SUPPORTING INFORMATION

Kershaw F, Carvalho I, Loo J, Pomilla C, Best P, Findlay K, Cerchio S, Collins T, Engel M, Minton G, Ersts P, Barendse J, Kotze PGH, Razafindrakoto Y, Ngouessono S, Meÿer M, Thorton M, and Rosenbaum HC. Multiple processes drive genetic structure of humpback whale (*Megaptera novaeangliae*) populations across spatial scales.

Procedure for checking genotype errors

First, automation was introduced whenever possible during PCR setup and manipulation of genomic DNA or PCR products. Negative controls were run at the PCR step to control for exogenous contamination. Two reference samples of known allele size were added to each amplification and subsequent analyses to standardize scoring. Scoring was automated in GENEMAPPER, and allele sizing was successively checked by hand. Samples that yielded ambiguous allele peaks were repeated a second time. Genotyping error was checked for the samples by re-amplifying and re-typing 15% of the total, chosen at random. In order to detect errors in our dataset, such as identifying possible non-amplified alleles (null alleles), large allele dropout, and scoring errors due to stutter peaks we used the programs DROPOUT v1.3 (McKelvey and Schwartz 2005) and MICRO-CHECKER v.2.2.3 (Van Oosterhout et al. 2004). Missing allelic data averaged 0.3% across all loci.

Overview of Discriminant Analysis of Principle Components (DAPC) methodology

The following overview has been adapted from Jombart et al. (2010). Principle Component Analysis (PCA) enables the identification of genetic structures in very large data sets within negligible computational time and the absence of assumptions about the underlying population genetics model. However, PCA does not provide group assessment and would require *a priori* definition of clusters to study population structure. In contrast, Discriminant Analysis (DA) is a multivariate method that defines a model in which genetic variation is partitioned into a between-group and a within-group component, and which maximizes the first while minimizing the second. This method therefore provides the best *discrimination* of individuals into pre-defined groups. However, DA requires the number of variables (alleles) to be less than the number of observations (individuals) and assumes uncorrelated variables.

Discriminant Analysis of Principle Components (DAPC) is a new method developed by Jombart et al. (2010) that relies on data transformation using PCA as a prior step to DA, which ensures that the variable submitted to the DA are perfectly uncorrelated, and that their number is less than that of analyzed individuals. The method assigns individuals to clusters and provides a visual assessment of between-population genetic structure. When group priors are unknown, the method employs K-means clustering of principle components to identify groups of individuals. The best-supported number of clusters is assessed using the Bayesian Information Criteria (BIC).

Selection of number of PC axes retained in DAPC

The number of PC axes that explain the largest amount of total genetic variability in the data set while achieving the best discrimination between populations was determined using the optima.a.score function. All discriminant analyses (DA) axes were retained to capture the maximum amount of variability within the data set (Warmuth *et al.* 2012).

Sequential K-means clustering in DAPC

The number of genetic clusters in the data set was estimated without *a priori* population information using sequential K-means clustering (Legendre & Legendre 1998; see Supplementary Materials). The Bayesian Information Criterion (BIC) was used to determine the optimal number of clusters by selecting the value of K after which the BIC either increased or decreased by a minimal amount (Warmuth *et al.* 2012). Structure was also tested for each sequential value of K for K = 1 - 20 by examining individual assignment plots.

Sub-sampling protocol for MIGRATE

Given the large number of individuals in our sample, their unequal distribution among populations, and the fact that including more individuals does not necessarily improve estimates but only increases computation time due to the augmented complexity of the genealogies (Beerli 1998), we chose to sub-sample our data set prior to analysis. The data set was randomly sub-sampled without replacement so that a maximum of 50 samples were included for each population (Pomilla 2005). We checked the consistency of results between repeated runs for two different sub-sets of data.

Sub-sampling protocol for BAYESASS

Due to inconsistencies in the results from initial runs, which seemed to be due to the disproportionate sample size of BSB1 and BSC3, we randomly sub-sampled these two populations without replacement resulting in 150 individuals from each population being included in the final data set. The analysis was conducted on two different random sub-sets and the results were compared for consistency.

