ELECTRONIC APPENDIX

This is the Electronic Appendix to the article

Against the current: an inter-oceanic whale migration event

by

Cristina Pomilla and Howard C. Rosenbaum

Biol. Lett. (doi:10.1098/rsbl.2005.0351)

ELECTRONIC APPENDIX

2. MATERIALS AND METHODS

a. DNA extraction and molecular analyses

A total of 1,202 skin samples were collected from free-ranging humpback whales in the Indian Ocean off the northeast coast of Madagascar (n=722) from 1996 to 2001, and in the South Atlantic Ocean off the coast of Gabon (n=480) in years 2001 and 2002 (Table 1 of this Appendix), mostly using the biopsy dart procedure (Lambertsen 1987) or as sloughed skin when available. Genomic DNA was extracted using standard Phenol/Chloroform extraction protocol (Sambrook et al. 1989) or with the DNAeasy tissue kit (Qiagen).

Eleven cetacean di-, tri- and tetra- nucleotide microsatellite loci were selected from literature: 199/200, 417/418 and 464/465 (Schlötterer et al. 1991); EV1Pm, EV37Mn, EV94Mn and EV96Mn (Valsecchi & Amos 1996); and GATA028, TAA031, GATA053 and GATA417 (Palsbøll et al. 1997). One primer of each pair was labeled with a fluorescent tag on the 5′ end and polymerase chain reaction (PCR) for all samples was carried out in a 20μl or 10μl volume with the following conditions: 50mM KCl, 10mM Tris-HCl pH8.8, 2.5-3.5mM $MgCl_2$, 200μM of each dNTP, 0.4μM of each primer, 0.025 U/μl *Taq* Gold polymerase (Perkin-Elmer). Amplifications were completed after optimization of published annealing temperatures and profiles. Pooled PCR products were loaded with the addition of an internal standard ladder on an ABI 377 or 3700 DNA analyzer (Applied Biosystems). The allele size in base pairs was identified with the software GENESCAN ANALYSIS and GENOTYPER 2.1 (Applied Biosystems).

Sex determination was carried by PCR amplification followed by *Taq*I digestion of

the ZFX/ZFY region of the sex chromosomes, and amplification of the SRY region located on the Y chromosome (Palsbøll et al. 1992). To characterize maternal lineages we used a 486-basepair fragment containing the majority of variable nucleotide positions in the mitochondrial DNA (mtDNA) control region of humpback whales (Baker et al. 1993). Amplification protocols for this fragment and analyses of haplotype diversity are described in Rosenbaum et al. (2004).

All the analyses for the two identified matching samples were repeated in duplicate starting from the DNA extraction step.

b. Statistical analyses

Deviation from Hardy-Weinberg equilibrium (HWE) was evaluated for each locus and each population using a probability test based on a Markov chain Monte Carlo (dememorization 1000 iterations, batches 1000, iterations per batch 10000, Guo & Thompson 1992) as implemented in GENEPOP 3.4 (Raymond & Rousset 1995). The same software was used to investigate linkage disequilibrium (LD) between each pair of loci for each population in a total of 110 tests (dememorization 1000 iterations, batches 100, iterations per batch 10000). Significance levels (p=0.05) for departure from HWE and for LD were corrected for simultaneous comparisons with the sequential Bonferroni test (Rice 1989).

The average probability of different individuals in the populations sharing the same genotype by chance (Probability of Identity, PI) was estimated using the software API-CALC 1.0 (Ayres & Overall 2004; Waits et al. 2001). This software implements an algorithm corrected for the presence of relatives and sub-structure in the population. For this estimate the samples from both populations were pooled together and corrected for

F_{ST}=0.0021 (Pomilla et al. 2004). Since it is more likely that two relatives will share the same genotype, we also calculated, for the specific case of the match between Madagascar and Gabon the more conservative PI for siblings (PI_{sib}) as advised by Waits et al. (2001) and PI for parent-offspring (PI_{pof}). Although a relationship of siblings is very unlikely in humpback whales (Clapham & Palsbøll 1997), PI_{sib} is more conservative than PI_{pof}, which would be the next closer kin relationship. These were calculated for each locus (for heterozygote genotypes $PI_{sib}=(1+p_i+p_j+2p_ip_j)/4$ and $PI_{pof}=(p_i+p_j)/2$; for homozygote genotypes $PI_{sib}=(1+2p_i+2p_i^2)/4$ and $PI_{pof}=p_i$; p_i is the frequency of the ith allele and p_j is the frequence of the jth allele), and then multiplied across loci.

The hypothesis of mother-offspring relationship between the samples BA-00-S-200 (offspring) and BA-00-S-201 (mother) was tested with a maximum likelihood estimator implemented in KINSHIP 1.3.1, a program for testing hypotheses of pedigree relationships between pairs of individuals (Goodnight & Queller 2000). Given the hypothesis, KINSHIP uses the population allele frequencies and the genotypes of the two individuals under consideration to calculate the likelihood that this genotype combination could have been produced by the relationship as specified. The likelihood for the primary hypothesis and a null hypothesis are calculated, and the ratio between them (primary/null) is reported. A high value of the ratio favours the primary hypothesis and a low value rejects it in favour of the null hypothesis. The primary hypothesis of mother-offspring relationship was tested against the null hypothesis of half-sibling relationship. The significance level of the obtained ratio was calculated by simulating 1000 pairs of individuals using the hypothesis settings and the observed allele frequencies and determining the ratio needed to reject the null hypothesis with p = 0.05, 0.01 and 0.001.

