MOLECULAR ECOLOGY

Molecular Ecology (2017) doi: 10.1111/mec.13943

Multiple processes drive genetic structure of humpback whale (*Megaptera novaeangliae*) populations across spatial scales

FRANCINE KERSHAW,*[1] INÊS CARVALHO,†‡ JACQUELINE LOO,§ CRISTINA POMILLA,¶ PETER B. BEST,** KEN P. FINDLAY,** SALVATORE CERCHIO,††[2] TIM COLLINS,†††‡‡ MARCIA H. ENGEL,§§ GIANNA MINTON,‡‡ PETER ERSTS,¶¶ JACO BARENDSE,** P. G. H. KOTZE,*** YVETTE RAZAFINDRAKOTO,†††[3] SOLANGE NGOUESSONO,‡‡‡ MICHAEL MEŸER,*** MEREDITH THORTON** and HOWARD C. ROSENBAUM†††§§§

*Columbia University, 116th Street and Broadway, New York, NY 10027, USA, †Population and Conservation Genetics Group, Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156, Oeiras, Portugal, ‡Centre for Environmental and Marine Studies (CESAM), Universidade de Aveiro, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal, §Department of Biology, New York University, 100 Washington Square, New York, NY 10012, USA, ¶Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK, **Mammal Research Institute, University of Pretoria, c/o Iziko South African Museum, P.O. Box 61, Cape Town 8000, South Africa, ††Wildlife Conservation Society, Ocean Giants Program, 2300 Southern Blvd., Bronx, NY 10460-1099, USA, ‡‡Environment Society of Oman, P.O. Box 3955, PC 112, Ruwi, Sultanate of Oman, §§Humpback Whale Project/Humpback Whale Institute, Rua Barão do Rio Branco, 125, Caravelas, Bahia, Brazil, ¶¶Center for Biodiversity and Conservation, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA, ***Department of Environmental Affairs, Branch Oceans and Coasts, Private Bag x2, Roggebaai 8012, Cape Town, South Africa, †††Wildlife Conservation Society-Madagascar Program, 2300 Southern Blvd., Bronx, NY 10460-1099, USA, ‡‡‡Agence Nationale des Parcs Nationaux, Batterie 4, BP 20379, Libreville, Gabon, §§§Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA*

Abstract

Elucidating patterns of population structure for species with complex life histories, and disentangling the processes driving such patterns, remains a significant analytical challenge. Humpback whale (*Megaptera novaeangliae*) populations display complex genetic structures that have not been fully resolved at all spatial scales. We generated a data set of nuclear markers for 3575 samples spanning the seven breeding stocks and substocks found in the South Atlantic and western and northern Indian Oceans. For the total sample, and males and females separately, we assessed genetic diversity, tested for genetic differentiation between putative populations and isolation by distance, estimated the number of genetic clusters without *a priori* population information and estimated rates of gene flow using maximum-likelihood and Bayesian approaches. At the ocean basin scale, structure is governed by geographical distance (IBD $P < 0.05$) and female fidelity to breeding areas, in line with current understanding of the drivers of broadscale population structure. Consistent with previous studies, the Arabian Sea breeding stock was highly genetically differentiated (F_{ST} 0.034–0.161; $P < 0.01$ for all comparisons). However, the breeding stock boundary between west South Africa and east Africa was more porous than expected based on genetic differentiation, cluster and geneflow analyses. Instances of male fidelity to breeding areas and relatively high rates of dispersal for females were also observed between the three substocks in the western Indian Ocean.

Correspondence: Francine Kershaw, Fax: (212) 727 1773; E-mail: fkershaw@nrdc.org; and Howard C. Rosenbaum, E-mail: hrosenbaum@wcs.org

[1] Present address: Natural Resources Defense Council, 40 West 20th Street, New York, NY 10011, USA

[2] Present address: New England Aquarium, Central Wharf, Boston, MA 02110-3399, USA

[3] Present address: COSAP Sahamalaza Miaro Dugong, C/O Madagascar National Parks Sahamalaza, Maromandia, Madagascar

The relationships between demographic units and current management boundaries may have ramifications for assessments of the status and continued protections of populations still in recovery from commercial whaling.

Keywords: behaviour, humpback whale, International Whaling Commission, population genetics, Southern Hemisphere, wildlife management

Received 11 March 2016; revision received 1 October 2016; accepted 16 November 2016

Introduction

The field of molecular ecology has contributed significant insights into patterns of population structure for a broad range of terrestrial and marine species (e.g. Wang *et al.* 2009; Mendez *et al.* 2010; Kormann *et al.* 2012). However, understanding patterns of population structure for species with complex life histories, and the processes driving those patterns, remains a significant challenge. Genetic population structure (i.e. the spatial and temporal distribution of allele frequencies) may be influenced by a variety of interacting processes, including behavioural and ecological responses (Andrews *et al.* 2010; Piou & Prévost 2012), environmental conditions (Kormann *et al.* 2012) and microevolutionary factors such as genetic drift and gene flow (Gaggiotti *et al.* 2009), all of which operate against a background of phylogeographic history (Muscarella *et al.* 2011). Disentangling the processes influencing population patterns therefore requires an integrative analytical approach (Gaggiotti *et al.* 2009).

The genetic architecture of migratory species is often complex due to the evolution of behaviours related to reliance on ephemeral patches of breeding and foraging habitat, such as group cohesion and hysteresis (or 'memory') effects (Guttal & Couzin 2010). At regional scales, population-level fidelity to breeding and feeding areas may be a primary driver of genetic structure in these species (Guttal & Couzin 2010); however, at local scales there may be a more nuanced interplay of processes. For instance, genetic divergence between colonies of Cook's petrel (*Pterodroma cookii*) is linked to habitat specialization that segregates different populations during the nonbreeding season (Rayner *et al.* 2011), and spinner dolphins in Hawaii (*Stenella longirostris*) exhibit two alternative social strategies associated with different levels of gene flow between social groups (Andrews *et al.* 2010). Analyses of biparentally inherited molecular markers can help to shed light on how patterns of population structure may be influenced by processes operating across different spatial and temporal scales (Amaral *et al.* 2012a,b).

One of the best-studied migratory marine species is the humpback whale (*Megaptera novaeangliae*), which migrates annually from low-latitude breeding areas to high-latitude feeding areas (Gambell 1976). Humpback whale genetic structure at the ocean basin scale is driven by a combination of maternal fidelity to feeding areas and natal philopatry to breeding areas (Baker *et al.* 1998, 2013). Patterns of migratory fidelity for baleen whales result from the close dependency of a first-year calf on its mother during the first complete annual migration, and thus vertical cultural transmission of migratory route and destinations (Baker *et al.* 1987, 2013; Alter *et al.* 2009; Valenzuela *et al.* 2009; Barendse *et al.* 2013). This mechanism of information transfer, from mother to calf, contrasts with natal philopatry found in the majority of other migratory marine species, such as sea turtles and sharks, which is likely driven by environmental cues or genetic inheritance (Shamblin *et al.* 2012; Baker *et al.* 2013; Feldheim *et al.* 2014). However, as observed for other migratory baleen whale species in both hemispheres (Alter *et al.* 2012; Kershaw *et al.* 2013), genetic studies of humpback whales continue to reveal more complex structure at finer spatial scales than accounted for in current stock designations (e.g. Rosenbaum *et al.* 2009; Carvalho *et al.* 2014; Schmitt *et al.* 2014), indicating that other behavioural mechanisms may be driving humpback whale genetic structure at the population and subpopulation levels.

Demographically discrete humpback whale 'breeding stocks' are designated by the International Whaling Commission (IWC) globally for assessment purposes. Some of these stocks have been divided into further 'substocks' due to genetic differentiation in samples taken from within the stock being suggestive of a level of demographic independence (IWC 2006). In the Southern Hemisphere, these include Breeding Stock A (BSA) and BSB in the southwest and southeast Atlantic, respectively, and BSC in the southwest Indian Ocean. A fourth, termed the Arabian Sea humpback whale (ASHW) population, is often discussed in the context of Southern Hemisphere stocks, even though it is in the Northern Hemisphere (Fig. 1). BSA shows relatively little diversity or genetic substructure (Cypriano-Souza *et al.* 2010); however, direct movements and song similarity between BSA and the west and east coasts of Africa indicate some degree of broadscale connectivity (Darling & Sousa-Lima 2005; Stevick *et al.* 2011), the significance of which is not yet fully understood. BSB has been partitioned into two substocks (BSB1 and BSB2;

