population as a whole should have sustained a hunt of 1,152 animals a year indefinitely. Yet, data from Smith and Reeves (2010), Fig. 1 show that there has never been a recorded catch of North Atlantic humpback whales that is this high: there were only two periods of time of a few years each when the total taken was above 400 animals per year. Even though the above estimate of sustainable yield is very crude, it demonstrates the large discrepancy between the catch record and the reproductive capacity of North Atlantic whale populations.

There are several possible explanations for these discrepancies between hunting and population growth. One alternative is that the carrying capacity of the ocean to support Atlantic humpback whales might have increased 2–3 fold during the 20th century (an assumption that has not yet been supported by data or theory), and that the maximum reproductive rate of humpback whales in the 19th century was extremely low (Punt et al. 2006). Alternatively, the models would be improved if catch rates were about twice as high as suggested by Smith and Reeves (2010). Given a higher rate of catch (about 43,000–69,000 over the course of the hunt instead of 29,000, (Punt et al. 2006, Table 1), the carrying capacity of humpback whales in the North Atlantic is estimated to be about 72,000–117,000 (Punt et al. 2006, Table 7). In general, for the ‘alternate baseline’ scenario that Punt et al. (2006) favor, models that suggest higher original estimates fit the data better (e.g. the negative log-likelihood values ( $-\ln L$ ) are closer to zero, (Punt et al. 2006, Table 6). These factors suggest that a larger historical number of humpback whales in the North Atlantic would better fit the catch data, the mathematical models and the genetic data.

The summary of these various threads of evidence is that estimates of historical abundance of North Atlantic humpback whales from catch and genetic data are converging, but remain about 2–3 fold apart. Older casual catch-based estimates of original population size from before 1990 (10,000–20,000) have been superseded by population models allowing for enhanced catch rates (20,000–46,000, Punt et al. 2006). The single locus mtDNA only genetic estimate moved from 240,000 to 150,000, after correcting for mutation rate and the addition of multiple nuclear loci resulted in an estimate of  $\sim 112,000$  (95 % CI 45,000–235,000). Further declines are possible if whale reproductive life times are vastly higher than currently supposed. Regardless of the generation length employed, the lower 95 % confidence limits on multi-locus estimates of long-term population size are now much closer to the range of the population models. Further closure of these differences may depend on population trajectories of whales during climate cycles, the development of population models that correctly reflect past population trajectories, and an enhanced view of whale generation times.

**Acknowledgments** We thank Barry Nickel with his help with the creation of Fig. 1. This work was supported by a grant from the Lenfest Ocean Program (#2004-001492-023).

## References

- Alter SE, Palumbi SR (2009) Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome B in three species of baleen whales. *J Mol Evol* 68:97–111
- Alter SE, Rynes E, Palumbi SR (2007) DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proc Natl Acad Sci USA* 104:15162–15167
- Alter ES, Rosenbaum HC, Postma LD, Whitridge P, Gaines C, Weber D, Egan MG, Lindsay M, Amato G, Dueck L, Brownell RL, Heide-Jørgensen M-P, Laidre KL, Caccone G, Hancock BL (2012) Gene flow on ice: the role of sea ice and whaling in shaping Holarctic genetic diversity and population differentiation in bowhead whales (*Balaena mysticetus*). *Ecol Evol* 2:2895–2911
- Baker CS, Clapham PJ (2004) Modeling the past and future of whales and whaling. *Trends Ecol Evol* 19:365–371
- Baker CS, Perry A, Bannister JL, Winrich MT, Abernethy RB, Camlambodikis J, Lien J, Lambertsen RH, Urban Ramirez J, Vasquez O, Clapham PJ, Alling A, O’Brien SJ, Palumbi SR (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc Natl Acad Sci USA* 90:8239–8243
- Beerli P (2009) How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? In: Population genetics for animal conservation. Cambridge University Press, The Pitt Building, Trumpington St, Cambridge, pp. 42–79
- Best PB (2011) A note on the age at sexual maturity of humpback whales. *J Cetacean Res Manage* 71–73
- Braham HW (1984) The status of endangered whales: an overview. *Mar Fish Rev* 46:2–6
- Chittleborough RG (1965) Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). *Aust J Mar Freshw Res* 16:33–128
- Clapham PJ, Palsboll PJ, Pastene LA, T. S, Walloe L (2005) Annex S: estimating pre-whaling abundance. *J Cetacean Res Manage* 7 (Suppl.), 386–387
- Emerson BC (2007) Alarm bells for the molecular clock? No support for Ho et al.’s model of time-dependent molecular rate estimates. *Syst Biol* 56
- Estes JA, Demaster DP, Doak DF, Williams TM, Brownell RLJ (eds) (2006) Whales, whaling and ocean ecosystems. California University Press, Berkeley
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinf (Online)* 1:47–50
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Gabriele CM, Lockyer C, Straley JM, Jurasz CM, Kato H (2009) Sighting history of a naturally marked humpback whale (*Megaptera novaeangliae*) suggests ear plug growth layer groups are deposited annually. *Mar Mammal Sci* 26:443–450
- Ho S, 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
- Holt SJ, Mitchell ED (2004) Counting whales in the North Atlantic. *Science* 303:39–40
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9:1322–1332
- Hudson RR (1991) Gene genealogies and the coalescent process. In: Oxford Surveys in Evolutionary Biology (eds. Futuyma D, Antonovics J). Oxford University Press, New York
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164
- IWC (2002) Annex H: Report of the Sub-committee on the comprehensive assessment of North Atlantic humpback whales. *J Cetacean Res Manage* 4 (Suppl.):230–260