To estimate migration rates between Madagascar and Gabon we used a maximum-likelihood framework based on coalescence theory, implemented in the program MIGRATE 2.0.3 (Beerli & Felsestein 1999; Beerli & Felsestein 2001). Specifically, the parameters estimated by MIGRATE for diploid bi-parentally inherited markers as microsatellites are M (m/μ) and Θ $(4N_e\mu)$ where m is the immigration rate, μ the mutation rate, and N_e the effective population size. The product ΘM results in the number of immigrants per generation $4N_e m$ here reported simply as $N_e m$. MIGRATE assumes that the sampled populations represent all populations, the likelihood of which is rare. Missing populations (ghost populations) can create the appearance of migration between populations that do not actually exchange migrants (Beerli 2004). To minimize the effect of ghost populations, beside Gabon and Madagascar, the analysis included seven additional sampled localities within the wintering regions of the western (one site) and eastern (one site) South Atlantic, and of the Southwestern (four sites) and northern (one site) Indian Oceans. Methodology and results of this analysis are described in (Pomilla 2005).

Table 1. Years of sampling and number of samples collected from individual (N) humpback whales in Gabon and Madagascar. Number of known female (F) and male (M) individuals is also shown.

Sampling Site	Years	Samples	N	F	M
Iguela and Gamba, Gabon	2001	181	150	39	111
	2002	299	255	98	156
Antongil Bay, Madagascar	1996	98	72	19	52
	1997	53	45	13	31
	1998	91	68	17	49
	1999	149	113	34	76
	2000	144	112	38	71
	2001	187	154	41	113

REFERENCES

Ayres, K. L. & Overall, D. J. 2004 api-calc 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives. *Molec. Ecol. Notes* **4**, 315-318.

Baker, C. S., Perry, A., Bannister, J. L., Abernathy, R. B., Weinrich, M. T., Abernathy, B., Calambokidis, J., Lien, J., Lambertsen, R., Urban, J., Clapham, P., Ailing, A., O'Brien, S. J. & Palumbi, S. R. 1993 Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8239-8243.

Beerli, P. & Felsestein, J. 1999 Maximum-Likelihood Estimation of Migration Rates and Effective Population Numbers in Two Populations Using a Coalescent Approach. *Genetics* **152**, 763-773.

Beerli, P. & Felsestein, 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 USA* **98**, 4563-4568.

Beerli, P. 2004 Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology* **13**, 827-836.

Clapham, P. J. & Palsbøll, P. J. 1997 Molecular analysis of paternity shows promiscuous mating in female humpback whales. *Proceedings of the Royal Society of London - series B* **264**, 95-98.

Goodnight, K. F. & Queller, D. C. 2000 KINSHIP: Rice University.

Guo, S. & Thompson, E. 1992 Performing the exact test of Hardy Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372.

Lambertsen, R. H. 1987 A biopsy system for large whales and its use for cytogenetics. *J. Mamm.* **68**, 443-445.

Palsbøll, P. J., Bérubé, M., A.H., L. & Jørgensen, H. 1997 Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Molec. Ecol.* **6**, 893-895.

Palsbøll, P. J., Vader, A., Bakke, I. & Raafat El-Gewely, M. 1992 Determination of gender in cetaceans by the polymerase chain reaction. *Can. J. Zool.* **70**, 2166-2170.

Pomilla, C. 2005 *Genetic stucture of humpback whale (Megaptera novaeangliae) populations on Southern Hemisphere wintering grounds*. Ph.D Thesis. pp. 185. New York (US): New York University.

Pomilla, C., Best, P. B., Findlay, K., Kotze, P. J. H., Engel, M. H., Barendse, J. & Rosenbaum, H. C. 2004 Population structure of Southern Hemisphere humpback whales from wintering Regions A, B, and C based on nuclear microsatellite variation, pp. 12: Paper SC/56/SH4 presented to the IWC the Scientific Committee, (unpublished).

Raymond, M. & Rousset, F. 1995 GENEPOP (Version 1.1): population genetic software for exact tests and ecumenicism. *J. Heredity* **86**, 248-249.

Rice, W. R. 1989 Analyzing table of statistical tests. *Evolution* **43**, 223-225.

Rosenbaum, H. C., Pomilla, C., Leslie, M., Best, P. B., Findlay, K. P., Engel, M. H., Minton, G. & Collins, T. 2004 MtDNA diversity and population structure of humpback whales from their wintering areas in the Indian and South Atlantic Ocean (Breeding regions A, B, C and X), pp. 10: Paper SC/56/SH3 presented to the IWC Scientific Committee (unpublished).

Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989 *Molecular Cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

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

Valsecchi, E. & Amos, W. 1996 Microsatellite markers for the study of cetacean populations. *Molec. Ecol.* **5**, 151-156.

Waits, L., Luikart, G. & Taberlet, P. 2001 Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molec. Ecol.* **10**, 249-256.