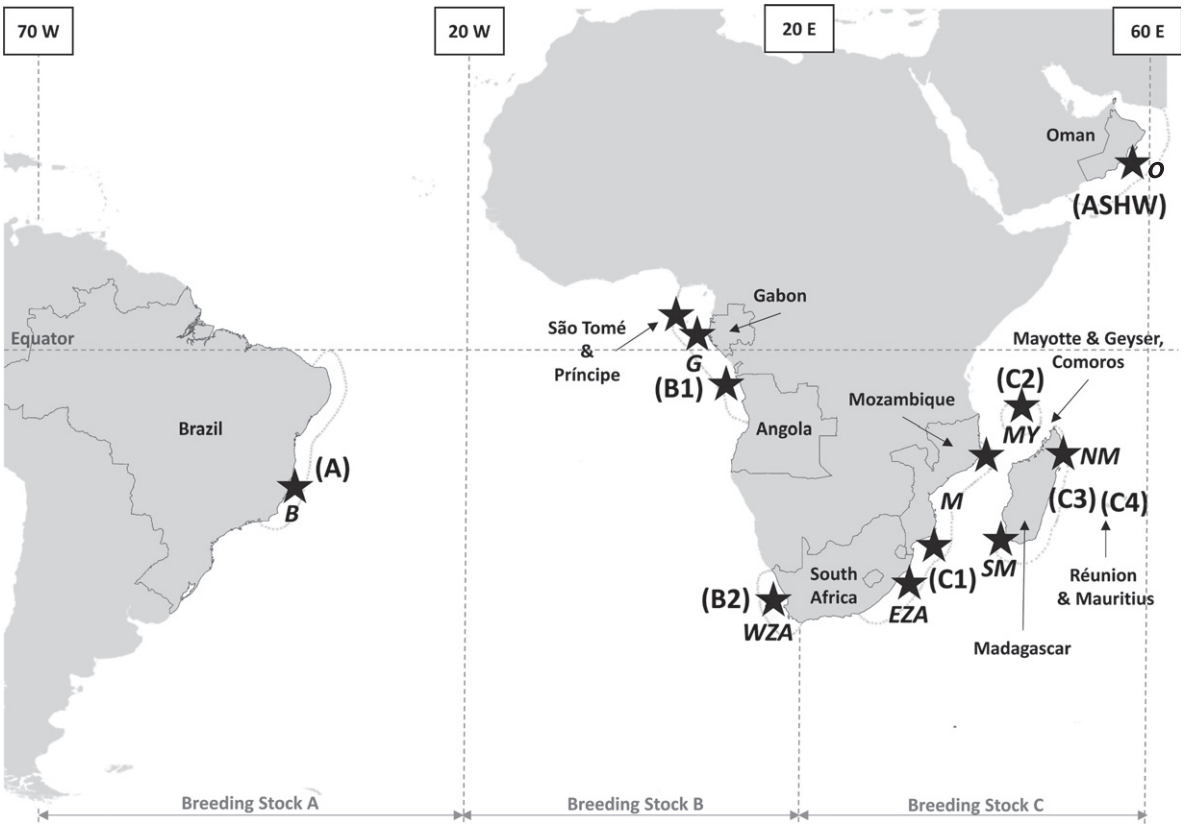

Fig. 1 Map showing sampling locations for the humpback whale breeding stocks and substocks analysed in this study. The location of breeding stocks and substocks are indicated by white shading and labelled in parentheses. Sampling locations are indicated by stars and labelled as follows: *B*, Abrolhos Bank, Brazil; *G*, comprising samples from Iguela and Gamba, Gabon, Cabinda region, Angola and São Tomé & Príncipe; *WZA*, Cape Columbine, West South Africa; *EZA*, Richard's Bay, East South Africa; *M*, comprising samples from Cabo Inhaca and Mozambique Island, Mozambique; *MY*, Mayotte and Geyser-Zelee, Comoros Archipelago; *SM*, Tulear, Southwest Madagascar; *NM*, Antongil Bay, Northeast Madagascar; *O*, Gulf of Masirah and Dhofar, Oman.

Fig. 1); however, the spatial and temporal boundaries of each substock remain unclear (Rosenbaum *et al.* 2009; Carvalho *et al.* 2014). Direct movement and song similarity have also been observed between BSB and BSC (Pomilla & Rosenbaum 2005; G.A. Carvajal, M.L. Rekdahl, E.C. Garland, T. Collins, Y. Razafindrakoto & H.C. Rosenbaum unpublished data), possibly indicating demographically meaningful levels of connectivity between the South Atlantic and western Indian Ocean. Differences in levels of migrant exchange and records of individual movements between the four substocks of BSC (BSC1-C4) also suggest genetic structure may be more complex than currently considered (Rosenbaum *et al.* 2009; Ersts *et al.* 2011; Fossette *et al.* 2014). In contrast, the ASHW has been established as the only known nonmigratory humpback whale population globally and is known to be small (approximately 80–200 individuals) and extremely isolated (Rosenbaum *et al.* 2009; Minton *et al.* 2011; Pomilla & Amaral *et al.* 2014).

A more complete understanding of patterns of humpback whale population structure using nuclear microsatellite markers, and an exploration of the potential processes underlying those patterns, has therefore not yet been achieved at multiple spatial scales. Addressing this knowledge gap is essential, not only to better understand the ecology and evolutionary biology of the species, but also to directly inform the conservation and management of humpback whales at a level that reflect demographically discrete populations.

We therefore present an analysis of an extensive biparentally inherited genetic data set to further elucidate population genetic patterns across the south Atlantic and southwestern and northern Indian Ocean. We partition our analyses to undertake a detailed investigation of the influence of sex on dispersal and site fidelity on population genetic structure. This large data set of genotyped individuals enables the detection of low levels of interchange for both sexes not previously quantifiable from examination of haplotype frequencies alone due to the inadequate power to detect low levels of gene flow (i.e. <100 migrants per generation) (Baker *et al.* 2013). Therefore, our analysis also provides

additional resolution to the population substructure previously observed within BSB, off West Africa, and BSC, off East Africa and Madagascar (Rosenbaum *et al.* 2009; Carvalho *et al.* 2014).

Materials and methods

Laboratory protocols

Sample collection, DNA extraction and sex determination. A total of 3575 humpback whale genetic samples originating from multiyear collections across 12 sampling locations were used in this study (Table 1, Fig. 1; no genetic samples were available from BSC4). Skin tissues were mostly obtained using biopsy darts (Lambertson 1987), but also from sloughed skin and stranded specimens. Samples were preserved in 95% ethanol or salt-saturated 20% dimethyl sulfoxide solution (DMSO) and later stored at −20°C until processed. Total genomic DNA was extracted from the tissue samples using proteinase K digestion, followed by a standard phenol/chloroform extraction method (Sambrook *et al.* 1989) or using QIAamp Tissue Kit (QiaGen) following manufacturer's protocol. Sex determination was either carried out by polymerase chain reaction (PCR) amplifications followed by TaqI digestion of the ZFX/ZFY region of the sex chromosomes (Palsbøll *et al.* 1992) or using multiplex PCR amplification of the ZFX/ZFY sex-linked gene (Berube & Palsbøll 1996).

Microsatellite molecular analyses. Samples were genotyped at 10 microsatellite loci proven to be polymorphic for this species: GATA028, GATA053, GATA417 (Palsbøll *et al.* 1997), 199/200, 417/418, 464/465 (Schlötterer *et al.* 1991), EV1Pm, EV37Mn, EV94Mn, EV96Mn (Valsecchi & Amos 1996). The 5′-end of the forward primer from each locus was labelled with a fluorescent tag (HEX, 6-FAM, and TET, Qiagen-Operon; NED, Applied Biosystems, Inc). PCRs were carried out in a 20 μL volume with the following conditions: 50 mM KCl, 10 mM Tris-HCl pH8.8, 2.5–3.5 mM MgCl2, 200 μM of each dNTP, 0.4 μM of each primer and 0.025 U/μL Taq Gold polymerase (PerkinElmer). Amplifications were completed in an Eppendorf Gradient Mastercycler, after optimization based on published articles characterizing the loci (Schlötterer *et al.* 1991; Valsecchi & Amos 1996; Palsbøll *et al.* 1997). PCR products were loaded with the addition of an internal standard ladder (GS600 LIZ, ABI) on a 3730xl DNA Analyzer (Applied Biosystems, Inc). Microsatellite alleles were identified by their sizes in base pairs using the software GENEMAPPER v4.0 software (ABI). Specific guidelines were used during laboratory work and scoring procedures to reduce genotyping errors (Supporting Information).

The probability of different individuals and siblings sharing the same genotype by chance (probability of identity, P_{ID}, and P_{ID} for siblings, $P_{ID(sibs)}$, respectively) were estimated using Cervus (Kalinowski *et al.* 2007). The reciprocal of the sample size was used as the arbitrary cut-off below which the probability values are sufficiently small to conclude that matching genotypes belong to the same individual (Peakall *et al.* 2006). Duplicate samples were removed from subsequent analyses. Related individuals were retained in the sample

Table 1 Sample location, size and diversity indices for nine microsatellite loci across breeding grounds, migratory corridors and feeding grounds of humpback whales sampled in the South Atlantic and Western Indian Ocean. STP, São Tomé & Príncipe; N, sample size; M, number of males; *F*, number of females; *k*, mean number of alleles per locus; H_o, observed heterozygosity; H_e, expected heterozygosity. The sum of the number of males and females does not always match the total sample size because the sex of some individuals was indeterminant. Duplicate samples have been removed.

Breeding ground & breeding stock/substock	N	M	F	M: F	Years	*k*	H_o	H_e
(A) Southwestern Atlantic Ocean								
Abrolhos, Brazil	50	30	20	1.5: 1	1997–98	10.11	0.702	0.715
(B) Southeastern Atlantic Ocean								
(B1) Gabon, STP, Angola	1395	826	421	2: 1	1999–2006	12.89	0.732	0.735
(B2) West South Africa	204	95	103	1: 1.1	1990, 93, 95 2000–2009	11.33	0.740	0.737
(C) Southwestern Indian Ocean								
(C1) Mozambique & East South Africa	203	112	81	1.4: 1	1991 1997–2005	12.00	0.742	0.738
(C2) Mayotte & Geyser, Comoros	75	17	55	1: 3.2	1997–2003	10.44	0.723	0.735
(C3) Madagascar	1227	842	373	2.3: 1	1994 1996–2006	12.78	0.731	0.729
(ASHW) Northern Indian Ocean								
Oman	34	20	14	1.4: 1	2001–2002	6.44	0.706	0.678

as relatedness between humpback whales is, in general, extremely low and unlikely to confound the results (Pomilla & Rosenbaum 2006; Kershaw 2015).

Data analysis

Diversity estimates. Genetic diversity was measured as the mean number of alleles per locus (K), observed heterozygosity (H_o) and expected heterozygosity (H_E) under Hardy–Weinberg assumptions (Nei 1987) using the program CERVUS v3.0 (Kalinowski *et al.* 2007). Departure of loci from Hardy–Weinberg (HW) assumptions was tested using Cervus, and genotypic disequilibrium (GD) between pairs of loci was assessed using FSTAT v1.2 (Goudet 1995).

Analysis of population structure. To test for spatial structure, samples were grouped into seven putative populations, corresponding to the breeding stocks and substocks delineated by the International Whaling Commission (IWC) (Table 1, Fig. 1). To explore the presence of sex-biased dispersal, we partitioned the data set into male and female subsamples and conducted the analyses described below on all three data partitions (the total combined sample, and the separate male and female samples). Partitioning the data in this way does not provide a true reflection of male versus female genetic differences, which would require a comparison of a male-specific chromosome versus mtDNA (e.g. Andrews *et al.* 2013; Schregel *et al.* 2015). Nonetheless, in the absence of male chromosomal data, the method we employed has proven useful for comparative insights into the behaviour of species that exhibit sex-specific differences in their life histories (e.g. Mendez *et al.* 2010; Alexander *et al.* 2016).