Selection of mixing parameters for BAYESASS analysis

Short MCMC chains were conducted (0.08% completion) to determine appropriate values for the mixing parameters for allele frequencies (Δ_A), inbreeding coefficients (Δ_F), and migration rates (Δ_M). Mixing parameters were chosen so that acceptance rates remained within the optimal range of 20-60% (Rannala 2007). Final mixing parameter values for each data partition were as follows: total sample, Δ_A=0.3, Δ_F=0.4, Δ_M=0.2; male and female samples, Δ_A=0.6, Δ_F=0.8, Δ_M=0.4.

Table S1: Pairwise fixation index values obtained between humpback whale breeding stocks and substocks for F_{ST}, R_{ST}, and Jost's *D*. Values are shown for the total sample, and males and females, separately. * indicates statistical significance at *p<0.05, **p<0.01, ***p<0.001.

	Total			**Male**			**Female**		
	F_{ST}	R_{ST}	Jost's *D*	F_{ST}	R_{ST}	Jost's *D*	F_{ST}	R_{ST}	Jost's *D*
A/B1	0.004**	0.001	0.011	0.004*	0.001	0.008	0.004	-0.002	0.018
A/B2	0.007**	0.003	0.026**	0.005*	-0.001	0.013	0.008*	0.006	0.040*
A/C1	0.008***	0.002	0.027**	0.006*	0.006	0.012	0.009*	0.000	0.030*
A/C2	0.010***	0.004	0.024	0.008	0.012	-0.001	0.009*	0.002	0.021
A/C3	0.007***	0.006	0.025**	0.005*	0.006	0.012*	0.007*	0.001	0.038**
A/ASHW	0.055***	0.076***	0.161**	0.044***	0.088***	0.119***	0.065***	0.056*	0.181***
B1/B2	0.001*	0.000	0.005*	0.001	-0.002	0.000	0.000	-0.001	0.001
B1/C1	0.002***	0.000	0.010**	0.002*	-0.001	0.005*	0.001	0.008*	0.006
B1/C2	0.002*	0.004	0.005	0.001	0.006	-0.009	0.000	0.001	-0.004
B1/C3	0.001***	0.001*	0.006**	0.001*	0.001	0.005***	0.001*	0.001	0.006***
B1/ASHW	0.046***	0.056***	0.155**	0.041***	0.062**	0.107***	0.049***	0.033*	0.175***
B2/C1	0.001*	0.000	0.007	0.000	-0.003	-0.004	0.002	0.007	0.009*
B2/C2	0.002	0.003	0.008	-0.002*	0.006	-0.019	0.000	-0.002	0.005
B2/C3	0.000	0.000	0.002	-0.001***	-0.001	-0.004	0.000	-0.001	0.004
B2/ASHW	0.045***	0.056***	0.151**	0.041***	0.063**	0.100***	0.044***	0.030	0.158***
C1/C2	0.002	0.000	0.004	-0.001	0.009	-0.017	-0.001	-0.001	-0.006
C1/C3	0.001*	0.000	0.002	0.000	-0.001	-0.002	0.001	0.004	0.008*
C1/ASHW	0.037***	0.048***	0.125**	0.034***	0.070***	0.080***	0.038***	0.012	0.131***
C2/C3	0.001	0.001	0.002	-0.001	0.003	-0.013	0.000	-0.002	-0.003
C2/ASHW	0.047***	0.025*	0.135**	0.048***	0.013	0.102***	0.046***	0.004	0.145***
C3/ASHW	0.044***	0.047***	0.144**	0.034***	0.054**	0.099***	0.000***	0.025	0.154***