- IWC (2003) Report of the subcommittee on the comprehensive assessment of humpback whales. *J Cetacean Res Manage* 5 (Suppl.):293–323
- Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR (2009) Big and Slow: phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). *Mol Biol Evol* 11:2427–2440
- Johnson JH, Wolman AA (1984) The humpback whale, *Megaptera novaeangliae*. *Mar Fish Rev* 46:30–37
- Katona SK, Beard JA (1990) Population size, migrations and feeding aggregations of the humpback whale (*Megaptera novaeangliae*) in the western North Atlantic Ocean. Report of the International Whaling Commission, 295–305
- Klinowska M (1991) Dolphins. Porpoises and Whales of the World, The IUCN Red Data Book IUCN
- Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–770
- Lubick N (2003) New count of old whales adds up to big debates. *Science* 301:451
- Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat Genet* 15:47–56
- Mackintosh NA (1965) The stocks of whales. Fishing News (Books) Ltd., London
- Mitchell E, Reeves RR (1983) Catch history, abundance, and present status of Northwest Atlantic humpback whales. *Rep Int Whal Commn* 5:153–212
- Myers RA, Worm B (2003) Rapid worldwide depletion of predatory fish communities. *Nature* 423:280–283
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY
- Nunney L, Elam DR (1994) Estimating the effective population size of conserved populations. *Conserv Biol* 8:175–184
- Olavarria C, Baker CS, Garrigue C, Poole M, Hauser N, Caballero S, Florez-Gonzalez L, Brasseur M, Bannister J, Capella J, Clapham P, Dodemont R, Donoghue M, Jenner C, Jenner MN, Moro D, Oremus M, Paton D, Rosenbaum H, Russell K (2007) Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. *Mar Ecol Prog Ser* 330:257–268
- Palsboll PJ, Clapham PJ, Mattila D, Larsen F, Sears R, Siegmund HR, Sigurjonsson J, Vasquez O, Arctander P (1995) Distribution of mtDNA haplotypes in North Atlantic humpback whales—the influence of behavior on population structure. *Mar Ecol Prog Ser* 116:1–10
- Palumbi SR (1995) Nucleic acids II: The polymerase chain reaction In: Hillis D, Moritz C, Mable BK (eds) *Molecular systematics*. Sinauer Associates Inc., Sunderland
- Palumbi SR, Roman J (2007) The history of whales read from DNA. In: Estes JA (ed) *Whales, whaling and ocean ecosystems*. University of California Press, Berkeley
- Pauly D, Christensen V, Dalsgaard J, Froese R, Jr. FT (1998) Fishing down Marine food webs. *Science* 279:860–863
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Punt AE, Friday NA, Smith TD (2006) Reconciling data on the trends and abundance of North Atlantic humpback whales within a population modelling framework. *J Cetacean Res Manage* 8:145–159
- Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science* 301:508–510
- Rosenbaum HC, Pomilla C, Mendez M, Leslie MS, Peter BB, Findlay KP, Minton G, Ersts PJ, Collins T, Engel MH, Bonatto SL, KDPG H, Meyer M, Barendse J, Thornton M, Razafindrakoto Y, Ngouesso S, Vely M, Kiszka J (2009) Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *PLoS ONE* 4:e7318
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* 3:380–390
- Rozas J, Sanchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR (2010) Are Antarctic minke whales unusually abundant because of 20th century whaling? *Mol Ecol* 19:281–291
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491
- Scott MJ, Goble DD, Wiens JA, Wilcove DS, Bean M, Male T (2005) Recovery of imperiled species under the Endangered Species Act: the need for a new approach. *Front Ecol Environ* 3:383–389
- Smith TD, Pike DG (2009) The enigmatic whale: the North Atlantic humpback. *NAMMCO Sci Publ* 7:161–178
- Smith TD, Reeves RR (2010) Historical catches of humpback whales, *Megaptera novaeangliae*, in the North Atlantic Ocean: estimates of landings and removals. *Mar Fish Rev* 72:1–43
- Smith TD, Allen J, Clapham PJ, Hammond PS, Katona S, Larsen F, Lien J, Mattila D, Palsboll PJ, Sigurjonsson J, Stevick PT, Oien N (1999) An ocean-basin-wide mark-recapture study of the North Atlantic humpback whale (*Megaptera novaeangliae*). *Mar Mammal Sci* 15:1–32
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Stevick PT, Oien N, Mattila D (1999) Migratory destinations of humpback whales from Norwegian and adjacent waters: evidence for stock identity. *J Cetacean Res Manage* 1:147–152
- Stevick PT, Allen J, Clapham PJ, Friday N, Katona SK, Larsen F, Lien J, Mattila DK, Palsboll PJ, Sigurjonsson J, Smith TD, Oien N, Hammond PS (2003) North Atlantic humpback whale abundance and rate of increase four decades after protection from whaling. *Mar Ecol Prog Ser* 258:263–273
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Taylor BL, Chivers SJ, Larese J, Perrin WF (2007) Generation length and percent mature estimates for IUCN assessments of cetaceans. National Marine Fisheries Service, Southwest Fisheries Science Center
- Valsecchi E, Palsboll PJ, Hale P, Glockner-Ferrari D, Ferrari M, Clapham PJ, Larsen F, Mattila D, Sears R, Sigurjonsson J, Brown M, Cockeron P, Amos B (1997) Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). *Mol Biol Evol* 14:355–362
- Waples RS (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Mol Ecol Resour* 10:785–796
- Winn HE, Reichley NE (1985) Humpback whale—*Megaptera novaeangliae* (Borowski, 1781). In: Ridgway SH, Harrison RG (eds) *Handbook of marine mammals*. Academic Press, London and Orlando, pp 241–273
- Zerbini AN, Clapham PJ, Wade PR (2010) Assessing plausible rates of population growth in humpback whales from life-history data. *Mar Biol (Berlin)* 157:1225–1236