Pairwise genetic differentiation was estimated by counting the number of different alleles between each pair of genotypes, the equivalent of estimating weighted F_{ST} over all loci (Weir & Cockerham 1984), and by counting the sum of the square number of repeat differences between two haplotypes, the equivalent of estimating R_{ST} (Slatkin 1995). Estimations were made from 1000 permutations at the 0.05 significance level using ARLEQUIN v3.5 (Excoffier & Lischer 2010). The statistic Jost's *D* (Jost 2008) was estimated using the DEMETICS package (Gerlach *et al.* 2010) in R. Jost's *D* has been shown to produce a more accurate measure of differentiation when using highly polymorphic microsatellite loci (Jost 2008). An analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) was conducted in ARLEQUIN v 3.5 to assess hierarchical population structure. *F*- and *R*-statistics were computed at three levels that considered differences: (i) among breeding stocks, (ii) among substocks within breeding stocks and (iii) within substocks. Estimations were made from 1000 permutations at the 0.05 significance level. No correction for multiple tests was applied to significant levels of pairwise comparisons (Narum 2006).

To evaluate whether isolation by distance (IBD) can explain the genetic patterns in the study area, we tested for potential correlations between the pairwise genetic (F_{ST} and Jost's *D*) and geographical distances, for the total sample and males and females separately, using Mantel tests in IBD v.3.23 (Jensen *et al.* 2005). The significance of these tests was assessed through 30 000 random permutations of the variables. Geographical distance were calculated in ARCMAP v.10.3 (ESRI) using two alternative methods: (i) Euclidean distance between sampling sites and ii) movement between sampling sites along the Southern Ocean convergence zone (−60°S) to more realistically reflect species movement behaviour and current understanding that humpback whale connectivity between stocks may most often occur as a result of longitudinal movements and mixing in the Southern Ocean (Amaral & Loo *et al.* 2016; Supporting Information). Rejection of the null hypothesis of a negative or flat slope for the correlation between variables is used as evidence for IBD. After initial tests, ASHW was removed from the analysis due to the potential for its long-term isolation from the other breeding stocks and substocks to obscure or skew levels of IBD.

To infer the number of genetic clusters in our data set without *a priori* designation of populations, we analysed individual multilocus genotypes using the program STRUCTURE v2.3.3 (Pritchard *et al.* 2000), via the University of Oslo Bioportal (Kumar *et al.* 2009). We performed five independent iterations of $K = 2$–10 for 5 000 000 Markov Chain Monte Carlo (MCMC) generations with a 500 000 burn-in period, assuming correlated allele frequencies (gamma distribution with mean 0.01 and standard deviation 0.05). Separate runs were performed with and without admixture and a sample location prior (LOCPRIOR). A two-cluster scenario was chosen as the minimum number because when population structure is expected to be low, the scenario $K = 1$ may be disproportionately favoured, reducing the likelihood of all other scenarios to zero and resulting in a loss of overall resolution (Pomilla 2005). We selected the most probable value of K based on the average maximum estimated log-likelihood of P(X|K) and the ΔK method (Evanno *et al.* 2005), where optimum *K* has the highest rate of change in log probability in the data between successive *K* values (i.e. ΔK). All calculations were conducted using STRUCTURE Harvester (Earl & vonHoldt 2012). Clusters were aligned using CLUMPP v 1.1.2 (Jakobsson & Rosenberg 2007) and graphically displayed using the program DISTRUCT v 1.1 (Rosenberg 2004).

As an independent test of population structure, we performed a discriminant analysis of principal components (DAPC) (Jombart *et al.* 2010) on individual allele frequencies using the adegenet package in R (Jombart 2008; Supporting Information). DAPC has been shown to recover complex patterns of population subdivision and has proved robust to deviations from HW equilibrium and GD because it does not rely on an underlying genetic model (Jombart *et al.* 2010). To assess the genetic distinctiveness of each breeding stock, the proportion of correct reassignment of each individual to its putative population was computed (Supporting Information).

Measures of migration rates and gene flow. We estimated relative effective population size (θ) and levels of historical gene flow ($M = m/\mu$), where m represents the immigration rate and μ the mutation rate, using the maximum-likelihood algorithm implemented in MIGRATE v3.5.1 (Beerli & Felsenstein 2001). To address the issue of unequal sample sizes between locations, we chose to subsample our data set prior to analysis (Beerli 1998; Supporting Information). We used Brownian motion approximation to obtain initial parameter values and implemented a complete pairwise migration matrix model of gene flow between all breeding stocks. The final Markov chain scheme consisted of 20 short chain searches (50 000 trees sampled, 500 trees recorded) followed by three long-chain searches (5 000 000 trees sampled, 50 000 trees recorded) after a burn-in period of 10 000 genealogies. The final long-chain searches were averaged over ten independent runs and across subsamples. To aid visualization, results were transformed: $N_e m_T = ((1-(1/N_e m)^2)*100$.

We estimated the magnitude and direction of contemporary gene flow among populations using BAYESASS v 3.0 (Wilson & Rannala 2003). To address inconsistencies in the results from initial runs, we again elected to subsample our data set (Supporting Information). Apart from the mixing parameters, all other options were left at their default settings (Supporting Information). The final Markov chain scheme comprised 50 000 000 iterations including a 2 500 000 burn-in period, and a sampling rate of 100. Results were averaged over the five independent runs, and across both random subsamples, if convergence was achieved.

Results

Sample description

The 3575 genetic samples analysed were determined to represent 3188 different whales (hereafter, 'total sample'; individuals resighted in the same sampling location were removed; Table 1). Average probability of identity (P_{ID}) for the total sample was small enough to exclude duplicate individuals with high confidence ($P_{ID} = 1.95 \times 10^{-12}$; $P_{ID(sibs)} = 9.2 \times 10^{-5}$; reciprocal of sample size = 2.5×10^{-4}). Sex was determined for 3045 individuals, 1978 males and 1067 females, resulting in an overall proportion of 1.8:1 males to females (Table 1). Proportionally greater numbers of males were sampled within most breeding stocks, likely due to a sampling bias resulting from breeding behaviour differences between the sexes (Smith *et al.* 1999). Conversely, there were almost equal numbers of males and females sampled within BSB2, and a strong female bias in BSC2, with more than three times the number of females sampled than males (Table 1; Fig. 1).

Genetic diversity

All ten microsatellites were highly polymorphic, ranging from four alleles per locus (EV1Pm) to 28 (GATA417). No significant differences were found between the observed heterozygosity (H_o) and the expected heterozygosity (H_e) under Hardy–Weinberg assumptions. Two loci, GATA028 and GATA053, were in significant GD ($P < 0.01$). The least polymorphic locus, GATA053, was removed from subsequent analyses (Weir 1990). Values of observed and expected heterozygosity were relatively high across all breeding stocks ($H_o = 0.702$–0.742; $H_e = 0.678$–0.738) and the mean number of alleles per locus ranged from 6.44 (ASHW) to 12.89 (BSB1), although Oman was an outlier with 6.44 while the Southern Hemisphere stocks ranged from 10.11 (BSA) to 12.89 (BSB1; Table 1). Diversity estimates for BSB2 ($n = 204$, $k = 11.33$) and BSC1 ($n = 203$, $k = 12.00$) were disproportionately high relative to sample size.

Population structure

The AMOVA showed that genetic variance was best explained within the substock level for all sample partitions (for the total sample, males and females, $F_{ST} = 0.003$, $P < 0.001$; Table 2). Significant variation was also observed for the total sample among substocks within breeding stocks ($F_{ST} = 0.001$, $P < 0.01$). Significant variation among breeding stocks (i.e. the highest level of organization) was found for females ($F_{ST} = 0.003$, $P < 0.003$), but not males or the total sample.

Pairwise genetic differentiation estimates were found to be low but significant for a number of comparisons: F_{ST} estimates ranged from 0 to 0.065, R_{ST} from 0 to 0.088 and Jost's D from 0 to 0.181 (Table S1,

Table 2 Analysis of hierarchical variance (AMOVA) results obtained using *F*- and *R*-statistics at three levels for the total sample (n = 3188), and males (n = 1978) and females (n = 1067) separately

Sample	Source of variation	% var	*F*-statistics	% var	*R*-statistics
Total	Among breeding stocks	0.24	F_{CT} = 0.0024	0.32	R_{CT} = 0.0032
	Among substocks within breeding stocks	0.08	F_{SC} = **0.0008****	−0.01	R_{SC} = −0.0001
	Within substocks	99.68	F_{ST} = **0.0032*****	99.69	R_{ST} = **0.0031*****
Male	Among breeding stocks	0.27	F_{CT} = 0.0027	0.42	R_{CT} = 0.0043
	Among substocks within breeding stocks	0.01	F_{SC} = 0.0001	−0.12	R_{SC} = −0.0012
	Within substocks	99.72	F_{ST} = **0.0028*****	99.70	R_{ST} = **0.0030***
Female	Among breeding stocks	0.31	F_{CT} = **0.0031***	0.22	R_{CT} = 0.0022
	Among substocks within breeding stocks	0.02	F_{SC} = 0.0002	0.05	R_{SC} = 0.0005
	Within substocks	99.67	F_{ST} = **0.0033*****	99.73	R_{ST} = 0.0027

Bold type indicates statistical significance at *P < 0.05, **P < 0.01, ***P < 0.001.

Supporting information). ASHW proved the most highly differentiated across all fixation indices and partitions, and BSA showed relatively high genetic differentiation for F_{ST}; however, results for BSA were more variable for Jost's D (Table 3). BSB1 also showed strong differentiation from all other breeding stocks for F_{ST} in the total sample, a result generally supported by Jost's D; however, this relationship is less pronounced when segregated by sex (Table 3). At the local scale (i.e. within breeding stocks), the relationships between BSB2 and the substocks of BSC are less clear. BSB2 is significantly differentiated from BSC1 for F_{ST}, and Jost's D suggests this is driven by the female sample (Table 3). For males, BSB2 was significantly differentiated from BSC2 and BSC3 but only for F_{ST} (Table 3). Within BSC, BSC1 and BSC3 showed significant differentiation for the total sample, which appears to be driven by females. All other comparisons between BSB2 and BSC, and within BSC, were not significant for any indices (Table 3; Table S1, Supporting information).

Isolation by distance was evidenced by statistically significant correlations between the genetic distance and Euclidean geographical distance for all data partitions and fixation indices (Table 4). These findings were consistent for the geographical distance scenario based on movement behaviour in the Southern Ocean except for the female sample (Table 4).