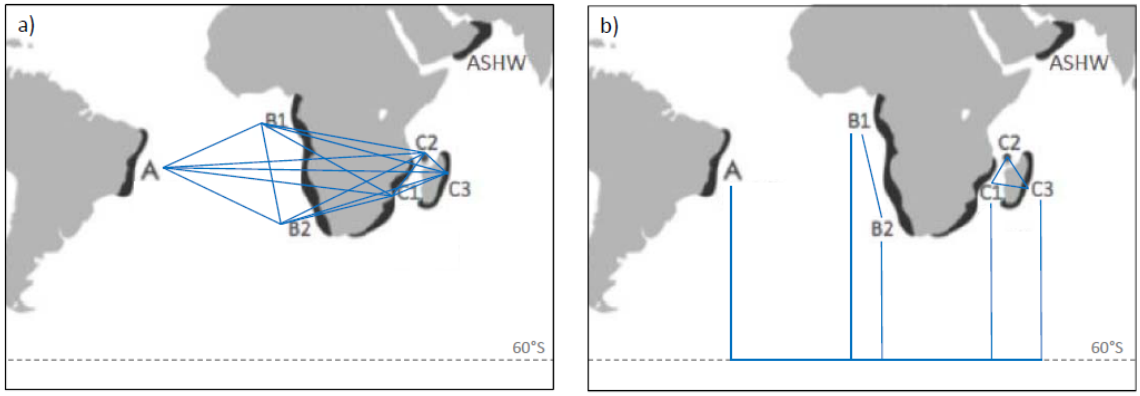

Figure S1: Two methods of calculating geographical distance for the isolation by distance (IBD) analysis: a) Euclidean geographical distance between all sample sites, and b) geographical distance based on current understanding of humpback whale longitudinal movements and mixing in the Southern Ocean. Geographical distance was calculated in kilometers (km) using ArcMap v. 10.3. ASHW was excluded from the IBD analysis due to its long-term isolation from the other breeding stocks and substocks.

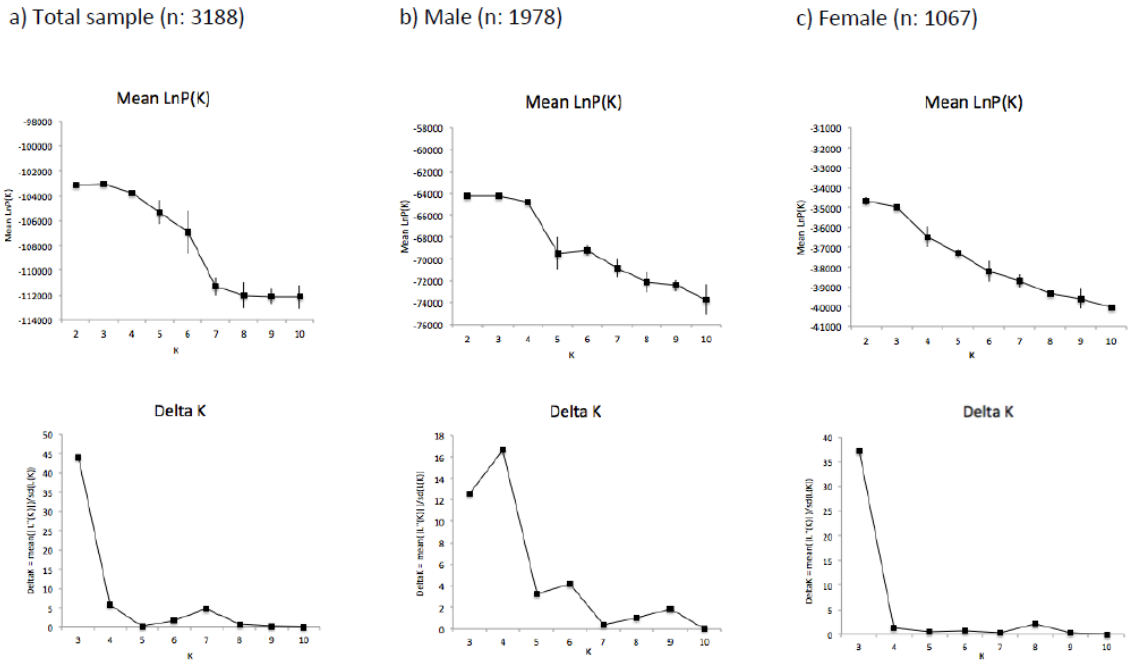

Figure S2: Mean LnP(K) and Delta K (ΔK) plots for the STRUCTURE outputs for the a) total sample, b) males, and c) females. For mean LnP(K), variance is indicated by error bars.