Genetic structure based on individual allele frequencies without *a priori* designation of populations was detected by STRUCTURE when a location prior was used with correlated allele frequencies and no admixture. No convergence was attained without a location prior, or when admixture was used. The ln P(K) and ΔK values did not clearly discriminate whether the optimal number of clusters was K = 3 or K = 4 (Fig. S1, Supporting information). However, the individual assignment plots clearly show K = 3 (Fig. 2) as the most likely for all data partitions (see Fig. S2, Supporting information for K = 4 plots). The clusters primarily correspond to the South Atlantic (BSA and BSB1), western Indian Ocean (BSC) and the northern Indian Ocean (ASHW; Fig. 2a–

Table 3 Significance values for pairwise fixation indices obtained between humpback whale breeding stocks and substocks for F_{ST} and Jost's D. F_{ST} values are shown above the diagonal, Jost's D below the diagonal. Results are shown for the total sample (n = 3188), and males (n = 1978) and females (n = 1067), separately.

	A	B1	B2	C1	C2	C3	ASHW
Total							
A	–	**	**	***	***	***	***
B1	ns	–	*	***	*	***	***
B2	**	*	–	*	ns	ns	***
C1	**	**	ns	–	ns	*	***
C2	ns	ns	ns	ns	–	ns	***
C3	**	**	ns	ns	ns	–	***
ASHW	**	**	**	**	**	**	–
Male							
A	–	*	*	*	ns	*	***
B1	ns	–	ns	*	ns	*	***
B2	ns	ns	–	ns	*	***	***
C1	ns	*	ns	–	ns	ns	***
C2	ns	ns	ns	ns	–	ns	***
C3	*	***	ns	ns	ns	–	***
ASHW	***	***	***	***	***	***	–
Female							
A	–	ns	*	*	*	*	***
B1	ns	–	ns	ns	ns	*	***
B2	*	ns	–	ns	ns	ns	***
C1	*	ns	*	–	ns	ns	***
C2	ns	ns	ns	ns	–	ns	***
C3	**	***	ns	*	ns	–	***
ASHW	***	***	***	***	***	***	–

*Statistical significance at *P < 0.05, **P < 0.01, ***P < 0.001. Estimations of significance were made from 1000 permutations at the 0.05 significance level. Shaded F_{ST} values indicate a statistically significant result for mitochondrial DNA data, adapted from Rosenbaum *et al.* (2009). See Table S1 for fixation index values and R_{ST} results.

Table 4 Summary results for the IBD tests. The [$P(r \leq 0)$], slope values (r), and correlation coefficients (R^2) of the correlations between genetic (F_{ST} and Jost's D) and (i) Euclidean geographical distance and (ii) geographical distance based on movement behaviour in the Southern Ocean (km) are presented. The slope between F_{ST} and km is expressed as F_{ST}/1000 km. See Supporting Information for a description of the two geographical distance scenarios. Due to its long-term isolation from the other breeding stocks and substocks, ASHW was not included in the analyses.

	(i) Euclidean geographical distance			(ii) Geographical distance based on movement in the Southern Ocean		
	[$P(r \leq 0)$]	r (slope)	R^2	[$P(r \leq 0)$]	r (slope)	R^2
Total sample						
F_{ST}	**0.01**	0.858	0.737	**0.05**	0.557	0.311
Jost's D	**0.01**	0.871	0.759	**0.02**	0.609	0.370
Male						
F_{ST}	**0.02**	0.877	0.770	**0.03**	0.645	0.417
Jost's D	**0.03**	0.653	0.428	**0.02**	0.666	0.443
Female						
F_{ST}	**0.03**	0.841	0.707	0.06	0.575	0.331
Jost's D	**0.03**	0.786	0.618	0.06	0.550	0.303

Significant [$P(r \leq 0)$] values are shown in bold type ($P < 0.05$).

c). BSA does not appear substantially different in composition from BSB1 for the total sample or for males; however, it does appear to be less admixed for females. At the local scale, BSB2 appears more genetically similar to BSC than to BSB1 for all data partitions. The assignment plots also show evidence of population substructure within BSC when sampling location is considered (Fig. 2). The males sampled from southwest Madagascar, for example, appear to be highly differentiated from those sampled in northeast Madagascar (Fig. 2).

The DAPC was not able to resolve the number of genetic clusters due to the genetic similarity of the samples in the absence of a location prior. The first two principal components of the DAPC analysis explained 94.18% of the variance in allele frequencies for the total sample (118 PCs retained), 68.61% for the male sample (115 PCs retained) and only 29.98% for the female sample (43 PCs retained, Fig. 3). For the total sample, the first principal component shows separation between breeding stocks that reflects their longitudinal distribution, at least for BSA, BSB1 and ASHW; the remaining breeding stocks show a significant degree of overlap (Fig. 3a). The second principal component clearly shows the differentiation of ASHW, possibly based on latitude. The longitudinal gradient is not clear for the male sample (Fig. 3b); however, BSB2 also shows some separation on PC2 in addition to ASHW. The strongest differentiation of ASHW is observed on PC1 for females (Fig. 3c), whereas PC2 describes the longitudinal separation of BSA and BSB1.

The proportion of individuals correctly assigned to their original putative breeding stock by the DAPC was highest for BSB1 (65–67%), BSC3 (54–68%) and ASHW (65–74%; Fig. 3d–f). The reassignment percentages for ASHW offer a useful benchmark to compare against other stocks as reassignment was expected to be 100% due to its long-term isolation (Pomilla & Amaral *et al.* 2014). This indicates that a 65–75% reassignment percentage is adequate to indicate genetic distinctiveness. Despite its geographical distance from the other breeding stocks, BSA showed relatively low reassignment success (20–30%). At the local scale within BSC, reassignment to BSB2 (2–9%), BSC1 (3–7%) and BSC2 (5–35%) performed particularly poorly. Individuals from the breeding stocks and substocks with the smaller sample sizes were primarily assigned to the much larger BSB1 and BSC3 samples regardless of sampling locality (Fig. S3, Supporting information).

Geneflow estimation

Historical gene flow (N_em) was estimated to occur to some degree between all pairwise breeding stock comparisons with little bias in directionality of movements (Fig. 4a–c). For the total sample, migration estimates ranged from $N_em_T = 1.553$ (BSB1 to ASHW) to $N_em_T = 22.345$ (BSC2 to BSC3, Fig. 4a). BSA showed relatively high exchange for all comparisons apart from ASHW. The highest estimates occurred from BSB2 to BSB1 and BSC, and also within BSC. Estimates for ASHW were the lowest of all comparisons; however, some estimates between ASHW and BSC remained $N_em_T > 10$. For males, estimates ranged from $N_em_T = 0.391$ (BSC2 to ASHW) to $N_em_T = 23.341$ (BSC3 to BSC1, Fig. 4b) and approximated the same pattern as

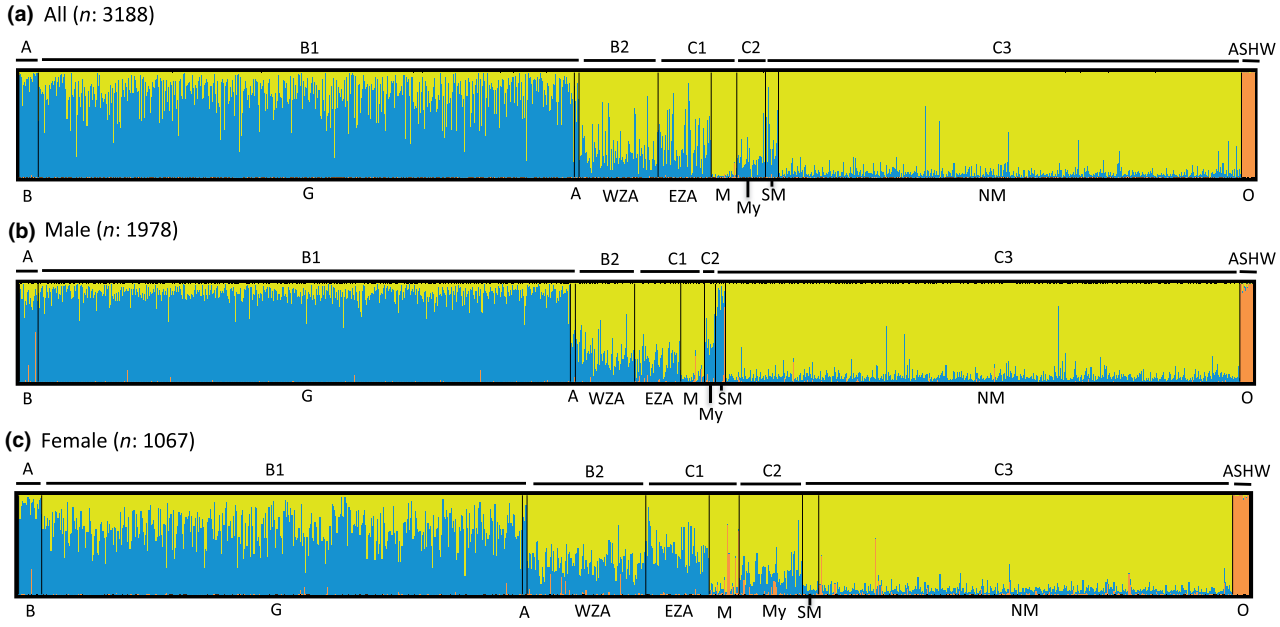

Fig. 2 Distribution of three genetic clusters estimated using STRUCTURE for (a) the total sample, (b) males and (c) females. Vertical lines are partitioned into coloured segments showing the proportion of each individual assigned to each K cluster. Breeding stocks are indicated above each figure and sampling locations are below.

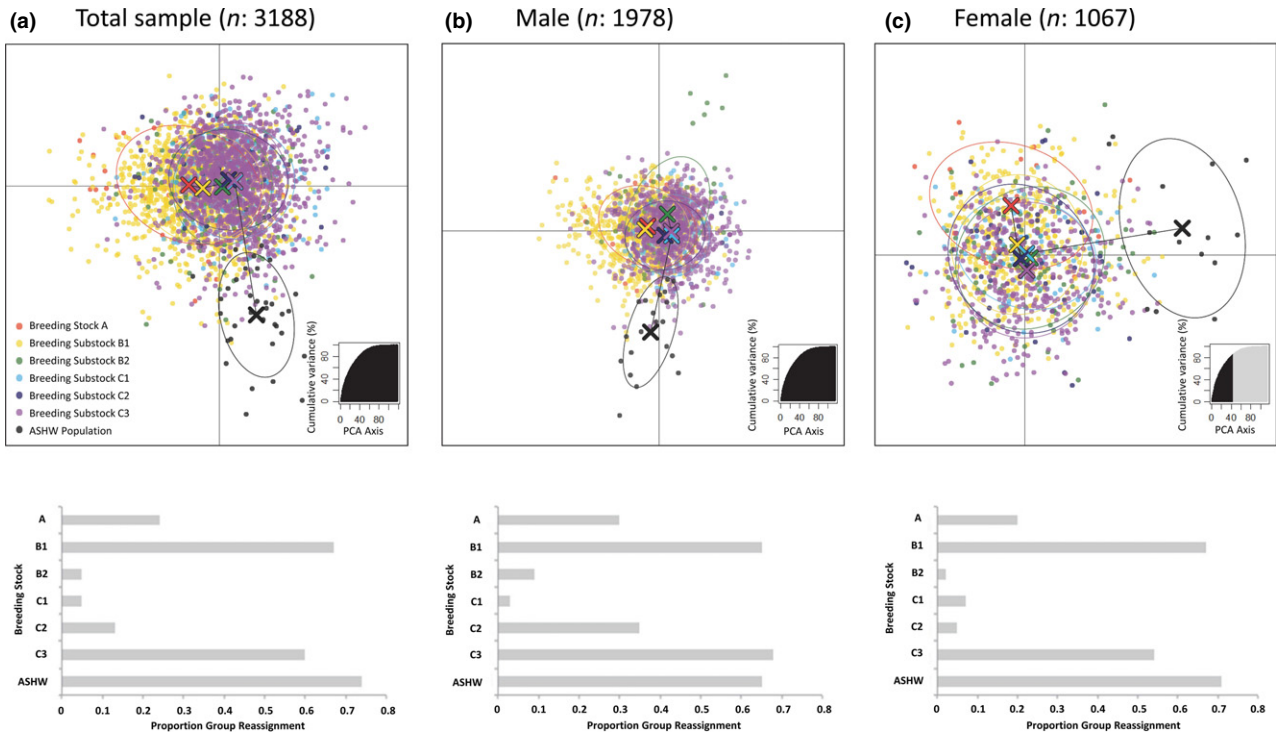

Fig. 3 Discriminant analysis of principal components (DAPC) scatterplots showing the genetic structure between humpback whale breeding stocks for (a) the total sample, (b) males and (c) females. Key describes the colours attributed to each breeding stock and substock and inertia ellipses describe the general distribution of points. Eigenvalues for each PC axis are shown (PC1, vertical; PC2, horizontal). The number of PCA axes retained in each DAPC analyses is shown in the bottom-right inset (black bars). Bar charts show the proportion of reassignment of each individual to its original putative breeding stock (group).

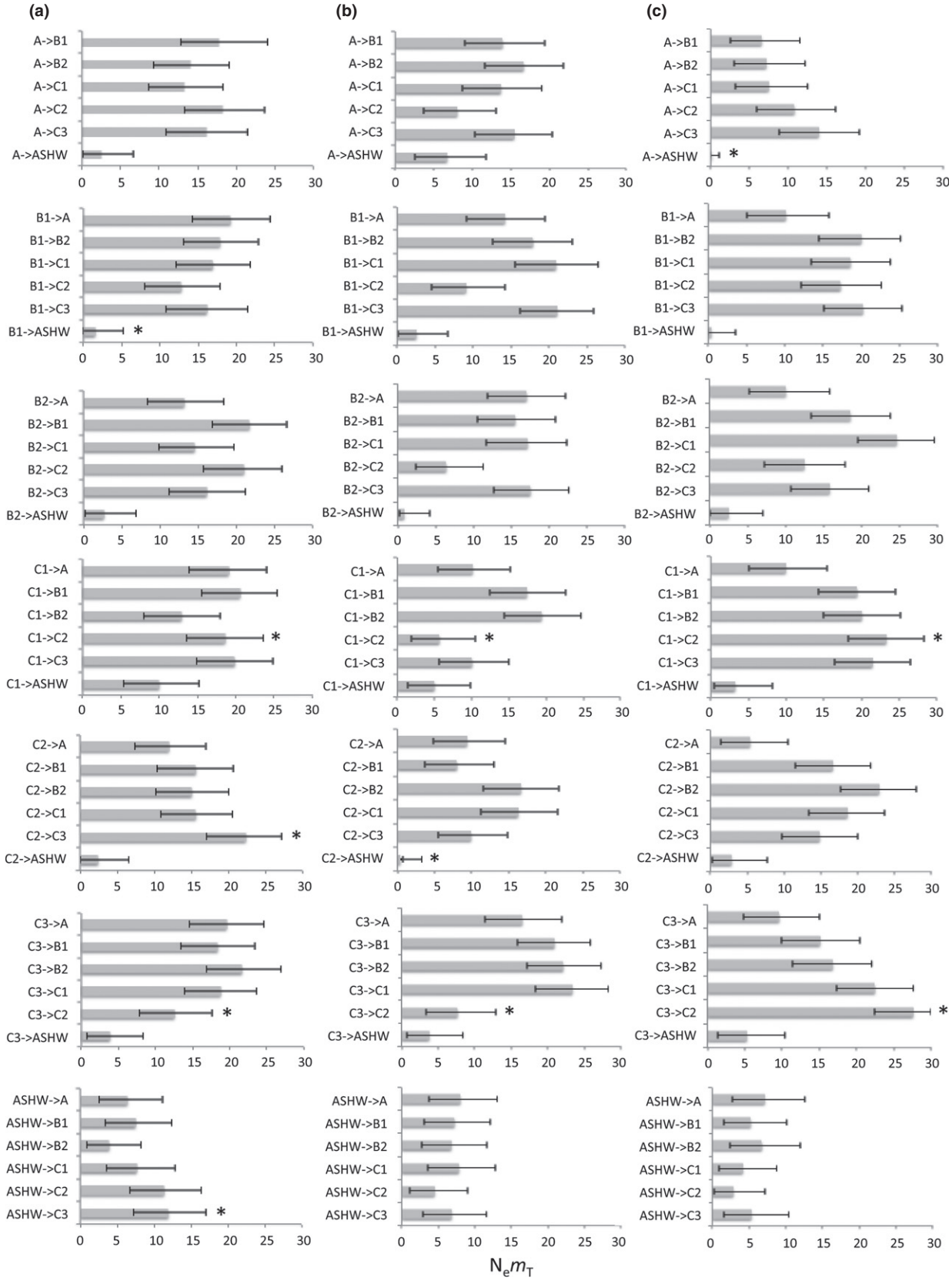

Fig. 4 Magnitude and directionality of historic gene flow between breeding stocks. The historic estimated number of migrants per generation ($N_e m_T$) exchanged between breeding stocks is shown for (a) the total sample, (b) males and (c) females, as estimated using MIGRATE. Asterisks highlight key results discussed in the main text.

the total sample; however, a stronger westward bias was evident for some comparisons. For females, estimates ranged from $N_e m_T = 0.148$ (BSA to ASHW) to $N_e m_T = 24.618$ (BSC1 to BSC2, Fig. 4c) with high rates of multidirectional exchange ($N_e m_T \approx 20$) estimated between BSB and BSC (Fig. 4c). Notably, westward exchange for females from BSC3 to BSC2 was more than three times that of males (female $N_e m_T = 27.565$; male $N_e m_T = 7.669$). Overall, females showed less exchange between BSA and the other breeding stocks than males (Fig. 4b–c).

Levels of contemporary gene flow (proportion of migrants, *M*) estimated using BayesAss were less informative as many of the pairwise comparisons did not achieve convergence: ~31% for the total sample, ~33% for males and ~40% for females (Fig. S4a–c, Supporting information). For the total sample, *M* ranged from $M = 2.7^{10-3}$ (BSB2 to BSC2) to $M = 29.7^{10-3}$ (BSB2 to BSB1, Fig. S4a, Supporting information). Comparisons of BSA with all other breeding stocks indicate an eastward bias in migration. Estimates for males were generally higher than females (Fig. S4b–c, Supporting information) with directional gene flow being more evident, notably from BSC2 to BSC3 ($M = 135^{10-3}$). Estimates for females support a strong eastward bias from BSA to the other breeding stocks, particularly for BSA to BSB1, with an eastward estimate ~58 times that of the westward (Fig. S4c, Supporting information).

Discussion

Hierarchical assessment of population structure

This first examination of the diversity and differentiation of nine microsatellite loci for more than 3000 individual humpback whales from across the South Atlantic and western Indian Ocean suggests that a hierarchy of processes is likely to be driving patterns of genetic population structure at different spatial scales. Such a hierarchy reflects the interplay between phylogeographic and ecological processes evident in other behaviourally complex mammals (Wolf *et al.* 2007; VanderWaal *et al.* 2014).

At regional scales (i.e. between ocean basins), processes of isolation by distance (Wright 1943) and phylogeographic history appear to be the primary drivers of genetic structure. Similar to other highly migratory whale species (Alter *et al.* 2012; Torres-Florez *et al.* 2014), pairwise genetic differentiation estimates were found to be low but significant for a number of comparisons (Table 3). Breeding stocks showed differentiation along the longitudinal axis consistent with a model of isolation by distance (Table 4) and previous studies suggesting the long-term isolation of ASHW (Pomilla & Amaral *et al.* 2014). This finding was consistent across all data partitions when Euclidean geographical distance was considered and for males and the total sample for geographical distance based on movement behaviour in the Southern Ocean (Table 4). Both BSA and BSB1 appear distinct from BSB2 and the substocks of BSC, and the ASHW population is clearly differentiated (Fig. 3). Pairwise comparisons of genetic differentiation for nine nuclear introns also revealed regional population structure between Brazil, Gabon and Madagascar (Ruegg *et al.* 2013).