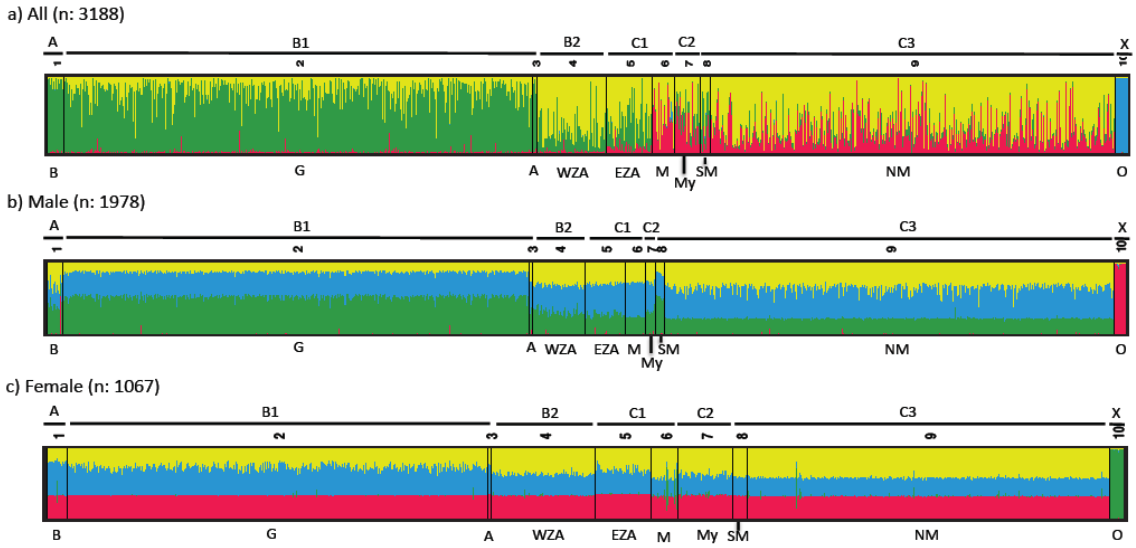

Figure S3: Distribution of 4 genetic clusters estimated using STRUCTURE for a) the total sample, b) males, and c) females. Vertical lines are partitioned into colored segments showing the proportion of each individual assigned to each K. Breeding stocks are indicated above each figure and sampling locations are below (B, Brazil; G, Gabon; A, Angola; WZA, West South Africa; EZA, East South Africa; M, Mozambique; My, Mayotte & Comoros; SM, South Madagascar; NM, North Madagascar; O, Oman).

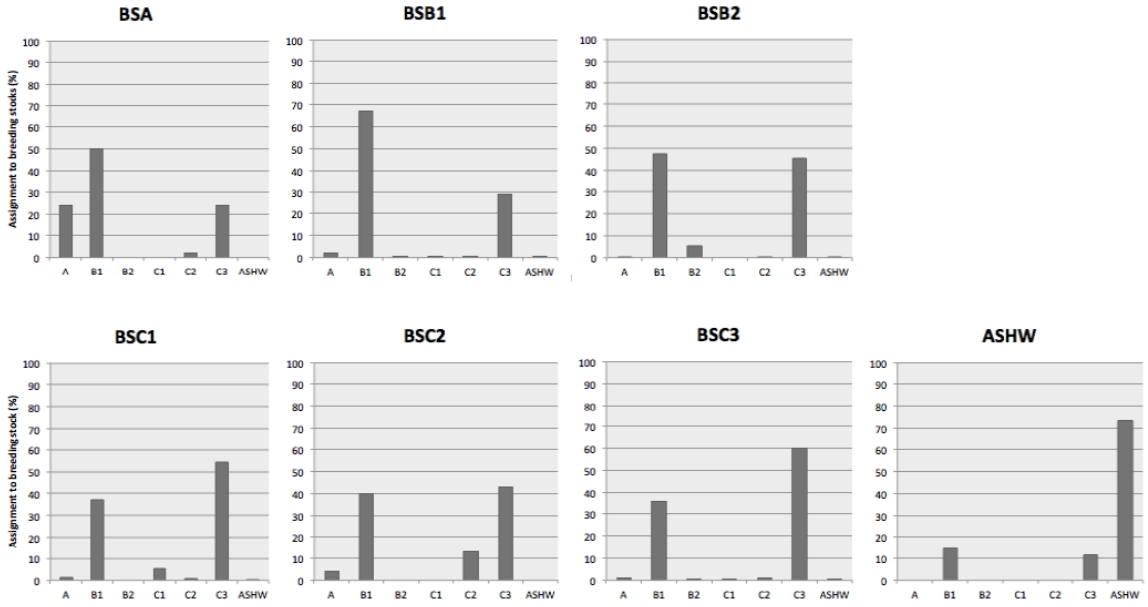

Figure S4: Distribution of individual reassignment of each breeding stock and substock by the DAPC.