At the subregional scale (i.e. within ocean basins), previous examinations of the distribution of humpback whale mtDNA haplotypes indicate strong differentiation between humpback whale breeding stocks for females and less so for males, supporting a model of maternally directed philopatry to breeding areas, due to culturally transmitted hysteresis (or 'memory') of specific locations, combined with male-biased gene flow (Palsbøll *et al.* 1995; Baker *et al.* 1998, 2013; Rosenbaum *et al.* 2009). Our findings of significant genetic differentiation (statistical significance of AMOVA for females among breeding stocks, and not for males, Table 2; comparison of Jost's *D* fixation index for males and females, Table 3; evidence of relatively higher levels of structuring in DAPC, Fig. 3) and limited gene flow for females (comparison between BSA and BSB, Fig. 4b–c) provide some support to the assertions that this model represents the primary driver of population genetic structure (Palsbøll *et al.* 1995; Baker *et al.* 1998, 2013) at this scale. Correlations between genetic distance and geographical distance based on movement behaviour were not found to be statistically significant for females (Table 4). This may indicate that there are other factors in addition to geographical distance driving female genetic structure, such as philopatric behaviour. These patterns are similar to those observed for the southern right whale (*Eubalaena australis*; Carroll *et al.* 2011), where combined genetic and isotope analyses indicate that site fidelity to feeding areas is culturally transmitted along matrilineal lines, providing a mechanism for maintaining genetic population structure (Valenzuela *et al.* 2009; Vighi *et al.* 2014).

The model of maternally directed philopatry and male-biased dispersal was not generalizable at all

spatial scales, however, and our findings indicate that even at the subregional scale within ocean basins there may be additional factors influencing genetic structure. Historic and contemporary gene flow estimates for females between BSA and BSB and BSC (Fig. 4c,f), in concert with recent capture–recapture records (Stevick *et al.* 2011), suggest that long-distance movements by females should be afforded more consideration as a driver of circumglobal genetic variation, as well as in terms of enhancing the capacity for population expansion and habitat recolonization (Bohrer *et al.* 2005). Transitioning to local scales (i.e. within breeding stocks), we observed additional divergence from this model as complex patterns of isolation and connectivity appeared as the norm for both sexes, suggesting an array of interacting processes, such as groups selecting or migrating through different habitats at different times (such as observed for BSB1 and BSB2, where males and females may be migrating at different times that may have led to variability in the statistical significance of the fixation indices for the total sample ($P < 0.05$) compared with males ($P > 0.05$) and females ($P > 0.05$); Table 3; Carvalho *et al.* 2014), or instances of male fidelity to breeding areas (such as in South Madagascar; Fig. 2), may be responsible for driving population patterns.

Effect of spatiotemporal variation in migratory behaviour on population substructure

Earlier genetic evidence have been interpreted to support the existence of two demographically discrete substocks (i.e. BSB1 and BSB2) off West Africa (Rosenbaum *et al.* 2009); with additional data over time, an alternative hypothesis proposes that BSB1 and BSB2 represent two temporal 'ends' of a single population (BSB) widely distributed in space and time (Van Waerebeek *et al.* 2013; Carvalho *et al.* 2014; Rosenbaum *et al.* 2014). In our study, we detected significant genetic differentiation between BSB1 and BSB2 for the total sample for F_{ST} and Jost's D ($P < 0.05$); however, neither the male nor female partitions were found to be significant (Table 3) consistent with previous findings of subtle temporal population substructuring based on sex (Carvalho *et al.* 2014). This may, at least in part, be related to different migratory groups undertaking coastal *versus* oceanic routes to and from the breeding areas off Gabon (Elwen *et al.* 2014; Rosenbaum *et al.* 2014). These patterns reflect evidence of weak, but some genetic differentiation between neighbouring breeding areas, similar to what has been detected in other expansive ranges for humpback whales along the northeastern and northwestern coasts of Australia (Schmitt *et al.* 2014).

Our results provide an additional line of evidence for the observed genetic differentiation of BSB2, namely its potential relationship with BSC1 off East South Africa. This is demonstrated by the lack of a statistically significant pairwise F_{ST} value between BSB2 and BSC1 when males and females were tested separately (Table 3). In addition, BSB2 is more genetically similar to BSC than BSB1, particularly for males (Fig. 2). Our results also demonstrate that BSB2 exhibits a high level of admixture (Fig. 2) and low reassignment probabilities (Fig. 3a–c), which are consistent with a migratory population comprising whales from different breeding stocks, including BSC1. While the lack of differentiation observed between BSB2 and BSC1 may be due to retained shared ancestral polymorphism (Rosenbaum *et al.* 2009) or differences in sample size, evidence for contemporary exchange between populations on the west and east coasts of Africa during the sampling period (i.e. since 1990) increasingly suggests recent interoceanic migration (Pomilla & Rosenbaum 2005; Rosenbaum *et al.* 2009; Carvalho *et al.* 2014), and genotypic matches have revealed direct movements of two males between BSB2 and BSC1 and another two males between BSB1 and BSC3, and between feeding areas associated with BSB and BSC in the Antarctic (IWC 2009; Kershaw 2015; Amaral & Loo *et al.* 2016).

Migratory overlap and sex-specific differences drive genetic complexity

Substock-scale genetic patterns of structure and connectivity for Breeding Stock C appear to be highly complex and challenging to generalize in terms of maternal and paternal influence or directionality. Significant genetic differentiation was found between BSC1 and BSC3 for the combined male and female sample ($F_{ST} = 0.001$; $P < 0.05$), and this difference appears to be driven by the female data set (Jost's $D = 0.001$; $P < 0.05$; Table 3). In contrast, no significant genetic differentiation was detected between BSC1 and BSC2 or between BSC2 and BSC3 (Table 3). However, we did detect subtle differences between these patterns for males and females. No differentiation was detected for males between any of the BSC substocks, supporting the general model of male-biased gene flow between populations that would result in the erosion of signals of genetic differentiation. However, estimates of historical gene flow showed no clear pattern in directionality and were found to be particularly high for females, calling the general model of maternally driven natal philopatry to breeding areas into dispute (Fig. 4a–c; Palsbøll *et al.* 1995; Baker *et al.* 1998, 2013). Estimations of gene flow also suggest higher levels of potential exchange between BSC1-C3 for both females and males (Fig. 4) than previously detected (IWC 2009).

It is possible that the BSC substocks have diverged from one another relatively recently or may have remained consistently 'fluid' (Marko & Hart 2011). This latter suggestion is consistent with previous hypotheses regarding the presence of three migratory streams within the southwestern Indian Ocean, one along the east coast of southern Africa (BSC1), one along the Madagascar ridge (BSC3), and possibly a third through the central Mozambique Channel (BSC2; Best *et al.* 1998). It is possible, however, that this third stream comprises wide-ranging animals from coastal Africa and Madagascar (Best *et al.* 1998). Our findings of the lack of genetic differentiation between BSC1 and BSC2, and BSC2 and BSC3, combined with high geneflow estimates, support the assertion that BSC2 may represent a mixed migratory stream of wide-ranging animals from BSC1 and BSC3. Recent photo-identification studies (Ersts *et al.* 2011), satellite telemetry data (Fossette *et al.* 2014; Cerchio *et al.* in press) and genotypic matches (Kershaw 2015) show that there is indeed considerable movement between BSC2 and BSC3, which are geographically close to one another relative to distances humpback whales are capable of travelling. In addition, long-distance movements between northeastern Madagascar (BSC3) and coasts of Kenya and Somalia in northern BSC1 appear more frequent than previously supposed and may even represent a second, more northern migratory stream between BSC1 and BSC3 (Fossette *et al.* 2014; Cerchio *et al.* in press). Photographic recaptures and satellite telemetry data also suggest relatively substantial interchange between BSC3 and Réunion (BSC4; IWC 2012; Globice, unpublished data), but levels of genetic connectivity have yet to be assessed. These findings demonstrate that while substock differentiation exists in the southwestern Indian Ocean, the extent of movements, exchange and connectivity between areas requires better understanding (Rosenbaum *et al.* 2009; Ersts *et al.* 2011). Similar substructure scenarios exist for humpback populations in the Hawaiian Archipelago and within the wintering region off the coast of Mexico, suggested to result from mating occurring during migration (Cerchio *et al.* 1998; Calambokidis *et al.* 2001).

Importantly, we detected additional genetic structure within BSC3 indicating that further discrete substructure or discrete demographic units may be present. The small sample of males ($n = 17$) from the southwest of Madagascar (BSC3) shows greater levels of admixture than the large number of whales sampled in Antongil Bay to the north (Fig. 2). A number of factors apart from population structure could be driving these observed differences in allele frequencies between sampling locations, such as nonrepresentative sampling from different years or disparities in sample size (Marko & Hart 2011), so until larger, more representative samples are available, conclusions drawn from these results should be considered with caution. However, given our increasing understanding of the behavioural complexity (e.g. alternative migration routes) of humpback whale populations in this region and others (Cerchio *et al.* 1998; Ersts *et al.* 2011; Carvalho *et al.* 2014; Rosenbaum *et al.* 2014), further investigation into the genetic structure of the BSC substocks is required.

Implications for management at multiple scales

These results for Southern Hemisphere humpback whales provide one of the largest genetic data sets for the great whales, and the most definitive evidence to date related to these IWC substocks. We view these new results as the best evidence to date that advances understanding of population complexity for humpback whales in these regions, especially until further or new evidence provides additional or finer-scale resolution concerning the number of demographically discrete population units. Given the range of contemporary anthropogenic impacts potentially affecting whale populations and important breeding habitat in some of these regions (Martins *et al.* 2013; Pomilla & Amaral *et al.* 2014; Rosenbaum *et al.* 2014), and the ongoing reassessment of the species' conservation status (such as the recent down-listing of many humpback whale management units under the Endangered Species Act (ESA) in the United States; Federal Register 2016), the accurate identification of demographically discrete populations is paramount to the effective management of this species. While some of these stocks and substocks are not necessarily threatened with extinction (Bettridge *et al.* 2015), they are still undergoing recovery from depletion due to commercial whaling, such as breeding sub-stocks within BSB and BSC (IWC 2011).

The regional genetic structure between ocean basins detected by the microsatellite analyses of isolation by distance presented here, and previous studies of nuclear introns (Ruegg *et al.* 2013) and the mitochondrial control region (Rosenbaum *et al.* 2009), is generally consistent with current designations of breeding stocks A, B, C and ASHW, by the International Whaling Commission. In particular, our findings support recommendations that the ASHW population be attributed international conservation priority in the light of its extreme isolation and regional distinctiveness, and increasing levels of anthropogenic development occurring in the Arabian Sea (Pomilla & Amaral *et al.* 2014). Notably, a number of our analyses provided new evidence that the typically presumed biological or management boundaries of BSB and BSC may be more 'porous'

than previously assumed. In particular, the lack of genetic differentiation and relatively high levels of mixing between BSB2 and the substocks of BSC suggest higher levels of connectivity than currently accounted for in current management designations. These findings could also have an impact on estimates of abundance and recovery levels. Our findings further indicate that an array of ecological drivers, including sex-specific site fidelity and dispersal behaviour, are likely responsible for the complex patterns of genetic structure observed at a local scale within breeding stocks. This evidence of local-scale demographic independence supports the need for substock delineation; however, the apparent complexity of the processes driving local-scale structure and levels of connectivity also underscores that substock differentiation and relationships might require further evaluation, particularly when new evidence becomes available.

While these results provide a considerable degree of resolution concerning differentiation and exchange rates, there are still areas of uncertainty. Further investigation will be necessary to disentangle which processes are operating at each hierarchical spatial scale. Application of high-power genomic data is needed to resolve the subtle boundaries between BSB2 and the substocks of BSC. Mixed-stock analysis may also provide a useful tool to understand levels of mixing between different stocks and substocks on feeding areas and, if the data set were of sufficient power, to better understand the composition of substocks thought to represent mixed migratory streams, such as BSB2 and BSC2. For the highly endangered ASHW population, further examination of relatedness and levels of inbreeding, and immunogenetic assessments of disease risk, could provide vital information for conservation and management. Additionally, models of habitat suitability and physiology (e.g. energy requirements) may prove useful complementary tools for predicting whale movements, and therefore the extent of mixing of different breeding stocks and substocks.

Even with the significant sample size used in our study, rare among studies of highly mobile and migratory marine species, we were unable to definitively resolve population structure and connectivity for humpback whales in the study region. This may be an unrealistic expectation given the levels of sampling of multiple populations across different ocean basins and oceanographic regimes. This was also, at least in part, due to the limited power of our data set of nine microsatellite loci. Genomic studies of nonmodel organisms, including marine mammals, are becoming increasingly prevalent (Cammen *et al*. 2016) and are now providing the analytical power required to resolve subtle genetic population structures and associated drivers for marine species across taxa, including with more limited sample sizes (e.g. Dierickx *et al.* 2015; Foote *et al.* 2016). Even with these advances, it is clear that genetics alone does not provide a 'silver bullet' to understanding population structure in migratory species. Rather, an integrative approach encompassing genetic, behavioural and environmental data is required. Efforts to integrate different types of data sets are now emerging (Rittschof & Robinson 2014; Selkoe *et al.* 2016). Further methodological development for integrating data in a way that is ecological meaningful and the continued forging of collaborations between researchers provide promising avenues for enhancing our understanding of the evolutionary and ecological mechanisms underlying genetic population structure and will also provide information essential for conservation and management.

Acknowledgements

This work would not be possible without the hard work and dedication of all country staff, multitude of interns, boat pilots and other field staff who contributed to collect data over these years. We are grateful to Dr. George Amato, Rebecca Hersh and other members of the Sackler Institute for Comparative Genomics. IC also thanks the Director of Environment Dr. Arlindo Carvalho and the Portuguese Embassy in São Tomé and Príncipe for logistical support. We would like to thank the Ministère des Eaux et Forêts and the Direction de la Pêche et Ressources Halieutiques de Madagascar, Institut Halieutique et des Sciences Marines de Toliara and Madagascar National Parks for permission to conduct this research. We are thankful to Daanie Bennet of Advantage cruises for the significant collection of sloughed skin provided to the BSC1 collection and the Department of Environmental Affairs for funding three humpback cruises on the *RV Algoa* (Mozambique, KwaZulu-Natal and the Wild coast) and a fourth cruise on the RV Ellen Khuzwayo (off the west coast of South Africa). The research would not be possible without the gracious financial support of the American Museum of Natural History, RAPAC – Projet Espèces Phares, and the Wildlife Conservation Society. IC was supported by a grant from the Fundação para a Ciência e Tecnologia (SFRH/BPD/97566/2013). Funding to HCR was provided by numerous individuals and foundations. We dedicate this work to the memory and career of Dr. Peter B Best. One of the pioneers in the field, he supported these regional collaborative efforts for many decades and continually challenged us in order to produce the most robust scientific findings and interpretations.

References

Alexander A, Steel D, Hoekzema K *et al.* (2016) What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Molecular Ecology*, **25**, 2754–2772.

Alter SE, Ramirez SF, Nigenda S, Ramirez JU, Bracho LR, Palumbi SR (2009) Mitochondrial and nuclear genetic variation across calving lagoons in eastern North Pacific gray whales (*Eschrichtius robustus*). *Journal of Heredity*, **100**, 34–46.

Alter SE, Rosenbaum HC, Postma LD *et al.* (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*). *Ecology and Evolution*, **2**, 2895–911.

Amaral AR, Beheregaray LB, Bilgmann K *et al.* (2012a) Influences of past climatic changes on historical population structure and demography of a cosmopolitan marine predator, the common dolphin (genus *Delphinus*). *Molecular Ecology*, **21**, 4854–4871.

Amaral AR, Beheregaray LB, Bilgmann K *et al.* (2012b) Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (genus *Delphinus*). *PLoS ONE*, **7**, e31482.

Amaral AR, Loo J, Jarris H *et al.* (2016) Complex population structure uncovered in feeding aggregations of humpback whales in the Southern Oceans. *Marine Biology*, **163**, 132.

Andrews KR, Karcmarski L, Au W *et al.* (2010) Rolling stones and stable homes: social structure, habitat diversity and population genetics of Hawaiian spinner dolphin (*Stenella longirostris*). *Molecular Ecology*, **19**, 737–748.

Andrews KA, Perrin WF, Oremus M *et al.* (2013) The evolving male: spinner dolphin (*Stenella longirostris*) ecotypes are divergent at Y chromosome but not mtDNA or autosomal markers. *Molecular Ecology*, **22**, 2408–2423.

Baker CS, Perry A, Herman LM (1987) Reproductive histories of female humpback whales *Megaptera novaeangliae* in the North Pacific. *Marine Ecology Progress Series*, **41**, 103–114.

Baker CS, Flores-Gonzalez L, Abernethy B *et al.* (1998) Mitochondrial DNA variation and maternal gene flow among humpback whales of the Southern Hemisphere. *Marine Mammal Science*, **14**, 721–737.

Baker CS, Steel D, Calambokidis J *et al.* (2013) Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. *Marine Ecology Progress Series*, **494**, 291–306.

Barendse J, Best PB, Carvalho I, Pomilla C (2013) Mother knows best: Occurrence and associations of resighted humpback whales suggest maternally derived fidelity to a southern hemisphere coastal feeding ground. *PLoS ONE*, **8**, e81238.

Beerli P (1998) Estimation of migration rates and population sizes in geographically structured populations. In: *Advances in Molecular Ecology* (ed. Carvalho G), pp. 291–303. NATO Science Series A: Life Sciences, IOS Press, Amsterdam.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Berube M, Palsbøll PJ (1996) Identification of sex in cetaceans by multiplexing with three ZFX and ZFY specific primers. *Molecular Ecology*, **5**, 283–287.

Best PB, Findlay KP, Sekiguchi K, *et al.* (1998) Winter distribution and possible migration routes of humpback whales *Megaptera novaeangliae* in the southwest Indian Ocean. *Marine Ecology Progress Series*, **162**, 287–299.

Bettridge S, Baker CS, Barlow J *et al.* (2015) Status review of the humpback whale (Megaptera novaeangliae) under the Endangered Species Act. NOAA-TM-NMFS-SWFSC-540, March 2015, pp. 241.

Bohrer GI, Nathan RA, Volis S (2005) Effects of long-distance dispersal for metapopulation survival and genetic structure at ecological time and spatial scales. *Journal of Ecology*, **93**, 1029–1040.

Calambokidis J, Steiger GH, Straley JM *et al.* (2001) Movements and population structure of humpback whales in the North Pacific. *Marine Mammal Science*, **17**, 769–794.

Cammen KM, Andrews KA, Carroll EL *et al.* (2016) Genomic methods take the plunge: Recent advances in high-throughput sequencing of marine mammals. *Journal of Heredity*, **107**, 481–495.

Carroll E, Patenaude NJ, Alexander AM *et al.* (2011) Population structure and individual movement of southern right whales around New Zealand and Australia. *Marine Ecology Progress Series*, **432**, 257–268.

Carvalho I, Loo J, Collins T *et al.* (2014) Does temporal and spatial segregation explain the complex population structure of humpback whales on the coast of West Africa? *Marine Biology*, **161**, 805–819.

Cerchio S, Crudele L, Zerbini AN, *et al.* (In press) Satellite telemetry of humpback whales off Madagascar reveals insights on breeding behavior and long range movements within the Southwest Indian Ocean. *Marine Ecology Progress Series*, doi: 10.3354/meps11951

Cerchio S, Gabriele CM, Norris T, Herman LM (1998) Movements of humpback whales between Kauai and Hawaii: implications for population structure and abundance estimation in the Hawaiian Islands. *Marine Ecology Progress Series*, **175**, 13–22.

Cypriano-Souza AL, Fernandez GP, Lima-Rosa CAV, Engel MH, Bonatto SL (2010) Microsatellite genetic characterization of the humpback whale (*Megaptera novaeangliae*) breeding ground off Brazil (Breeding Stock A). *Journal of Heredity*, **101**, 189–200.

Darling J, Sousa-Lima R (2005) Songs indicate interaction between humpback whale (*Megaptera novaeangliae*) populations in the western and eastern South Atlantic Ocean. *Marine Mammal Science*, **21**, 557–566.

Dierickx EG, Shultz AJ, Sato F, Hiraoka T, Edwards SV (2015) Morphological and genomic comparisons of Hawaiian and Japanese Black-footed Albatrosses (*Phoebastria nigripes*) using double digest RADseq: implications for conservation. *Evolutionary Applications*, **8**, 662–678.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Elwen SH, Tonachella N, Barendse J *et al.* (2014) Humpback whales off Namibia: occurrence, seasonality and a regional comparison of photographic catalogs and scarring. *Journal of Mammalogy*, **95**, 1064–1076.

Ersts PJ, Pomilla C, Kiszka J *et al.* (2011) Observations of individual humpback whales utilising multiple migratory destinations in the south-western Indian Ocean. *African Journal of Marine Science*, **33**, 333–338.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–20.

Excoffier L, Lischer HEL (2010) Arlequin suite ver. 3.5: a new series of programs to perform population genetic analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 546–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Federal Register (2016) Endangered and threatened species; Identification of 14 Distinct Population Segments of the humpbak whales (*Megaptera novaeangliae*) and revision of species-wide listing. *Federal Register*, **81**, 62259–62320.