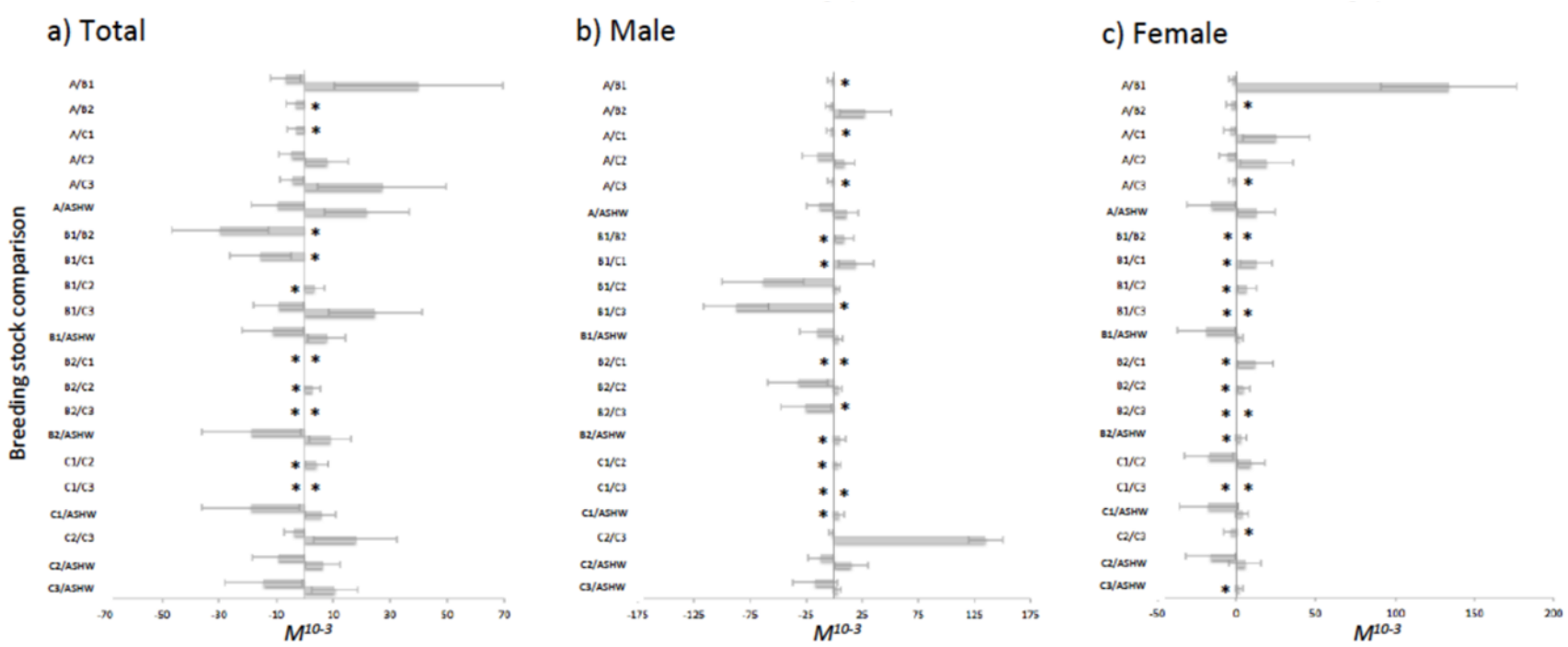
a) Total
b) Male
c) Female
Breeding stock comparison
A/B1
A/B2
A/C1
A/C2
A/C3
A/ASHW
B1/B2
B1/C1
B1/C2
B1/C3
B1/ASHW
B2/C1
B2/C2
B2/C3
B2/ASHW
C1/C2
C1/C3
C1/ASHW
C2/C3
C2/ASHW
C3/ASHW
M10-3

Figure S5: Magnitude and directionality of contemporary gene flow as estimated using BayesAss. The estimated proportion of migrants from one population to another are shown for a) the total sample; b) males; and c) females. Note the varying magnitudes of *M* for each data partition on the horizontal axes. Left bars indicate a westerly migration direction between the two breeding stocks; right bars indicate an easterly direction. Asterisks indicate comparisons where no convergence was reached. Results were transformed to aid visualization (see Materials and Methods).