Feldheim KA, Gruber SH, DiBattista JD *et al.* (2014) Two decades of genetic profiling yields first evidence of natal philopatry and long-term fidelity to parturition sites in sharks. *Molecular Ecology*, **23**, 110–117.

Foote AD, Vijay N, Ávila-Arcos MC *et al.* (2016) Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications*, **7**, 11693.

Fossette S, Heide-Jørgensen M-P, Jensen MV *et al.* (2014) Humpback whale (*Megaptera novaeangliae*) post breeding dispersal and southward migration in the western Indian Ocean. *Journal of Experimental Marine Biology and Ecology*, **450**, 6–14.

Gaggiotti OE, Bekkevold D, Jørgensen HB *et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing the genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**, 2939–2951.

Gambell R (1976) World whale stocks. *Mammal Review*, **6**, 41–53.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on G(ST) and D: forget G(ST) but not all of statistics!. *Molecular Ecology*, **19**, 3845–3852.

Goudet J (1995) Fstat version 1.2: a computer program to calculate Fstatistics. *Journal of Heredity*, **86**, 485–486.

Guttal V, Couzin ID (2010) Social interactions, information use, and the evolution of collective migration. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16172–16177.

IWC (2006) Annex H: report of the sub-committee on other Southern Hemisphere Whale Stocks. *Journal of Cetacean Research and Management*, **8**(Suppl.), 151–170.

IWC (2009) Annex H: report of the sub-committee on other Southern Hemisphere Whale Stocks. *Journal of Cetacean Research and Management*, **11**, 220–247.

IWC (2011) Report of the workshop on the comprehensive assessment of Southern Hemisphere humpback whales. *Journal of Cetacean Research and Management (Special Issue)*, **3**, 1–50.

IWC (2012) Annex H: report of the sub-committee on other Southern Hemisphere Whale Stocks. *Journal of Cetacean Research and Management*, **13**(Suppl.), 1–73.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics*, **6**, 13. v.3.2.3 http://ibdws.sdsu.edu/

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jost L (2008) G(ST) and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099–1106.

Kershaw F (2015) *Understanding the evolution of two species of highly migratory cetacean at multiple scales and the potential value of a mechanistic approach*. PhD Thesis. Columbia University, New York.

Kershaw F, Leslie MS, Collins T *et al.* (2013) Population differentiation of 2 forms of Bryde's whales in the Indian and Pacific Oceans. *Journal of Heredity*, **104**, 755–764.

Kormann U, Gugerli F, Ray N, Excoffier L, Bollman K (2012) Parsimony-based pedigree analysis and individual-based landscape genetics suggest topography to restrict dispersal and connectivity in the endangered capercaillie. *Biological Conservation*, **152**, 241–252.

Kumar S, Skjæveland A, Orr RJS *et al.* (2009) AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics*, **10**, 357.

Lambertson RH (1987) A biopsy system for large whales and its use for cytogenetics. *Journal of Mammalogy*, **68**, 443–445.

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, **26**, 448–456.

Martins CCA, Andriolo A, Engel MH, Kinas PG, Saito CH (2013) Identifying priority areas for humpback whale conservation at Eastern Brazilian Coast. *Ocean & Coastal Management*, **75**, 63–71.

Mendez M, Rosenbaum HC, Subramaniam A, Yackulic C, Bordino P (2010) Isolation by environmental distance in mobile marine species: molecular ecology of franciscana dolphins at their southern range. *Molecular Ecology*, **19**, 2212–2228.

Minton G, Collins T, Findlay KP *et al.* (2011) Seasonal distribution, abundance, habitat use and population identity of humpback whales in Oman. *Journal of Cetacean Research and Management*, Special Issue, **3**, 186–198.

Muscarella RA, Murray KL, Ortt D, Russell AL, Fleming TH (2011) Exploring demographic, physical, and historical explanations for the genetic structure of two lineages of Greater Antillean bats. *PLoS ONE*, **6**, e17704.

Narum R (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783–787.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Palsbøll PJ, Vader A, Bakke I, Raafat El-Gewely M (1992) Determination of gender in cetaceans by the polymerase chain reaction. *Canadian Journal of Zoology*, **70**, 2166–2170.

Palsbøll PJ, Clapham PJ, Mattila DK *et al.* (1995) Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. *Marine Ecology Progress Series*, **116**, 1–10.

Palsbøll PJ, Bérubé M, Larsen AH, Jørgensen H (1997) Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Molecular Ecology*, **6**, 983–985.

Peakall R, Ebert D, Cunningham R, Lindenmayer D (2006) Mark-recapture by genetic tagging reveals restricted movements by bush rats (*Rattus fuscipes*) in a fragmented landscape. *Journal of Zoology*, **268**, 207–216.

Piou C, Prévost E (2012) A demo-genetic individual-based model for Atlantic salmon populations: model structure, parameterization and sensitivity. *Ecological Modelling*, **231**, 37–52.

Pomilla C (2005) Genetic structure of humpback whale (Megaptera novaeangliae) populations on Southern Hemisphere wintering grounds [PhD Thesis]. New York University, New York, NY.

Pomilla C, Rosenbaum HC (2005) Against the current: an inter-oceanic whale migration event. *Biology Letters*, **1**, 476–9.

Pomilla C, Rosenbaum HC (2006) Estimates of relatedness in groups of humpback whales (*Megaptera novaeangliae*) on two wintering grounds of the Southern Hemisphere. *Molecular Ecology*, **15**, 2441–2555.

Pomilla C, Amaral AR, Collins T *et al.* (2014) The world's most isolated and distinct whale population? Humpback whales of the Arabian Sea. *PLoS ONE*, **9**, e114162.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rayner MJ, Hauber ME, Steeves TE *et al.* (2011) Contemporary and historical separation of transequatorial migration between genetically distinct seabird populations. *Nature Communications*, **2**, 332.

Rittschof CC, Robinson GE (2014) Genomics: moving behavioral ecology beyond the phenotypic gambit. *Animal Behavior*, **92**, 263–270.

Rosenbaum HC, Pomilla C, Mendez M *et al.* (2009) Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *PLoS ONE*, **4**, e7318.

Rosenbaum HC, Maxwell S, Kershaw F, Mate B (2014) Long-range movement of humpback whales and their overlap with anthropogenic activity in the South Atlantic Ocean. *Conservation Biology*, **28**, 604–615.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Ruegg K, Rosenbaum HC, Anderson EC *et al.* (2013) Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. *Conservation Genetics*, **14**, 103–114.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schlötterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature*, **354**, 63–65.

Schmitt NT, Double MC, Jarman SN *et al.* (2014) Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. *Marine Mammal Science*, **30**, 221–241.

Schregel J, Eiken HG, Grøndahl FA *et al.* (2015) Y chromosome haplotype distribution of brown bears (*Ursus arctos*) in Northern Europe provides insight into population history and recovery. *Molecular Ecology*, **24**, 6041–6060.

Selkoe KA, Cassify ACD, Crandal E *et al.* (2016) A decade of seascape genetics: contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, **554**, 1–19.

Shamblin BM, Bjorndal KA, Bolten AB *et al.* (2012) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. *Molecular Ecology*, **21**, 2330–2340.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Smith TD, Allen J, Clapham PJ *et al.* (1999) An ocean-basin-wide mark-recapture study of north Atlantic humpback whale (*Megaptera novaeangliae*). *Marine Mammal Science*, **15**, 1–32.

Stevick PT, Neves MC, Johansen F *et al.* (2011) A quarter of a world away: female humpback whale moves 10,000 km between breeding areas. *Biology Letters*, **7**, 299–302.

Torres-Florez JP, Hucke-Gaete R, LeDuc R *et al.* (2014) Blue whale population structure along the eastern South Pacific Ocean: evidence of more than one population. *Molecular Ecology*, **23**, 5998–6010.

Valenzuela LO, Sironi M, Rowntree VJ, Seger J (2009) Isotopic and genetic evidence for culturally inherited site fidelity to feeding grounds in southern right whales (*Eubalaena australis*). *Molecular Ecology*, **18**, 782–791.

Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations. *Molecular Ecology*, **5**, 151–156.

Van Waerebeek K, Djiba A, Krakstad J *et al.* (2013) New evidence for a South Atlantic stock of humpback whales wintering on the Northwest African continental shelf. *African Zoology*, **48**, 177–186.

VanderWaal KL, Wang H, McCowan B, Fushing H, Isbell LA (2014) Multilevel social organization and space use in reticulated giraffe (*Giraffa camelopardalis*). *Behavioral Ecology*, **25**, 17–26.

Vighi M, Borrell A, Crespo EA, *et al.* (2014) Stable isotopes indicate population structuring in the Southwest Atlantic population of right whales (*Eubalaena australis*). *PLoS ONE*, **9**, e100024.

Wang IJ, Savage WK, Shaffer HB (2009) Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Molecular Ecology*, **18**, 1365–1374.

Weir (1990) *Genetic Data Analysis. Methods for Discrete Population Genetic Data*, 377 pp. Sinauer Associates, Inc, Sunderland, MA.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wolf JBW, Mawdsley D, Trillmich F, James R (2007) Social structure in a colonial mammal: unravelling hidden structural layers and their foundations by network analysis. *Animal Behaviour*, **74**, 1293–1302.

Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–138.

Data accessibility

Sampling locations, year of sampling, sex of individuals and microsatellite genotypes: Dryad doi:10.5061/dryad.h7db0.

F.K. and H.C.R. designed the research. F.K. performed analyses and wrote the manuscript with contributions from H.C.R., I.C., J.L., C.P., P.B.B., K.P.F., S.C., T.C., M.H.E., G.M., P.E., J.B., P.G.H.K., Y.R., S.N., M.M. and M.T. contributed data. All authors had the opportunity to review and offer comments on the manuscript.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Two methods of calculating geographical distance for the IBD analysis.

Fig. S2 Mean LnP(K) and Delta K (ΔK) plots for the STRUCTURE outputs.

Fig. S3 Distribution of 4 genetic clusters estimated using STRUCTURE.

Fig. S4 Distribution of individual reassignment by the DAPC.

Fig. S5 Magnitude and directionality of contemporary gene flow by BayesAss.

Table S1 Pairwise fixation indices values between breeding stocks and substocks.

Appendix S1 Additional methodological information.